

Trabajo Fin de Grado

Facultad de Ciencias Experimentales



Grado en Química

Clonación y expresión de las enzimas
Endoglucanasa M1 y M2 de *bacterias termófilas*.

Cloning and expression of the *Endoglucanase*
enzymes M1 and M2 from thermophilic bacteria.

Nerea Rodríguez Pardo
Curso académico 2020-2021
Julio 2021

Tutores
Dr. Josefa María Clemente Jiménez
Dr. Felipe Rodríguez Vico

Área de Bioquímica y Biología Molecular
Departamento de Química y Física

UNIVERSIDAD DE ALMERÍA



Clonación y expresión de las enzimas Endoglucanasa M1 y M2 de bacterias termófilas.

Memoria del Trabajo Fin de Grado en Química presentada por
Nerea Rodríguez Pardo

Almería, 10 de Julio de 2021

Fdo: Nerea Rodríguez Pardo

Fdo: Josefa María Clemente Jiménez

Fdo: Felipe Rodríguez Vico

Agradecimientos

Este trabajo supone fin a una etapa muy dura, pero a la vez muy enriquecedora tanto en el ámbito profesional como a nivel personal.

En primer lugar, me gustaría agradecerles a mis tutores Josefa María Clemente Jiménez, Felipe Rodríguez Vico, Lellys Mariela Contreras Moyeja y Francisco Javier Las Heras Vázquez, por todo su esfuerzo, dedicación y paciencia conmigo en todos estos meses.

Me gustaría hacer mención especial a mi grupo de amigos que he encontrado dentro de la carrera, comenzamos siendo compañeros de estudio y se han convertido en una pequeña familia.

Como no, agradecerse a mis padres y a mi abuela, sin vuestro apoyo, cariño y comprensión durante estos años, no hubiera podido llegar hasta aquí

1. RESUMEN.....	1
2. ABSTRACT.....	1
3. INTRODUCCIÓN.....	3
3.1. Biorremediación.	3
3.2. Biomasa, naturaleza y composición.....	4
3.2.1. Biorremediación enzimática.	4
3.3. Endoglucanasa.....	5
3.4. Organismos termófilos.....	6
3.5. Aplicaciones Industriales de la celulosa.	7
3.6. Clonación de genes.....	8
3.6.1. Clonación del producto obtenido por la reacción en cadena de la polimerasa (PCR)	8
3.6.2. Vectores de subclonado pTOPO.....	9
3.7. Enzimas de restricción.	10
3.8. Sistemas de expresión.	11
3.8.1. Cepas procariotas utilizadas en el sistema de expresión.....	13
3.8.2. Vectores de expresión.....	14
3.8.2.1. Vector pET.....	15
3.8.2.2 Vector pJOE4036.1.....	15
4. OBJETIVOS.....	16
5. MÉTODOS EXPERIMENTALES.....	17
5.1. Extracción de ADN <i>G. kaustophilus</i> CECT4264 y <i>G.stearothermophilus</i> CECT43 y 49 a través del método de ebullición.....	17
5.2. Preparación de oligos.....	17
5.3. Amplificación de los genes mediante PCR.	17
5.4. Electroforesis en gel de agarosa.	18
5.5. Purificación de fragmentos de ADN a partir de gel.	18
5.6. Preparación de células competentes.	19
5.7. Sub-Clonaje en pTOPO.	19
5.8. Crecimiento de <i>E. coli</i> DH5 α con los plásmidos de expresión.....	20
5.9. Extracción de plásmidos.	20
5.10. Corte con enzimas de restricción.....	21
5.11. Unión de fragmentos de ADN	22
5.12. Expresión de proteínas: crecimiento e inducción de las células transformadas.....	22
5.13. Ruptura celular.....	22
5.14. Electroforesis de proteínas mediante SDS-PAGE.....	23
5.15. Ensayo de actividad Endoglucanasa.....	24

6. RESULTADOS Y DISCUSIÓN.....	25
6.1. Búsqueda de secuencias de posibles Endoglucanas.....	25
6.2. Diseño de los oligonucleótidos para la amplificación de los posibles genes de M1 y M2 de <i>G. kaustophilus</i> CECT4264 y <i>G.stearothermophilus</i> CECT43 y 49.....	29
6.3. Extracción de ADN de <i>Geobacillus kaustophilus</i> CECT4264 y <i>Geobacillus stearothermophilus</i> CECT43 y 49.....	31
6.4. Amplificación mediante PCR de los genes M1 y M2.....	31
6.5. Clonación de los genes en el vector pTOPO.....	34
6.6. Construcción de los plásmidos de expresión.....	38
6.7. Análisis de las secuencias.....	41
6.8. Expresión de las proteínas recombinantes.....	42
6.9. Medida de la actividad de las proteínas recombinantes.....	44
7. CONCLUSIONES.....	47
8. BIBLIOGRAFÍA.....	48

ANEXOS

ANEXO 1

1. RESUMEN

La biorremediación es la principal vía de degradación de ambientes contaminados mediante el uso de microorganismos o biomoléculas. Se han buscado genes con probable actividad endoglucanasa, usándose como molde el genoma de *Geobacillus* sp. WSUCF1. Se identificaron los posibles genes y se diseñaron los oligonucleótidos necesarios para la amplificación y posterior clonaje molecular a partir de ADN de *Geobacillus kaustophilus* CECT4264 y *Geobacillus stearothermophilus* CECT43 y 49. Este proceso se llevó a cabo mediante la amplificación del fragmento por PCR y su posterior clonaje en los plásmidos de expresión pET21b+ y pJOE4036.1. Se aislaron, analizaron y expresaron los genes que codifican para la Endoglucanasa M1 y M2 (pJMC120 y pJMC122). Por último, los genes clonados de las Endoglucanasas M1 y M2 se expresaron en *E. coli* C43 y en *E. coli* BL21, siendo el peso de estas proteínas activas de 42 y 40 kDa, respectivamente.

Palabras clave: Reacción en cadena de la polimerasa (PCR), *Geobacillus kaustophilus* (*Gk*) y *Geobacillus stearothermophilus* (*Gs*), *Escherichia coli* (*E. coli*), Endoglucanasa.

2. ABSTRACT

Bioremediation is the main way of degradation of contaminated environments through the use of microorganisms or biomolecules. Genes with probable endoglucanase activity have been searched, using the *Geobacillus* sp. WSUCF1 genome as a template. The possible genes were identified and the necessary oligonucleotides were designed for the amplification and subsequent molecular cloning from DNA of *Geobacillus kaustophilus* CECT4264 and *Geobacillus stearothermophilus* CECT43 and 49. This process was carried out by amplifying the fragment by PCR and its subsequent cloning in the expression plasmids pET21b+ and pJOE4036.1. The genes coding for Endoglucanase M1 and M2 (pJMC120 and pJMC122) were isolated, analyzed and expressed. Finally, the cloned genes for endoglucanase M1 and M2 were expressed in *E. coli* C43 and *E. coli* BL21, the weight of these active proteins being 42 and 40 kDa, respectively.

Keywords: Polymerase chain reaction (PCR), *Geobacillus kaustophilus* (*Gk*) and *Geobacillus stearothermophilus* (*Gs*), *Escherichia coli* (*E. coli*), Endoglucanase.

3. INTRODUCCIÓN

3.1. Biorremediación.

Siempre se ha creído que la Tierra sería una fuente inagotable de recursos, sin embargo, el estado actual de los recursos y de la Tierra, nos permite observar que esto no es así. Por ejemplo, el uso de fertilizantes químicos en la agricultura, aunque mejora el rendimiento y productividad de los cultivos también genera grandes cantidades de nitrógeno que son perjudiciales para el suelo y el ecosistema (Sonal y Reeta, 2013).

Existen varios métodos por los cuales se sustituyen los espacios contaminados de uso agrícola, siendo la práctica más habitual llevar el suelo contaminado a un vertedero y sustituirlo por sustrato nuevo. Sin embargo, esto puede crear riesgos importantes durante la excavación, manipulación y transporte de material peligroso, siendo además caro y muy difícil. También se han utilizado técnicas como incineración a alta temperatura y uso químico para descomposición (por ejemplo, decoloración catalizada por bases y oxidación UV) (Sonal y Reeta, 2013).

Para superar estos inconvenientes, una mejor perspectiva es destruir completamente el contaminante, o transformarlo en sustancias biodegradables. Este enfoque puede lograrse mediante el uso de una técnica conocida como biorremediación.

Según Van Dillewijn (Van Dillewijn y col., 2007) la "Biorremediación" se define como el proceso mediante el cual varios agentes biológicos, principalmente microorganismos degradan el medio ambiente contaminado en compuestos menos tóxicos. La primera patente de este proceso no fue publicada hasta el año 1974 (Prescott y col., 2002), en el que se utilizó una cepa de *Pseudomonas putida* para degradar el petróleo. En 1991, se sumaron más de 70 géneros microbianos que degradaban compuestos del petróleo y casi un número igual se ha añadido a la lista en las dos décadas sucesivas (Sonal y Reeta, 2013).

Las principales ventajas de la biorremediación sobre el tratamiento convencional incluyen: bajo costo, alta eficiencia, minimización de lodos químicos y biológicos, selectividad para metales, sin necesidad de nutrientes adicionales y la posibilidad de recuperación de metales.

La biorremediación puede ocurrir por sí sola en la naturaleza (atenuación natural o biorremediación intrínseca) o se puede estimular mediante la adición de fertilizantes para la mejora de biodisponibilidad en el medio.

Existen una serie de tecnologías aplicadas únicamente a este campo, entre las que pueden estar la bioestimulación, bioventilación, biolixiviación, biorreactor, compostaje, cultivo de la tierra, fitorremediación y por último la rizofiltración (Sonal y Reeta, 2013).

En los últimos años los microorganismos han mostrado grandes capacidades para su aplicación en biorremediación de suelos, aguas subterráneas, y entornos de lodos activados. Esto se debe en gran medida al desarrollo de las técnicas de biología molecular que han permitido la amplificación, alteración y modificación de los genes diana que codifican las enzimas de las vías metabólicas (Melchor Partida, 2014).

3.2. Biomasa, naturaleza y composición.

La biomasa está compuesta por un conjunto heterogéneo de materia orgánica, siendo la más abundante la de origen vegetal, la cual está compuesta en su mayoría por celulosa, lignina y hemicelulosa. La celulosa se trata de un polisacárido lineal compuesto por más de diez mil moléculas de D-glucosa, unidas entre sí por enlaces β -1,4, siendo la unidad de repetición la celobiosa. Aunque presenta resistencia a la degradación enzimática, su gran homogeneidad es una ventaja, ya que provoca la despolimerización completa en un solo producto, la glucosa (Horn y col., 2012). La hemicelulosa, se trata de un polisacárido no celulósico, que presenta diferentes grados de polimerización. El polisacárido más abundante es el xilano, y se encuentra sobre todo en pastos. La hemicelulosa generalmente es más fácil de degradar enzimáticamente que la celulosa, pero existen ciertas estructuras oligoméricas que son recalcitrantes debido a la compleja ramificación y patrones de acetilación. En la figura 1 se enumeran las enzimas implicadas en la degradación de celulosa y hemicelulosa (Santosh y col., 2020). Por último, encontramos la lignina, la cual es un polímero relativamente hidrófobo y aromático que consta de tres monolignoles que están metoxilados en varios grados. Las enzimas que actúan sobre la lignina son en su mayoría oxidorreductasas dependientes de cofactores, lo que implica que su uso en la industria será bastante caro (Horn y col., 2012).

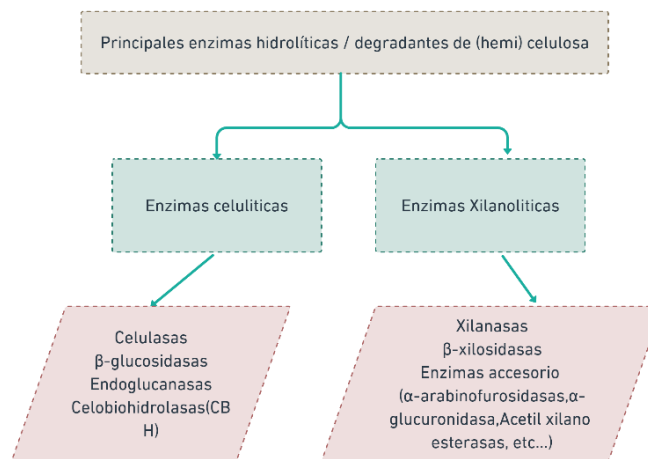


Figura 1: Enzimas hidrolíticas primarias que participan activamente en la degradación de celulosa / hemicelulosa (Santosh y col., 2020).

3.2.1. Biorremediación enzimática.

La biorremediación enzimática consiste en utilizar la capacidad catalizadora, eficaz, estable y degradativa de las enzimas, con el fin de eliminar sustancias nocivas para el medio ambiente (Copley, 2009). Se ha demostrado que las enzimas tienen mecanismos para degradar gran cantidad de compuestos, como por ejemplo compuestos halogenados (Melchor Partida, 2014). Al ser catalizadores biológicos facilitan la conversión de sustratos en productos favoreciendo condiciones que disminuyen la energía de activación de la reacción. Las enzimas aisladas de microorganismos no pierden sus propiedades, ya que pueden ser más eficaces cuando están puras.

Como todo proceso, el uso de enzimas en biorremediación tiene una serie de ventajas y desventajas que se resumen en la siguiente tabla (Tabla 1):

Tabla 1: Ventajas y desventajas del uso de enzimas en biorremediación (Melchor Partida, 2014).

Ventajas	Desventajas
Transforman la estructura y propiedades de los contaminantes hasta disminuirlos o eliminarlos.	Dificultad en el aislamiento y/o purificación a nivel técnico y económico.
Presentan promiscuidad de sustrato.	Necesita puesta a punto individual, así como proporciones en el caso de cócteles enzimáticos.
Se pueden utilizar en altas concentraciones de contaminante.	
Estabilidad, ya que no son consumidas durante la catálisis, actúan como catalizadores.	

Las enzimas potencialmente útiles para la biorremediación deben tener una serie de características como son: Promiscuidad de sustrato, estabilidad, eficacia y bajo coste. Dos enzimas importantes que se utilizan para biorremediar los contaminantes persistentes organoclorados son la Glutación-S-transferasa y la Catecol I 2,3-dioxigenasa (Melchor Partida, 2014).

3.3. Endoglucanasa.

La celulosa es un polímero largo, compuesto por más de 10,000 moléculas de glucosa, su proceso de degradación comienza con la degradación en subunidades muy pequeñas que podrán penetrar en las células microbianas y ser metabolizadas. Las celulasas son un grupo muy diverso de enzimas que catalizan la hidrólisis de los enlaces glucosídicos β -1,4 que mantiene unidas las cadenas de celulosa. El proceso de degradación de estos enlaces glucosídicos y por ende la degradación de celulosa, está relacionado con la acción de 3 tipos de enzimas, que son las β -1,4-endoglucanasas, β -1,4-exoglucanasas y β -glucosidasas. Estas enzimas son producidas por bacterias, hongos, actinomicetos y protozoos, incluyéndose los géneros *Bacillus*, *Cellulomonas*, *Clostridium*, *Corynebacterium*, *Cytophaga*, *Polyangium*, *Pseudomonas*, *Sporocytophaga*, *Vibrio*, *Micromonospora*, *Nocardia*, *Streptomyces* y *Streptosporangium* (Viteri y col., 2015). Las diversas formas de celulosa pueden ser en parte una consecuencia de la variación en la presencia de módulos de unión de carbohidratos (CBM) que están unidos covalentemente a los dominios catalíticos de éstas (Horn y col., 2012).

Además de la degradación de la celulosa en los residuos sólidos orgánicos, las celulasas han sido utilizadas en la producción de alimentos, ácidos orgánicos, azúcares fermentables, etanol, bebidas, textiles, detergentes, papel, pulpa y tintas. Aunque la fuente de celulasas utilizadas en la industria está representada principalmente por hongos, se han desarrollado muchos estudios con el fin de aislar y caracterizar

bacterias productoras de celulasas con un mayor grado de especificidad y efectividad (Viteri y col., 2015).

Las endoglucanasas se encargan de hidrolizar aleatoriamente los enlaces glicosídicos internos, que darán como resultado una rápida disminución de la longitud del polímero y en consecuencia un aumento de la concentración de azúcar (de forma reductora) (Berlemont y Gerday, 2011).

La actividad endoglucanasas en bacterias del género *Geobacillus*, no se identificó hasta 2004, que fue cuando se reportó que la bacteria *G. thermoleovorans* T4, era capaz de secretar una endocelulosa termoestable, que hidroliza diversos sustratos entre ellos la carboximetilcelulosa (CMC) (Tai y col., 2004). Fueron los mismos autores los que demostraron que *G. stearothermophilus*, era capaz de hidrolizar celulosa cristalina (Tai y col., 2004). La utilización de organismos termófilos y por ello sus enzimas afines, han aumentado en los últimos años debido a que las enzimas termoestables, al trabajar a altas temperaturas, tienen mayor capacidad de penetrabilidad en la biomasa (Paes y O'Donohue, 2006). Las grandes ventajas del uso de las enzimas termoestables provienen de su elevado tiempo de almacenamiento, su tolerancia a disolventes orgánicos, bajo riesgo de contaminación, así como su baja pérdida de actividad al trabajar a temperaturas próximas a su temperatura de desnaturalización (Turner y col., 2007).

3.4. Organismos termófilos.

Los microorganismos termófilos tienen la capacidad de crecer a altas temperaturas, incluso a temperaturas superiores a la temperatura de ebullición del agua, esto implica que las enzimas que producen estos organismos estarán adaptadas para trabajar a altas temperaturas. Esto no quiere decir necesariamente que las enzimas termófilas sean capaces de mantenerse estables a esas altas temperaturas, ya que existen múltiples factores, como pueden ser otros componentes celulares o la presencia de algunos solutos que impidan el despliegue térmico (Berlemont y Gerday, 2011).

Uno de los organismos que llega a temperaturas más extremas es la arquea *Pyrodictium occultum*, la cual es comúnmente conocida como cepa 121 ya que crece a temperaturas tan altas como 121 °C (Berlemont y Gerday, 2011). Los microorganismos termófilos no se utilizan a menudo en procesos biotecnológicos. Sin embargo, sus enzimas extracelulares, principalmente en combinación con tecnologías de producción recombinante han recibido mucha atención en investigación debido a su posible incorporación en procesos biocatalíticos que se ejecutan a altas temperaturas.

A pesar de todo este esfuerzo, el uso de enzimas termófilas es aún bastante limitado debido principalmente al hecho de que muestran una baja actividad específica a temperaturas moderadas debido al alto valor de su energía de activación y también porque aunque se hayan clonado y expresado una gran variedad de enzimas termoestables, los rendimientos siguen siendo muy bajos debido a las diferencias en el uso de codones, a un plegamiento inadecuado en *E. coli* y al hecho que muchas enzimas son heterooligómeros complejos tales como celulosomas o requieren cofactores unidos covalentemente (Berlemont y Gerday, 2011).

Geobacillus steatothermophilus, es una bacteria del filo *Firmicutes*, gram positiva y con forma de bacilo. Es una bacteria termófila aeróbica formadora de esporas elipsoidales, que se pueden encontrar ampliamente distribuida por el suelo, sedimentos oceánicos y manantiales calientes. Su temperatura máxima de crecimiento es de 65-75 °C, siendo la mínima 40 °C y la temperatura optima es de 55 °C. El pH óptimo para su crecimiento es de 5,2. Dicha especie es la responsable del amargor de los alimentos. Las esporas de *Geobacillus steatothermophilus* se utilizan como indicador biológico de un correcto proceso de esterilización de los productos en las industrias alimentarias (Nitrigual Matamala, 2012).

Geobacillus kaustophilus HTA426, al igual que *Geobacillus steatothermophilus* es una bacteria termófila aeróbica, *Firmicutes*, gram positiva y con forma de bacilo. Se aisló del sedimento de las aguas profundas de las fosas de las marianas (Suzuki y Ken-Ichi., 2012), su temperatura máxima de crecimiento es de 42-74 °C, siendo la temperatura optima de 60 °C (Megumi y col., 2018).

3.5. Aplicaciones Industriales de la celulosa.

La celulosa se ha convertido en un polímero de potencial utilización industrial, gracias a su disponibilidad y su alta degradabilidad biológica. La biomasa lignocelulósica presenta una gran ventaja, como es la elevada resistencia, proporcionándole múltiples usos en diversas industrias y procesos industriales encargados de aprovechar la energía resultante de la degradación de las paredes de celulosa de las plantas. A continuación, se describen algunos procesos industriales en los que participan la celulosa y las enzimas que la degradan:

- La industria papelera es una de las principales demandantes de celulosa y de las enzimas que la degradan, ya que la acción de estas sobre la materia prima mejora el drenaje pulpar, lo que conlleva un ahorro de la madera y reducción del gasto de productos químicos (Bajpai, 1999; Verma., y col., 2017).
- Aditivos para piensos de aves de ganadería, con los que se mejora la ganancia de peso y la eficiencia a la conversión al mejorar la digestibilidad de los piensos, ya que las enzimas celulolíticas ayudan a la reducción de la viscosidad intestinal degradando los polímeros del pienso que son los encargados de la absorción del agua (Bedford y Classen, 1992).
- Uso en la industria de tejidos, como mejora de la resistencia a la rotura y a la alta compresibilidad (Aarstad y col., 2017).
- Uso en el proceso de reciclaje de la basura, ya que añadiendo cantidad de nanofibras de celulosa sobre la matriz polimérica se produce una mejora de las propiedades mecánicas y propiedades barrera a los gases de los films alimentarios (Eduardo Espinosa, 2019).
- Las fibras de nanocelulosa extraídas de las paredes celulares de las plantas se usan en la industria alimentaria y en cosmética para la generación de emulsiones comestibles (Long y col., 2018).

3.6. Clonación de genes.

Un clon se define como un grupo de células o macromoléculas todas idénticas a una célula o macromolécula original ancestral, mientras que si hablamos en el contexto de la biotecnología nos referiremos a la clonación como el proceso mediante el cual se obtiene una molécula capaz de replicar un fragmento de ADN. Es por ello que el clonaje de fragmentos de DNA permite tener cantidades infinitas de éste a partir de una sola molécula (Perera y col., 2002)

La tecnología empleada en el clonaje permite la construcción de nuevas moléculas de ADN, mediante secuencias de distinto origen, conociéndose el producto como ADN recombinante. La posibilidad de clonar es debida a la propiedad que tiene los plásmidos de seguir el desarrollo normal a pesar de haber modificado su material genético (Ávila y col., 1995; Perera y col., 2002). Es decir, autorreplicarse siempre y cuando tenga el origen de replicación y así poder formar muchas copias del propio plásmido, junto con todas las secuencias introducidas en él.

Un vector (plásmido) de clonaje posee un lugar en el cual el ADN foráneo (ADN no perteneciente a la célula de origen), se puede insertar sin interrumpir la función del vector. El fragmento de interés del ADN foráneo generado se une a la zona de clonaje del vector, previo corte del plásmido, dando lugar a un plásmido quimérico circular que al introducirse en la célula huésped se multiplicará generando un gran número de copias (Perera y col., 2002). Los plásmidos poseen genes que presentan resistencia a antibióticos, y permite diferenciar las células que contienen al plásmido de aquellas que no lo contienen. Los tamaños de los plásmidos pueden variar entre 3-5 kb y el fragmento que pueden albergar puede variar entre 0,5 y 5 kb (Ávila y col, 1995).

El ADN recombinante se construye, utilizando las mismas enzimas de restricción, tanto en el vector como en el ADN diana, produciéndose extremos romos o cohesivos. El restablecimiento de enlaces entre el vector y el fragmento se hace con ayuda de una enzima denominada ligasa, la cual se encarga de unir los fragmentos de ADN (Ávila y col, 1995).

3.6.1. Clonación del producto obtenido por la reacción en cadena de la polimerasa (PCR).

En el año 1984, Kary B. Mullis (Mullis, 1991), quien sería nombrado posteriormente Premio Nobel en Química (1993), aplicó la actividad de la enzima polimerasa a la tecnología del ADN recombinante, diseñando la Reacción en Cadena de la Polimerasa (PCR de su acrónimo en inglés *Polymerase Chain Reaction*). Ha significado un gran desarrollo científico e impacto en lo que fue el Proyecto Genoma Humano y actualmente está siendo utilizada como técnica de rutina para diagnosticar enfermedades, clonar y secuenciar genes. Dos claros ejemplos han sido en medicina forense para la identificación de muestras biológicas o en el trasplante de órganos, donde se utiliza para establecer un grado de semejanza de los HLA (*Human Leucocyte Antigen*), ya que en un trasplante de órgano depende de la semejanza entre los HLA de los donantes y del receptor (Ávila y col, 1995; Garibyan y Avashia, 2013).

Para poder llevarse a cabo una PCR, se requiere la presencia de:

1. ADN molde, donde se encuentra el fragmento de ADN a amplificar.
2. Los cebadores (oligonucleótidos) que son fragmentos cortos de ADN con una secuencia definida y complementaria al ADN molde.
3. Nucleótidos, que corresponden a las 4 bases que forman el ADN (Adenina, Timina, Citosina y Guanina).
4. ADN polimerasa, es la encargada de añadir los nucleótidos a partir de los cebadores, de forma complementaria a la hebra molde, para que al final se obtengan múltiples copias de doble hebra de una zona concreta (producto de PCR).

La mezcla de reactivos se introduce en un tubo de ensayo o placa de 96 pocillos y luego se colocan en un termociclador que permite que se produzcan ciclos repetidos de amplificación del ADN en tres pasos básicos.

El primer paso produce la separación de la doble hebra de la molécula de ADN, se mantiene a 95 °C durante 15 segundos. En un segundo paso se produce la hibridación de los cebadores a 55-60 °C, donde cada uno de ellos se une al extremo 3' del ADN a amplificar. Por último, se produce la síntesis de ADN donde la disolución resultante del segundo paso se calienta hasta 72 °C, que es la temperatura óptima para la actividad de la polimerasa. La síntesis de cada hebra de ADN que se desea amplificar se lleva a cabo en la dirección 5' → 3'. Estas 3 etapas, constituyen un único ciclo. La reacción consiste en repetir estas tres etapas (1 ciclo) muchas veces sin alterar la mezcla, variando únicamente la temperatura que tiene cada etapa (95, 55, 72 °C). Tras n ciclos se produce la amplificación de la secuencia deseada de 2^n .

Una vez llevada a cabo la amplificación de un fragmento de ADN mediante la reacción en cadena de la polimerasa (PCR), éste se puede insertar en un plásmido siguiendo varias técnicas específicas.

Existen al menos 3 estrategias para ello:

1) Introducción de sitios de restricción al principio y final del fragmento amplificado, donde el producto de PCR y el vector de clonación seleccionado, son digeridos por las enzimas de restricción, generándose así extremos complementarios que posteriormente serán ligados. 2) La estrategia T/A, en la que el vector necesita únicamente extremos cohesivos 3' de una sola base, seguidamente estos plásmidos llamados vectores-t, pueden clonar diferentes fragmentos de PCR. 3) Subclonado en TOPO el cual al ser empleado en este TFG se detallará a continuación (Patel, 2009).

3.6.2. Vectores de subclonado pTOPO.

La clonación en plásmidos pTOPO es una técnica utilizada en Biología Molecular en la que los fragmentos de ADN se clonan en vectores específicos sin el uso de ADN ligasas. Los elementos presentes en los plásmidos de la serie TOPO, son secuencias funcionales de ácidos nucleicos (promotores, colas de fusión, terminación, secuencias, etc.), unidos a la Topoisomerasa, que cataliza la unión de los elementos funcionales a los productos de PCR en una reacción rápida que se realiza a temperatura ambiente (Sambrook y col., 2012).

La clave para ese proceso es la enzima Topoisomerasa (Figura 2), que funciona como enzima de restricción y como ligasa. Su función biológica es la de cortar y volver a unir el ADN en el proceso de replicación.

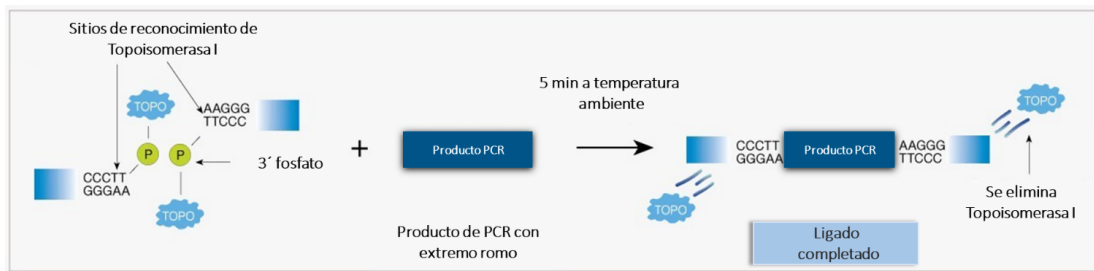


Figura 2: Representación gráfica del sistema de expresión TOPO y su acción junto con la Topoisomerasa I (Shuman, 1991, 1994).

✓ Topoisomerasa I

La topoisomerasa I, fue aislada del virus *vaccinia* y al igual que las topoisomerasas de tipo IB, se encargan de relajar el superenrollamiento del ADN rompiendo y uniendo de forma transitoria una doble hebra de ADN.

Esta ruptura se desarrolla de manera específica sobre la secuencia 5' (C/T) CCTT, creando un intermedio fosfotirosil-ADN estable. Este intermedio puede actuar de 2 formas: 1) catalizar el ligado de la cadena doble para volver a su forma original, o bien, 2) Crear un ADN recombinado, formando un enlace fosfodiéster con un ADN aceptor (fragmento a unir) con un OH en extremo 5' complementario a la del complejo donador (Sambrook y col., 2012).

3.7. Enzimas de restricción.

Una enzima de restricción puede sufrir una serie de cambios conformacionales a la hora de reconocer el ADN, lo que facilita la unión de la enzima de forma mucho más específica.

Por definición las endonucleasas de restricción forman parte de los sistemas de restricción-modificación (RM), que proviene de la actividad de las endonucleasas y las metiltransferasas. Se han encontrado que existen 3 tipos de sistemas de MR y se clasifican según la composición de sus subunidades, requerimiento de cofactor y modo de acción (Pingoud y Jeltsch, 2001). Durante el presente estudio se han utilizados 3 enzimas de restricción del tipo II:

- *Bam*HI es una endonucleasa que se une a la secuencia de reconocimiento bicatenario 5'-G/GATCC-3' y escinde estas secuencias justo después de la 5'-guanina en cada hebra (Pingoud y Jeltsch, 2001).

- *NdeI*, es una enzima de restricción que reconoce a la secuencia de ADN bicatenario 5'-CA/TATG-3'. Esta enzima es útil para generar cadenas de ADN, que contengan un inicio ATG, es por ello que es usada en muchos vectores de expresión, como es el caso de pET (Watson y col., 1982).
- *XhoI* enzima de restricción que reconoce la secuencia de ADN bicatenario 5'-C/TCGAG-3'. Estas endonucleasas catalizan la ruptura del ADN para dar fragmentos bicatenarios específicos con 5'-fosfatos terminales (Theriault y col., 1985).

3.8. Sistemas de expresión.

En la actualidad, existe un gran interés por la obtención proteínas puras y solubles, que además sean biológicamente activas para su uso en proyectos de investigación, con fines terapéuticos, de diagnóstico o industriales. La purificación de proteínas directamente de sus fuentes naturales, en la mayoría de los casos no cumple los requerimientos de cantidad, calidad, facilidad de aislamiento o rentabilidad económica del proceso.

A mediados de la década de los 70, se desarrolló la tecnología del ADN recombinante, y fue entonces cuando se marcó el comienzo de la era moderna de la biotecnología. Dado el alto conocimiento en genómica, proteómica y bioinformática, el número de proteínas producidas por vía recombinante se ha incrementado exponencialmente.

El objetivo esencial de los sistemas de expresión es la capacidad de producción de altos rendimientos y proporcionar una expresión estable. Las propiedades de estos sistemas están determinadas por la combinación de una serie de organismos específicos junto con el gen deseado. Es por ello, que se deben probar sistemas de expresión que se ajusten a las diferentes condiciones del o de los ensayos (Wegerer y col., 2008; González y Fillat, 2018).

Actualmente existe una amplia variedad de sistemas de expresión, para proteínas recombinantes. Estos se clasifican en 2 grandes grupos; sistemas eucariota y procariota. Dentro de estos últimos, *Escherichia coli* es el hospedador más ampliamente utilizado, tanto a nivel de laboratorio como a escala industrial. Su alta velocidad de crecimiento y rendimiento celular unido a la capacidad de expresar elevados niveles de proteínas heterólogas, en ocasiones hasta un 30% del contenido proteico total, hacen de esta enterobacteria una auténtica biofábrica (González y Fillat, 2018).

Los vectores de expresión en los sistemas procariotas, se obtiene comúnmente por manipulación genética. A fin de asegurar niveles óptimos de síntesis de la proteína de interés, existe una gran selección de promotores regulables, los cuales están clasificados en 2 grandes grupos, controlados positiva o negativamente (Wegerer y col., 2008). Sin embargo, no todos los genes pueden expresarse de manera eficiente y generar un alto contenido de proteínas eucariotas funcionales en *E. coli*. Algunos de los factores que afectan la expresión son: 1) las características estructurales de la secuencia a expresar, 2) la estabilidad y eficiencia de la traducción del ARNm, 3) la facilidad en el plegamiento de ciertas proteínas y su degradación en la célula huésped y 4) toxicidad del uso de codones de la proteína en el huésped. Una de las reglas más importantes para una buena inducción es estandarizar las condiciones y el medio de

crecimiento celular, sin embargo, no existe ninguna regla empírica hasta la fecha que ayude a regularlo.

Los vectores de expresión se destacan por poseer diferentes elementos necesarios para la posterior replicación en la célula hospedadora y la transcripción del gen clonado. Estos vectores tienen diferentes zonas entre las que se encuentran el origen de replicación, promotor, sitio de unión al ribosoma, secuencia codificadora, terminadores de la transcripción, así como los marcadores de resistencia a antibióticos los que facilitaran la selección de los recombinantes y garantizará la estabilidad del plásmido (Figura 3). Como complemento a la estructura del vector, se puede encontrar un regulador que permitirá la unión directa o indirecta sobre el promotor, modulándose así la actividad y permitiendo condicionar los niveles de expresión según los intereses de síntesis de la proteína en cuestión.

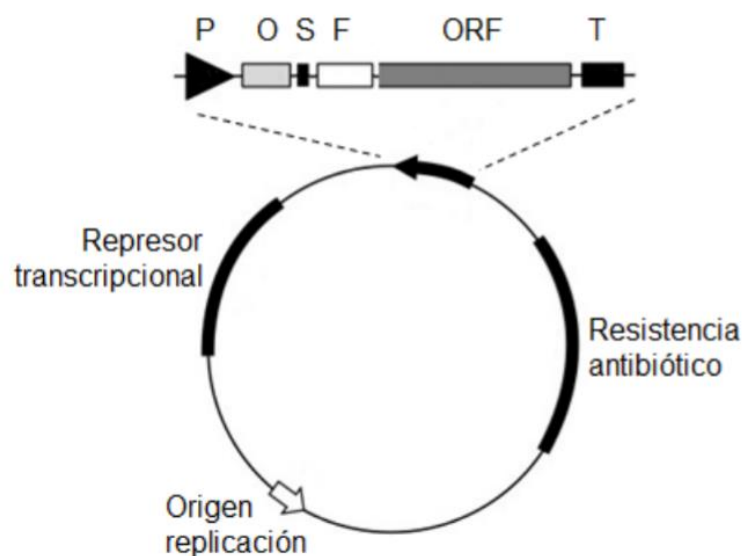


Figura 3: Elementos básicos que debe presentar un vector plasmídico para la expresión recombinante en *Escherichia coli*. P: promotor, O: operador, S: sitio de unión al ribosoma, F: secuencia de fusión, ORF: secuencia codificante del gen de interés, T: terminador (González y Fillat, 2018).

Para llevar a cabo la expresión de proteínas codificadas en un plásmido de expresión, será necesario el uso de una cepa hospedadora. Existen diferentes genotipos celulares y cada uno de ellos es específico para una función. Cabe destacar que las cepas de *E. coli* empleadas en la producción de proteínas recombinantes han sido manipuladas genéticamente para eliminar los genes implicados en los mecanismos de patogenicidad (lo cual asegura que son hospedadores seguros para la producción a gran escala). La célula hospedadora, deberá carecer de las principales proteasas naturales, mantener la estabilidad del vector de expresión y conferir los elementos genéticos necesarios para una expresión eficiente y regulable (González y Fillat 2018).

3.8.1. Cepas procariotas utilizadas en el sistema de expresión.

- *E. coli* BL21

Esta cepa se ha utilizado ampliamente para expresar una variedad de proteínas nativas y heterólogas. La cepa BL21 de *E. coli* es deficiente en proteasa Lon y OmpT y, por lo tanto, es adecuada para la sobreexpresión de proteínas (Yadava y col., 2005). Esta cepa se utiliza cuando se requiere grandes cantidades de proteína, ya que el rendimiento es mayor partiendo de un mismo volumen (Vanatalu y col., 1993).

- *E. coli* BL21 (DE3)

La cepa *E. coli* BL21 (DE3), deriva de la cepa BL21, y por ello posee todas sus características sumadas a las que añade secuencias del profago λ DE3. En su genoma está el gen de la ARN polimerasa de T7 bajo el control del promotor lacUV5 cuya expresión se induce por IPTG. Estas cepas se usan para la expresión recombinante de genes bajo el control de promotores T7, T7-lac y promotores reconocidos por la ARN polimerasa de *E. coli*. Además, presenta características procedentes de su progenitor como la deficiencia en proteasa Lon y OmpT.

Para poder entender bien el funcionamiento de este sistema de expresión, es necesario conocer el sistema de expresión genética basado en la ARN polimerasa del bacteriófago desarrollado por Studier y Moffat en 1986. La ARN polimerasa de T7 es altamente selectiva para sus propios promotores, algo que no ocurre naturalmente en *E. coli*. Se observó como una cantidad relativamente pequeña de ARN polimerasa de T7 a partir de una copia del gen de la polimerasa de T7 era suficiente para dirigir la transcripción en alto nivel de un promotor de T7 en un plásmido multicopia. Tal transcripción puede ocurrir varias veces alrededor del plásmido sin terminar, y puede ser tan activa que la transcripción por la ARN polimerasa de *E. coli* se ve disminuida considerablemente. Se observó como la velocidad de síntesis de proteínas a partir de ARN mensajero específico, producido por la ARN polimerasa de T7 depende principalmente de la eficiencia de su traducción.

Por tanto, se puede decir que esta cepa es una cepa de gran expresión en la que el ARN mensajero se traduce de forma eficiente, y puede acumular las proteínas de interés a más del 50% de la proteína celular, en poco tiempo.

- *E. coli* DH5- α

La cepa DH5- α de *E. coli* proviene de la cepa Hoffman-Berling 1100 (Meselson y Yuan, 1968). Se caracteriza por permitir la expresión de grandes plásmidos, es auxótrofa para la arginina y la tiamina y no puede metabolizar la lactosa. Su fenotipo es relajado, es decir, permite la síntesis de ARN, aunque no sintetizan proteínas. Es deficiente para la reparación del ADN y sensible a los rayos UV, lo cual disminuye la aparición de recombinaciones no deseadas de ADN clonado. Además, es resistente al ácido nalidíxico y deficiente en endonucleasa A, lo cual evita la digestión no específica de dicha endonucleasa en el ADN, obteniéndose preparaciones más limpias y mejores resultados en sus posteriores aplicaciones.

- *E. coli* JM109

La cepa *E. coli* JM109 se describe como carente de la endonucleasa I, con supresión de codones de stop (UAG) por inserción de glutamina (necesario para el crecimiento de algunos fagos). Requiere tiamina y como deriva de la cepa DH5- α , de fenotipo relajado, es sensible a la luz ultravioleta y carece de los sistemas de reparación del ADN (evitando la formación de recombinaciones indeseadas del ADN clonado). Tiene una mutación en la ADN girasa que le confiere resistencia al ácido nalidíxico. Puede utilizarse con el fago M14 para clonación, así como para la selección de colonias blanco/azul (α -complementación). Finalmente, al ser deficiente en la expresión de proteasa Lon, se reduce la proteólisis de las proteínas expresadas haciéndola una buena cepa para la expresión de proteínas recombinantes (González y Fillat, 2018). Por tanto y gracias a todas estas características, las cepas de *E. coli* JM109, se han convertido en una excelente cepa para la expresión en proteínas recombinantes (Hanaham, 1985; Grant y col., 1990).

- *E. coli* C43 (DE3)

Dicha cepa fue descrita por Miroux y Walker (1996). La cepa *E. coli* C43 (DE3), al igual que la C41 (DE3), son derivadas de la *E. coli* BL21 (DE3), la cual contiene el profago λ DE3, que se encarga de transformar el gen de la ARN polimerasa T7 bajo el control del promotor lac UV5, inducible por IPTG.

E. coli C43 (DE3) es muy utilizada para la expresión de proteínas tóxicas y proteínas de membrana de diversos organismos, tanto eucariotas como procariontes, incluyéndose insectos, mamíferos y plantas (González y Fillat, 2018).

Se ha determinado en varios estudios que para la expresión en proteínas heterólogas las cepas de C43 y C41 son claramente superiores a su cepa parental BL21 (DE3). Esto se debe a la mayor estabilidad de los plásmidos en C43 (DE3), presentando estas cepas una transformación del 100% respecto al 62% observado en BL21 y un 81% de expresión frente a un 54% de BL21(DE3). Es por ello, que gracias al estudio de Dumon-Seignovert (Dumon-Seignovert y col., 2004), se ha dejado clara la superioridad de esta cepa para la expresión de proteínas tanto de membrana como tóxicas, con respecto a su predecesor BL21 (DE3).

3.8.2. Vectores de expresión.

Los elementos que forman parte del vector de expresión incidirán directamente en la eficiencia de la expresión del gen clonado. Aunque existe una gran variedad de vectores de expresión comerciales, en muchas ocasiones resulta de interés la creación de un nuevo vector o la modificación de vectores disponibles en el laboratorio, a fin de alcanzar un determinado resultado (González y Fillat, 2018)

3.8.2.1. Vector pET.

Los vectores pET fueron construidos originalmente por Studier (Studier y Moffat, 1986), aunque hoy en día existen nuevos derivados, desarrollados por la empresa Novagen, con características mejoradas que permiten un subclonaje, detección y purificación de las proteínas diana clonadas.

Los genes diana se clonan en plásmidos pET bajo el control de fuertes señales de transcripción y (opcionalmente) de traducción del bacteriófago T7. La expresión de éste es inducida por una ARN polimerasa de T7 en la célula huésped. Esta ARN polimerasa de T7 es tan selectiva y activa, que casi todos los recursos de la célula se utilizan en expresión génica diana del fragmento deseado.

Unos de los beneficios más importantes de este sistema es la capacidad que tiene para mantener los genes diana silenciados transcripcionalmente cuando se encuentra en estado no inducido, evitando así la síntesis de escape.

Los plásmidos se transfieren a hospedadores de expresión que contienen una copia del gen de la ARN polimerasa de T7 bajo control *lacUV5* (derivados (DE3)), y la expresión es inducida por la adición de IPTG (Bok y col., 1998).

Se pueden utilizar muchas estrategias para subclonar una región de ADN que codifica una proteína en un vector pET para expresión. Una de las más convenientes es utilizar los sitios de restricción únicos que se encuentran en la región de clonación múltiple (policlonado) de los vectores pET para clonar insertos (fragmento de ADN de interés) en una orientación definida utilizando dos sitios diferentes.

3.8.2.2. Vector pJOE4036.1.

El plásmido pJOE4036.1 se trata de un derivado del plásmido pJOE3065, conteniendo ahora el gen de *LacZ α* más el promotor *lac* (Altenbuchner, comunicación personal) utilizado para la expresión de la azoreductasas recombinantes en *E. coli*. Los genes clonados en este vector se inducen con el carbohidrato L-ramnosa.

Podemos decir que este plásmido ha sido utilizado para la expresión de proteínas por parte de nuestro grupo, como podría ser la enzima hidantoín-racemasa II de *Agrobacterium tumefaciens* C58), así como por otros grupos, como una hipotética proteína transmembrana que se asocia a granulos de polihidroxibutirato en *Ralstonia eutropha* H16 (Martínez-Rodríguez y col., 2004; Pfeiffer y col., 2011).

4. OBJETIVOS

Los objetivos de este trabajo fin de grado son:

1. Búsqueda de genes a partir de organismos termófilos utilizando como base la secuencia de ADN del genoma de *Geobacillus* sp. WSUCF1.
2. Amplificación de los posibles genes M1 y M2 de *Geobacillus kaustophilus* CECT4264 y *Geobacillus stearothermophilus* CECT 43 y 49.
3. Clonación de los genes que codifican para las endoglucanasas M1 y M2 en el vector de subclonado pTOPO y posterior clonación en los vectores de expresión BBM35 (pJOE4036.1) y BBM48 (pET21b+), respectivamente.
4. Análisis de la secuencia de los nuevos genes aislados y clonados.
5. Expresión de las proteínas recombinantes Endoglucanasas M1 y M2 en diferentes cepas de *E. coli*.
6. Determinación de la actividad de las Endoglucanasas recombinantes M1 y M2.

5. MÉTODOS EXPERIMENTALES

5.1. Extracción de ADN *G. kaustophilus* CECT4264 y *G. stearothermophilus* CECT43 y 49 a través del método de ebullición.

Para la extracción de ADN por el método de ebullición se preparan tubos de 1,5 mL con 200 μ L de H₂O Ultrapura (Servicios Técnicos, Universidad de Almería). Se toma una colonia aislada de la placa de cada cepa y se disuelve por pipeteo en el agua en cada tubo. A continuación, se mantiene 5 minutos en un baño de agua a 100 °C (Precistern, J.P Selecta Precistern Laboquimia, La Rioja, España), y se dejan enfriar durante 5 minutos en el congelador (Liebherr Premium no Frost, Suiza). Seguidamente se centrifugan a 13.000xg (Sigma 1-14, Alemania) durante 5 minutos, pasando el sobrenadante a un tubo nuevo para su cuantificación en Nano-Drop (Nanodrop 2000S Spectrophotometer, Thermo Scientific™, Madrid) y posterior uso como ADN molde en la reacción de amplificación (Reacción en Cadena de la Polimerasa, PCR).

5.2. Preparación de oligos.

Cada uno de los genes se debe amplificar con 2 oligonucleótidos que actuarán como cebadores en la reacción de PCR.

Los oligonucleótidos se adquieren de la casa comercial Metabion (Metabion Internacional AG, Alemania) siendo suministrados como polvo liofilizado. Para su uso se realizan dos diluciones, la primera de almacenaje y la segunda de trabajo. En ambos casos se utiliza agua ultrapura autoclavada (MiliQ). La concentración de almacenaje es 100 μ M y la de trabajo 20 μ M. Para la dilución de almacenaje se emplean los siguientes volúmenes de agua:

1. EngluM1Gka 5; 262 μ L
2. EngluM1Gka 3; 281 μ L
3. EngluM2Gka 5; 233 μ L
4. EngluM2Gka 3; 254 μ L
5. EngluM2Gste 5; 200 μ L
6. EngluM2Gste 3; 268 μ L

Para la de trabajo se diluye 5 veces hasta un volumen final de 50 μ L.

5.3. Amplificación de genes mediante PCR.

Las reacciones de PCR se llevan a un volumen final de 12,5 μ L. Para ello se mezclan 6,25 μ L del Kit Clone Amp HiFi PCR Premix (Takara Bio Inc, Shiga, Japan) que contiene la ADN polimerasa y dNTPs en el tampón optimizado, además se añade 0,25 μ L de cada oligonucleótido (concentración final 0,4 μ M), 2 μ L de ADN molde, el cual se añade en 2 concentraciones diferentes: directo (sin diluir) e indirecto (diluido a la mitad). Por último, se completa hasta el volumen final con 3,25 μ L de agua. Antes de realizar la PCR, se debe comprobar la concentración de ADN molde en un espectrofotómetro de gota (Nanodrop 2000S spectrophotometer, Thermo Scientific™, Madrid) ya que en el kit TAKARA exige que la concentración de ADN sea menor de 100 ng.

La reacción se lleva a cabo utilizando un termociclador 2720 (Applied Biosystems, ThermoScientific, Madrid), siguiendo el protocolo TAKARA (Takara Bio Inc): un ciclo inicial de desnaturalización de 1 minuto a 98 °C, 35 ciclos de amplificación que incluyen una desnaturalización de 10 segundos a 98 °C, seguida por una fase de hibridación de cinco segundos a 55 °C, y 20 segundos de elongación a 72 °C. Pasados los 35 ciclos se mantiene la muestra durante 2 minutos a 72 °C, y una vez han acabado todos los ciclos se puede guardar a 4 °C.

5.4. Electroforesis en gel de agarosa.

Se prepara el gel de agarosa al 1%, para ello es necesario disolver 0,5 g de agarosa (Agarosa D-1 Low EEO, Conda Madrid, España) en 50 mL de tampón TAE 1X (Tris-Base 0,04 M, ácido acético glacial 0,02M, EDTA 0,001 M, pH 8,5). Esta disolución se debe calentar en el microondas (Zanussi Jetwave, Italia), ya que la agarosa no es soluble a temperatura ambiente y se disuelve por encima de 65 °C. El matraz debe estar tapado con un vaso de precipitado para evitar que evapore y a su vez permitir el calentamiento de la mezcla casi hasta ebullición. Antes de añadirlo en el soporte del gel, se deja enfriar. Se le añaden 50 µL de SYBR green (Invitrogen, Thermo Scientific™, Estados Unidos), se agita suavemente y se vierte sobre el porta geles donde se deja enfriar y gelificar.

Una vez se ha solidificado el gel, se coloca en la cubeta de electroforesis (Bio-Rad, Estados Unidos) y seguidamente se cargan las muestras. Para poder colocar las muestras en el gel será necesario mezclar los productos de PCR (10 µL) con 4 µL de tampón de carga (1,6 mL glicerina, 0,02% azul de bromofenol y 2,4 mL tampón TAE 1X).

Se inicia la electroforesis conectándola a una fuente de alimentación (Bio-Rad, Power-Rad, Estados Unidos) a 100 V, y se deja avanzar las muestras en el campo eléctrico durante aproximadamente 20 minutos.

Transcurrido el tiempo, el gel se coloca en un transiluminador (Dar Reader Transiluminator, Clare Chemical Research, Estados Unidos), que nos permitirá visualizar el/los tamaño/s y calidad de las muestras de ADN.

5.5. Purificación de fragmentos de ADN a partir de gel.

La purificación a partir de gel es una técnica que se realiza para recuperar fragmentos de ADN del gel de agarosa después de la separación electroforética.

Para ello se debe extraer el fragmento de gel con una punta especial de pipeta, que recorta una determinada banda del gel y se pesa en un tubo de 1,5 mL.

Para la purificación del fragmento escindido del gel, se ha seguido el protocolo del kit GeneJET Kit (GeneJET Kit, Thermo Scientific™, Madrid, España). Por cada 100 mg de fragmento gel extraído, se añaden 100 µL de tampón de unión (Binding Buffer). A continuación, se coloca en un termo bloque (LabNet Accublock Mini, LabNet, Madrid, España), a una temperatura entre 50-60 °C durante 10 minutos, mezclando suavemente en intervalos de 3 min, para que el gel se funda. Pasado el tiempo la disolución cambia a color amarillo.

Una vez la disolución tiene color amarillo, se traspasa toda la mezcla a una columna suministrada en el kit, y se coloca en un tubo de 1,5 mL, donde se centrifuga (Sigma

1-14, Alemania) durante 1 min a 12.000xg, descartando el volumen eluido. A continuación, se añade 100 µL de tampón de unión y se vuelve a centrifugar otro minuto eliminando de nuevo el volumen eluido. Seguidamente se añaden 700 µL de tampón de lavado (Wash Buffer) y se vuelve a centrifugar, en este caso se hacen 2 centrifugaciones de 1 min, descartando el volumen eluido en cada ciclo. Por último, se lleva la columna a un tubo nuevo se añaden 50 µL de agua en el centro de la columna, se centrifuga y el eluido de la columna se cuantifica en un espectrofotómetro de gota (Nanodrop 2000S spectrophotometer, Thermo Scientific™, Madrid).

Se etiquetan los tubos correctamente, indicando la concentración de obtenida, y se guardan a -20 °C (Liebherr Premium no Frost, Suiza).

5.6. Preparación de células competentes.

A partir de una colonia de *Escherichia coli*, se prepara un cultivo de células en 50 mL de LB (Tripton 10 g/L, NaCl 10 g/L, extracto de levadura 5 g/L), junto con 250 µL de Solución I (MgCl₂ 1 M, MgSO₄ 1 M) y se dejan crecer en un incubador a 37 °C (EC240/PP/C Ingeniera Climas S.L. Barcelona, España) durante toda la noche. Al día siguiente se toma 1 mL del cultivo y se añade a 50 mL de LB conteniendo 50 µL de Solución I y se deja incubar a 37 °C, en agitación constante entre una hora y media y 2 horas, hasta que la densidad óptica a 600 nm oscile en torno a 0,3-0,4.

El cultivo se coloca en un baño de hielo durante 5-10 minutos. A partir de este momento es imprescindible que las células se mantengan siempre en frío. Seguidamente se transfiere a tubos de centrifugación que deberán estar previamente enfriados y esterilizados. Se centrifugan durante 7 minutos a 4 °C a 7000 rpm (Beckman J2-21, Estados Unidos). El rotor (Beckman JA-20, Estados Unidos) también debe ser previamente enfriado metiéndolo en la cámara fría. Tras la centrifugación, se descarta el sobrenadante y el sedimento se resuspende con 4 mL de Solución II (CaCl₂ 60 mM, Glicerol 15%, Pipes pH 7 10 mM) enfriada con anterioridad. Las células se incuban en hielo durante una hora y se vuelven a centrifugar durante 7 minutos a 7000 rpm, descartando nuevamente el sobrenadante. Las células sedimentadas se resuspenden con ayuda de 4 mL de la Solución II y se incuban en hielo durante 10 minutos. Una vez transcurrido el tiempo, las células se transfieren a tubos de 1,5 mL nuevos, previamente enfriados en volúmenes de 100 µL o 200 µL. Se etiquetan correctamente y se guardan a -80 °C (Revco PLUS -86 °C Freezers Thermo Fisher Scientific, Estados Unidos), hasta su uso.

5.7. Sub-Clonaje en pTOPO.

Para poder llevar a cabo el sub-clonaje de fragmentos amplificados, se prepara la reacción de clonado con TOPO seguido de una transformación de células competentes.

Para la reacción de TOPO se debe preparar una mezcla de reacción a un volumen final de 6 µL que contiene: 3 µL ADN producto de PCR extraído de gel, 1 µL de agua, 1 µL de solución salina (TOPO PCR Cloning, ThermoFisher Scientific, Madrid) y 1 µL del vector de sub-clonaje pTOPO (TOPO PCR Cloning, ThermoFisher Scientific, Madrid). La mezcla se deja incubar durante 5 minutos a temperatura ambiente, seguido de 1 minuto en hielo.

A continuación, se transforman células competentes para lo que se necesita un vial que contiene aproximadamente 50-100 μL de células competentes (ver apartado 5.6) y que deberán permanecer en hielo. Se añaden 2 μL de la mezcla de reacción pTOPO a las células competentes y se deja 10 min en hielo, mezclando suavemente cada 4 minutos. Seguidamente se aplica un choque térmico a 42 °C durante 30 segundos y se vuelven a incubar en hielo otros 2 min.

Posteriormente se añaden 250 μL de medio SOC (0,5% Extracto de levadura, 2,0% Triptona, 10 mM NaCl, 2,5 mM KCl, 10 mM MgCl_2) y se coloca en el agitador de tubos (CMV-rom Fröbel Labortechnik GmbH, Alemania) durante 55 minutos a 37 °C. Transcurrido el tiempo, se vierten 10 y 100 μL en dos placas de medio LB sólido (Triptona 10 g/L, NaCl 10 g/L, Extracto de Levadura 5 g/L, Agar Bacteriológico 5 gr/L, con Kanamicina 50 mg/mL) de cada una de las transformaciones. Para su distribución homogénea en toda la placa se utiliza un asa de Digrafsky (VWR, Llinars del Vallès, Barcelona, España). Las transformaciones se incuban toda la noche a 37 °C (EC240/PP/C Ingeniera Climas S.L. Barcelona, España).

Todo este proceso se realiza bajo esterilidad utilizando una campana de flujo laminar (Laminar Flow Workstation Telstar, Madrid, España), previamente desinfectada. Al día siguiente se cuentan las colonias resultantes y se separan, de forma ordenada, en otra placa con ayuda de un asa de siembra. Se dejan incubar durante toda la noche a 37 °C en un incubador (EC240/PP/C Ingeniera Climas S.L. Barcelona, España) antes de su análisis.

5.8. Crecimiento de *E. coli* DH5 α con los plásmidos pBBM35 o pBBM48.

Para llevar a cabo la clonación de fragmentos es necesario realizar un cultivo de la bacteria que contenga el plásmido hospedador. Para ello se añaden 20 mL de medio de cultivo LB (Triptona 10 g/L, NaCl 10 g/L, Extracto de Levadura 5 g/L, 100 $\mu\text{g}/\text{mL}$ Ampicilina), en un matraz Erlenmeyer de 100 mL.

Con ayuda de un asa de siembra estéril se toma una colonia aislada de *E. coli* DH5 α , crecida en placas de LB sólido (Triptona 10 g/L, NaCl 10 g/L, Levadura 5 g/L, agar bacteriológico 5 gr/L y ampicilina 100 $\mu\text{g}/\text{mL}$), conteniendo alguno de los plásmidos de expresión pBBM35 (pJOE4036.1, comunicación personal) o pBBM48 (pET21+ Novagen, Merck, Darmstadt, Alemania). Se deja crecer en agitación constante a 250 rpm (Agitador GFL,3005, Burgwedel, Alemania), durante toda la noche a 37 °C y al día siguiente se lleva a cabo la extracción de plásmido siguiendo el protocolo del apartado 5.9.

5.9. Extracción de plásmidos.

Para la extracción de plásmido se utilizó el kit Macherey-Nagel (NucleoSpin Plasmid EasyPure, Alemania). A partir del cultivo crecido, obtenido en el apartado anterior, se precipitan 4 mL de dicho cultivo, centrifugando a 13.000xg (Sigma 1-14, Alemania) durante 1 min. Una vez se tiene el sedimento de los 4 mL, se resuspende con 150 μL del tampón A1 del kit con ayuda de agitación en vórtex (Heirdolph Reax 2000, Alemania). Una vez resuspendidas las células, se añaden 250 μL de tampón A2 y se mezcla por inversión 5 veces. A continuación, se incuba durante 2 minutos a temperatura ambiente, realizándose la lisis de las células. Una vez transcurrido los 2 minutos, se añaden 350 μL de tampón A3 y se mezclan por inversión de forma suave

hasta que la preparación se vuelva incolora, apareciendo un precipitado blanco que se centrifuga durante 3 minutos a 12.000 g (Sigma 1-14, Alemania). Se pasa el sobrenadante a una columna para volver a centrifugarlo descartando el filtrado. Se añaden 450 μL de tampón AQ y se centrifuga 2 veces, 1 minuto a 12.000xg. Por último, se añade 50 μL de agua destilada a la columna y se recoge el ADN, por centrifugación, en un tubo de 1,5 mL nuevo. Finalmente, se cuantifica en espectrofotómetro de gota (Nanodrop 2000S Spectrophotometer, ThermoScientific™, Madrid), la concentración de plásmido obtenida.

Los tubos con el plásmido se etiquetan correctamente, incluyendo concentración, fecha, nombre y se guardan a $-20\text{ }^{\circ}\text{C}$ (Liebherr Premium no Frost, Suiza).

5.10. Corte con enzimas de restricción.

El análisis de los fragmentos clonados en los plásmidos se realiza mediante el corte con enzimas de restricción. Según el plásmido utilizado, se utilizarán unas enzimas de restricción diferentes, pero siendo el protocolo de actuación el mismo. En el caso de la clonación en el plásmido pBBM35, las enzimas de corte para el clonado son *NdeI* y *BamHI*. En el caso de la clonación en el plásmido pBBM48, las enzimas de corte son *BamHI* y *XhoI*.

Se prepara la mezcla a un volumen final de 10 μL conteniendo ADN plasmídico 4 μL , enzimas de restricción 0,5 μL (Thermo Scientific™, Lituania), Tampón 10X 1 μL y 4,5 μL de agua. Las muestras se incuban (EC240/PP/C Ingeniera Climas S.L. Barcelona, España) a $37\text{ }^{\circ}\text{C}$ entre 4 – 12 horas. A continuación, se muestra una tabla con las enzimas de restricción y los tampones utilizados en cada caso (Tabla 2):

Tabla 2: Tampones utilizados para el corte con las enzimas de restricción.

Gen clonado	Corte 5'	Corte 3'	Tampón
Endoglucanasa M1	<i>BamHI</i>	<i>NdeI</i>	Tampón Tango
Endoglucanasa M2	<i>BamHI</i>	<i>XhoI</i>	Tampón H

En el caso de los plásmidos de expresión será necesario linealizarlos de igual modo cortando con las mismas enzimas de restricción que se hayan incluido para la amplificación de los correspondientes insertos. Los fragmentos se visualizan en una electroforesis en geles de agarosa. Una vez se comprueba que se ha producido el corte, se realiza una extracción de gel (ver apartado 5.5), de forma que finalmente se obtenga por un lado el fragmento de interés y por otro el plásmido abierto y se procederá a la unión mediante una ligasa. En el caso de la endoglucanasa M1, el plásmido de expresión es pBBM35 y en el caso de la endoglucanasa M2, el plásmido de expresión es pBBM48.

Este protocolo se utiliza tanto para analizar las clonaciones como para preparar los fragmentos y los plásmidos de expresión.

5.11. Unión de fragmentos de ADN.

Para poder llevar a cabo el ligado del plásmido de expresión y el inserto, es necesario utilizar la enzima llamada ligasa. Para ello se mezclan 4 μL ADN plasmídico abierto, 10 μL de inserto, 2 μL tampón, 1 μL de ligasa y 3 μL de agua.

Una vez están listas las muestras se dejan en el frigorífico 4 °C durante toda la noche, los tubos se deben colocar en gradillas no cubiertas.

Al día siguiente, se toman 5 μL del producto de ligado para transformar 100 μL de células competentes *E. coli* DH5- α , siguiendo el protocolo anteriormente detallado.

Una vez transformadas las células se separarán y analizarán mediante análisis de restricción, eligiendo la colonia que contiene plásmido con inserto para su extracción y secuenciación. La secuenciación se realiza en el Servicio de Secuenciación de la Universidad de Almería, llevándose a cabo por el personal técnico de dicho servicio.

5.12. Expresión de proteínas: crecimiento e inducción de las células transformadas

La obtención de células competentes de cada cepa y su transformación se realiza según los protocolos descritos anteriormente. Una vez transformadas y seleccionadas las colonias que contienen al plásmido de interés, se realiza un precultivo de 5 mL de medio de cultivo LB con ampicilina (100 $\mu\text{g}/\text{mL}$), en un matraz Erlenmeyer de 100 mL. Cada transformante se deja crecer 12 horas a 37 °C (EC240/PP/C Ingeniera Climas S.L. Barcelona, España), con agitación constante a 250 rpm (Agitador orbital, GFL 3005, Genesys SL Madrid, España). A continuación, se inoculan 0,5 mL del precultivo en un matraz de 250 mL conteniendo 50 mL de LB con ampicilina (100 $\mu\text{g}/\text{mL}$) y se deja crecer a 37 °C en agitación constante hasta alcanzar una absorbancia entre 0,4 – 0,6 medida a 600 nm en un espectrofotómetro Helios γ (Thermo-Electron Corporation, Thermo Scientific™, Madrid, España), lo que supone unas 4 horas.

Una vez alcanzada la densidad óptica deseada, los cultivos de diferentes cepas de *E. coli* transformadas con pJMC104, pJMC106 y pJMC120 se añadió ramnosa a una concentración final de 0,2% mientras que para *E. coli transformadas con pJMC108* y pJMC122 se indujo con 32 μL de IPTG (isopropil- β -D-1-tiogalactopiranosido) siendo la concentración final de 0,5 mM. Todos los cultivos se dejaron induciendo en agitación constante a 32 °C, durante 12 horas, para conseguir la expresión del gen. Antes de comenzar la inducción se tomó 1 mL de cada cultivo sin inducir para su posterior análisis mediante electroforesis en condiciones desnaturalizantes. Esta alícuota se centrifugó a 13.000xg eliminando el sobrenadante y el sedimento se conservó a -20 °C hasta su utilización.

5.13. Ruptura celular.

Las células inducidas se centrifugan a 5000 r.p.m. (rotor JA 14, Beckman J2-21, Estados Unidos) durante 25 min a 4 °C. Se elimina el sobrenadante por decantación y las células inducidas sedimentadas se guardan en un bote cerrado -20 °C.

Las células sedimentadas se resuspenden con 10 mL de tampón de Lavado TALÓN (Tris-HCl 50 mM, NaCl 300 mM pH 7) en agitador magnético y ayudado de una micropipeta. A continuación, se colocan en hielo y se rompen con ayuda de un

sonicador (Ultraschall Prozessor UP 200S, Dr Hielscher, Alemania). Se realizan 6 ciclos de sonicación de 1 minuto a una potencia del 60% y un pulso de 0,5. Entre ciclo y ciclo se descansó un 1 minuto manteniendo las células en hielo durante todo el proceso.

Del extracto celular se toma una alícuota 500 μ L y se centrifuga a 10.000xg (Sigma 1-14, Alemania) durante 10 minutos, guardando los restos celulares (sedimento) y sobrenadante de forma independiente a -20 °C. El resto de extracto celular se centrifugó a 10.000xg, guardando el sobrenadante para su análisis y descartando los restos celulares sedimentados.

5.14. Electroforesis de proteínas mediante SDS-PAGE.

La electroforesis permite visualizar y determinar el peso molecular de una proteína en función de la movilidad en un campo eléctrico y se puede realizar en condiciones nativas o desnaturizantes. En nuestro caso se realizó en condiciones desnaturizantes, es decir, en presencia de SDS (dodecilsulfato sódico), lo que permite separar y caracterizar los monómeros de las proteínas.

Se limpiaron los cristales, entre los que se formarán los geles, con abundante agua, etanol y acetona para eliminar restos de grasa. A continuación, se montaron en el sistema de electroforesis, dejando un espacio entre ellos (espaciadores que definen el grosor del gel).

El gel separador se realiza al 12% acrilamida/bisacrilamida, mezclando los siguientes compuestos: 10,7 mL de agua, 7,5 mL de acrilamida 40% (Amresco), 4 mL de metilen-bisacrilamida 2% (Amresco), 7,5 mL de Tris-HCl 1,5 M, pH 8,8 (Amresco), 0,3 mL SDS 10% (Serva), 7,5 μ L de TEMED (Amresco) y 505 μ L de persulfato amónico al 22% (AppliChem).

Se vierte la disolución entre los 2 cristales hasta una altura de 4 cm, aproximadamente por debajo del borde superior (si se comprueba con el peine puesto, la distancia debe ser 1 cm por debajo del peine). Debe verterse cuidadosamente, de forma que no se formen burbujas, además se debe hacer de forma rápida para evitar la polimerización en el vaso. Se añaden 0,2 mL de isopropanol para definir el borde del gel y eliminar ondulaciones y posibles burbujas.

A continuación, se preparan 10 mL del gel apilador al 4% acrilamida/bisacrilamida, con la siguiente composición: 2,8 mL de agua, 0,6 mL de acrilamida 40% (Amresco), 0,32 mL de metilen-bisacrilamida 2% (Amresco), 1 mL de Tris-HCl 1,5 M, pH 8,8 (Amresco), 1 mL de Tris-HCl 0,5 M, pH 6,8 (Amresco), 0,03 mL SDS 10% (Serva), 2 μ L de TEMED (Amresco) y 100 μ L de persulfato amónico al 60% (AppliChem).

Cuando el gel separador polimeriza completamente, se elimina el isopropanol. A continuación, se vierte el gel apilador, hasta cubrir todo el espacio entre los cristales. Inmediatamente después se coloca el peine para definir los pocillos, donde posteriormente se colocarán las muestras, con cuidado de que no quede ninguna burbuja.

Una vez ha polimerizado por completo, se retira el peine cuidadosamente y se rellenan los pocillos con tampón Tris pH 8,8 y posteriormente se retira con ayuda de una pipeta. Esta acción se repite 2 veces con el objetivo de limpiar y eliminar los restos de SDS que hayan podido quedar en los pocillos.

Para preparar las muestras de electroforesis, se deben desnaturizar, para ello se mezclan con 8 μ L de tampón de carga que está compuesto por: Tris /HCl 0,125 M, pH

6,5 (Amresco), SDS 4% (Serva), Glicerol 20% (Fluka), Azul de bromofenol 0,1% (Serva). A continuación, se añaden 2 μL de β -mercaptoetanol (Fluka) y las muestras se calientan en un baño a 80 °C durante 3 min y por último se enfrían en hielo.

Una vez preparadas las muestras se colocan en los diferentes pocillos formados. Finalmente se coloca los geles en la cubeta de electroforesis y se cubrieron con 300 mL de Tampón de Electrodo (Tris-HCl 0,025 M, pH 8,3 (Roche), Glicocola 0,192 M (AppliChem) y SDS 0,1% (Serva)). Se conectan los electrodos a la fuente de alimentación, asegurando que los polos coincidan correctamente y se aplica un voltaje de 100 V. Al terminar el proceso, se apaga la fuente, se sacan los vidrios de la cubeta y se extrae el gel. Se lavan para eliminar el exceso de SDS y se retiran los cristales con mucho cuidado, manteniendo el gel hidratado en todo momento.

Para la visualización de las proteínas, los geles se sumergen en una disolución de tinte con la siguiente composición: 0,1% Reactivo azul de Comassie BBRC 250 (Sigma), 40% Metanol (Panreac), 10% Ácido Acético (Merk), 50% Agua destilada.

En esta disolución se dejará durante 24 horas, hasta que el gel se haya teñido de color azul completamente. Transcurrido este tiempo, la disolución se cambia a una disolución de desteñido: 40% Metanol, 10% Ácido Acético, 50% Agua destilada.

Se mantiene la disolución de desteñido, reemplazándola si es necesario, hasta que todo el gel quede transparente, exceptuando las proteínas, cuya unión con el azul de Comassie es irreversible, apareciendo bandas horizontales que corresponden a las bandas correspondientes con los monómeros de las proteínas.

5.15. Estudio de la actividad Endoglucanasa.

La puesta a punto del ensayo por el método del ácido dinitrosalicílico (DNSA se realiza con la enzima comercial celulasa de *Penicillium funiculosum* (SigmaAldrich, Estados Unidos) midiendo su actividad del tipo CMCasa según las condiciones estándares de la casa comercial.

Para confirmar la actividad enzimática de los sobrenadantes provenientes de los cultivos inducidos primeramente se realizan varios blancos. El blanco de enzima, que contiene 90 μL de tampón acetato sodico (50 mM, pH 5) junto con 10 μL del extracto enzimático. El blanco de sustrato que contiene 50 μL de CMC al 1% (disuelta en tampón acetato sodico 50 mM, pH 5).

El ensayo estándar de actividad CMCasa de los sobrenadantes se realiza en 100 μL de volumen final (50 μL de CMC al 1%, 40 μL de tampón acético (50 mM, pH 5) y 10 μL del extracto enzimático). La reacción enzimática se incuba por 30 minutos a 40 °C. Finalizado el tiempo de incubación la reacción se detiene por la adición de 100 μL del reactivo 3,5-ácido dinitrosalicílico, se calienta durante 5 minutos a 100 °C. Finalmente se añadieron 660 μL de H_2O para estabilizar el color y se midió la absorbancia a 540 nm.

Para cuantificar la actividad de la proteína es necesario realizar una recta patrón, para la que se va a utilizar un stock de xilosa de concentración inicial 10 $\mu\text{g}/\mu\text{L}$. Se prepararon concentraciones de xilosa entre 0-250 μg , en un volumen final de 100 μL . Una vez elaboradas las distintas disoluciones se añade 100 μL del reactivo 3,5-ácido dinitrosalicílico, se calienta durante 5 minutos a 100 °C. Finalmente se añadieron 660 μL de H_2O para estabilizar el color y se midió la absorbancia a 540 nm. De la representación gráfica de absorbancia frente a concentración se obtiene el valor del coeficiente de extinción molar para este producto de reacción.

6. RESULTADOS Y DISCUSIÓN

6.1. Búsqueda de secuencias de posibles Endoglucanas.

Tal y como se ha indicado en el apartado de Objetivos, en este trabajo se han buscado Endoglucanasas de organismos termófilos, en concreto en *Geobacillus stearothermophilus* y *Geobacillus kaustophilus*. El microorganismo *Geobacillus* sp. WSUCF1, aislado en 2010, ha mostrado capacidad para degradar celulosa (Rastogi y Bhalla, 2010; Bhalla y col., 2013). Esta capacidad se le ha atribuido por poseer las enzimas Endoglucanasas M1 y M2 (Dass y col., 2018). En la búsqueda de estas enzimas en organismos termófilos y tras la secuenciación del genoma completo de *Geobacillus* sp. WSUCF1 (Bhalla y col., 2013), se han empleado las secuencias de M1, con número de acceso ATCO01000152.1 y M2, con número de acceso ATCO01000242.1 de dicho genoma como secuencias modelo (Figuras 1 y 2 del Anexo I).

A partir de las secuencias M1 y M2 se realizó un análisis BLAST (Basic Local Alignment Search Tool) (Altschul y col., 1990; Boratyn y col., 2019), que consiste en la búsqueda de secuencias similares depositadas en el GenBank (Altschul y col., 1997). De entre las encontradas se seleccionaron los genes presentes en los organismos *Geobacillus kaustophilus* y *Geobacillus stearothermophilus*, por tener dichas especies microbianas en el nuestro laboratorio (Figura 4).

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident
<input type="checkbox"/>	Geobacillus sp. GHH01 complete genome	2161	2161	100%	0.0	100.00%
<input type="checkbox"/>	[Bacillus] caldolyticus strain NEB414 chromosome complete genome	2156	2156	100%	0.0	99.91%
<input type="checkbox"/>	Geobacillus lituanicus strain N-3 complete genome	2134	2134	100%	0.0	99.57%
<input type="checkbox"/>	Geobacillus stearothermophilus strain B5 chromosome complete genome	2128	2128	100%	0.0	99.49%
<input type="checkbox"/>	Geobacillus stearothermophilus strain DSM 458 complete genome	2111	2111	100%	0.0	99.23%
<input type="checkbox"/>	Geobacillus sp. JS12 complete genome	2106	2106	100%	0.0	99.15%
<input type="checkbox"/>	Geobacillus sp. Y412MC52 complete genome	2050	2050	100%	0.0	98.29%
<input type="checkbox"/>	Geobacillus sp. Y412MC61 complete genome	2050	2050	100%	0.0	98.29%
<input type="checkbox"/>	Geobacillus sp. C56-T3 complete genome	2045	2045	100%	0.0	98.21%
<input type="checkbox"/>	Geobacillus stearothermophilus 10 complete genome	2039	2039	100%	0.0	98.12%
<input checked="" type="checkbox"/>	Geobacillus kaustophilus NBRC 102445 chromosome complete genome	2034	2034	100%	0.0	98.03%
<input type="checkbox"/>	Geobacillus thermoleovorans strain FJAT-2391 complete genome	2034	2034	100%	0.0	98.03%
<input type="checkbox"/>	Geobacillus thermoleovorans strain SGAir0734 chromosome	2028	2028	100%	0.0	97.95%
<input type="checkbox"/>	Geobacillus thermoleovorans strain KCTC 3570 complete genome	2028	2028	100%	0.0	97.95%
<input type="checkbox"/>	Geobacillus thermoleovorans strain ARTRW1 chromosome complete genome	2023	2023	100%	0.0	97.86%
<input type="checkbox"/>	Geobacillus thermoleovorans CCB_US3_UF5 complete genome	2023	2023	100%	0.0	97.86%
<input checked="" type="checkbox"/>	Geobacillus kaustophilus HTA426 DNA complete genome	2023	2023	100%	0.0	97.86%

Figura 4: Secuencias que presentan una mayor similitud a la Endoglucanasa M1 de *Geobacillus* sp. WSUCF1 tras enfrentarla al banco de genes mediante la herramienta BLAST.

Tras enfrentar la secuencia de M1 mediante BLAST (Figura 4) al GenBank aparecieron dos secuencias de *G. kaustophilus* con alta homología, NRBC 102445 (98%) y HTA426 (97,86%). El gen de *Geobacillus* sp. tiene una secuencia de 1170 nucleótidos (nt) e hibridan 1147 nt de NRBC102445 y 1145 nt de HTA426. Estas secuencias están localizadas en el genoma de:

- *G. kaustophilus* NRBC102445 entre los nucleótidos 3344229 y 3345398. Se indica en el genoma como "M42 family peptidase".
- *G. kaustophilus* HTA426 entre los nucleótidos 2735400 y 2736569. Se indica en el genoma como "hypothetical conserved protein".

Cuando se compararon ambas secuencias encontramos que eran totalmente iguales (Figura 3 del Anexo I).

También se analizó la homología de secuencia de M1 con la encontrada en *G. stearothermophilus* B5, presentando un 99% de homología. Por tanto, si comparamos *G. kaustophilus* NRBC102445 y *G. stearothermophilus* B5 para la Endoglucanasa M1 (Figura 4 del Anexo I), se comprueba que tienen una alta homología, siendo las zonas de comienzo (Figura 5) y final (Figura 6) 100% homólogos, por lo que se usaron la misma pareja de cebadores para ambos genes.

```

EnGluM1GkaNRBC102445  gtggcgaatgattcattttgccaatcgtcagccttttcaagtataataaaaaagtggaca  60
EnGluM1GsteB5         gtggcgaatgattcattttgccaatcgtcagccttttcaagtataataaaaaagtggaca  60
*****

```

Figura 5: Secuencia del comienzo del fragmento M1 para *G. stearothermophilus* y *G. kaustophilus*.

```

EnGluM1GkaNRBC102445  aaagtgaacaactgacgtttgacgaataa  1170
EnGluM1GsteB5         aaagtgaacaactgacgtttgacgaataa  1170
*****

```

Figura 6: Secuencia de la parte final del fragmento M1 para *G. stearothermophilus* y *G. kaustophilus*.

Para el caso de Endoglucanasa M2 se realizaron los mismos pasos utilizando la herramienta BLAST, (Figura 7), en el que aparecieron las 3 secuencias anteriores con una alta homología, *G. kaustophilus* NRBC102445 (95,5%) y *G. kaustophilus* HTA426 (95,5%) y *G. stearothermophilus* B5 (86%).

<input checked="" type="checkbox"/>	Geobacillus sp. JS12. complete genome	1945	1945	100%	0.0	99.17%
<input checked="" type="checkbox"/>	[Bacillus] caldolyticus strain NEB414 chromosome. complete genome	1934	1934	100%	0.0	98.98%
<input checked="" type="checkbox"/>	Geobacillus stearothermophilus 10. complete genome	1810	1810	99%	0.0	96.94%
<input checked="" type="checkbox"/>	Geobacillus sp. Y412MC52. complete genome	1810	1810	99%	0.0	96.94%
<input checked="" type="checkbox"/>	Geobacillus sp. Y412MC61. complete genome	1810	1810	99%	0.0	96.94%
<input checked="" type="checkbox"/>	Geobacillus sp. C56-T3. complete genome	1788	1788	99%	0.0	96.57%
<input checked="" type="checkbox"/>	Geobacillus thermoleovorans strain FJAT-2391. complete genome	1735	1735	100%	0.0	95.65%
<input checked="" type="checkbox"/>	Geobacillus thermoleovorans strain ARTRW1 chromosome. complete genome	1724	1724	100%	0.0	95.46%
<input checked="" type="checkbox"/>	Geobacillus kaustophilus NBRC 102445 chromosome. complete genome	1724	1724	100%	0.0	95.46%
<input checked="" type="checkbox"/>	Geobacillus thermoleovorans strain SGAir0734 chromosome	1724	1724	100%	0.0	95.46%
<input checked="" type="checkbox"/>	Geobacillus thermoleovorans CCB_US3_UF5. complete genome	1724	1724	100%	0.0	95.46%
<input checked="" type="checkbox"/>	Geobacillus kaustophilus HTA426 DNA. complete genome	1724	1724	100%	0.0	95.46%
<input checked="" type="checkbox"/>	Geobacillus thermoleovorans strain KCTC 3570. complete genome	1718	1718	100%	0.0	95.37%
<input checked="" type="checkbox"/>	Geobacillus sp. GH01. complete genome	1718	1718	100%	0.0	95.37%
<input checked="" type="checkbox"/>	Geobacillus thermocatenulatus strain KCTC 3921. complete genome	1629	1629	100%	0.0	93.89%
<input checked="" type="checkbox"/>	Geobacillus stearothermophilus strain B5 chromosome. complete genome	1157	1157	99%	0.0	86.07%

Figura 7: Secuencias que presentan una mayor similitud a la Endoglucanasa M2 de *Geobacillus sp. WSUCF1* tras enfrentarla al banco de genes mediante la herramienta BLAST.

Estas secuencias están localizadas en el genoma de:

- *G. kaustophilus* NRBC102445 en la región del genoma desde los nucleótidos 3459220 hasta 3460299. En la leyenda consta como "M42 family peptidase"
- *G. kaustophilus* HTA426 en la región del genoma desde los nucleótidos 2848352 hasta 2849431. En la leyenda consta como hypothetical conserved protein".
- *G. stearothermophilus* B5 en la región del genoma desde los nucleótidos 2649394 hasta 2650470. En la leyenda consta como "M42 family peptidase".

De nuevo, cuando se comparan las 2 secuencias para la Endoglucanasa M2 de *Geobacillus kaustophilus* se establece que son totalmente iguales (Figura 5 Anexo I).

Al comparar las secuencias de *G. kaustophilus* NRBC102445 y *G. stearothermophilus* B5 para la secuencia correspondiente a una posible Endoglucanasa M2 (Figura 6 Anexo I) se comprueba que tienen cierta homología, pero las regiones de inicio (Figura 8) y final (Figura 9) no están conservadas. Por tal motivo, deberemos usar una pareja de cebadores distinta para cada uno de los genes.

```

EnGluM2GsteB5          atgaatcaagaacgctccaatttttcaacgttgaccgagctccgggggcaccggg 60
EnGluM2GkaNRBC102445  atgaataaggacacgctgcaattattccaacgttgaccgagcttccaggcgcaccgggt 60
EnGluM2GkaHTA426      atgaataaggacacgctgcaattattccaacgttgaccgagcttccaggcgcaccgggt 60
***** * ** ***** ** ** ***** ** ** ***** ** ** *****

```

Figura 8: Secuencia del comienzo del fragmento M2 para *G. stearothermophilus* B5 y *G. kaustophilus* NRBC102445 y HTA426.

```

EnGluM2GsteB5          gagctcgtcaaacggtgacgaccggcgcacggtgagcgcgatcgccaaaacagctga 1075
EnGluM2GkaNRBC102445  gagctcgtcaaacggtgacgaccggacaacggtcgagacgattcgccaaaacagctga 1077
EnGluM2GkaHTA426      gagctcgtcaaacggtgacgaccggacaacggtcgagacgattcgccaaaacagctga 1077
***** ***** * ***** ** ** ***** *****

```

Figura 9: Secuencia del final del fragmento M2 para *G. stearothermophilus* B5 y *G. kaustophilus* NRBC102445 y HTA426.

Una vez se han descrito y encontrado las secuencias de interés, fue necesario establecer las enzimas de restricción que no cortan al gen del interés. Para ello se utilizó la aplicación WebCutter (<http://heimanlab.com/cut2.html>).

Las enzimas que no cortan a los fragmentos equivalentes a M1 y M2 de *G. kaustophilus* NRBC102445, HTA426 y *G. stearothermophilus* B5 se muestran en las figuras 7 a 10 del Anexo I. A modo de ejemplo se indica, en la figura 10 (equivalente a la figura 9 del Anexo I), la correspondiente a la posible M2 de *G. kaustophilus* NRBC102445.

The following endonucleases were selected but don't cut this sequence:

AatI, AatII, AccI13I, AccI16I, Acc65I, AccB1I, AccB7I, AccIII, AclNI, AflIII, AgeI, AhdI, Alw44I, AlwNI, AocI, ApaLI, AscI, AseI, AsnI, Asp700I, Asp718I, AspEI, AspI, AtsI, AviII, AvrII, BalI, BamHI, BanI, BanIII, BbeI, BbrPI, BbsI, Bbv16II, BcgI, BfrI, BglII, BlnI, BlpI, BpiI, BpmI, Bpu1102I, Bpu14I, BpuAI, Bsa29I, BsaAI, BsaI, BscI, Bse21I, BseAI, BseCI, BsePI, BshNI, BsiI, BsiMI, BsiWI, Bsp106I, Bsp119I, Bsp13I, Bsp1407I, Bsp1720I, Bsp19I, BspDI, BspEI, BspHI, BspMI, BspTI, BspXI, BsrDI, BsrGI, BssHII, BssSI, BssTI, Bst1107I, Bst98I, BstBI, BstEII, BstI, BstPI, BstSFI, BstSNI, BstX2I, BstXI, BstYI, BstZI, Bsu15I, Bsu36I, CciNI, CelII, Cfr9I, ClaI, CpoI, Csp45I, CspI, CvnI, DraI, DraIII, DrdI, EagI, Eam1105I, EclHKI, EclXI, Eco105I, Eco130I, Eco147I, Eco255I, Eco31I, Eco32I, Eco52I, Eco57I, Eco64I, Eco72I, Eco81I, Eco91I, EcoNI, EcoO65I, EcoRI, EcoRV, EcoT14I, EcoT22I, EheI, ErhI, Esp1396I, FseI, FspI, GsuI, HindIII, HpaI, KasI, Kpn2I, KpnI, LspI, MfeI, MflI, MluI, MluNI, Mph1103I, MroI, MscI, MspCI, MunI, NarI, NcoI, NheI, NotI, NsiI, NspV, PacI, PaeR7I, Pfl23II, PflMI, PinAI, PmaCI, Pme55I, PmeI, PmlI, Ppu10I, PpuMI, PshBI, Psp5II, PspAI, PspALI, PspEI, PspLI, PstI, PstNHI, RcaI, RsrII, SapI, SbfI, ScaI, SexAI, SfcI, SfiI, Sfr274I, SfuI, SgfI, SgrAI, SmaI, SmiI, SnaBI, SpeI, SplI, SrfI, Sse8387I, SseBI, SspBI, SspI, StuI, StyI, SunI, SwaI, Tth111I, Van91I, Vha464I, VneI, VspI, XbaI, XcmI, XhoI, XhoII, XmaI, XmaIII, XmnI, Zsp2I

Figura 10: Enzimas que no cortan al fragmento de una posible M2 de *G. kaustophilus* NRBC102445 tras utilizar la aplicación WebCutter. Como puede verse al fragmento no lo cortan las enzimas BamHI y XhoI (este análisis de restricción se ha realizado a todas las secuencias y se encuentran en las Figuras 7 a 10 del Anexo I).

Una vez conocidas las enzimas de restricción que no cortan a los fragmentos a clonar, debemos saber qué enzimas cortan al plásmido receptor en el sitio de clonación. Para ello nos fijamos en los mapas de restricción de pBBM35 (pJOE4036.1) (Figura 11) y pBBM48 (pET21b+) (Figura 12).

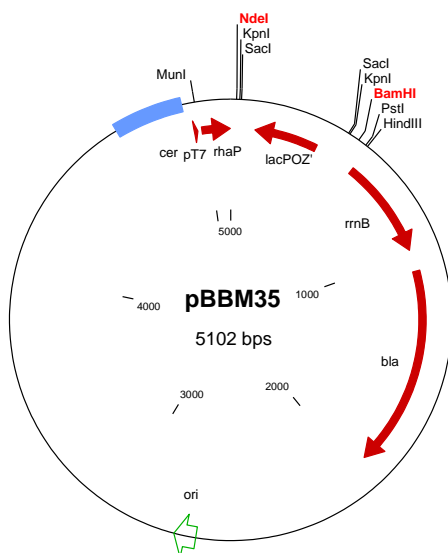


Figura 11: Mapa del plásmido pBBM35 donde aparecen los sitios de reconocimiento de enzimas de restricción para realizar el clonado. Aparecen en rojo los sitios de corte para NdeI y BamHI, que son las seleccionadas para clonar los fragmentos homólogos a M1. La cola de 6 histidinas está codificada entre los sitios BamHI y HindIII.

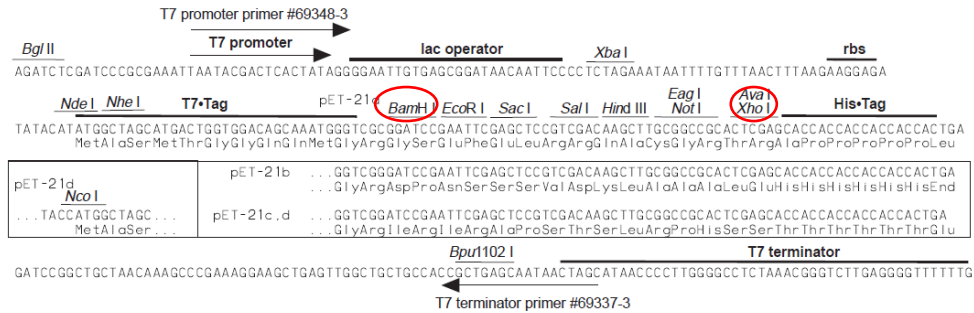


Figura 12: Esquema de la zona de policlonado del plásmido pBBM48. Los sitios de corte seleccionados para clonar los fragmentos homólogos a M2 son BamHI y XhoI. La cola de 6 histidinas está contenida en el plásmido después del sitio XhoI.

Para el caso de la Endoglucanasa M1 tanto para *G. kaustophilus* como *G. stearothermophilus*, el plásmido de clonado fue pBBM35, siendo las enzimas seleccionadas BamHI y de NdeI, ya que estas no cortan al gen (Figura 7 y 8 Anexo I), pero sí al plásmido (Figura 11).

Para el caso de la Endoglucanasa M2, tanto de *G. kaustophilus* como de *G. stearothermophilus* no ha sido posible clonarla en pBBM35 porque la enzima NdeI corta ambos fragmentos. Por tal motivo el clonado se realizó en pBBM48, donde se pueden utilizar las enzimas de restricción BamHI y XhoI que no cortan al fragmento, como puede verse (Figuras 9 y 10 Anexo I) y abren el plásmido para su posterior ligado (Figura 12).

6.2. Diseño de los oligonucleótidos para la amplificación de los posibles genes de M1 y M2 de *G. kaustophilus* CECT4264 y *G. stearothermophilus* CECT43 y 49.

Para la amplificación de los diferentes fragmentos de ADN correspondientes a los posibles genes M1 y M2 de los organismos termófilos que posee el laboratorio, *G. kaustophilus* CECT4264 y *G. stearothermophilus* CECT43 y 49, se han de diseñar los oligonucleótidos (oligos) para la Reacción en Cadena de la Polimerasa (PCR).

Como se ha visto, las secuencias de ADN en el caso de la Endoglucanasa M1 coinciden tanto para *G. stearothermophilus* como para *G. kaustophilus*, quedando el inicio de la secuencia como aparece en la figura 5. De esta forma en el oligo del extremo 5' se añadió el sitio de corte NdeI (CATATG) (Figura 7 Anexo I). En realidad, añadimos CAT y se sustituye el primer triplete GTG por ATG, quedando por tanto CATATG. Es por ello que la secuencia del oligo de M1 de 5' queda como:

Oligo *EnGluM1Gka5*:



Para el caso del extremo 3', la secuencia de la parte final del fragmento M1 para *G. stearothermophilus* y *G. kaustophilus* es la que aparece en la figura 6. Es necesario eliminar el codón de STOP (TAA), ya que la cola de histidina y el STOP están incluidos en el plásmido. Se añade el sitio de corte BamHI (GGATCC) a la secuencia

del oligo para que pueda ser clonado el fragmento en el plásmido. Una vez preparada la secuencia con el sitio de corte se hace la reversa complementaria, quedando el oligo de M1 de 3' como:

Final de secuencia:

5'-AAAGTGAAACAACACTGACGTTTGACGAA**TAA**-3'

Secuencia sin STOP y sitio *Bam*HI:

5'-AAAGTGAAACAACACTGACGTTTGACGAA**GGATCC**-3'

Oligo *EnGluM1Gka3*:

5'-**GGATCC**TTCGTCAAACGTCAGTTGTTTCACTTT-3'

Para el caso de Endoglucanasa M2, a la hora de diseñar el oligo 5', tanto de *G. kaustophilus* como de *G. stearothermophilus*, vemos que la secuencia se corta con la enzima *Nde*I (Figuras 9 y 10 Anexo I), ya que no aparece en listado de enzimas que NO cortan al fragmento, por lo que la clonación se hizo en el plásmido de expresión pBBM48 (pET21b+). La secuencia del oligo lleva el sitio de reconocimiento para la enzima *Bam*HI, que está en el plásmido y no corta al fragmento M2, además de la secuencia del principio del fragmento, empezando por el ATG. La comparación completa de ambas secuencias M2 para *G. kaustophilus* se encuentra en la figura 5 del Anexo I y también puede verse el principio de la secuencia en la figura 8.

De esta forma el oligo para 5' para M2 fue:
Oligo *EnGluM2Gka5*:

5'-**GGATCC**GTGACAGGAAAGAGGAATGGAATAAGGACA-3'

Para el extremo 3', se actúa de la misma forma que para el caso de la Endoglucanasa M1, en primer lugar, se selecciona el final del gen (Figura 9). En segundo lugar, se elimina de la secuencia el codón de STOP (TGA), se añade el sitio de corte *Xho*I (CTCGAG), ya que no corta al fragmento M2 de *G. kaustophilus* (Figura 9 Anexo I) y el sitio de corte es el último inmediatamente anterior a la cola de 6 histidinas (Figura 12). El oligo de 3' se obtiene realizando la reversa complementaria, quedando la secuencia como:

Final de la secuencia:

5'-GTTGAGGCGATCCGCCAAAACAGC**TGA**-3'

Secuencia sin STOP y sitio *Bam*HI:

5'-GTTGAGGCGATCCGCCAAAACAGC**CTCGAG**-3'

Oligo *EnGluM2Gka3*:

5'-**CTCGAG**GCTGTTTTGGCGAATCGTCTCGAC-3'

Para el diseño los oligos de PCR para el fragmento de M2 de *Geobacillus stearothermophilus* B5, seguimos el mismo protocolo en el diseño de los dos oligos para M1, teniendo en cuenta la secuencia del fragmento (Figura 10 Anexo I) y que el plásmido utilizado es pBBM48 (Figura 12), utilizando las enzimas *Bam*HI y *Xho*I.

Siendo la secuencia de los dos oligos:

Oligo **EnGluM2Gste5**:

5'-GGATCCGTGACAGGAAAGAGGAATGAATGAAGAAA-3'

Mientras que para el oligo 3':

Final de secuencia:

5'-GTCGAGACGATTCGCCAAAACAGCTGA-3'

Secuencia sin STOP y sitio *Xho*I5'

5'-GTCGAGACGATTCGCCAAAACAGCCTCGAG-3'

Oligo **EnGluM2Gka3**:

5'-CTCGAGGCTGTTTTGGCGAATCGTCTCGAC-3'

6.3. Extracción de ADN de *G. kaustophilus* CECT4264 y *G. stearothermophilus* CECT43 y 49.

Las muestras de ADN genómico de *G. kaustophilus* CECT4264 y *G. stearothermophilus* CECT43 y 49 se prepararon tal y como se describe en el apartado 5.1. de Métodos Experimentales. Se obtuvieron las concentraciones de 25,1 ng/μL para *G. kaustophilus* CECT4264 y 18,9 y 44,5 ng/μL para *G. stearothermophilus* CECT43 y 49, respectivamente.

6.4. Amplificación mediante PCR de los genes M1 y M2.

Una vez preparados los oligos, tanto para M1 como para M2 y extraído el ADN de *Geobacillus kaustophilus* CECT4264 y *Geobacillus stearothermophilus* CECT43 y 49 a través del método por ebullición, se pasó a realizar la Reacción en Cadena de la Polimerasa (PCR), donde se producen 2ⁿ copias del fragmento de interés a partir de una sola copia, protocolo descrito en el apartado 5.3. de Métodos Experimentales.

Para poder visualizar los productos tras la PCR se realizó una electroforesis en gel de agarosa al 1% y la posterior extracción en gel (apartados 5.4. y 5.5. de Métodos Experimentales).

La figura 13 representa el gel agarosa al 1% de la amplificación de las Endoglucanasas M1 y M2 de *G. kaustophilus* CECT4264. Las calles 1 y 2 corresponden a la amplificación del gen M1, utilizando ADN molde sin diluir (25,1 ng/μL) y diluido a la mitad (12,2 ng/μL). La calle 4 corresponde al marcador de Peso

Molecular (Plus Ladder). Las calles 6 y 7 corresponden a la amplificación del gen M2 sin diluir y diluido.

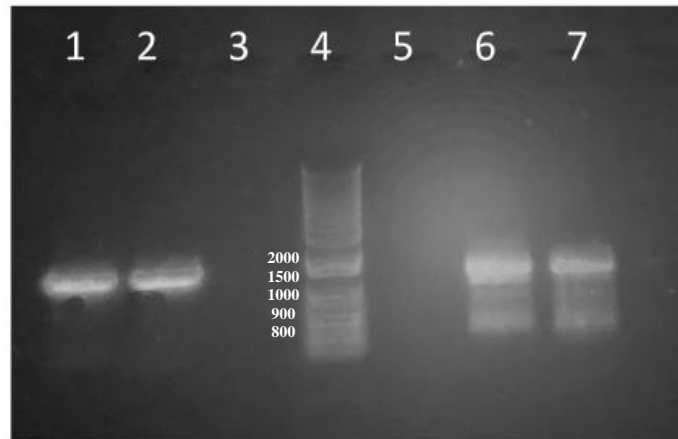


Figura 13: Gel de electroforesis de los productos de la PCR de M1 y M2 de *G. kaustophilus* CECT4264. Calle 1: amplificado de GkM1(sin diluir). Calle 2: amplificado de GkM1(diluido). Calle 4: marcador de peso molecular, Calle 6 amplificado de GkM2 (sin diluir). Calle 7: amplificado de GkM2 (diluido).

El producto de PCR obtenido para M1 generó una única banda de peso molecular por debajo de la banda de 1500 pares de bases (pb) del marcador, lo que concuerda con el peso molecular obtenido a partir de su secuencia (1170 pb). Sin embargo, el producto de PCR para M2 generó 2 bandas de peso molecular de 1000 y 800 pb, aproximadamente. El peso molecular para M2 a partir de la secuencia es de 1102 pb, de manera que el producto de PCR correcto es el de la banda de mayor tamaño.

La purificación de los productos de PCR se realizó según el apartado 5.5 de Métodos Experimentales. Tras extraer el fragmento del gel, se verificó que no quedaran restos del mismo (Figura 14). Se continuó con el protocolo de purificación y las muestras se guardaron a -20°C .



Figura 14: Gel de electroforesis tras haber extraído las bandas correspondientes a los productos de PCR de M1 y M2 de *G. kaustophilus* CECT4264.

Se realizó, de igual forma, la PCR para la Endoglucanasa M1 de *Geobacillus stearothermophilus* CECT43 y 49 a dos concentraciones (Figura 15 A y B), obteniéndose un producto de PCR a una masa molecular entre las bandas de 1000 y 1500 pb del marcador de peso molecular. Las bandas correspondientes a cada producto de PCR fueron extraídas del gel de agarosa y purificadas tal y como se ha descrito para el fragmento M1 de *G. kaustophilus* CECT4264, almacenándose a -20°C hasta su uso. Por último, se debe indicar que ambas amplificaciones dieron un único producto de PCR con una masa similar a la obtenida a partir de la secuencia (1170 pb).

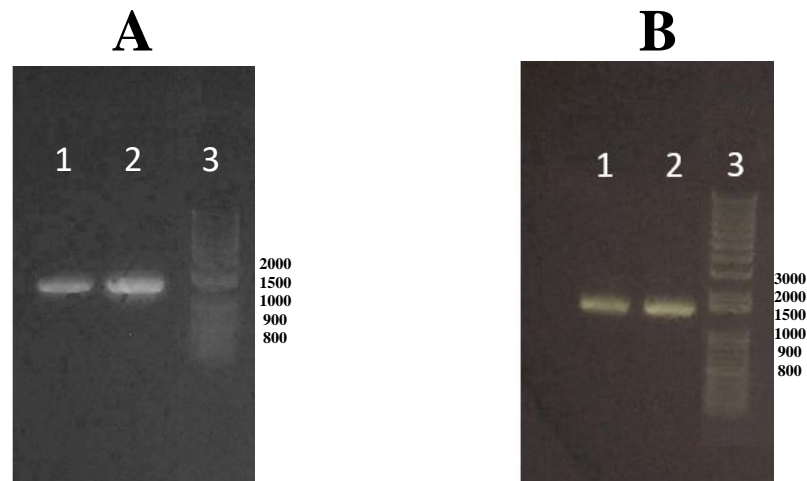


Figura 15: Gel de electroforesis de los productos de PCR de M1 de *G. stearothermophilus* CECT43 y 49. A) Calle 1: amplificado de Gs43M1 (18,9 ng/ μL). Calle 2: amplificado de Gs43M1 (11,6 ng/ μL). Calle 3: marcador de peso molecular. B) Calle 1: amplificado de Gs49M1 (44,5 ng/ μL), calle 2: amplificado de Gs49M1 (21,1 ng/ μL), Calle 3: marcador de peso molecular.

La amplificación por PCR del fragmento M2 de *Geobacillus stearothermophilus* CECT43 y 49 se muestra en la figura 16. Se obtuvo un fragmento entre las bandas de 1000 y 1500 pb del marcador de peso molecular en ambos casos, próximo al valor obtenido a partir de su secuencia (1102 pb). Al igual que los anteriores productos de PCR, se extrajeron las bandas del gel y se purificaron según lo descrito anteriormente y las muestras purificadas se almacenaron a -20°C hasta su utilización.

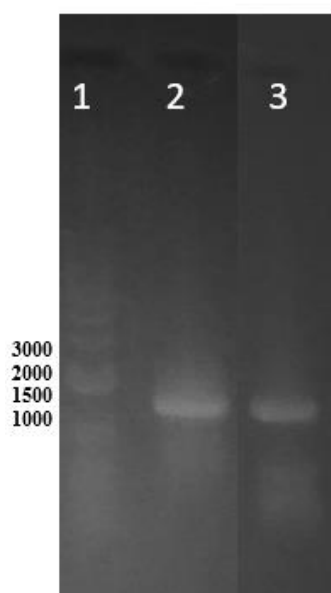


Figura 16: Gel de electroforesis de los productos de la PCR de M2 de *G. stearothermophilus* CECT43 y 49. Calle 1: marcador de peso molecular, calle 2: amplificado de Gs43M2 (11,6 ng/ μ L), calle 3: amplificado de Gs49M2 (21,1 ng/ μ L).

Una vez extraídas todas las muestras de los diferentes geles y siguiendo el protocolo 5.5. de Métodos Experimentales, se cuantificó en un espectrofotómetro de gota “NanoDrop” la concentración de ADN amplificado en cada caso, obteniéndose las concentraciones descritas en la tabla 3.

Tabla 3: Concentración de ADN amplificado mediante PCR para los fragmentos M1 y M2 en los organismos *G. kaustophilus* CECT4264 y *G. stearothermophilus* CECT 43 y 49.

Fragmento amplificado	[ADN amplificado] (ng/ μ L)
Endoglucanasa M1 <i>G. kaustophilus</i> CECT4264	39,5
Endoglucanasa M1 <i>G. stearothermophilus</i> CECT43	37,7
Endoglucanasa M1 <i>G. stearothermophilus</i> CECT49	11,3
Endoglucanasa M2 <i>G. kaustophilus</i> CECT4264	9
Endoglucanasa M2 <i>G. stearothermophilus</i> CECT43	15,7
Endoglucanasa M2 <i>G. stearothermophilus</i> CECT49	9,2

6.5. Clonación de los genes en el vector pTOPO.

Una vez amplificados y purificados de gel los fragmentos de interés se clonaron en el plásmido de sub-clonaje pTOPO (Figura 17) siguiendo el protocolo 5.7. del apartado Métodos Experimentales, en el que se describe la clonación y posterior transformación de células competentes.

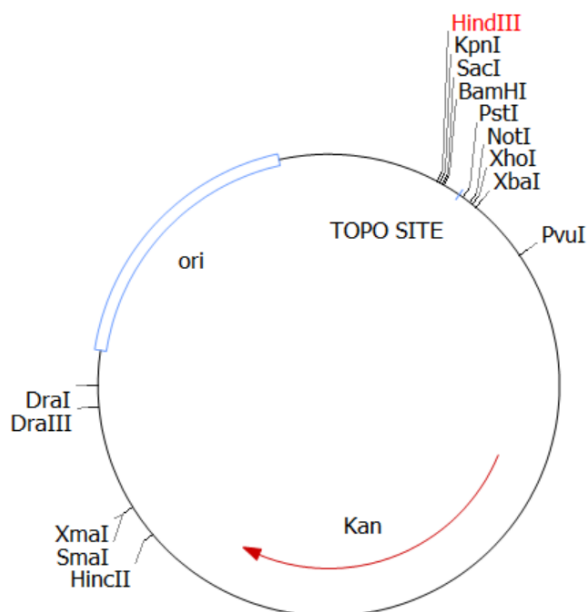


Figura 17: Esquema del vector pTOPO en el que aparecen los sitios de corte. Realizado con el programa Clone Manager 9.0.

La cepa hospedadora utilizada para la transformación fue *E. coli* DH5 α por carecer de las principales proteasas naturales, mantener la estabilidad del vector de expresión y conferir los elementos genéticos necesarios para una expresión eficiente y regulable (González y Fillas, 2018). A nivel de rendimiento en la producción de plásmidos recombinantes, esta cepa posee gran capacidad para expresar plásmidos y además es deficiente en endonucleasa, lo cual evita la digestión no específica de ADN, obteniéndose preparaciones limpias y mejores resultados en sus posteriores aplicaciones (Grant y col., 1990). Una vez transformadas las colonias se dejaron crecer en placas de LB con kanamicina como antibiótico, durante 12 horas a 37 °C. Posteriormente, dichos transformantes se separaron y ordenaron. En la Tabla 4 se recoge el número de transformantes que se obtuvieron para cada uno de los plásmidos recombinantes derivados de pTOPO. En la figura 11 del Anexo I, se muestran los esquemas de los vectores para cada una de la Endoglucanasas en estudio.

Tabla 4: Número de transformantes obtenidos con los plásmidos de subclonaje conteniendo los genes para las enzimas M1 y M2.

Enzima	Plásmido sub-clonaje	Transformantes
<i>GkM1</i>	pJMC103	38
<i>Gs43M1</i>	pJMC105	241
<i>Gs49M1</i>	pJMC119	14
<i>GkM2</i>	pJMC107	260
<i>Gs43M2</i>	pJMC109	25
<i>Gs49M2</i>	pJMC121	17

De cada tipo de transformante se separaron 10 colonias en placas de LB con kanamicina, utilizando plantillas cuadradas y numeradas. Cada una de las cuadrículas contenía una cepa evitando contacto entre ellas. Se dejaron crecer durante toda la noche a 37 °C y posteriormente se realizaron cultivos de 5 mL de las 4 primeras colonias de cada uno de los plásmidos. Los cultivos se crecieron en medio LB

conkanamicina dejándolos en agitación constante durante 12 horas a 37 °C. Se realizó la extracción de plásmido siguiendo el protocolo descrito en el apartado 5.9. de Métodos Experimentales y se realizó un análisis de restricción con la enzima *EcoRI* para comprobar si los plásmidos contenían un fragmento con el tamaño esperado para nuestros genes de interés. El sitio de policlonação de pTOPO permite obtener dos fragmentos cortando con una sola enzima de restricción ya que tiene dos sitios repetidos para algunas enzimas como es el caso de *EcoRI*. El corte con esta enzima dio lugar a dos fragmentos para pJMC103: 1194 pb y 3501 pb y también dos fragmentos para pJMC107: 1120 pb y 3500 pb. En el caso de no contener el fragmento de interés la digestión de pTOPO daría lugar a dos fragmentos: 18 pb y 3501, pero en este caso la banda de 18 pb no se verá y el resultado real sería de una única banda en el gel de electroforesis (Figura 18).

Las calles correspondientes a los cortes de los plásmidos pJMC103 (*Gkm1*) dieron lugar al resultado esperado para los cuatro plásmidos purificados, mientras que en el caso de pJMC107 (*Gkm2*) solamente se obtuvieron resultados positivos para los clones 2 y 4. En el caso pJMC107.1 no se obtuvo plásmido de la preparación y el corte de pJMC107.3 dió lugar a unas bandas cuyos tamaños no correspondía con los tamaños esperados. Se seleccionaron pJMC103.2 y pJMC107.2 para los genes que codificaban para M1 y M2 de *Geobacillus kaustophilus* CECT4264.

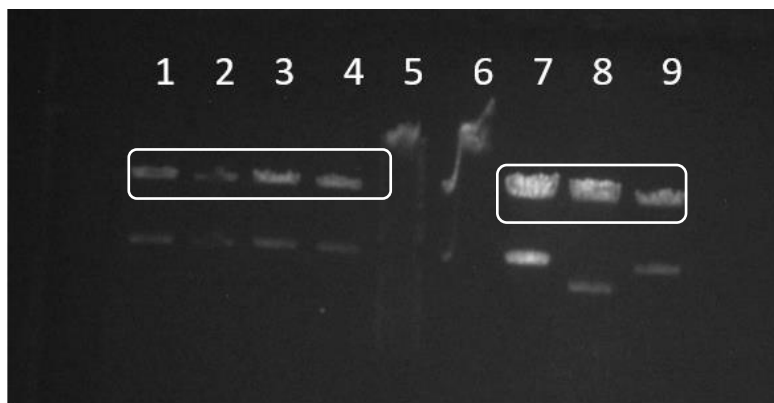
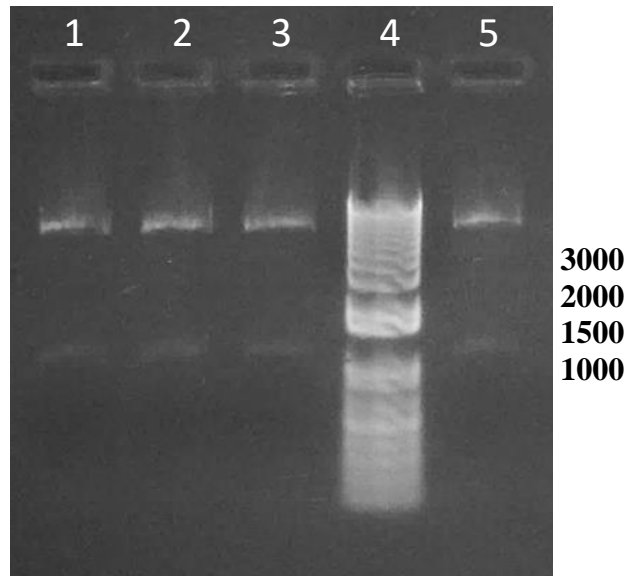


Figura 18: Comprobación de la presencia de los genes que codifican para las Endoglucanasas M1 y M2 de *Geobacillus kaustophilus* CECT4264 en los diferentes plásmidos recombinantes. Calle 1: pJMC103.1 *EcoRI*. Calle 2: pJMC103.2 *EcoRI*. Calle 3: pJMC103.3 *EcoRI*. Calle 4: pJMC103.4 *EcoRI*. Calle 5: Marcador. Calle 6: pJMC107.1 *EcoRI*. Calle 7: pJMC107.2 *EcoRI*. Calle 8: pJMC107.3 *EcoRI*. Calle 9: pJMC107.4 *EcoRI*.

El análisis de los genes M1 y M2 de los genomas de *G. stearothermophilus* CECT43 y 49 se analizaron cortando con las enzimas *NdeI/BamHI* para pJMC105 (*Gs43M1*) y pJMC119 (*Gs49M1*) y *BamHI/XhoI* para pJMC109 (*Gs43M2*) y pJMC121 (*Gs49EndGM2*). En este caso no se utilizó *EcoRI* porque el gen de la Endoglucanasa M1 de *G. stearothermophilus* CECT43 tenía un sitio de corte para esta enzima y se utilizó la misma estrategia para todos los genes de *G. stearothermophilus*.

En la figura 19 se puede observar el análisis de un transformante de cada uno de los plásmidos (pJMC105, pJMC119, pJMC109 y pJMC121). Los resultados esperados eran tres bandas de 43 pb, 1171 pb y 3481 pb para los genes M1 de ambas cepas, y 43 pb, 45 pb, 1096 pb y 3473 pb en el caso de M2 de ambas cepas. De todos los posibles fragmentos solamente los de mayor tamaño son las que se pueden visualizar en los geles al 1% de agarosa (Figura 29). Se seleccionaron pJMC105.7, pJMC119.14, pJMC109.5 y pJMC121.7 por presentar una banda aproximada de 1100 pb.

Figura 19: Comprobación de la presencia de los genes que codifican para las Endoglucanasas M1 y M2 de *G. stearothermophilus* CECT43 y 49 en los diferentes plásmidos recombinantes. Calle 1: pJMC105.7 digerido con NdeI/BamHI. Calle 2: pJMC119.4 digerido con NdeI/BamHI. Calle 3: pJMC109.5 digerido con BamHI/XhoI. Calle 4: Marcador de Peso Molecular. Calle 5: pJMC121.7 digerido por BamHI/XhoI.



Una vez seleccionados los transformantes, realizamos una nueva extracción de ADN plasmídico y cortamos con las enzimas que flanquean los genes de interés. En esta electroforesis se incluyeron los plásmidos de expresión pBBM35 y pBBM48 que también se separaron una vez digeridos con las correspondientes enzimas de restricción. (Figura 20).

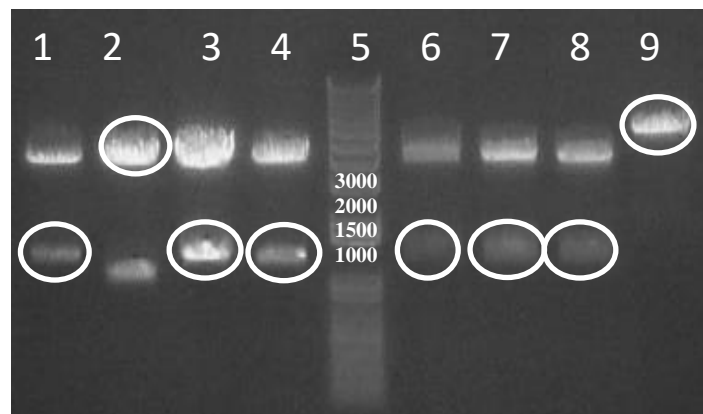


Figura 20: Electroforesis de ADN de los diferentes productos de corte. Calle 1: pJMC103.2 digerido con NdeI/BamHI. Calle 2: pBBM35 digerido con NdeI/BamHI. Calle 3: pJMC105.7 digerido con NdeI/BamHI. Calle 4: pJMC119.4 digerido con NdeI/BamHI. Calle 5: Marcador de Peso Molecular. Calle 6: pJMC107.2 digerido con BamHI/XhoI. Calle 7: pJMC109.5 digerido con BamHI/XhoI. Calle 8: pJMC121.7 digerido con BamHI/XhoI. Calle 9: pBBM48 digerido con BamHI/XhoI.

Resuelta la electroforesis se cortaron en gel los fragmentos de interés (marcados en círculo en la figura 20), en el caso de los plásmidos de la serie pJMC las bandas correspondientes a los fragmentos de menor tamaño y en el caso de la serie pBBM la banda de mayor tamaño que corresponde a los plásmidos abiertos.

6.6. Construcción de los plásmidos de expresión.

Una vez obtenidos los plásmidos de expresión abiertos y los fragmentos correspondientes a cada una de las Endoglucanasas, éstos se unieron mediante la acción de una ADN ligasa siguiendo el protocolo descrito en el apartado 5.11. de Métodos Experimentales. Los plásmidos recombinantes resultantes se utilizaron para transformar células competentes de *E. coli* DH5 α , obteniendo el número de colonias que aparece en la tabla 5 (Figura 12 Anexo I). Los plásmidos resultantes se analizaron por restricción tal y como en el caso de los plásmidos de subclonaje. En este caso los fragmentos resultantes eran de 1169 pb y 4623 pb para los plásmidos pJMC104, pJMC106 y pJMC120 digeridos con *NdeI/BamHI*, y 1096 pb y 4623 pb para los plásmidos pJMC108, pJMC110 y pJMC122 digeridos con *BamHI/XhoI*

Tabla 5: Número de clones obtenidos al transformar con los plásmidos de expresión conteniendo las enzimas M1 y M2.

Enzima	Plásmido expresión	Transformantes
<i>GkM1</i>	pJMC104	10
<i>Gs43M1</i>	pJMC106	81
<i>Gs49M1</i>	pJMC120	239
<i>GkM2</i>	pJMC108	403
<i>Gs43M2</i>	pJMC110	20
<i>Gs49M2</i>	pJMC122	60

Los plásmidos se enviaron a secuenciar a los Servicios Técnicos de la Universidad de Almería. Se analizaron las secuencias de ADN y en el caso de pJMC104, tras utilizar cebadores de secuenciación para el fragmento M1, se obtuvo un fragmento de 1167 pb (Figura 21A) con una homología del 100% con la secuencia teórica de *Geobacillus kaustophilus* NRBC102445 (Figura 13 del Anexo I). Respecto a M1 de *Geobacillus stearothermophilus* CECT43 para pJMC106, la secuenciación no fue exitosa, ya que se obtuvo un fragmento pequeño que no hibridaba con la secuencia molde y por tanto fue descartada. La secuenciación de ambos extremos del fragmento M1 de *Geobacillus stearothermophilus* CECT49 clonado en pJMC120 proporcionó una secuencia de 1167 nucleótidos con tan solo 7 nucleótidos de diferencia respecto a la secuencia teórica de *Geobacillus stearothermophilus* B5 (marcados en verde en la figura 21B), que se tradujo solamente en 3 aminoácidos de diferencia (Figura 14 del Anexo I).

ATGGCGAATGATTCATTTTGCCAATCGTCAGCCTTTTCAAGTATAATAAAAAAGTGGACAAAACCGGAAAGGGGTG
 CCAAGATGGCGAAGTTGGACGAAACGTTGACGATGCTGAAAGCGCTGACCGATGCGAAAAGGCGTCCCGGGCAATGA
 GCGGGAAGCGCGCGATGTGATGAAGACATACATAGCTCCGTATGCGGATGAAGTGACAACGGATGGCTCGGCAGC
 TTGATTGCCAAAAAAGAAGGGAAATCGGGCGGGCCGAAAGTGTATGATCGCCGGCCATTTGGATGAAGTCGGCTTTA
 TGGTGACGCAAAATCGACGACAAAGGATTCATCCGCTTCCAAACGCTTGGCGGATGGTGGAGCCAAGTGATGCTCGC
 CCAGCGCTGACGATCGTGACGAAAAAAGGCGACATACCGGCGTCATCGGTTGAAAGCCGCCGCACATTCTGCCG
 CCGGAGGCGCGCAAAAAACCGGTGAAATCAAAGATATGTTTCATCGACATCGGCGGACAAGCCCGGAGGAAGCGA
 TGAGTGGGGCGTCCGCCCGGGCGATATGATCGTGCCGATTTTTGAATTTACGGTATTGAACAATGAAAAATGCT
 GCTCGGAAAGCATGGGACAACCGGATCGGCTGTGCGGTGCGCATCGATGTGCTCAAGCAGCTGAAAGGCGTCGAC
 CATCCAAACACGGTATACGGCGTCGGCACGGTGCAGGAAGAAGTCGGCTTGC CGCGCGCGCACGGCCGCCCAAT
 TCATTAGCCAGATATCGCCTTTGCTGTTGACGTTCGGCATTGCCGCGACACGCCGGCGTGTGCGAAAAAGAAGC
 GATGGGCAAACTCGGCGCGGGCCCGCACATCGTCTGTACGACGCAACGATGGTGTGCGACCGCGGCTTGC GCGAA
 TTTGTCATCGAAGTGGCGGAAGAGCTCAACATTCGTAACATTTTGTATGCCATGCCAGGCGGCGGTACGGACGCGG
 GAGCGATTCACTTAACCGGCATCGCGGTTCCGTGCTCACGATTGCCATCCCGACGCGCTACATCCACTCGCACGC
 CGCCATTTTGCACCGGACGACTACGAAAAACCGGTCAAGCTGCTTGTGAGGTGATCAAACCGGCTTGACGCCGAC
 AAAGTGAACAACACTGACGTTTGACGAA

Figura 21A: Secuencia nucleotídica del fragmento M1 *G. kaustophilus* CECT4264 clonado en pJMC104, mostrándose en gris el codón de inicio y presentado un 100% de homología con respecto a la secuencia teórica *G. kaustophilus* NRBC102445.

ATGGCGAATGATTCATTTTGCCAATCGTCAGCCTTTTCAAGTATAATAAAAAAGTGGACAAAACCGGAAAGGAGTG
 CCAAGATGGCGAAGTTGGACGAAACGTTGACGATGCTGAAAGCGCTGACCGATGCGAAAAGGCGTCCCGGGCAATGA
 GCGGGAAGCGCGCGATGTGATGAAGACATACATAGCTCCGTATGCGGATGAAGTGACAACGGATGGCTCGGCAGC
 TTGATTGCAAAAAAGAAGGGAAATCGGGCGGGCCGAAAGTGTATGATCGCCGGCCATTTGGATGAAGTCGGCTTTA
 TGGTGACGCAAAATCGACGACAAAGGCTTCATCCGCTTCCAAACGCTTGGCGGATGGTGGAGCCAAGTGATGCTCGC
 CCAGCGCTGACGATCGTGACGAAAAAAGGCGACATACCGGCGTCATCGGTTGAAAGCCGCCGCACATTCTGCCG
 CCGGAGGCGCGCAAAAAACCGGTGAAATCAAAGATATGTTTCATCGACATCGGCGGACAAGCCCGGAGGAAGCGA
 TGAGTGGGGCGTCCGCCCGGGCGATATGATCGTGCCGATTTTTGAATTTACGGTATTGAACAATGAAAAATGCT
 GCTCGGAAAGCATGGGACAACCGGATCGGCTGTGCGGTGCGCATCGATGTGCTCAAGCAGCTGAAAGGCTCGAC
 CATCCAAACACGGTATACGGCGTCGGCACGGTGCAGGAAGAAGTCGGCTTGC CGCGCGCGCACGGCCGCCCAAT
 TCATTAGCCAGATATCGCCTTTGCGGTGGATGTCGGCATTGCCGCGATACACCGGGCGTGTGCGAAAAAGAAGC
 GATGGGCAAACTCGGCGCGGGCCCGCACATCGTCTGTACGACGCGACGATGGTGTGCGACCGCGGCTTGC GCGAA
 TTCGTCATCGAAATGGCGGAAAGCTGAACATTCGTAACATTTTGTATGCCATGCCAGGCGGCGGTACGGATGCGG
 GAGCGATTCACTTAACCGGCAGCGCGGTTCCGTGCTCACGATCGCCATCCCGACGCGCTACATCCACTCGCACGC
 CGCCATTTTGCATCGCGACGACTACGAGAACACGGTCAAGCTGCTTGTGCGAAGTGTCAAACCGGCTTGACGCCGAC
 AAAGTGAACAACACTGACGTTTGACGAA

Figura 21B: Secuencia nucleotídica del fragmento M1 *G. stearothermophilus* CECT49 clonado en pJMC120, mostrándose en gris el codón de inicio y en verde los 7 nucleótidos de diferencia respecto a la secuencia teórica *G. stearothermophilus* B5.

En el caso de la proteína M2, los plásmidos de expresión eran pJMC108 (gen amplificado a partir de *G. kaustophilus* CECT4264), pJMC110 (gen amplificado a partir de *G. stearothermophilus* CECT43) y pJMC122 (gen amplificado a partir de *G. stearothermophilus* CECT49). La secuencia obtenida a partir del plásmido pJMC108, no hibridó con el gen teórico al hacer el análisis por homología de secuencia mediante BLAST por lo que fue descartado. Por otro lado, las secuencias de pJMC110 y pJMC122 resultaron ser iguales. La figura 22, muestra la secuencia del fragmento M2 de *G. stearothermophilus* CECT49, la secuencia tuvo un tamaño de 1101 nt, mostrando una diferencia de 143 nt lo que se traduce en 27 aminoácidos diferentes con respecto a la teórica de *G. stearothermophilus* B5 (Figura 16 Anexo I).

```

ATGAATGAGGACGCTGCAATTTTC CAAACGTTGACCGAGCTCCAGGCGCACCGGGTATGAACATCCGGTGC
GCAGTTTTATGCGCAAGAGCTCGCAAATACGCCGATGACATGTGCAAGACCGCTCGGCAGCTTTTCGGCGT
CAAACGCGGTGATGAGCGCGGCCCGACGGTGATGGTCGCCGGCCATATGGATGAGTCCGGAATTTATGGTGACAGCG
ATCACCGATCACGGCATGATTCGTTTCAGCCGCTAGGAGGCTGGTGGGATCAAGTGCTGCTCGCCAGCGCGTGC
AGATTATCACGGACGGAGGGCCGGTCATCGGCGTATCGGCTCGATTCCGCCGCAATTTGCTTGATGACGAGCAGCG
CAA AAAACCGATGGAGATCAAAAACATGTTGATTGACATCGGAGCGGAAAGCCGTGAGATGCCGACGAGTCCGGC
ATTCGCCCCGGCCAGCCGATCGTGCCGGTCAGCCGTTTACTTTGCTTGC CAACGGAAAAACCGTGATGGCAAAA
CGTGGGACAAACGGTTTGGCTGCGGATTGGCGATGAGCTGTTAAAAGAGCTGAAAGGGCAAAACACTCCGAACGT
CTTGACGCAAGGCGGACCGTGCAGGAAGAGGTTCGGGCTGCGTGGGGCCAGACGGCGGCTTCGATGATCAACCCG
GACATCTTTTTCGCCCTCGAAGCAAGCCCGGC AACGATATGTCGGCGACAAGCAGGCGTTCCGCCATATCGGCA
AAGGAGCGCTATCCGCTTGTACGACCGGACGATGATACGCACCGCGGCATGCGCGAATTTGTGCTTGATACGGC
GGAACGATCGGCTCCGTACCAATTTTTTCATTTCGCCAGCGGGGGAACGACCGCGGCGGTGCATATCGCC
AACCGCGGCTGCGCTCGGCGGTGATCGGCATTTCCGCCCGCTACATTACATACGCATGCGTTCGATCATTATGTCG
ACGATTACGCGCGGCGCAAACAGCTCATCGTTGAGCTCGTCAAACGGTGCACCGGACAAACGGTGAGACGATTCG
CCAAAACAGCCTCGAGCACCACCACCACCACCTGA

```

Figura 22: Secuencia nucleotídica del fragmento M2 de *G. stearothermophilus* CECT49 clonado en pJMC122, mostrándose en gris el codón de inicio y en verde los 143 nt de diferencia respecto a la secuencia teórica de *G. stearothermophilus* B5.

Para finalizar con el análisis de ambas secuencias estas se tradujeron hasta aminoácidos obteniendo las secuencias presentadas en las figuras 23 y 24. La secuencia aminoacídica de la proteína Gs49M1 (Figura 23), presenta 389 residuos con una masa molecular de 43kDa, mientras que la proteína Gs49M2 contenida pJMC122 (Figura 24), presenta 359 residuos con una masa molecular de 42kDa.

El tamaño descrito en bibliografía para las diferentes Endoglucanasas varía desde 37 kDa para *Bacillus licheniformis* (Liu y col., 2018) hasta 72 kDa para la aislada de *Clostridium thermocellum* (Jantasaeng y col., 2019). Nuestro resultado está en consonancia con el tamaño descrito para *Geobacillus* sp. 70PC53 que mostró un tamaño de 43 kDa (Ng y col., 2009). Estos autores destacan la versatilidad de utilizar enzimas con alta estabilidad térmica que sumado a las pocas Endoglucanasas descritas hasta la fecha dentro del género *Geobacillus*, evidencian la gran importancia de estas nuevas enzimas encontradas en *Geobacillus stearothermophilus* CECT49.

```

MANDSFCQSSAFSSI IKKWTKPERSAKMAKLDLTLMLKALTDAGVPGNEREARDVMKTYIAPYADEVTTDGLGS
LIAKKEGKSGGPKVMIAGHLDEVGFMVTQIDDKGFI RFQTLGGWWSQVMLAQRVTIVTKKGDVTGVI GSKPPHILP
PEARKKPVEIKDMFIDIGATSREEAMEWGV RPDIMVYPYFEFTVLNNEKMLLAKAWDNRIGCAVAIDVLKQLKGV
DHPNTVYGVGTVQEEVGLRGARTAAQFI QPDI AFVDVGIAGDTPGVSEKEAMGKLGAGPHIVLYD ATMVSHRGLRE
FVIELAEELNIPYHFDAMPGGTDAGAIHLTGS GVP SLTIAIPTRYI HSHAAI LHRDDYENTVKLLVEVIKRLDAD
KVKQLTFDE

```

Figura 23: Secuencia aminoacídica de la proteína Gs49M1. Se resalta el primer aminoácido metionina sombreada en gris.

```

MNEDTLQLFQTLTELPGAPGYEHPVRQFMRKELEKYADDIVQDRLGSLFGVKRGDEAGPTVMVAGHMDEVA FMVTA
ITDNGMIRFQPLGGWWDQVLLSQRVQIITDGGPVI GVI GSI PPHLLDDEQRKKPMEMKNMLIDIGAESREDAERVG
VRPGQPIVPVSPFTVLANGKTVMAKAWDNRF GCGLAIELLKELNGETV PNVLYAGATVQEEVGLRGAQTAAAMINP
DIFFALEASPANDMSGDKQAFGHIGK GALIRLYDRTMI THRGMREFVLDTAETIGV P YQFFI SPGGGTDAGR VHIA
NRGVPSAVIGICARYIHTHASI IHIDDYAAAKQLIVELVKRCDRRTTVETIRQNS

```

Figura 24: Secuencia aminoacídica de la proteína Gs49M2. Se resalta el primer aminoácido metionina sombreada en gris.

Las secuencias de aminoácidos de M1 y M2 de *G. stearothermophilus* CECT49 se compararon con las secuencias aminoacídicas de las ambas Endoglucanasas de *G. stearothermophilus* B5. En la figura 14 del Anexo I se pone en evidencia la elevada

homología de secuencia para M1, ya que ambas presentan un 99% de similitud, diferenciándose únicamente en 3 aminoácidos. Sin embargo, para la Endoglucanasa M2 la homología de secuencia es menor, existiendo 27 aminoácidos diferentes lo que supone una homología del 92,5% (Figura 16 Anexo I).

6.7. Análisis de las secuencias.

Como sabemos *E. coli* es una bacteria ampliamente utilizada en la expresión de proteínas recombinantes gracias a su facilidad de manejo y crecimiento, así como el buen conocimiento de su genética. Es por ello que el uso óptimo de codones en las secuencias a expresar aumenta la velocidad de traducción. La presencia de codones raros en genes altamente expresados puede reducir la velocidad de traducción e incluso alterar la traducción de otros genes que utilizan el mismo codón (Frumkina y col., 2018; Carmignotto y col., 2019).

En 1990, gracias al estudio S. Anderson (Anderson y col., 1990) se estableció que la tasa promedio de traducción de los codones bacterianos era de 12 aminoácidos por segundo. Por ejemplo, si se compara la tasa del codón GAA con el codón GAG, siendo ambos codones comunes para Glu, el triplete GAA se traduce 3 veces más rápido que la cadena de GAG (Anderson y col., 1990). Es por ello que se deduce que la composición de nucleótidos de la secuencia que codifica nuestra proteína podría afectar drásticamente a su expresión.

Con ayuda del programa Caltor (<https://people.mbi.ucla.edu/sumchan/caltor.html>) se analizaron las secuencias de ambas Endoglucanasas, en busca de codones raros en su secuencia.

Para el caso de la Endoglucanasa M1 de *G. stearothermophilus* CECT49 (pJMC120) se encontraron 29 codones raros de 389, es decir, un 7,5%, distribuidos en toda la secuencia del gen (Figura 17 Anexo I, Tabla 6), mientras que para la secuencia de la Endoglucanasa M2 de *G. stearothermophilus* CECT49 (pJMC122), se encontraron 37 codones raros de 367, es decir, un 10%, distribuyéndose de nuevo a lo largo de toda la secuencia del gen (Figura 18 Anexo I, Tabla 7).

Tabla 6: Codones presentes en la secuencia de la enzima Endoglucanasa M1 de *G. stearothermophilus* CECT49. Se han encontrado un total de 29 codones raros uniformemente distribuidos en la secuencia genómica

Aminoácido	Codón raro	Frecuencia de aparición	Posición
Arginina	CGG	3	52, 211, 374
	AGG	1	241
Glicina	GGA	2	119, 330
	GGG	2	83, 87
Isoleucina	AUA	3	15, 16, 62
Treonina	ACG	18	34, 36, 71, 104, 116, 131, 134, 195, 232, 238, 250, 295, 327, 335, 343, 348, 366, 386

Tabla 7: Codones presentes en la secuencia de la enzima Endoglucansa M2 de *G. stearothermophilus* CECT49 (pJMC122). Se han encontrado un total de 37 codones raros uniformemente distribuidos en la secuencia genómica.

Aminoácido	Codón raro	Frecuencia de aparición	Posición
Arginina	CGG	6	26, 182, 263, 300, 345, 348
	CGA	1	150
Glicina	GGA	8	71, 89, 107, 141, 171, 186, 255, 295
	GGG	7	46, 108, 115, 207, 215, 218, 299
Leucina	CUA	1	88
Prolina	CCC	1	155
Treonina	ACG	13	5, 11, 60, 78, 105, 221, 264, 267, 278, 281, 322, 350, 353

Sharp y Li en 1987 propusieron el “índice de adaptación de codones, (CAI)”. Este índice varía de 0 a 1, siendo el valor de 1, si el gen utiliza codones usados frecuentemente en el sistema heterólogo de expresión o 0 si no los utiliza (Sharp y Li, 1987). Al calcular del índice de adaptación de los codones (Codon Adaptation Index, CAI) de las secuencias de estudio se obtuvo un valor de 0,74 para M1 y 0,71 M2. Si sumamos estos índice CAI a los porcentajes de codones raros presentes en las secuencias, cercanos al 10% indicando una posible dificultad a la hora de expresar ambas proteínas en *E. coli*.

6.8. Expresión de las proteínas recombinantes.

Debido a que de la cepa *G. kaustophilus* CECT4264 únicamente se ha verificado la secuencia del fragmento M1, y además de *Geobacillus stearothermophilus* CECT43 solamente se obtuvo secuencia de M2, se decidió trabajar a partir de esta etapa con *Geobacillus stearothermophilus* CECT49 ya se había obtenido la secuencia tanto de M1 como de M2. Se procedió a la expresión en varias cepas de *E. coli* (C43, JMC109, y BL21) con los plásmidos pJMC120 y pJMC122

Para el crecimiento, inducción de las células recombinantes y su posterior ruptura el protocolo a seguir fue el descrito en los apartados 5.12. y 5.13. de Métodos Experimentales. Para evaluar qué cepa de *E. coli* era la óptima produciendo proteína recombinante tanto para M1 como M2 se realizaron electroforesis de poliacrilamida en condiciones desnaturalizantes, con dodecil sulfato sódico (SDS-PAGE).

A partir del estudio de secuencia aminoacídica conocemos la masa molecular teórica para el monómero de la enzima M1, que es de 43 kDa (pJMC120) y de 42 kDa para M2 (pJMC122), ambas de *G. stearothermophilus* CECT49. De esta manera podemos visualizar la proteína recombinante en una SDS-PAGE. Se realizó un estudio preliminar de inducción con la concentración de inductor estándar 0,2% de ramnosa para expresar M1 contenida en el plásmido pJMC120 y 0,5 mM de IPTG para inducir la expresión de M2 contenida en el plásmido pJMC122. En la figura 25 se observan las bandas de proteínas correspondientes a los cultivos de tres cepas de *E. coli* no inducidas e inducidas, y en este último caso, precipitado y sobrenadante inducido. Las condiciones de inducción fueron 32 °C durante 12 horas. En todos los casos la proteína quedaba mayoritariamente en el precipitado.

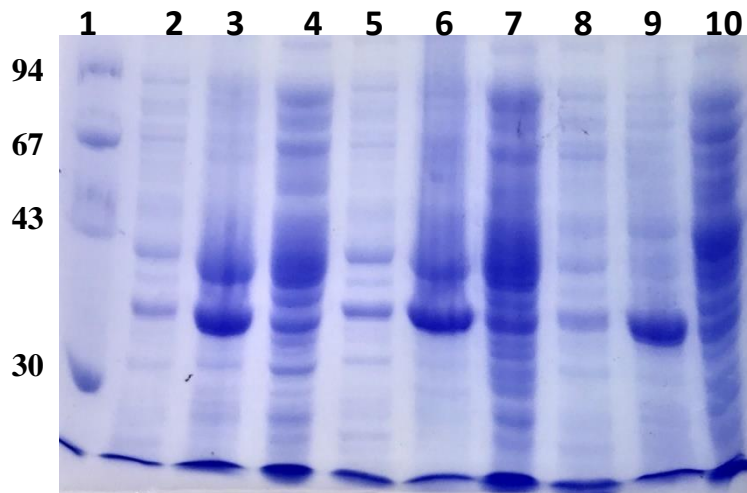


Figura 25: Calle 1: Marcador de Masa Molecular. Calle 2: Cultivo no inducido *E. coli* BL21 pJMC120. Calle 3: *E. coli* BL21 pJMC120 precipitado inducido. Calle 4: *E. coli* BL21 pJMC120 sobrenadante inducido. Calle 5: Cultivo no inducido *E. coli* C43 pJMC120. Calle 6: *E. coli* C43 pJMC120 precipitado inducido. Calle 7: *E. coli* C43 pJMC120 sobrenadante inducido. Calle 8: Cultivo no inducido *E. coli* JM109 pJMC120. Calle 9: *E. coli* JM109 pJMC120 precipitado inducido. Calle 10: *E. coli* JM109 pJMC120 sobrenadante inducido.

Siguiendo la misma estrategia, se ha estudiado la expresión de M2 contenida en el plásmido pJMC122, pero en este caso la inducción fue a 37 °C manteniendo el tiempo de inducción toda la noche. Los resultados de la expresión en las tres cepas de bacterias se recogen en la figura 26.

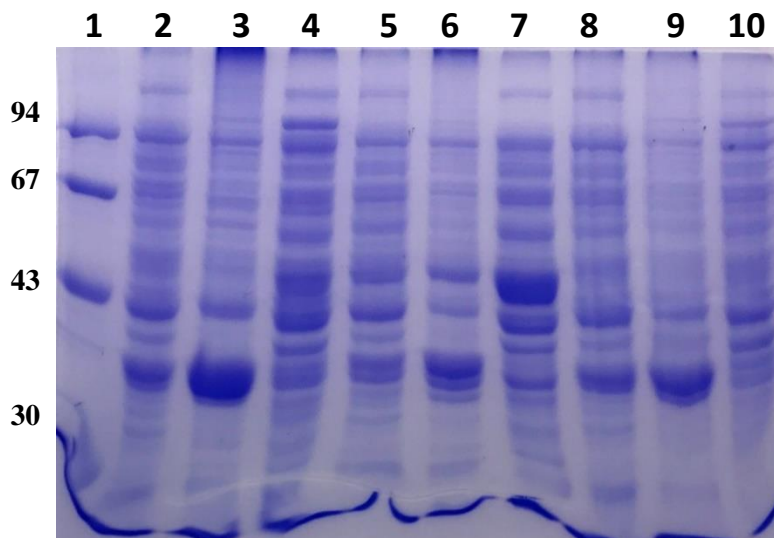


Figura 26: Calle 1: Marcador de Masa Molecular. Calle 2: Cultivo no inducido *E. coli* C43 pJMC122. Calle 3: *E. coli* C43 pJMC122 precipitado inducido. Calle 4: *E. coli* C43 pJMC122 sobrenadante inducido. Calle 5: Cultivo no inducido *E. coli* JM109 pJMC122. Calle 6: *E. coli* JM109 pJMC122 precipitado inducido. Calle 7: *E. coli* JM109 pJMC122 sobrenadante inducido. Calle 8: Cultivo no inducido *E. coli* BL21 pJMC122. Calle 9: *E. coli* BL21 pJMC122 precipitado inducido. Calle 10: *E. coli* BL21 pJMC122 sobrenadante inducido.

Como pudimos comprobar, mayoritariamente quedaba la proteína expresada en el precipitado. Estos resultados corroboran los análisis tanto de índice CAI como de presencia de codones raros en la secuencia de ambos genes. Será necesario realizar nuevas expresiones utilizando condiciones más relajadas ya sea bajando la temperatura de inducción o la concentración de inductor. Por otro lado, también se pueden utilizar cepas capaces de expresar codones raros como puede ser *E. coli* Rossetta.

6.9. Medida de la actividad de las proteínas recombinantes.

Para poder determinar la actividad de las enzimas expresadas, se llevó a cabo un análisis sobre CMC (Carboximetilcelulosa). Para la medida de la actividad se siguió el protocolo descrito en el apartado 5.15. de Métodos Experimentales.

En primer lugar, se realizó una recta patrón poder cuantificar la actividad fue necesario realizar una recta patrón para determinar el coeficiente de extinción molar del producto de la reacción, en nuestro caso xilosa. Se utilizó una concentración inicial de 10 mg/mL de xilosa y se prepararon muestras de entre 0-250 µg, en un volumen final de 100 µL, para las que finalmente se medirá su absorbancia, obteniéndose los resultados de la tabla 8 y figura 27:

Tabla 8: Concentraciones para la construcción de la recta patrón de concentraciones crecientes del producto de la reacción (xilosa) medida a 540 nm.

Concentración (µg)	Abs 540 nm (y)	mg/mL (x)
0	0	0
25	0,188	0,25
50	0,426	0,5
75	0,763	0,75
100	0,952	1
125	1,165	1,25
150	1,329	1,5
175	1,573	1,75
200	1,743	2
225	1,824	2,25
250	2,056	2,5

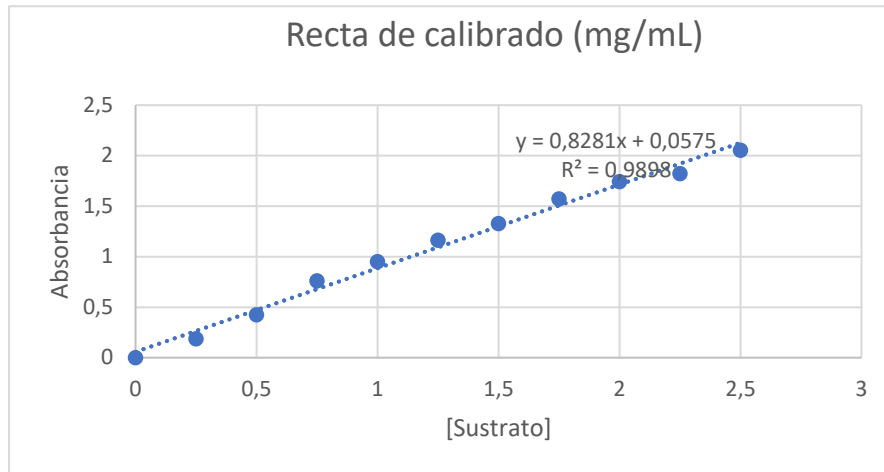


Figura 27: Recta patrón de concentraciones crecientes del producto de la reacción (xilosa) medida a 540 nm.

Del resultado de la recta patrón obtuvimos un valor de ϵ de $0,83 \text{ mL} \cdot \text{cm}^{-1} / \text{mg}$. La reacción enzimática se llevó a cabo incubando $10 \mu\text{L}$ de cada uno de los extractos celulares durante 30 minutos a 40°C sobre el sustrato CMC. De forma cualitativa pudimos comprobar que para M1 la mejor actividad se encontró en las cepas inducidas en C43 mientras que para M2 se obtuvieron los mejores resultados en BL21. En la figura 28 se puede apreciar un incremento en la intensidad de color en la cepa C43 pJMC120 y BL21 pJMC122, no apreciando diferencias con respecto al control en el caso de JM109, para ninguno de los plásmidos.

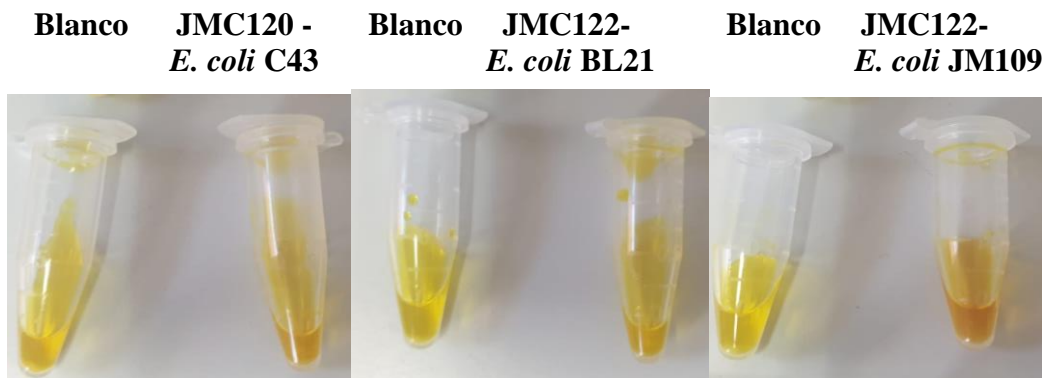


Figura 28: Representación colorimétrica de la actividad Endoglucanasa de las diferentes cepas de *E. coli* conteniendo los plásmidos JMC120 y pJMC122

Los resultados de la medida de actividad se recogen en la tabla 9 y tabla 10. Como podemos observar la expresión en la cepa C43 es la que mostró mejores resultados cuando se expresó la enzima Endoglucanasa M1 (pJMC120), siendo BL21 la mejor cepa para la expresión de M2 (pJMC122).

Tabla 9: Evaluación de la actividad Endoglucanasa M1 expresada en diferentes cepas de *E. coli*.

Plásmido	Cepas de <i>E. coli</i>	Absorbancia	[Azúcar Reductor obtenido] (mg/mL)
pJMC120	C43	0,142	0,17
	JM109	0,068	0,08
	BL21	0,071	0,08

Tabla 10: Evaluación de la actividad Endoglucanasa M2 expresada en diferentes cepas de *E. coli*.

Plásmido	Cepas de <i>E. coli</i>	Absorbancia	[Azúcar Reductor obtenido] (mg/mL)
pJMC122	BL21	0,160	0.19
	JM109	0,064	0.08
	C43	0,070	0.08

Estos estudios de actividad para las Endoglucanasas M1 y M2 de *Geobacillus stearothermophilus* CECT49 son preliminares, siendo el punto de partida para la puesta a punto de la expresión y purificación a gran escala de esta nueva enzima en futuros trabajos.

7. CONCLUSIONES

1. Mediante el estudio de homología de secuencia se han encontrado 2 posibles Endoglucanasas, M1 y M2 en los organismos termófilos *Geobacillus kaustophilus* CECT 4264 y *Geobacillus stearothermophilus* CECT43 y 49.
2. Utilizando la técnica de la Reacción en Cadena de la Polimerasa se han logrado amplificar los fragmentos de ADN codificantes para M1 y M2 en ambas especies de *Geobacillus*. La clonación en el vector de subclonado pTOPO facilitó la posterior inserción de los fragmentos amplificados en los vectores de expresión pJOE4036.1 y pET21b+.
3. La secuenciación de los 6 fragmentos clonados fue positiva para 4 de ellos, mostrando M1 de *G. kaustophilus* CECT4264 un 100% de homología con la secuencia teórica y M1 de *G. stearothermophilus* CECT49 únicamente 3 aminoácidos de diferencia respecto a la descrita. Las secuencias M2, fueron positivas las dos de *G. stearothermophilus* (CECT43 y 49), mostrando una misma secuencia. En este caso la homología para M2 disminuyó sensiblemente respecto a la secuencia teórica presente en el GenBank, ya que se encontraron 27 aminoácidos diferentes.
4. Se ha abordado la expresión preliminar de las Endoglucanasas M1 y M2 de *G. stearothermophilus* CECT49 en diferentes hospedadores de *E. coli*. Las electroforesis de proteínas de M1 y M2, mediante SDS-PAGE, muestran la presencia de proteína a 43kDa en la fracción soluble, pero también en el precipitado, poniendo de manifiesto la necesidad de optimizar las condiciones de expresión.
5. Por último, se evaluó la posible actividad de las Endoglucanasas M1 y M2 de *G. stearothermophilus* CECT49 presente en la fracción soluble de las células recombinantes de *E. coli* sobre el sustrato carboximetil celulosa. La mayor actividad para M1 se encontró en la cepa C43, mientras que para M2 la mejor cepa de expresión fue BL21.

8. BIBLIOGRAFÍA

- Aarstad, O., Heggset, E.B., Pedersen, I.S., Bjørnø, S.H., Syverud, K., Strand, B.L. (2017). Mechanical Properties of Composite Hydrogels of Alginate and Cellulose Nanofibrils. *MDPI*. 9 (378): 2-19.
- Ávila, J.T., De Vera, A.M., Hernández, C., y Martín Vasallo, P. (1995). Análisis del DNA. II: Clonar, secuenciar y PCR. *Nefrología*. 15(2): 31-38.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215: 403-410.
- Altschul, S.F., Madden, T. L., Schäffer, A. A., Zheng Zhang, J. Z., Miller, W., Limpan, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Volumen 25(17)*: 3389–3402.
- Anderson, S., Kurland, C. (1990). Codon Preferences in Free-Living Microorganisms. *Microbiol Rev.* 54: 198-210.
- Bajpai, P. (1999). Application of enzymes in the pulp and paper industry. *Biotechnol. Prog.* 15: 147-157.
- Bedford, M.R., Classen, H.L. (1992). The influence of dietary xylanase on intestinal viscosity and molecular weight distribution of carbohydrates in rye-fed-broiler chick. *Biotechnol. Prog.* 7: 361-370.
- Berlemont, R., Gerday, C. (2011). *Comprehensive Biotechnology*. ISBN: 9780080885049.
- Bhalla, A., Kainth, A.S., Sani, R.K. (2013). Draft genome sequence of lignocellulose-degrading thermophilic bacterium *Geobacillus* sp. strain WSUCF1. *Genome Announc.*1(4): e00595-13.
- Bok, J.D., Yernool, D.A., Eveleigh, D.E. (1998). Purification, Characterization, and Molecular Analysis of Thermostable Cellulases CelA and CelB from *Thermotoga neapolitana*. *Appl. Environ. Microb.* 64 (12): 4774–4781.
- Boratyn, G.M., Thierry-Mie, J., Thierry-Mieg, D., Busby, B., Madden, T.L. (2019). Magic-BLAST, an accurate RNA-seq aligner for long and short reads. *BMC Bioinformatics*. 20(1): 405.
- Carmignotto, G., Azzozzi, A. (2019). On the expression of recombinant Cas9 protein in *E. coli* BL21 and BL21(DE3). *Biotechnol.* 306: 62-70.
- Copley, S.D. (2009). Evolution of Efficient Pathways for Degradation of Anthropogenic Chemicals. *Nat. Chem. Biol.* 5(8): 559–566.
- Dass, M.J.A., Nijse, B., Van de Weijer, A.H.P., Groenendaal, B.W.A. J., Janssen, F., Van der Oost, J., Van Kranenburg, R. (2018). Engineering *Geobacillus thermodenitrificans* to introduce cellulolytic activity; expression of native and heterologous cellulase genes. *BMC Biotechnol.* 18(42): 1-13.
- Dumon-Seignovert, L., Cariot, L., Vuillard, L. (2004). The toxicity of recombinant proteins in *Escherichia coli*: a comparison of overexpression in BL21 (DE3), C41(DE3), and C43(DE3). *Protein Expres Purif.* 37: 203–206.
- Eduardo Espinosa, V. (2019). Lignanofibras de celulosa (Infc) a partir de residuos agroindustriales madereos. [Tesis de licenciatura]. Universidad de Córdoba.
- Frumkiba, I., Lajoie, M., Gregg, C., Hornung, G., Church, G., Pipel, Y. (2018). Codon usage of highly expressed genes affects proteome-wide translation efficiency. *PNAS*. 11. E4940-E4949.
- Gariyban, L., Avashia, N. (2013). Research Techniques Made Simple: Polymerase Chain Reaction (PCR). *NIH*. 133(3): 1-3.
- González, A., Fillat, M.F. (2018). Aspectos metodológicos de la expresión de proteínas recombinantes en *Escherichia coli*. *REB*. 37(1): 14-27.
- Grant, S., Jessee, J., Bloom, F., Hanahan, D. (1990). Differential plasmid Rescue From Transgenic Mouse DNAs *Escherichia coli* Methylation-Restriction Mutants. *Proc. Natl Acad Sci USA*. 87: 4645-4649.

- Hanahan, D. (1985). Techniques for transformation of *E. coli*. DNA cloning: A practical approach. Ed. Glover. D.M. ISBN: 097946187.
- Horn, S. J., Vaaje-Kolstad, Svein Jarl, G., Westereng, B., Eijsink, V. (2012). Novel enzymes for the degradation of cellulose. *Biotechnol Biofuels*. 5(45): 1-4.
- Jantasaeng, O., Thaenkudrua, P., Tan J.S., Phapugrangkul, P. (2019). Expression of the Endoglucanase Gene in *Escherichia coli* for the Second-generation Bioethanol Production. *J. Food Appl. Biosc.* 7: 27-35.
- Liu, Y., Guo, H., Wu, Y., Qin, W. (2018) Purification and characterizations of a novel recombinant *Bacillus velezensis* endoglucanase by aqueous two-phase system. *Bioresour. Bioprocess.* 5:19 OpenAcces. 5(45): 1-4.
- Long, L., Tian, D., Zhai, R., Li, X., Zhang, Y., Hu, J., Wang, F., Saddler, J. (2018). Thermostable xylanase-aided two-stage hydrolysis approach enhances sugarrelease of pretreated lignocellulosic biomass. *Bioresour Technol.* 257: 334-338.
- Martínez-Rodríguez, S., Las Heras Vazquez, F.J., Clemente Jiménez J.M., Rodríguez Vico, F. (2004). Biochemical characterization of a novel hydantoin racemase from *Agrobacterium tumefaciens* C58. *Biochimie.* 86(2): 77-81.
- Megumi Miyano, K.T., Shu Ishikawa, K.M., Arribas, A.M., Wilfried, J. J.M., y Ken-Ichi, Y. (2018). A novel method for transforming the thermophilic bacterium *Geobacillus kaustophilus*. *Microb. Cell Fact.* 17(127): 2-4.
- Melchor Partida, G.N. (2014). Obtención de enzimas recombinantes para biorremediación. [Tesis doctoral]. Universidad autónoma de Nayarit.
- Meselson, M., Yuan, R. (1968). RNA Restriction Enzyme from *E. coli*. *Nature.* 217: 1110-1114.
- Miroux, J.B., Walker, E. (1996). Over-production of Proteins in *Escherichia coli*: Mutant Hosts that Allow Synthesis of some Membrane Proteins and Globular Proteins at High Levels. *J. Mol Biol.* 260: 289-298.
- Mullis, K (1991). The polymerase chain reaction in an anemic mode: How to avoid cold oligodeoxyribonucleic acid fusion. *PCR Methods Appl.* 1: 1-5.
- Nitrigual Matamala, A.W., (2012). Determinación de los parámetros de Resistencia Térmica de *Geobacillus stearothermophilus* ATCC7953 bajo condiciones de Calentamiento No Isotérmico [Tesis de licenciatura]. Universidad Austral de Chile.
- Ng, I.S., Li, C.W., Yeh, Y.F., Chen, P.T., Chir, J.L., Ma, C.H., Yu, S.M., Ho, T.H.D., Tong, C.G. (2009). A novel endo-glucanase from the thermophilic bacterium *Geobacillus* sp. 70PC53 with high activity and stability over a broad range of temperatures. *Extremophiles* 13: 425-435.
- Paes, G., O'Donohue, M.J. (2006). Engineering increased thermostability in the thermostable GH-11 xylanase from *Thermobacillus xylanilyticus*. *J. Biotechnol.* 125: 338-350.
- Patel, B. (2009). Simple, Fast, and Efficient Cloning of PCR Products with TOPO® Cloning Vectors. *Bio Techniques.* 46(6): 2-6.
- Perera, J., Tormo, Antonio., García, J.L., (2002). Ingeniería Genética; Preparación, análisis, manipulación y clonaje de DNA. Ed: Síntesis. Vol 1. ISBN: 9788477389644.
- Pfeiffer, D., WhanI, A., Jendrossek, D. (2011). Identification of a multifunctional protein, PhaM, that determines number, surface to volume ratio, subcellular localization, and distribution to daughter cells of poly(3-hydroxybutyrate), PHB, granules in *Ralstonia eutrophia* H16. *Mol. Microbiol.* 82(4): 936-951.
- Pingoud, A., Jeltsch, A. (2001). Structure and function of type II restriction endonucleases. *Nucleic Acids Res.* 29(18): 3705-3727.
- Prescott, L.M., Harley, J.P. and Klein, D.A. (2002). *Microbiology: Food and Industrial Microbiology*. ISBN: 9786071503398.
- Rastogi, G., Bhalla, A. (2010). Characterization of thermostable cellulases produced by *Bacillus* and *Geobacillus* strains. *Bioresour. Technol.* 101: 8798-806.

- Sambrook, J., Fritish, E., Maniatis, T. (2012). *Molecular Cloning*. Cold Spring Harbor Lab Pres. Volume 2: 206-227. ISBN: 1936113422.
- Santosh, T., Jitendra, M., Naveen, A., Priya, M., Hui, L., O'Hair, J., Sarabjit, B., Suping, Z. (2020). Microbial cellulolytic enzymes: diversity and biotechnology with reference to lignocellulosic biomass degradation. *Rev. Environ. Sci. Biotechnol.* 19: 621–648.
- Sharp, P., Li, W. (1987). The codon adaptation index a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 15(3): 1281-1295.
- Shuman, S. (1991). Recombination Mediated by Vaccinia Virus DNA Topoisomerase I in *Escherichia coli* is Sequence Specific. *Proc. Natl. Acad. Sci. USA.* 88: 10104-10108.
- Shuman, S. (1994). Novel Approach to Molecular Cloning and Polynucleotide Synthesis Using Vaccinia DNA Topoisomerase. *J. Biol. Chem.* 269: 32678-32684.
- Sonal, B., Reeta, K. (2013). Bioremediation: A sustainable Tool for Environmental Management. *Annu. Res. Rev. Biol.* 3(4): 974-993.
- Studier, F., Moffatt, B.A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* 189: 113-130.
- Suzuki, H., y Ken-Ichi, Y. (2012). Genetic Transformation of *Geobacillus kaustophilus* HTA426 by Conjugative Transfer of Host-Mimicking Plasmids. *J. Mol. Biol.* 22(9): 1279–1287.
- Tai, S.K., Lin, H.P., Kuo, J., Liu, J.K. (2004). Isolation and characterization of a cellulolytic *Geobacillus thermoleovorans* T4 strain from sugar refinery wastewater. *Extremophiles.* 8: 345–349.
- Therriault, G., Roy, P.H., Howard, K.A., Benner, K.A., Brooks, J.E., Waters, A.F., Gingeras, T.R. (1985). Nucleotide sequence of the PaeR7 restriction/modification system and partial characterization of its protein products. *Nucleic Acids Res.* 13 (23): 8841-8461.
- Turner, P., Mamo, G. y Karlsson, E.N. (2007). Potential and utilization of thermophiles and thermostable enzymes in biorefining. *BMC.* 6(9): 1-23.
- Van Dillewijn, P., Caballero, A., Paz, J.A., González-Pérez, M.M., Oliva, J. M., Ramos, J.L. (2007). Bioremediation of 2,4,6-trinitrotoluene under field conditions. *Environ. Sci. Technol.* 41: 1378-1383.
- Vanatalu, K., Paalme, T., Vilu, R., Jüneman, R., May, R., Rühl, M., Wadzack, J., Nierhaus, K. (1993). Large-scale preparation of fully deuterated cell components Ribosomes from *Escherichia coli* with high biological activity. *Eur. J. Biochem.* 216: 315-321.
- Verma, P.K., Bhardwaj, N.K., Singh, S.P. (2017). Upgradation of Recycled Pulp Using Endoglucanase Enzyme Produced by *Pycnoporus sanguineus*. *J. Chem. Eng.* 36(2): 191-201.
- Viteri Florez, P.A., Castillo Guerra, D.A., Viteri Rosero, S.E. (2015). Capacidad y diversidad de bacterias celulolíticas aisladas de tres hábitats tropicales en Boyacá. Colombia. ISSN: 0120-2812: 362-363.
- Watson, R.J., Schildkraut, I., Qiang, B.Q., Martin, S.M., Visentin, L.P. (1982). *NdeI*: a restriction endonuclease from *Neisseria denitrificans* which cleaves DNA at 5'-CATATG-3' sequences. *FEBS Lett.* 150(1): 114-116.
- Wegerer, A., Sun, T., Altenbuchner, J. (2008). Optimization of an *E. coli* L-rhamnose-inducible expression vector: test of various genetic module combinations. *BMC Biotechnol.* 8(2): 1-12.
- Yadava, R.S., Kumar, R., Yadava, P.K. (2005). Expression of lex A targeted ribozyme in *Escherichia coli* BL21 (DE3) cells. *Mol. Cell Biochem.* 271(1–2): 197–203.

ANEXOS

ANEXO 1

Secuencia de partida para la Endoglucanasa M1 de *Geobacillus sp.* WSUCF1, con número de accesoión ATCO01000152.1.

ATGCGCAATGATTCATTTTGCCAATCGTCAGCCTTTTCAAGTATAATAAAAAAGTGGACAAAACCGGAAAGGGGTGCC
 AAGATGGCGAAGTTGGACGAAACGTTGACGATGCTGAAAGCGCTGACCGATGCGAAAGGCGTCCCGGGCAATGAGCGG
 GAAGCGCGGATGTGATGAAGACATACATAGCTCCGTATGCGGATGAAGTGACAACGGATGGCCTCGGCAGCTTGATC
 GCCAAAAAAGAAGGAAATCGGGCGGGCCGAAAGTGATGATCGCCGGCCATTTGGATGAAGTCGGCTTTATGGTGACG
 CAAATCGACGACAAAGGATTCATCCGCTTCCAAACGCTTGCGGATGGTGGAGCCAAGTAATGCTCGCCCAGCGCGTG
 ACGATCGTGACGAAAAAAGCGACATCACCGCGCTCATCGGTTTGAAGCCGCCGCACATTCGCGCCGGAGGCGCGC
 AAAAAACCGGTGGAATCAAAGATATGTTTATCGACATCGCGCGACAAGCCGCGAGGAAGCGATGGAGTGGGGCGTC
 CGCCCGGGCGATATGATCGTGCCGATTTTGAATTTACGGTATTGAACAATGAAAAATGCTGCTCGCGAAAGCATGG
 GACAACCGGATCGGCTGTGCGGTGCGCATCGATGTGCTAAGCAGCTGAAAGGCGTCGACCATCCAAACACGGTATAC
 GCGTTCGGCACGGTGCAGGAAGAAGTTCGGCTTGCGCGGGCGCGCACGGCCGCCAATTCATTCAGCCGGATATCGCC
 TTTGCGGTGGATGTGCGCATTTGCCGGGATACACCGGGCGTGTGCGAAAAAGAAGCGATGGGCAAACTCGGCGCCGGC
 CCGCACATCGTCTGTACGACGCGACGATGGTGTGCGACCGCGGCTTGCGGAATTCGTATCGAAGTGGCGGAAGAG
 CTGAACATTCGGTACCATTTTGATGCCATGCCGGGCGGGTACGGATGCGGGAGCGATTATTTAACGGGCAGCGGC
 GTTCCGTGCTCAGATCGCCATCCCAGCGCTACATCCACTCGCACGGCCCATTTTGCATCGCGACGACTACGAG
 AACACGGTCAAGCTGCTTGTGCAAGTGATCAAACGGCTTGACCGGACAAAGTGAAACAACCTGACGTTTGACGAA TAA

Figura 1: Secuencia de partida para la Endoglucanasa M1 de Geobacillus sp. WSUCF1, con número de accesoión ATCO01000152.1. Se marca en amarillo el codón de inicio del gen (ATG) y en rosa en el codón de terminación (TAA).

Secuencia de partida para la Endoglucanasa M2 de *Geobacillus sp.* WSUCF1, con número de accesoión ATCO01000242.1.

ATGATGAATGAGGACACGCTGCAATTATTCCAAACGTTGACCGAGCTTCCAGGCGCACCGGGTTATGAACATCCGGTG
 CGGCAGTTTATGCGCAAAGAGCTCGAGAAATACGCCGATGACATTGTGCAAGACCGCCTCGGGAGCCTTTTCGGCGTC
 AAACGCGCGGATGAGGCGGGCCCGACGGTGTGTTGCGGCCATATGGATGAAGTCGCATTTATGGTGACGGCGATC
 ACCGACAAATGGCATGATTCGGTTTCAGCCGCTCGGCGGCTGGTGGGATCAAGTGTCTCTCCAGCGCGTGCAGATT
 ATCACGGACGGAGGGCCGGTTCATCGGCGTATCGGGTTCGATTCGCCCGCATTTGCTTGATGAGGAGCAGCGCAAGAAA
 CCAATGGAGATGAAAAACATGTTGATTGACATCGGAGCGGAAAGCCGTGAAGATGCCGAACGAGTCGGCGTCCGCCCC
 GGTGAGCCGATCGTGCCGGTTCAGCCATTACGGTGTGGCCAACGGAAAAACCGTGATGGCAAAAGCGTGGGACAAT
 CGGTTTGGCTGCGGGTTGGCGATCGAGCTGTTAAAAGAGCTGAACGGCGAAACAGTGCCGAAACGCTTTGTACGCTGGA
 GCGACTGTGCAGGAAGAGGTTCGGGCTGCGTGGGGCCAGACGGCGGCGAGCATGATCAACCCGATATCTTTTCGCC
 TTGGAAGCCAGCCCGCCAAAGACATGTCCGGCGACAAGCAGCGCTTCGGCCATATCGGCAAGGGGCGCTCATCCGC
 TTGTACGACCGGACGATGATTACGACCGCGCATGCGCGAGTTTGTGCTTGACACCGCGGAAACGATTGGCGTCCCG
 TATCAGTTTTTCATTTCCCGAGCGGGGAAACCGACCGCGGGTCCATATCGCCAACCGCGCGTCCGTCGGCG
 GTGATCGGCATTTGCGCCGCTACATTACATCGCATGCGTTCGATCATTATCGATGATTACGCGCGGCGCAACAG
 CTCATCGTTGAGCTCGTCAAACGGTTCGACCGGACGACGGTTCGAGACGATTCGCCAAAACAGCTGA

Figura 2: Secuencia de partida para Endoglucanasa M2 de Geobacillus sp. WSUCF1, con número de accesoión ATCO01000242.1. Se marca en amarillo el codón de inicio del gen (ATG) y en rosa en el codón de terminación (TGA).

Comparación de las secuencias de las Endoglucanasas M1 de *Geobacillus kaustophilus* NRBC102445 y *Geobacillus kaustophilus* HTA426.

EnGluM1GkaNRBC102445	gtggcgaatgattcattttgccaatcgtcagccttttcaagtataataaaaaagtggaca	60
EnGluM1GkaHTA426	gtggcgaatgattcattttgccaatcgtcagccttttcaagtataataaaaaagtggaca	60
EnGluM1GkaNRBC102445	*****	
EnGluM1GkaNRBC102445	aaaccggaaaggggtgccaaagtggcgaagtggacgaaacgttgacgatgctgaaagcg	120
EnGluM1GkaHTA426	aaaccggaaaggggtgccaaagtggcgaagtggacgaaacgttgacgatgctgaaagcg	120
EnGluM1GkaNRBC102445	*****	
EnGluM1GkaNRBC102445	ctgaccgatgcaaaagcgtcccgggcaatgagcgggaagcgcgcgatgtgatgaagaca	180
EnGluM1GkaHTA426	ctgaccgatgcaaaagcgtcccgggcaatgagcgggaagcgcgcgatgtgatgaagaca	180
EnGluM1GkaNRBC102445	*****	
EnGluM1GkaNRBC102445	tacatagctccgtatgctggatgaagtgacaacggatggcctcggcagcttgattgcaaaa	240
EnGluM1GkaHTA426	tacatagctccgtatgctggatgaagtgacaacggatggcctcggcagcttgattgcaaaa	240
EnGluM1GkaNRBC102445	*****	
EnGluM1GkaNRBC102445	aaagaaggaaatcgggcggcgcaaaagtgatgatcgccggccatttggatgaagtggc	300
EnGluM1GkaHTA426	aaagaaggaaatcgggcggcgcaaaagtgatgatcgccggccatttggatgaagtggc	300
EnGluM1GkaNRBC102445	*****	
EnGluM1GkaNRBC102445	tttatggtgacgcaaatcgacgacaagagattcatccgcttccaaacgcttgccggatgg	360
EnGluM1GkaHTA426	tttatggtgacgcaaatcgacgacaagagattcatccgcttccaaacgcttgccggatgg	360
EnGluM1GkaNRBC102445	*****	
EnGluM1GkaNRBC102445	tggagccaagtgatgctcgccagcgcgtgacgatcgtgacgaaaaaagggacatcacc	420
EnGluM1GkaHTA426	tggagccaagtgatgctcgccagcgcgtgacgatcgtgacgaaaaaagggacatcacc	420
EnGluM1GkaNRBC102445	*****	
EnGluM1GkaNRBC102445	ggcgtcatcggttcgaagccgcgcacattctgcccggagggcgcgcaaaaaaccggtg	480
EnGluM1GkaHTA426	ggcgtcatcggttcgaagccgcgcacattctgcccggagggcgcgcaaaaaaccggtg	480
EnGluM1GkaNRBC102445	*****	
EnGluM1GkaNRBC102445	gaaatcaagatattgctcatcgacatcggcgcgacaagccgcgaggaagcgatggagtgg	540
EnGluM1GkaHTA426	gaaatcaagatattgctcatcgacatcggcgcgacaagccgcgaggaagcgatggagtgg	540
EnGluM1GkaNRBC102445	*****	
EnGluM1GkaNRBC102445	ggcgtccgcccggcgatgatgatcgctcgtatcttgaatttacggattgaaacaatgaa	600
EnGluM1GkaHTA426	ggcgtccgcccggcgatgatgatcgctcgtatcttgaatttacggattgaaacaatgaa	600
EnGluM1GkaNRBC102445	*****	
EnGluM1GkaNRBC102445	aaaatgctgctcgcgaaagcatgggacaacggatcggctgtgcccggcgcacatcgatgtg	660
EnGluM1GkaHTA426	aaaatgctgctcgcgaaagcatgggacaacggatcggctgtgcccggcgcacatcgatgtg	660
EnGluM1GkaNRBC102445	*****	
EnGluM1GkaNRBC102445	ctcaagcagtgaaagcgctcgaccatccaaacacggatatacggcgtggcagcggcgcaag	720
EnGluM1GkaHTA426	ctcaagcagtgaaagcgctcgaccatccaaacacggatatacggcgtggcagcggcgcaag	720
EnGluM1GkaNRBC102445	*****	
EnGluM1GkaNRBC102445	gaagaagtccgcttgccgcccgcgcacgcgcccgaattcattcagccagatatacgcg	780
EnGluM1GkaHTA426	gaagaagtccgcttgccgcccgcgcacgcgcccgaattcattcagccagatatacgcg	780
EnGluM1GkaNRBC102445	*****	
EnGluM1GkaNRBC102445	tttgctgttgacgtcggcattgcccggcgcacgcgcccgggtgtcggaaaaagaagcgatg	840
EnGluM1GkaHTA426	tttgctgttgacgtcggcattgcccggcgcacgcgcccgggtgtcggaaaaagaagcgatg	840
EnGluM1GkaNRBC102445	*****	
EnGluM1GkaNRBC102445	ggcaaacctcggcgcggcccgcacatcgtcttgtacgacgcaacgatgggtgctgcaccgc	900
EnGluM1GkaHTA426	ggcaaacctcggcgcggcccgcacatcgtcttgtacgacgcaacgatgggtgctgcaccgc	900
EnGluM1GkaNRBC102445	*****	
EnGluM1GkaNRBC102445	ggcttgccgcaatttgtcatcgaagtggcgaagagctcaacattccgtaccattttgat	960
EnGluM1GkaHTA426	ggcttgccgcaatttgtcatcgaagtggcgaagagctgaacattccgtaccattttgat	960
EnGluM1GkaNRBC102445	*****	
EnGluM1GkaNRBC102445	gccatgccaggcggcggtacggacgcccggagcgattcacttaaccggcatcggcgttccg	1020
EnGluM1GkaHTA426	gccatgccaggcggcggtacggacgcccggagcgattcacttaaccggcatcggcgttccg	1020
EnGluM1GkaNRBC102445	*****	
EnGluM1GkaNRBC102445	tcgctcagcattgccaatcccgcgcgctacatccactcgcacgcccattttgcaaccgc	1080
EnGluM1GkaHTA426	tcgctcagcattgccaatcccgcgcgctacatccactcgcacgcccattttgcaaccgc	1080
EnGluM1GkaNRBC102445	*****	
EnGluM1GkaNRBC102445	gacgactacgaaaaacacggtaagctgcttggtaggtgatcaaacggcttgacgcccagc	1140
EnGluM1GkaHTA426	gacgactacgaaaaacacggtaagctgcttggtaggtgatcaaacggcttgacgcccagc	1140
EnGluM1GkaNRBC102445	*****	
EnGluM1GkaNRBC102445	aaagtgaacaactgacgtttgacgaataa 1170	
EnGluM1GkaHTA426	aaagtgaacaactgacgtttgacgaataa 1170	
EnGluM1GkaNRBC102445	*****	

Figura 3: Comparación de las secuencias de la Endoglucanasa M1 de *Geobacillus kaustophilus* NRBC102445 y *Geobacillus kaustophilus* HTA426 que son iguales para M1.

Comparación de las secuencias de las Endoglucanasas M1 de Geobacillus kaustophilus NRBC102445 y Geobacillus stearothermophilus B5.

Table with 3 columns: Sequence identifier (EnGluM1GkaNRBC102445 / EnGluM1GsteB5), DNA sequence (with non-matching nucleotides highlighted in green), and position number (60, 120, 180, 240, 300, 360, 420, 480, 540, 600, 660, 720, 780, 840, 900, 960, 1020, 1080, 1140, 1170).

Figura 4: Comparación de las secuencias de las Endoglucanasas M1 de Geobacillus kaustophilus NRBC102445 y Geobacillus stearothermophilus B5, resaltando en verdes los nucleótidos no coincidentes.

Comparación de las secuencias de las Endoglucanasas M2 de *Geobacillus kaustophilus* NRBC102445 y *Geobacillus kaustophilus* HTA42.

EnGluM2GkaNRBC102445	atgaataaggacacgctgcaattattccaaacgcttgaccgagcttccaggcgccacgggt	60
EnGluM2GkaHTA426	atgaataaggacacgctgcaattattccaaacgcttgaccgagcttccaggcgccacgggt	60

EnGluM2GkaNRBC102445	tatgaacatccggtgcggcagtttatgcgccaagagctcgcaatacggcgatgacatt	120
EnGluM2GkaHTA426	tatgaacatccggtgcggcagtttatgcgccaagagctcgcaatacggcgatgacatt	120

EnGluM2GkaNRBC102445	gtgcaagaccgctcgggagccttttcggcgtcaaacgcggtgatgaggcggcccgacg	180
EnGluM2GkaHTA426	gtgcaagaccgctcgggagccttttcggcgtcaaacgcggtgatgaggcggcccgacg	180

EnGluM2GkaNRBC102445	gtgatggtcgccggccatattggatgaagtgcgatttatggtgacagcgatcacggatcac	240
EnGluM2GkaHTA426	gtgatggtcgccggccatattggatgaagtgcgatttatggtgacagcgatcacggatcac	240

EnGluM2GkaNRBC102445	ggcatgattcgtttccagccgctaggaggctggtgggatcaagtgcctgctcccagcgc	300
EnGluM2GkaHTA426	ggcatgattcgtttccagccgctaggaggctggtgggatcaagtgcctgctcccagcgc	300

EnGluM2GkaNRBC102445	gtgcagattatcaaggacgggagcggcgtcatcgcgctgatcggtcgattccgcccagc	360
EnGluM2GkaHTA426	gtgcagattatcaaggacgggagcggcgtcatcgcgctgatcggtcgattccgcccagc	360

EnGluM2GkaNRBC102445	ttgcttgatgaggagcagcgcaaaaaacggatggagatcaaaaaacatggtgattgacatc	420
EnGluM2GkaHTA426	ttgcttgatgaggagcagcgcaaaaaacggatggagatcaaaaaacatggtgattgacatc	420

EnGluM2GkaNRBC102445	ggagcggaaaagccgtgaagatgccgaaacgagtcggcattccgcccggccagccgatcgtg	480
EnGluM2GkaHTA426	ggagcggaaaagccgtgaagatgccgaaacgagtcggcattccgcccggccagccgatcgtg	480

EnGluM2GkaNRBC102445	ccggtcagcccgtttactttgcttgccaaacggaaaaacggatggcaaaaagcgtgggac	540
EnGluM2GkaHTA426	ccggtcagcccgtttactttgcttgccaaacggaaaaacggatggcaaaaagcgtgggac	540

EnGluM2GkaNRBC102445	aatcggtttgctgctgagattggcgatcgagctgttaaaagagctgaacggcgaaacagtg	600
EnGluM2GkaHTA426	aatcggtttgctgctgagattggcgatcgagctgttaaaagagctgaacggcgaaacagtg	600

EnGluM2GkaNRBC102445	ccgaacgtcttgtagcaggggacccgtgacgaaagaggtcggctgctggtgggcccag	660
EnGluM2GkaHTA426	ccgaacgtcttgtagcaggggacccgtgacgaaagaggtcggctgctggtgggcccag	660

EnGluM2GkaNRBC102445	acggcggcttcgatgatcaaccggacatctttttgcccctggaagcaagcccggccaac	720
EnGluM2GkaHTA426	acggcggcttcgatgatcaaccggacatctttttgcccctggaagcaagcccggccaac	720

EnGluM2GkaNRBC102445	gatatgtccggcgacaagcaggcgttcggccatattcgcaaaaggagcctcatccgcttg	780
EnGluM2GkaHTA426	gatatgtccggcgacaagcaggcgttcggccatattcgcaaaaggagcctcatccgcttg	780

EnGluM2GkaNRBC102445	tacgaccggacgatgattacgcaccggcgtgacgcaatttgcttgatagcggcggaa	840
EnGluM2GkaHTA426	tacgaccggacgatgattacgcaccggcgtgacgcaatttgcttgatagcggcggaa	840

EnGluM2GkaNRBC102445	acgatcggcgtcccgtaccagtttttccatttcccaggcggcggaaaccgacccggggcg	900
EnGluM2GkaHTA426	acgatcggcgtcccgtaccagtttttccatttcccaggcggcggaaaccgacccggggcg	900

EnGluM2GkaNRBC102445	gtgcatacggcaaccggcgtgcccgtcggcgggtgatcgccatttgcgcccgtacatt	960
EnGluM2GkaHTA426	gtgcatacggcaaccggcgtgcccgtcggcgggtgatcgccatttgcgcccgtacatt	960

EnGluM2GkaNRBC102445	catacgcattcgtcgatcattcatgctgacgattacgcccggcgaacacagctcatcgtt	1020
EnGluM2GkaHTA426	catacgcattcgtcgatcattcatgctgacgattacgcccggcgaacacagctcatcgtt	1020

EnGluM2GkaNRBC102445	gagctcgtcaaacgggtgcgaccggacaacggctgagacgattcgccaaaacagctga	1077
EnGluM2GkaHTA426	gagctcgtcaaacgggtgcgaccggacaacggctgagacgattcgccaaaacagctga	1077

Figura 5: Comparación de las secuencias de las Endoglucanasas M2 de *Geobacillus kaustophilus* NRBC102445 y *Geobacillus kaustophilus* HTA426, observándose que son totalmente iguales.

Comparación de las secuencias de las Endoglucanasas M2 de Geobacillus stearothermophilus B5 y Geobacillus kaustophilus NRBC102445.

Table with 3 columns: EnGluM2GsteB5, EnGluM2GkaNRBC102445, and sequence alignment with line numbers (60, 120, 180, 240, 300, 360, 420, 480, 540, 600, 660, 720, 780, 840, 900, 960, 1020, 1075, 1077).

Figura 6: Comparación de las secuencias de las Endoglucanasas M2 de Geobacillus stearothermophilus B5 y Geobacillus kaustophilus NRBC102445, resaltando en verde los nucleótidos no coincidentes.

Enzimas que no cortan al fragmento de *Geobacillus kaustophilus* NRBC102445 para Endoglucanasa M1.

AatI, Acc113I, Acc16I, Acc65I, AccB7I, AccIII, AclNI, AflIII, AflIII, AhdI, Alw44I, AlwNI, AocI, ApaI, ApaLI, AscI, AseI, AsnI, Asp700I, Asp718I, AspEI, AspI, AtsI, AviIII, AvrII, BalI, **BamHI**, BbrPI, BbuI, BcgI, BfrI, BglII, BlnI, BlpI, BpmI, Bpu1102I, BsaAI, BsaI, BsaMI, Bse21I, BseAI, BseRI, BsiI, BsiMI, BsiWI, BsmBI, BsmI, Bsp120I, Bsp13I, Bsp1407I, Bsp1720I, Bsp19I, BspEI, BspHI, BspLU11I, BspMI, BspTI, BsrGI, BssSI, BssTI, Bst98I, BstEII, BstI, BstPI, BstSFI, BstSNI, BstX2I, BstYI, Bsu36I, CciNI, CelII, CpoI, CspI, CvnI, DraI, DraII, DraIII, DrdI, Eam1105I, EclHKI, Eco105I, Eco130I, Eco147I, Eco255I, Eco31I, Eco57I, Eco72I, Eco81I, Eco91I, EcoNI, EcoO109I, EcoO65I, **EcoRI**, EcoT14I, EcoT22I, ErhI, Esp1396I, Esp3I, FauNDI, FseI, FspI, GsuI, **HindIII**, HpaI, Kpn2I, KpnI, MfeI, MflI, MluI, MluNI, Mph1103I, MroI, MscI, MspCI, MunI, Mval269I, NcoI, **NdeI**, NheI, NotI, NsiI, NspI, PacI, PaeI, PaeR7I, Pfl23II, PflMI, PmaCI, Pme55I, PmeI, PmlI, Ppu10I, PpuMI, PshAI, PshBI, Psp5II, PspEI, PspLI, PspOMI, **PstI**, PstNHI, RcaI, RsrII, SbfI, ScaI, SexAI, SfcI, SfiI, Sfr274I, SgfI, SmiI, SnaBI, SpeI, SphI, SplI, Sse8387I, SseBI, SspBI, SspI, StuI, StyI, SunI, SwaI, Tth111I, Van91I, Vha464I, VneI, VspI, XbaI, XcmI, **XhoI**, XhoII, XmnI, Zsp2I

Figura 7: Enzimas que no cortan al fragmento de Geobacillus kaustophilus NRBC102445, para Endoglucanasa M1.

Enzimas que no cortan al fragmento de *Geobacillus stearothermophilus* B5 para la Endoglucanasa M1.

AatI, AatII, Acc113I, Acc16I, Acc65I, AccB7I, AccIII, AclNI, AflIII, AflIII, AhdI, Alw44I, AlwNI, AocI, ApaI, ApaLI, AscI, AseI, AsnI, Asp700I, Asp718I, AspEI, AspI, AtsI, AviIII, AvrII, BalI, **BamHI**, BanII, BbrPI, BbuI, BfrI, BglII, BlnI, BlpI, BpmI, Bpu1102I, BsaAI, BsaI, BsaMI, Bse21I, BseAI, BseRI, BsiI, BsiMI, BsiWI, BsmBI, BsmI, Bsp120I, Bsp13I, Bsp1407I, Bsp1720I, Bsp19I, BspEI, BspHI, BspLU11I, BspMI, BspTI, BsrGI, BssSI, BssTI, Bst98I, BstEII, BstI, BstPI, BstSFI, BstSNI, BstX2I, BstYI, Bsu36I, CciNI, CelII, CpoI, CspI, CvnI, DraI, DraII, DraIII, Eam1104I, Eam1105I, EarI, Ecl136II, EclHKI, Eco105I, Eco130I, Eco147I, Eco24I, Eco255I, Eco31I, Eco57I, Eco72I, Eco81I, Eco91I, EcoICRI, EcoNI, EcoO109I, EcoO65I, EcoT14I, EcoT22I, ErhI, Esp1396I, Esp3I, FauNDI, FriOI, FseI, FspI, GsuI, **HindIII**, HpaI, Kpn2I, KpnI, Ksp632I, MfeI, MflI, MluI, MluNI, Mph1103I, MroI, MscI, MspCI, MunI, Mval269I, NcoI, **NdeI**, NheI, NotI, NsiI, NspI, PacI, PaeI, PaeR7I, Pfl23II, PflMI, PmaCI, Pme55I, PmeI, PmlI, Ppu10I, PpuMI, PshAI, PshBI, Psp124BI, Psp5II, PspEI, PspLI, PspOMI, **PstI**, PstNHI, RcaI, RsrII, **SacI**, SapI, SbfI, ScaI, SexAI, SfcI, SfiI, Sfr274I, SgfI, SmiI, SnaBI, SpeI, SphI, SplI, Sse8387I, SseBI, SspBI, SspI, SstI, StuI, StyI, SunI, SwaI, Tth111I, Van91I, Vha464I, VneI, VspI, XbaI, XcmI, **XhoI**, XhoII, XmnI, Zsp2I

Figura 8: Enzimas que no cortan al fragmento de Geobacillus stearothermophilus B5 para Endoglucanasa M1.

Enzimas que no cortan al fragmento de *Geobacillus kaustophilus* NRBC102445 para Endoglucanasa M2.

AatI, AatII, Acc113I, Acc16I, Acc65I, AccB1I, AccB7I, AccIII, AclNI, AflIII, AgeI, AhdI, Alw44I, AlwNI, AocI, ApaLI, AscI, AseI, AsnI, Asp700I, Asp718I, AspEI, AspI, AtsI, AvIII, AvrII, **BamHI**, BanI, BanIII, BbeI, BbrPI, BbsI, Bbv16II, BcgI, BfrI, BglIII, BlnI, BlpI, BpiI, BpmI, Bpu1102I, Bpu14I, BpuAI, Bsa29I, BsaAI, BsaI, BscI, Bse21I, BseAI, BseCI, BsePI, BshNI, BsiI, BsiMI, BsiWI, Bsp106I, Bsp119I, Bsp13I, Bsp1407I, Bsp1720I, Bsp19I, BspDI, BspEI, BspHI, BspMI, BspTI, BspXI, BsrDI, BsrGI, BssHII, BssSI, BssTI, Bst1107I, Bst98I, BstBI, BstEII, BstI, BstPI, BstSFI, BstSNI, BstX2I, BstXI, BstYI, BstZI, Bsu15I, Bsu36I, CciNI, CelII, Cfr9I, ClaI, CpoI, Csp45I, CspI, CvnI, DraI, DraIII, DrdI, EagI, Eam1105I, EclHKI, EclXI, Eco105I, Eco130I, Eco147I, Eco255I, Eco31I, Eco32I, Eco52I, Eco57I, Eco64I, Eco72I, Eco81I, Eco91I, EcoNI, EcoO65I, **EcoRI**, EcoRV, EcoT14I, EcoT22I, EheI, ErhI, Esp1396I, FseI, FspI, GsuI, **HindIII**, HpaI, KasI, Kpn2I, KpnI, LspI, MfeI, MflI, MluI, MluNI, Mph1103I, MroI, MscI, MspCI, MunI, NarI, NcoI, NheI, NotI, NsiI, NspV, PacI, Paer7I, Pfl123II, PflMI, PinAI, PmaCI, Pme55I, PmeI, PmlI, Ppu10I, PpuMI, PshBI, Psp5II, PspAI, PspALI, PspEI, PspLI, **PstI**, PstNHI, RcaI, RsrII, SapI, SbfI, ScaI, SexAI, SfcI, SfiI, Sfr274I, SfuI, SgfI, SgrAI, **SmaI**, SmiI, SnaBI, SpeI, SplI, SrfI, Sse8387I, SseBI, SspBI, SspI, StuI, StyI, SunI, SwaI, Tth111I, Van91I, Vha464I, VneI, VspI, XbaI, XcmI, **XhoI**, XhoII, XmaI, XmaIII, XmnI, Zsp2I

Figura 9: Enzimas que no cortan al fragmento de *Geobacillus kaustophilus* NRBC102445 para Endoglucanasa M2.

Enzimas que no cortan al fragmento de *Geobacillus stearothermophilus* B5 para Endoglucanasa M2.

AatI, AatII, Acc113I, Acc16I, Acc65I, AccBSI, AclNI, AflIII, AflIII, AhdI, Alw44I, AlwNI, AocI, ApaI, ApaLI, AscI, AseI, AsnI, Asp700I, Asp718I, AspEI, AspI, AtsI, AvIII, AvrII, **BamHI**, BanIII, BbrPI, BfrI, BglIII, BlnI, BlpI, BpmI, Bpu1102I, Bpu14I, Bsa29I, BsaAI, BsaBI, BsaI, BscI, Bse21I, Bse8I, BseCI, BsePI, BseRI, Bsh1365I, BsiWI, BsmBI, Bsp106I, Bsp119I, Bsp120I, Bsp1407I, Bsp1720I, Bsp19I, Bsp68I, BspDI, BspHI, BspLU11I, BspTI, BspXI, BsrBI, BsrBRI, BsrDI, BsrGI, BssHII, Bst1107I, Bst98I, BstBI, BstD102I, BstEII, BstI, BstPI, BstSFI, BstSNI, BstX2I, BstXI, BstYI, Bsu15I, Bsu36I, CciNI, CelII, ClaI, CpoI, Csp45I, CspI, CvnI, DraI, DraII, DraIII, DrdI, Eam1105I, EclHKI, Eco105I, Eco147I, Eco255I, Eco31I, Eco32I, Eco57I, Eco72I, Eco81I, Eco91I, EcoNI, EcoO109I, EcoO65I, **EcoRI**, EcoRV, EcoT22I, Esp3I, FseI, FspI, GsuI, **HindIII**, HpaI, KpnI, LspI, MamI, MflI, MluI, Mph1103I, MspCI, NcoI, NheI, NotI, NruI, NsiI, NspV, PacI, Paer7I, Pfl123II, PmaCI, Pme55I, PmeI, PmlI, Ppu10I, PpuMI, PshAI, PshBI, Psp1406I, Psp5II, PspEI, PspLI, PspOMI, **PstI**, PstNHI, PvuII, RcaI, RsrII, SbfI, ScaI, SexAI, SfcI, SfiI, Sfr274I, SfuI, SgfI, SmiI, SnaBI, SpeI, SplI, SrfI, Sse8387I, SseBI, SspBI, SspI, StuI, SunI, SwaI, Tth111I, Vha464I, VneI, VspI, XbaI, XcmI, **XhoI**, XhoII, XmnI, Zsp2

Figura 10: Enzimas que no cortan al fragmento de *Geobacillus stearothermophilus* B5 para Endoglucanasa M2.

Esquemas de los vectores pTOPO de las Endoglucanasas M1 y M2 de *Geobacillus kaustophilus* CECT4264 y de las Endoglucanasas M1 de *Geobacillus stearothermophilus* CECT43 y 49 en el que aparecen los sitios de corte.

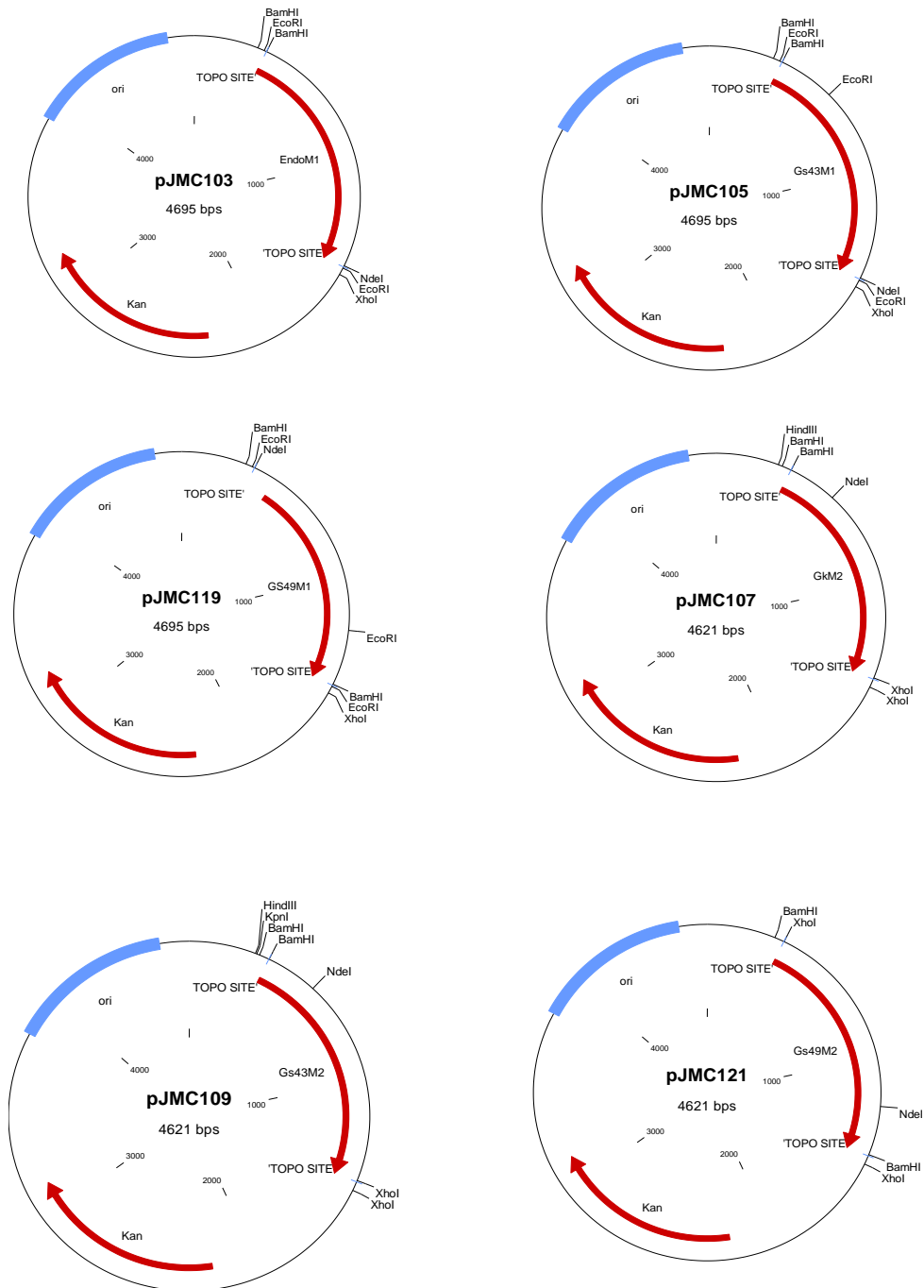


Figura 11: Esquemas del vector TOPO de las Endoglucanasas M1 y M2 de *Geobacillus kaustophilus* CECT4264 y de las Endoglucanasas M1 de *Geobacillus stearothermophilus* CECT43 y 49 en los que aparecen los sitios de corte. Realizado con el programa Clone Manager 9.0.

Esquemas de los vectores de las Endoglucanasas M1 y M2 de *Geobacillus kaustophilus* CECT4264 y *Geobacillus stearothermophilus* CECT43 y 49.

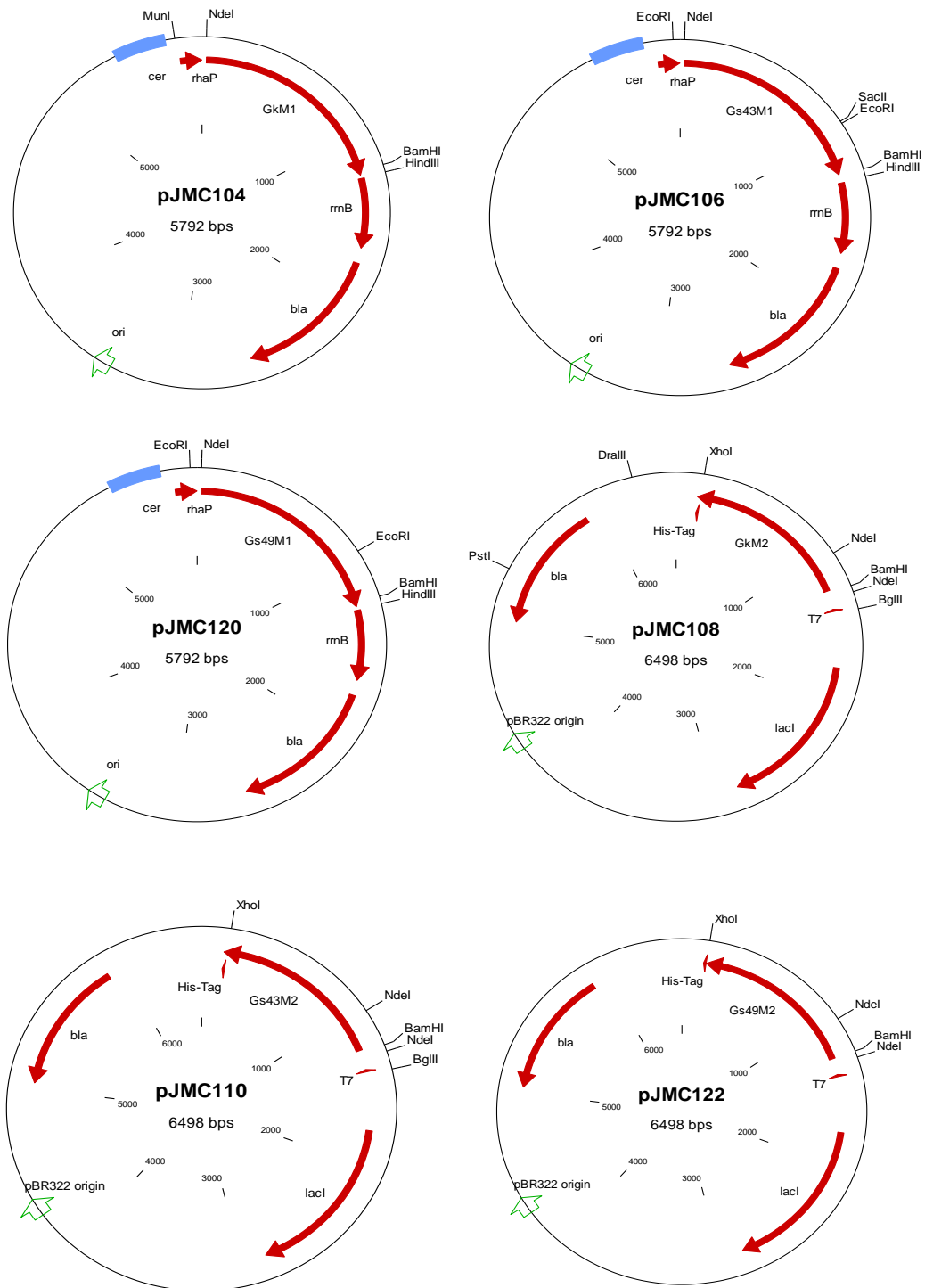


Figura 12: Esquemas de los vectores de las Endoglucanasas M1 y M2 de *Geobacillus kaustophilus* CECT4264 y *Geobacillus stearothermophilus* CECT43 y 49

Comparación de la secuencia nucleotídica de M1 de *G.stearotherophilus* CECT4264 y *G. stearotherophilus* NRBC102445.

GkEndoM1pJMC104	ATGGCGAATGATTTCATTTTGCCAATCGTCAGCCTTTTCAAGTATAATAAAAAAGTGGACA	60
EnGluM1GkaNRBC102445	gtggcgaatgattcatttttgccaatcgtcagccttttcaagtataataaaaaagtggaca	60

GkEndoM1	AAACCGGAAAGGGGTGCCAAGATGGCGAAGTTGGACGAAACGTTGACGATGCTGAAGCG	120
EnGluM1GkaNRBC102445	aaaccggaaaggggtgccaaagatggcgaaagtggacgaaacgcttgacgatgctgaagcgc	120

GkEndoM1	CTGACCGATGCGAAAGGCGTCCCGGGCAATGAGCGGGAAGCGCGGATGTGATGAAGACA	180
EnGluM1GkaNRBC102445	ctgaccgatgcgaaagcgtcccgggcaatgagcgggaagcgcgcatgtgatgaagaca	180

GkEndoM1	TACATAGTCTCCGTATGCGGATGAAGTGACAACGGATGGCCTCGGCAGCTTGATTGCCAAA	240
EnGluM1GkaNRBC102445	tacatagctccgatgctggatgaagtgacaacggatggcctcggcagcttgattgccaaa	240

GkEndoM1	AAAGAAGGAAATCGGGCGGGCCGAAAGTGTGATCGCCGGCCATTTGGATGAAGTCGGC	300
EnGluM1GkaNRBC102445	aaagaaggaaatcgggcgggccgaaagtgtgatcgccggccatttggatgaagtccgc	300

GkEndoM1	TTTATGGTGACGCAAAATCGACGACAAAGATTTCATCCGCTTCCAACCGCTTGGCGGATGG	360
EnGluM1GkaNRBC102445	tttatggtgacgcaaatcgaagcgaagatttcacccgcttccaacgcttggcggatgg	360

GkEndoM1	TGGAGCCAAGTGTGCTCGCCAGCGCGTGACGATCGTGACGAAAAAGGCGACATCACC	420
EnGluM1GkaNRBC102445	tggagccaagtgtgctcgccagcgcgtgacgatcgtgacgaaaaaggcgacatcacc	420

GkEndoM1	GGCGTCATCGGTTTGAAGCCGCGCACATTTGCGCCGGAGGCGCGCAAAAACCGGTG	480
EnGluM1GkaNRBC102445	ggcgtcatcggttcgaagccgcccacattctgcccggagggcgcgcaaaaaaccggtg	480

GkEndoM1	GAAATCAAAGATATGTTTATCGACATCGCGCGGACAAGCCGCGAGGAAGCGATGGAGTGG	540
EnGluM1GkaNRBC102445	gaaatcaaagatatgttcatcgacatcgcgcgacaagccgaggaagcgatggagtgg	540

GkEndoM1	GGCGTCCGCCCGGGCGATATGATCGTGCCTATTTGAATTTACGGTATTGAACAATGAA	600
EnGluM1GkaNRBC102445	ggcgtccgcccgggcgatgatcgctgcctatTTTGAATTTACGGTATTGAACAATGAA	600

GkEndoM1	AAAATGCTGCTCGGAAAGCATGGGACAACCGGATCGGCTGTGCGGTGCGCCATCGATGTG	660
EnGluM1GkaNRBC102445	aaaatgctgctcggaaagcatgggacaaccggatcggctgtgcggtcgcacatcgatgtg	660

GkEndoM1	CTCAAGCAGCTGAAAGGCGTCGACCATCCAACACGGTATACGGCGTGGCAGCGGTGACG	720
EnGluM1GkaNRBC102445	ctcaagcagctgaaaggcgtcgaccatccaacacggtatacggcgtggcagcggtgacg	720

GkEndoM1	GAAGAAGTCGGCTTGGCGGGCGCGCACGGCCGCCAATTCATTACAGCCAGATATCGCG	780
EnGluM1GkaNRBC102445	gaagaagtcggcttggcggcgcgcgcacggccgccaatTCATTACAGCCAGATATCGCG	780

GkEndoM1	TTTGCTGTGACGTCGGCATTGCCGGCAGACGCCGGGCGTGTGCGAAAAAGAAGCGATG	840
EnGluM1GkaNRBC102445	tttgctgtgacgctcggcattGCCGGCAGACGCCGGGCGTGTGCGAAAAAGAAGCGATG	840

GkEndoM1	GGCAAATCGGCGCGGCCGCACATCGTCTTGTACGACGCAACGATGGTGTGCGACCGC	900
EnGluM1GkaNRBC102445	ggcaaatcggcgcggccgcacatCGTCTTGTACGACGCAACGATGGTGTGCGACCGC	900

GkEndoM1	GGCTTGCGCAATTTGTTCATCGAAGTGGCGGAAGAGCTCAACATTCCTACCATTTTGAT	960
EnGluM1GkaNRBC102445	ggcttgcgcaatTTGTTCATCGAAGTGGCGGAAGAGCTCAACATTCCTACCATTTTGAT	960

GkEndoM1	GCCATGCCAGGCGGCGGTACGGACGCGGAGCGATTCACTTAACGGCATCGGCGTTCGG	1020
EnGluM1GkaNRBC102445	gccatgccaggcggcggtacggacgcgagcgattCACTTAACGGCATCGGCGTTCGG	1020

GkEndoM1	TCGCTCACGATTGCCATCCCAGCGCTACATCCACTCGACGCCCATTTTGACCCGC	1080
EnGluM1GkaNRBC102445	tcgctcacgattGCCATCCCAGCGCTACATCCACTCGACGCCCATTTTGACCCGC	1080

GkEndoM1	GACGACTACGAAAAACCGGTCAAGCTGCTTGTGAGGTGATCAAACGGCTTGACGCCGAC	1140
EnGluM1GkaNRBC102445	gacgactacgaaaaaccgggtcaagctgctTGTGAGGTGATCAAACGGCTTGACGCCGAC	1140

GkEndoM1	AAAGTGAACAACCTGACGTTTgacGAA---	1167
EnGluM1GkaNRBC102445	aaagtgaacaacctgacgTTTgacGAA---	1170

Figura 13: Comparación de la secuencia nucleotídica de M1 de *Geobacillus* *G.stearotherophilus* CECT4264 y *G. stearotherophilus* NRBC102445.

Comparación de las secuencias de aminoácidos de las Endoglucanasas M1 de *Geobacillus stearothermophilus* 49(pJMC120) y *Geobacillus stearothermophilus* B5.

Gs49EndoM1pJMC120	MANDSFCQSSAFSSI IKKWKTPERSAKMAKLDELTLMLKALTDAGKVPGNEREARDVMKT	60
EnGluM1GsteB5	MANDSFCQSSAFSSI IKKWKTPERSAKMAKLDELTLMLKALTDAGKVPGNEREARDVMKT	60

Gs49EndoM1pJMC120	YIAPYADEVTTDGLGSLIAKKEGKSGGPKVMIAGHLDEVGFMVTQIDDKGFIRFQTLGGW	120
EnGluM1GsteB5	YIAPYADEVTTDGLGSLIAKKEGKSGGPKVMIAGHLDEVGFMVTQIDDKGFIRFQTLGGW	120

Gs49EndoM1pJMC120	WSQVMLAQRVTIVTKKGDVTGVIGSKPPHILPPEARKKPVEIKDMFIDIGATSREEAMEW	180
EnGluM1GsteB5	WSQVMLAQRVTIVTKKGDVTGVIGSKPPHILPPEARKKPVEIKDMFIDIGATSREEAMEW	180

Gs49EndoM1pJMC120	GVRPGDMIVPYFEFTVLNNEKMLLAKAWDNRIGCAVAIDVLKQLKGVDPHTVYGVGTVQ	240
EnGluM1GsteB5	GVRPGDMIVPYFEFTVLNNEKMLLAKAWDNRIGCAVAIDVLKQLKGVDPHTVYGVGTVQ	240

Gs49EndoM1pJMC120	EEVGLRGARTAAQFIQPDIAFAVDVGIAGDTPGVSEKEAMGKLGAGPHIVLYDATMVSHR	300
EnGluM1GsteB5	EEVGLRGARTAAQFIQPDIAFAVDVGIAGDTPGVSEKEAMGKLGAGPHIVLYDATMVSHR	300

Gs49EndoM1pJMC120	GLREFVIELAEELNIPYHFDAMPGGGTDAGAIHLTGSVPSLTIAIPTRYIHSAAAILHR	360
EnGluM1GsteB5	GLREFVIEIAEKLNIPYHFDAMPGGGTDAGAIHLTGSVPSLTIAIPTRYIHSAAAILHR	360

Gs49EndoM1pJMC120	DDYENTVKLLVEVIKRLDADKVKQLTFDE	389
EnGluM1GsteB5	DDYENTVKLLVEVIKRLDADKVKQLTFDE	389

Figura 14: Comparación de las secuencias de aminoácidos de las Endoglucanasas M1 de *Geobacillus stearothermophilus* 49(pJMC120) y *Geobacillus stearothermophilus* B5. Se representa en verde los aminoácidos diferenciados en ambas secuencias.

Comparación aminoacídica de la Endoglucanasas M2 Endoglucanasas M2 de Geobacillus stearothermophilus CECT49 y Geobacillus stearothermophilus B5.

Gst49EndoM2	MNEETLQLFQTLTELPGAPGYEHPVRQFMRQELAKYADDTVQDRLGSEFGVKRGDEAGPT	60
EnGluM2GsteB5	MNEETLQLFQTLTELPGAPGYEHPVRQFMRQELAKYADEVVQDRLGSEFGVKRGDETGPT	60
	:**:*****:*****:*****:*****:*****:*****:*****	
Gst49EndoM2	VMVAGHMDEVGFMVTAITDNGMIRFQPLGGWVQVLLAQRVQIITDNGPVIQVIGSIPPH	120
EnGluM2GsteB5	VMVAGHMDEVGFMVTAITDNGMIRFQTLGGWVQVLLAQRVQIITDNGPVIQVIGSIPPH	120
	*****:*****:*****:*****:*****:*****:*****:*****:*****	
Gst49EndoM2	LLDEEQRKPKMKNMLIDIGAESRDAERVGIRPGQPIVPVSPFTLANGKTVMKAWD	180
EnGluM2GsteB5	LLDEEQRNKPMDLKNMLIDIGAESRDAERIGIRPGQPIVPVSPFTPLANGKTVMKAWD	180
	*****:***:*****:*****:*****:*****:*****:*****:*****	
Gst49EndoM2	NRFGCGLAIELLKELETPVNVLYAGATVQEEVGLRGAQTAAEMINPDIFFALEASPAN	240
EnGluM2GsteB5	NRFGCGLAIELLKELEQGAALPNVLYAGATVQEEVGLRGAQTAAATMINPDIFFALEASPAN	240
	*****:*****:*****:*****:*****:*****:*****:*****:*****	
Gst49EndoM2	DMSGDKQAFGHIGKGVRLYDRTMITHRGMREFVLDTAETLGVVPYQFFISPGGGTDAGR	300
EnGluM2GsteB5	DMTGDKQAFGHIGKGVRLYDRTMITHRGMREFVLDTAESLDVPYQFFISPGGGTDAGR	300
	:**:*****:*****:*****:*****:*****:*****:*****	
Gst49EndoM2	VHIANRGVPSAVIGICARYIHTHAIIHVDDYAAAKLIVELVKRCDETVETIRQNSLE	360
EnGluM2GsteB5	VHVANRGVPSAVIGICARYIHTHAIIHVDDYAAAKLIVELVKRCDRATVEAIRQNS--	358
	:**:*****:*****:*****:*****:*****:*****:*****	
Gst49EndoM2	HHHHHH	366
EnGluM2GsteB5	-----	358

Figura 16: Comparación aminoacídica de la Endoglucanasas M2 Endoglucanasas M2 de Geobacillus stearothermophilus CECT49 y Geobacillus stearothermophilus B5. Se representa en amarillo los aminoácidos diferenciados en ambas secuencias.

Representación de la disposición de los codones raros de la secuencia de la enzima M1 de *Geobacillus stearothermophilus* CECT49 (pJMC120).

```

atg gcg aat gat tca ttt tgc caa tcg tca gcc ttt tca agt ata ata aaa aag tgg aca
M A N D S F C Q S S A F S S I I K K W T 20
aaa ccg gaa agg agt gcc aag atg gcg aag ttg gac gaa acg ttg acg atg ctg aaa gcg
K P E R S A K M A K L D E T L T M L K A 40
ctg acc gat gcg aaa ggc gtc ccg ggc aat gag cgg gaa gcg cgc gat gtg atg aag aca
L T D A K G V P G N E R E A R D V M K T 60
tac ata gct ccg tat gcg gat gaa gtg aca acg gat ggc ctc ggc agc ttg att gcc aaa
Y I A P Y A D E V T T D G L G S L I A K 80
aaa gaa ggg aaa tcg ggc ggg ccg aaa gtg atg atc gcc ggc cat ttg gat gaa gtc ggc
K E G K S G G P K V M I A G H L D E V G 100
ttt atg gtg acg caa atc gac gac aaa ggc ttc atc cgc ttc caa acg ctt ggc gga tgg
F M V T Q I D D K G F I R F Q T L G G W 120
tgg agc caa gtg atg ctc gcc cag cgc gtg acg atc gtg acg aaa aaa ggc gat gtc acc
W S Q V M L A Q R V T I V T K K G D V T 140
ggc gtc atc ggt tcg aag ccg ccg cac att ctg ccg ccg gag gcg cgc aaa aaa ccg gtg
G V I G S K P P H I L P P E A R K K P V 160
gaa atc aaa gat atg ttc atc gac atc ggc gcg aca agc cgc gag gaa gcg atg gat tgg
E I K D M F I D I G A T S R E E A M E W 180
ggc gtc cgc ccg ggc gat atg atc gtg ccg tat ttt gaa ttt acg gta ttg aac aat gaa
G V R P G D M I V P Y F E F T V L N N E 200
aaa atg ctg ctc gcg aaa gca tgg gac aac cgg atc ggc tgt gcg gtc gcc atc gat gtg
K M L L A K A W D N R I G C A V A I D V 220
ctc aag cag ctg aaa ggc gtc gac cat cca aac acg gta tac ggc gtc ggc acg gtg cag
L K Q L K G V D H P N T V Y G V G T V Q 240
gaa gaa gtc ggc ttg cgc ggc gcg cgc acg gcc caa ttc att cag ccg gat atc gcc
E E V G L R G A R T A A Q F I Q P D I A 260
ttt gcg gtg gat gtc ggc att gcc ggc gat aca ccg ggc gtg tcg gaa aaa gaa gcg atg
F A V D V G I A G D T P G V S E K E A M 280
ggc aaa ctc ggc gcc ggc cac atc gtc ctg tac gac gcg acg atg gtg tcg cac cgc
G K L G A G P H I V L Y D A T M V S H R 300
ggc ttg cgc gaa ttc gtc atc gaa ttg gcg gaa gag ctg aac att ccg tac cat ttt gat
G L R E F V I E L A E E L N I P Y H F D 320
gcc atg ccg ggc ggc ggt acg gat gcg gga gcg att cat tta acg ggc agc ggc gtt ccg
A M P G G G T D A G A I H L T G S G V P 340
tcg ctc acg atc gcc atc ccg acg cgc tac atc cac tcg cac gcc gcc att ttg cat cgc
S L T I A I P T R Y I H S H A A I L H R 360
gac gac tac gag aac acg gtc aag ctg ctt gtc gaa gtg atc aaa cgg ctt gac gcc gac
D D Y E N T V K L L V E V I K R L D A D 380
aaa gtg aaa caa ctg acg ttt gac gaa
K V K Q L T F D E
389

```

Figura 17: Representación de la disposición de los codones raros de la secuencia de la enzima M1 de *Geobacillus stearothermophilus* CECT49 (pJMC120). La representación se realizó con las aplicaciones <https://web.expasy.org/translate/> y <https://people.mbi.ucla.edu/sumchan/caltor.html>.

**Representación de la disposición de los codones raros de la secuencia de la enzima Endoglucanasa M2 de *Geobacillus kaustophilus* CECT4264 (pJMC122).
representación de la disposición de los codones raros para pJMC122.**

```

atg aat gag gac acc ctg caa tta ttc caa acc ttg acc gag ctt cca ggc gca ccg ggt
M N E D T L Q L F Q T L T E L P G A P G 20
tat gaa cat ccg gtg cgg cag ttt atg cgc caa gag ctc gcg aaa tac gcc gat gac att
Y E H P V R Q F M R Q E L A K Y A D D I 40
gtg caa gac cgc ctc ggg agc ctt ttc ggc gtc aaa gcg ggt gat gag gcg ggc ccg acc
V Q D R L G S L F G V K R G D E A G P T 60
gtg atg gtc gcc ggc cat atg gat gaa gtc gga ttt atg gtg aca gcg atc acc gat cac
V M V A G H M D E V G F M V T A I T D H 80
ggc atg att cgt ttc cag cgc cta gga ggc tgg gat caa gtg ctg ctc gcc cag cgc
G M I R F Q P L G G W W D Q V L L A Q R 100
gtg cag att atc acc gac gga ggg ccg gtc atc ggc gtg atc ggg tcg att ccg ccg cat
V Q I I T D G G P V I G V I G S I P P H 120
ttg ctt gat gag gag cag cgc aaa aaa cgg atg gag atg aaa aac atg ttg att gac atc
L L D E E Q R K K P M E I K N M L I D I 140
gga gcg gaa agc cgt gaa gat gcc gaa cga gtc ggc att cgc ccc ggc cag ccg atc gtg
G A E S R E D A E R V G I R P G Q P I V 160
ccg gtc agc cgt ttt act ttg ctt gcc aac gga aaa acc gtg atg gca aaa gcg tgg gac
P V S P F T L L A N G K T V M A K A W D 180
aat cgg ttt ggc tgc gga ttg gcg atc gag ctg tta aaa gag ctg aac ggc gaa aca gtg
N R F G C G L A I E L L K E L N G E T V 200
ccg aac gtc ttg tac gca ggg gcg acc gtg cag gaa gag gtc ggg ctg cgt ggg gcc cag
P N V L Y A G A T V Q E E V G A L R G A Q 220
acc gcg gct tcg atg atc aac ccg gac atc ttt ttt gcc ctg gaa gca agc ccg gcc aac
T A A S M I N P D I F F A L E A S P A N 240
gat atg tcc ggc gac aag cag gcg ttc ggc cat atc ggc aaa gga gcg ctc atc cgc ttg
D M S G D K Q A F G H I G K G A L I R L 260
tac gac cgg acc atg att acc cac cgc ggc atg cgc gaa ttt gtg ctt gat acc gcg gaa
Y D R T M I T H R G M R E F V L D T A E 280
acc atc ggc gtc ccg tac cag ttt ttc att tcc cca ggc ggc gga acc gac gcc ggg cgg
T I G V P Y Q F F I S P G G G T D A G R 300
gtg cat atc gcc aac cgc ggc gtg ccg tcg gcg gtg atc ggc att tgc gcc cgc tac att
V H I A N R G V P S A V I G I C A R Y I 320
cat acc cat gcg tcg atc att cat gtc gac gat tac gcg gcg gcc aaa cag ctc atc gtt
H T H A S I I H V D D Y A A A K Q L I V 340
gag ctc gtc aaa cgg tgc gac cgg aca acc gtc gag acc att cgc caa aac agc ctc gag
E L V K R C D R T T V E T I R Q N S L E 360
cac cac cac cac cac tga
H H H H H H - 367

```

Figura 18: Representación de la disposición de los codones raros de la secuencia de la enzima Endoglucanasa M2 de *Geobacillus kaustophilus* CECT4264 (pJMC122). La representación se realizó con la aplicación <https://web.expasy.org/translate/> y <https://people.mbi.ucla.edu/sumchan/caltor.html>.

