



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

Generación y caracterización de una biblioteca
de secuencias promotoras y sitios de unión a
ribosoma en *Escherichia coli*.

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

BIÓLOGO

P R E S E N T A:

DAVID FERREIRA GÓMEZ

DIRECTOR DE TESIS: DR. GUILLERMO GOSSET LAGARDA



México, D.F.

2014



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

1. Datos del alumno

Ferreira
Gómez
David
55 29 65 35 16,
Universidad Nacional Autónoma de México
Facultad de Ciencias
Biología
306227415

2. Datos del tutor

Dr.
Guillermo
Gosset
Lagarda

3. Datos de sinodal 1

Dr.
Víctor Manuel
Valdés
López

4. Datos de sinodal 2

Dr.
José Luis
Puente
García

5. Datos de sinodal 3

Dra.
Claudia
Segal
Kischinevzy

6. Datos sinodal 4

Dr.
José Adelfo
Escalante
Lozada

7. Datos de tesis

Generación y caracterización de una biblioteca de secuencias promotoras y sitios de unión a ribosoma en *Escherichia coli*.

49 p.
2014

A mis padres, María Teresa Gómez y Raúl Ferreira

A mis hermanos, Adriana y Raúl

Agradecimientos

A mi familia, Raúl Ferreira Gómez, Adriana Ferreira Gómez, Raúl Ferreira Águila y María Teresa Gómez Barrios, por el apoyo incondicional que siempre me han brindado. No hubiera llegado a este punto sin ustedes.

Al Dr. Guillermo Gosset Lagarda, por haberme asesorado, recibido y apoyado en este trabajo. Con su paciencia y consejos aprendí las bases fundamentales del quehacer científico y el pensamiento crítico.

A la Q. I. Luz María Martínez Mejía por su asesoría y apoyo técnico durante toda mi estancia, además del trabajo que realizó durante la clonación y transformación de la biblioteca.

Al Dr. Paul Gaytan, por todo el apoyo ofrecido. Sin la ayuda técnica de la Unidad de Síntesis este trabajo no se hubiera podido realizar.

Al Q. I. Santiago Becerra por la ayuda técnica en la síntesis de oligonucleótidos.

Al Dr. José Adelfo Escalante, porque junto con mi tutor también me cedió parte importante de su tiempo durante el desarrollo y conclusión de este proyecto.

Al Dr. José Luis Puente, por haberme dado la oportunidad de llegar al Instituto de Biotecnología.

A la familia que encontré: Amira, Armín, Emilio, Gemma, Oscar y Pedraza. Por todos los momentos especiales que hemos compartido y por todos los que todavía nos faltan.

A mis amigos: Sofía, Oscar Chu, Mayra, Perlita, Adra, Gabriel, Dan, Andrea, Gustavo, Javi, Jordi, Nadia, Kenia, Paulina, Rodri, Daniel, Armando, Pablo, Omar, Josué, Natali, Fernanda, Monse, Marcela y Elizabeth.

A mis amigos y compañeros de laboratorio: Ale Mejía, René, Ale Vargas, Andrea, César, Beto, Andrés, Cinthia, Sonia, Uriel, Susy, Caheri, Manuel, Juan Carlos, Carina, Fabián, Diego, Mauricio, así como a Mechita, Aure, Geo y Ramón. Por haber hecho agradable mi estancia en el laboratorio.

A los miembros del jurado: Dra. Claudia Segal, Dr. Víctor Valdés, Dr. José Adelfo, Dr. José Luis Puente y Dr. Guillermo Gosset, por sus observaciones, ideas y comentarios que ayudaron a mejorar este trabajo

Este trabajo se realizó en el laboratorio número cuatro del Departamento de Ingeniería Celular y Biocatálisis del Instituto de Biotecnología de la UNAM (Cuernavaca, México), bajo la asesoría del Dr. Guillermo Gosset Lagarda. David Ferreira Gómez recibió apoyo para la realización de la tesis de licenciatura por parte del Instituto de Biotecnología y el CONACyT con la beca de ayudante de investigador.

**“La vida es un capítulo en el
sueño de la complejidad”**

A.C.

Índice

I.- Resumen.....	1
II.- Introducción.....	2
2.1 RNA polimerasa y el promotor consenso de σ^{70}	2
2.2 El ribosoma y el sitio de unión al ribosoma (RBS)	4
III.- Antecedentes.....	8
3.1 Evaluación de promotores y modelos propuestos para medir su eficiencia.....	8
3.1.1 Bibliotecas de promotores.....	9
3.1.2 Modelos propuestos.....	10
3.2 Evaluación de RNAm, RBS y modelos propuestos para medir su eficacia.....	11
3.2.1 Diferentes RBS y RNA mensajeros evaluados.....	11
3.2.2 Modelos propuestos para la traducción.....	12
3.3 La combinación de promotores y RBS en una biblioteca.....	14
IV.- Justificación.....	15
V.- Hipótesis.....	15
VI.- Objetivos.....	15
6.1 Objetivos generales.....	15
6.2 Objetivos particulares.....	15
VII.- Materiales y métodos.....	16
7.1 Cepa, plásmido y medios.....	16
7.2 Diseño del oligonucleótido molde.....	16
7.3 Generación de la biblioteca.....	17
7.4 Clonación de la biblioteca en el plásmido y transformación.....	18
7.4.1 Complementación de la hebra sencilla de oligonucleótido por PCR.....	18

7.4.2 Digestión del DNA de las candidatas y el plásmido receptor.....	19
7.4.3 Ligación de la biblioteca al plásmido.....	20
7.4.4 Transformación.....	21
7.5 Selección de candidatas.....	21
7.6 Caracterización de la biblioteca.....	22
7.6.1 Secuenciación de las candidatas.....	22
7.6.2 Determinación del crecimiento y fluorescencia.....	22
7.7 Análisis de resultados.....	23
VIII.- Resultados y discusión.....	27
8.1 Generación de la biblioteca.....	27
8.2 Caracterización de la biblioteca.....	30
8.2.1 Cinéticas de crecimiento y fluorescencia.....	30
8.2.2 Análisis de las secuencias.....	39
8.3 Análisis de la biblioteca.....	41
IX.- Conclusiones.....	45
X.- Perspectivas.....	45
XI.- Bibliografía consultada.....	46

I. RESUMEN

La supervivencia y éxito reproductivo de los seres vivos tiene relación con el empleo eficiente de su material genético. Esto puede interpretarse como la generación de los productos correctos en los momentos y cantidades adecuadas. *Escherichia coli* es uno de los organismos modelo más usados para entender de manera más amplia los procesos de transcripción y traducción, que son etapas fundamentales en la expresión genética.

El trabajo que aquí presento es un estudio enfocado a la generación de una herramienta que permite entender mejor las características que conforman las secuencias promotoras y las de sitio de unión al ribosoma; a partir de esta información se pueden crear nuevas estrategias experimentales en el campo de la ingeniería genética manipulando la expresión de cada gen. Esta herramienta es la producción y caracterización de un acervo de secuencias sintéticas. La síntesis del acervo partió de un diseño molde que reúne las secuencias del promotor consenso tipo σ^{70} y el sitio de unión a ribosoma del gen *lacZ*. Al molde se le hicieron inserciones aleatorias, originando así la diversidad de candidatas que conforman la biblioteca.

Cada candidata seleccionada tuvo una caracterización y análisis de resultados independiente. Para seleccionar a las candidatas se usaron tres métodos de tamizaje: el primero consistió en escoger únicamente las colonias que presentaran la proteína reportera (proteína verde fluorescente, GFP), el segundo y tercero se realizaron simultáneamente, la secuenciación única de cada candidata y el patrón de restricción esperado en el corte de cada plásmido junto con el tamaño esperado.

Luego de haber seleccionado a un grupo de colonias, se hicieron cinéticas de fluorescencia y de crecimiento en medio líquido Luria-Bertani en agitación, se tomaron muestras por 16 horas con intervalos de 2 horas entre cada muestra.

En total, se obtuvieron 14 secuencias, 7 secuencias por debajo de la expresión de la secuencia molde inicial y 6 por arriba. Los rangos de expresión relativos a la secuencia molde van de 0.17 hasta 7.12 veces. La mayor parte de las mutaciones se obtuvieron en el espacio que divide al promotor del sitio de unión a ribosoma.

Para analizar la correspondencia entre la teoría y lo observado de las candidatas, se usaron 2 modelos diferentes de predicción de fuerza, uno para la región promotora y otro para el sitio de unión a ribosoma. El modelo para promotores, está basado en la cantidad de energía requerida por la RNA polimerasa para unirse al promotor y comenzar la transcripción. Por otro lado, el modelo para los sitios de unión a ribosoma se enfoca en la eficiencia de traducción del mensajero, calculando la tasa de inicio de traducción a través de la dinámica del RNA ribosómico y el RNA mensajero.

La predicción realizada con los modelos no fue completamente acertada, este resultado puede deberse a efectos que no contemplan los modelos, como la vida media de los RNA mensajeros o la tasa de inicio de transcripción.

En perspectiva, el presente trabajo puede ser usado tanto para seguir ampliando los conocimientos de secuencias de DNA y expresión genética como para usar nuevas estrategias en la ingeniería genética que a su vez permitan a otras disciplinas de la biología seguir develando el funcionamiento de la vida.

II. INTRODUCCIÓN

Los más de 4 mil genes que tiene *E. coli* en el cromosoma se encuentran organizados en forma de operones o de manera independiente (Reznikoff William 1986), esto puede interpretarse como un resultado de la evolución en la expresión génica. Las primeras pares de bases de un gen, corresponden con frecuencia a las secuencias conocidas como promotores y sitios de unión a ribosoma (RBS) que son partes necesarias para obtener producto de los genes. En el sistema de expresión genético, el papel de estas secuencias de DNA están en los procesos de: el inicio de la transcripción, por ser la secuencia promotora un blanco de pegado de la holoenzima RNA polimerasa, y el inicio de la traducción, reclutando las subunidades del ribosoma mediante los Sitios de Unión a Ribosoma (RBS) (Shultzaberger et al. 2001; Zehua Chen 2007).

2.1. RNA polimerasa y el promotor consenso de σ^{70}

La RNA polimerasa (RNAPol) es la enzima encargada de sintetizar RNA mensajero (RNAm) a partir de una secuencia de DNA (Busby 2004) como se muestra en la **Figura 1**. Cuando esta enzima está completa se dice que está en su forma 'holoenzima' integrada por 6 subunidades: α_2 , β , β' , ω y σ (Cox, David L. Nelson 2009). El papel de cada subunidad es fundamental; a las subunidades α se les han encontrado dos funciones; la primera es dimerizar las subunidades β y β' , y su otra función es unirse al DNA en secuencias conocidas como elementos 'Upstream' o río arriba ubicadas antes del promotor (Gourse et al. 2000a), las subunidades β y β' son las encargadas de unirse al DNA y de sintetizar el RNA de transcripción (Korzheva 2000), la subunidad ω no tiene un rol directo en la transcripción, sin embargo, parece tener función como chaperona asistiendo a β' en su plegado, y por último, la subunidad σ es la encargada de posicionar a la polimerasa en sitios específicos para que inicie la transcripción de genes (Sofia Österberg 2011).

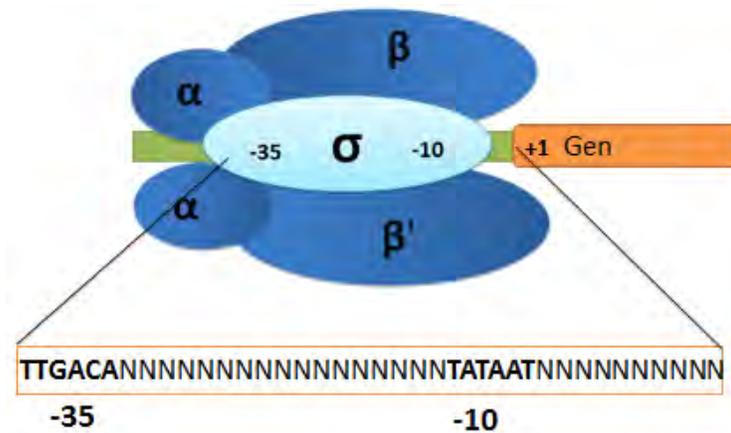


Figura 1: La unión de una holoenzima RNA polimerasa con una región promotora. Aunque en esta perspectiva no se pueda ver la subunidad ω , la imagen sí considera que se encuentra ahí dicha subunidad. La letra “N” significa que puede ser cualquiera de las 4 bases nitrogenadas del DNA.

Las subunidades α_2 , β , β' y ω son conocidas como núcleo de la enzima RNAPol, porque además de realizar los procesos de transcripción están juntas siempre, a diferencia de la subunidad σ . La frecuencia con la que se forma la holoenzima es dependiente de la concentración de la subunidad σ y de la afinidad que tenga el núcleo por la subunidad (Sofia Österberg 2011). *E. coli* cuenta con una variedad de 7 subunidades σ , cada subunidad dirige a la holoenzima a diferentes catálogos de genes (Sofia Österberg 2011; Rhodius et al. 2006; Kaczanowska & Rydén-Aulin 2007), la subunidad σ más abundante en la célula es la σ^{70} , en consecuencia es la que tiene mayores probabilidades de unirse a un núcleo de RNA (Grigorova et al. 2006).

Lo que guía la unión del DNA y la subunidad σ^{70} o cualquier otra subunidad σ para que empiece la transcripción, es una secuencia promotora (M. François Jacob 1964; Pribnow 1975; Rosenberg & Court 1979). σ^{70} reconoce en un promotor dos regiones hexaméricas conocidas como -35 y -10 por su posición relativa ante el sitio de inicio de transcripción, como se ilustra en la **figura 1**. El promotor tipo σ^{70} es abundante en *E. coli* (Gourse et al. 2000b), y se han alineado las bases más frecuentes que aparecen en el genoma de *E. coli*, el resultado es el promotor consenso (Zehua Chen 2007; Pribnow 1975; Hawley et al. 1983; Rossi et al. 1983; Grigorova et al. 2006).

El promotor consenso (para las subunidades σ^{70}) tiene por secuencia **5'-TTGACA-3'** y **5'-TATAAT-3'** en sus hexámeros -35 y -10 respectivamente, con un espacio intermedio promedio de 17 nucleótidos (Rossi et al. 1983; Zehua Chen 2007; Tumell & Harrison 1990). La transcripción es el primer paso en la expresión génica, por esto, se han realizado varios estudios sobre la variación original de bases en la

secuencia consenso (Rossi et al. 1983; Mijakovic et al. 2005; De Mey et al. 2007; Peter Ruhdal Jensen 1998; Recknagel & Rolf-Dieter 1994).

Se han propuesto modelos para la predicción de transcripción usando sólo la secuencia del promotor. Rhodius y Mutalik (2010) mencionan lo difícil que es predecir los niveles de transcripción para los promotores nativos tipo σ^{70} dentro del cromosoma, debido a que hay varias proteínas reguladoras, como los factores de transcripción que intervienen en el proceso. A pesar de este hecho, hay varios trabajos importantes en la predicción de fuerza de promotores nativos en cromosomas (Bintu et al. 2005).

Por otro lado, para promotores sintéticos hay más modelos disponibles (Recknagel & Rolf-Dieter 1994; Peter Ruhdal Jensen 1998; De Mey et al. 2007; Zehua Chen 2007; De Mey et al. 2010; Brewster et al. 2012). Las estrategias de análisis toman de referencia a la secuencia consenso, a partir de esta se hacen matrices de evaluación, asignando un grado de importancia al tipo de base (si eran purinas o pirimidinas) (Recknagel 1994) o a cada base individual (De Mey et al. 2007; Brewster et al. 2012) de acuerdo a su posición. La variedad de modelos nos ayudan a analizar una secuencia desde diferentes ángulos: algunos cualitativos (Peter Ruhdal Jensen 1998) que ayudan a identificar una secuencia promotora más eficiente que otra, y otros cuantitativos (De Mey et al. 2010; Brewster et al. 2012) que predicen los niveles de expresión de cada promotor.

También hay otros trabajos que consideran importante las regiones que flanquean a los hexámeros -10 y -35, particularmente la región de -10 al sitio de unión a ribosoma (Huerta & Collado-Vides 2003), por los procesos de elongación de la RNAPol y porque también podría evaluarse la efectividad de traducción del mensajero, sin embargo, estos sitios no tienen todavía un método de evaluación en los procesos de transcripción para su futura predicción.

2.2. El ribosoma y el sitio de unión al ribosoma (RBS)

La traducción es el paso inmediato después de la transcripción y también el siguiente punto necesario en la expresión genética (Laursen et al. 2005). El producto de la transcripción, el RNA mensajero (RNAm) se une al ribosoma para comenzar la síntesis de proteínas. La traducción en bacterias se divide en 3 etapas (Kaczanowska & Rydén-Aulin 2007):

1.- Iniciación, en esta etapa el RNAm y la subunidad pequeña del ribosoma interactúan hasta que hibridan la secuencia del RNA ribosómico 16S en su extremo 3' con el sitio de unión a ribosoma (RBS)

o también conocido como región Shine-Dalgarno (SD). Este paso se da por terminado cuando las 2 subunidades ribosomales (30S y 50S) se unen.

2.- Elongación, en este paso se complementan los codones con los RNAs de transferencia, dando lugar a la unión de varios aminoácidos por enlaces peptídicos.

3.- Terminación y reciclaje del complejo ribosomal, cuando el codón de término llega al complejo ribosomal, algunas proteínas ribosomales cortan la cadena peptídica por hidrólisis. Después de que se detiene la elongación y se separa la cadena peptídica, hay un desensamble del complejo ribosomal, esta separación permite que una nueva secuencia de RNAm se una a la subunidad pequeña para iniciar un nuevo ciclo de traducción.

Este proceso se encuentra esquematizado en la **Figura 2**.

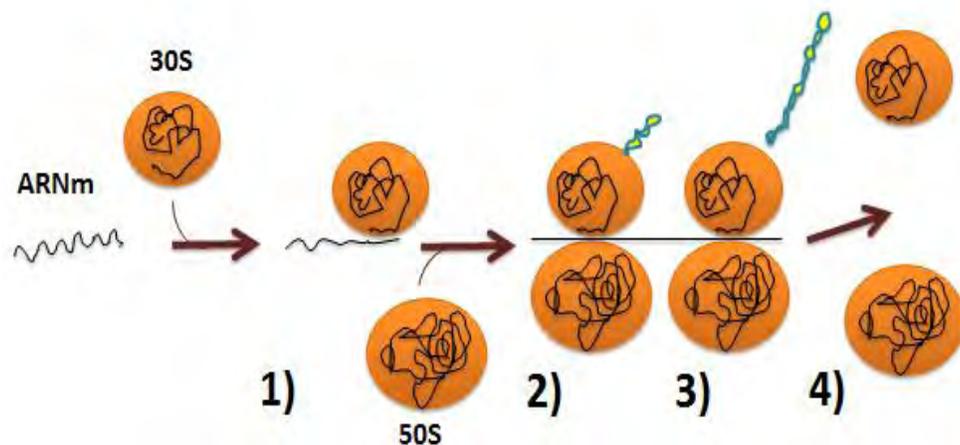


Figura 2: Proceso de traducción. **1)** Iniciación, **2)** elongación, **3)** terminación y **4)** reciclamiento

El ribosoma es la macromolécula encargada de la síntesis de proteínas, esta función la realizan dos subunidades (Cox, David L. Nelson 2009). Todo el complejo ribosómico tiene una tasa relativa de sedimentación de 70S en bacterias y un peso de 2.4 MDa, mientras que su subunidad más grande tiene 50S y 1.5 MDa, que en comparación con la más pequeña representa casi el doble, ya que la más pequeña posee 30S y 0.8 MDa (Laursen et al. 2005). Ambas subunidades están formados por dos tercios de ribonucleótidos (RNAr) y un tercio de proteínas; la subunidad 30S cuenta con 21 proteínas

diferentes y un tipo de RNAr, el 16S, que posee 1,542 pares de bases. Por otro lado, la subunidad mayor 50S tiene 33 proteínas y dos tipos de RNAr, 23S (2,904 pb) y 5S (120 pb) (Laursen et al. 2005).

El inicio de la traducción es la etapa que más determina la eficiencia de la síntesis de proteínas (Laursen et al. 2005; Hayashi et al. 2012; Gourse et al. 2000a; Pon & L. 1990). La traducción empieza cuando el complejo ribosómico está completo (Cox, David L. Nelson 2009), la formación del ribosoma depende de la secuencia del RNAm (Pon & L. 1990; Laursen et al. 2005; Hershey et al. 2012; Hayashi et al. 2012). Cada secuencia de RNAm tiene en su extremo 5' una zona conocida como Región de Inicio de la Traducción (RIT), esta se caracteriza por estar antes del codón de inicio de traducción (las tres bases del RNAm que indican el inicio de la síntesis) y tener al sitio de unión al ribosoma (Dalgarno 1975; Pon & L. 1990; Ringquist et al. 1992; Laursen et al. 2005; Salis et al. 2009), el RBS se aparea con la secuencia 16S (en su extremo 3') lo que permite el ensamblaje del complejo ribosómico y el comienzo de la síntesis de proteínas (Laursen et al. 2005; Shultzaberger et al. 2001).

El ensamblaje ribosómico empieza por el apareo de secuencias entre Shine-Dalgarno (SD) y el extremo 3' del RNA ribosómico 16S. La proteína más grande de la subunidad 30S, S1, es la encargada de facilitar la interacción entre el mensajero y la subunidad menor del ribosoma (Laursen et al. 2005; Delvillani et al. 2011a; Qu et al. 2012a), se une al RNAm por las bases río arriba de SD y posteriormente translocando al mensajero a un sitio cercano de las bases complementarias de SD (Delvillani et al. 2011b; Qu et al. 2012b) y se considera como terminada la parte de iniciación cuando se unen el codón de inicio con su anticodón (Laursen et al. 2005).

J. Shine y L. Dalgarno (1975) encontraron la secuencia complementaria del RNAr 16S su extremo 3', la secuencia es **5'-AGGAGGU-3'** para *E. coli*. Se pensaba en una relación directa entre complementariedad y productos de síntesis, sin embargo, actualmente hay varios estudios que demuestran que no hay una correspondencia lineal entre el número de bases que hibridan y la eficiencia de traducción (Shultzaberger et al. 2001; Salis et al. 2009; Qu et al. 2012b). La estabilidad en la estructura secundaria del RNAm y el número de bases que separan al codón de inicio de SD también influyen en la velocidad con la que ocurre la iniciación, ya que a pesar de que algunos mensajeros posean la secuencia ideal de hibridación con 16S, pierden esa ventaja de unión cuando la estructura en la cual está contenida SD es muy rígida o muy alejada del codón de inicio (Salis et al. 2009; Qu et al. 2012b; Shultzaberger et al. 2001).

Estudios recientes han concentrado la atención en encontrar secuencias sintéticas que permitan aumentar la tasa de inicio de traducción dando lugar al incremento de producto. Estos estudios tuvieron varios enfoques como: optimizar el espacio entre SD y el codón de inicio (Shultzaberger et al. 2001; J. Starmer, A.Stomp, M. Vouk 2006), aumentar la vida media de los mensajeros (Carrier & Keasling 1997; Carrier & Keasling 1999), mejorar las interacciones entre SD-16S (de Smit & van Duin 1994; Salis et al. 2009).

III. Antecedentes

El desarrollo de nuevas tecnologías y herramientas en la ingeniería genética ha dado lugar a nuevos planteamientos para conocer más a detalle los mecanismos que controlan a los organismos vivos. Un ejemplo donde se usa este desarrollo es en el estudio del metabolismo, porque a partir de estas herramientas se pueden experimentar nuevas condiciones fisiológicas, aumentando o disminuyendo el producto genético en cuestión. El resultado de la investigación metabólica en organismos modificados con frecuencia ayuda a bajar el costo de producción de compuestos químicos, materiales, combustibles o medicamentos, que son cada vez más solicitados por la industria (Arpino et al. 2013).

Como se menciona en la introducción, para dilucidar el proceso de gen a proteína hay que comprender la función de las secuencias promotoras y de sitios de unión a ribosoma que guían los pasos de transcripción y traducción respectivamente, una manera de estudiarlas es de forma aislada, es decir, sin agentes ajenos a estas secuencias que intervengan en la expresión (Arpino et al. 2013), por esta razón se prefiere usar promotores y RBS sintéticos en sistemas fáciles de medir, para evitar las acciones de los factores de regulación externos que poseen las secuencias nativas (Rhodius & Mutalik 2010). Para comprender la manera en que estas secuencias sintéticas intervienen en las dos etapas de la síntesis de proteínas, hay que tener una caracterización de cada secuencia, identificar cuáles son las particularidades de cada secuencia y la manera en que esas particularidades se ven reflejadas en el producto (Bintu et al. 2005).

Se han sugerido varios genes para reportar la cantidad de proteína producida, estos genes regularmente codifican para enzimas (midiendo la actividad catalítica) o para proteínas fluorescentes (midiendo la fluorescencia). Los genes que se usan con más frecuencia son: β -galactosidasa (*lacZ*), cloranfenicol acetil transferasa (*cat*), luciferasa (*luc*) y proteína verde fluorescente (*gfp*) (Lu et al. 2004); la última mencionada es más común por sus características de: no requerir sustrato para fluorescer, no presentar toxicidad a la célula y permitir un monitoreo en tiempo real (Lu et al. 2004).

3.1 Evaluación de promotores y modelos propuestos para medir su eficiencia

Después del descubrimiento de las secuencias -10 y -35 (Pribnow 1975; Rosenberg & Court 1979), se han realizado experimentos degenerando la secuencia consenso (resultando en bibliotecas de promotores) o comparando promotores nativos con los hexámeros consenso (Rossi et al. 1983; Lu et

al. 2004; Mijakovic et al. 2005; Santos & Stephanopoulos 2008a), con el fin de comprender mejor cómo estas secuencias actúan en la expresión genética y cómo podrían usarse para la ingeniería genética.

3.1.1 Bibliotecas de promotores

Con el aumento de la manipulación en los flujos y reacciones del metabolismo, se han incrementado las de bibliotecas de promotores (Santos & Stephanopoulos 2008b). Las bibliotecas de promotores son un acervo de secuencias sintéticas, a partir de una secuencia molde a la cual se hacen modificaciones como cambio de tamaño en la región espaciadora (Peter Ruhdal Jensen 1998) o la longitud de todo el promotor, incluyendo la región de los hexámeros (Alper Hal et al. 2006). El promotor consenso se usa con frecuencia como secuencia molde para hacer bibliotecas, por la característica de ser un promotor constitutivo y fuerte (Rossi et al. 1983).

Rossi *et al.* (1983) reportan uno de los primeros experimentos donde a partir del promotor consenso se hicieron variaciones en la región espaciadora, midiendo los niveles de RNAm. Este trabajo no sólo confirma la alta eficiencia ya reportada del promotor consenso con sus 17 bases de separación, sino que también constituye a uno de los primeros acervos que ya podrían clasificarse como biblioteca de promotores, a pesar de que sólo tuvo 4 sujetos.

La biblioteca generada por Jensen y Hammer (1998) fue diseñada con el propósito de regular la expresión genética en *Lactococcus lactis* para su eventual uso en la industria. El enfoque de las mutaciones en la librería era modificar cualquier base a lo largo de la secuencia consenso. A pesar de tener en un principio muchas candidatas, sólo se quedaron con 38 (pero de amplio rango de diferencias, de 0.3 a más de 2000 unidades relativas de expresión) que reportaban la actividad del gen reportero β -galactosidasa. Como resultados destacados, descubrieron que el espacio que había entre -10 y -35 era clave para modular la expresión, porque con sólo las candidatas con modificaciones en este sitio se cubre prácticamente todo el rango de expresión reportado en este trabajo. Por otro lado, le dieron más validez a su biblioteca usando el mismo vector (pAK80) pero en *E. coli*, los resultados no correlacionaron muy bien en comparación con los de *L. lactis*, probablemente por cambios en el fondo genético.

De manera similar, se han generado otras bibliotecas con miras a la ingeniería metabólica. En el trabajo de Alper *et al.* (2006) no sólo generan una biblioteca sino que también demuestran su uso para el control genético. Al igual que en el ejemplo pasado, a partir de una secuencia se hacen

variaciones degeneradas por medio de PCR propenso a errores (Zaccolo & Gherardi 1999). La diferencia con esta biblioteca es que en vez de usar al promotor consenso como secuencia de partida, usaron al promotor $P_{L-\lambda}$ (proveniente de un bacteriófago) que es constitutivo. Otra diferencia es que usaron al gen *gfp* para caracterizar la biblioteca. Tras 14 horas de continua medición de densidad óptica (DO), fluorescencia y RNAm, se concluyó que los niveles de proteína y de mensajeros correlacionan; después se usaron algunos de estos promotores para reemplazar en cromosoma de *E. coli*, donde se reporta que la fuerza de estos promotores se conserva.

3.1.2 Modelos propuestos

Una de las primeras aproximaciones que se hicieron para encontrar relación entre fuerza y secuencia de promotor es el trabajo de Weller y Recknagel (1994). Ellos usaron el teorema de Markov (Almagor 1983) para analizar la interdependencia de las bases en los promotores y la correlación que tenían de acuerdo a su fuerza. Los resultados demostraron que la proporción de pirimidinas y purinas en la secuencia del promotor entre más se parezca a la proporción del promotor consenso más fuerza tendrá el promotor aunque, este primer acercamiento sólo se realizó con 14 promotores, por lo que no les permitió obtener más información.

Con la misma técnica y el mismo promotor molde usados por Alper *et al.* en Jensen *et al.* (2006) generaron una biblioteca con la cual evaluaron las mutaciones relacionadas con el aumento o disminución de fuerza en el promotor. La estrategia usada para analizar la biblioteca fue alinear las secuencias y encontrar cuáles bases aparecían con más frecuencia en qué posiciones y en cuál grupo (promotores fuertes o promotores débiles); esta misma estrategia la siguieron De Mey *et al.* (2007) confirmando con su propia biblioteca que esos métodos sí podían predecir cuáles secuencias tenían más actividad. Sin embargo, estos últimos autores sugieren crear otro método para analizar las secuencias desde otro punto de vista, debido a que el primero sólo es cualitativo y necesita de la formación de grupos de promotores fuertes y débiles. Por lo anterior, el método de De Mey *et al.* (2007) consiste en hacer una regresión de mínimos cuadrados parciales con los datos de la biblioteca, ligando una matriz de variables independientes (los promotores) con una de variables dependientes (la intensidad de fluorescencia de cada uno). La ventaja de este método es que es cuantitativo y no tiene ningún paso subjetivo que pueda alterar el resultado. Este modelo fue posteriormente usado en De Mey *et al.* (2010) obteniendo los resultados esperados en el aumento de fuerza de sus promotores.

Por otro lado, también hay modelos que tienen por base la información adquirida de promotores nativos (Zehua Chen 2007) o la termodinámica de la transcripción (Brewster et al. 2012). El primer modelo se hizo con información recopilada (599 promotores), con ayuda de la teoría de la información distinguieron las posiciones y las bases más importantes en los promotores, sin embargo, este modelo no se comprobó en un sistema vivo. El modelo de termodinámica en la transcripción se realizó con sólo 18 promotores con mutagénesis dirigida, se elaboraron matrices (como las que usan en Kinney *et al.* (2010)) donde indicaban la secuencia y la energía requerida por σ^{70} para unirse al promotor, como conclusión se demostró que su modelo sí predijo un aproximado de los niveles de expresión.

3.2 Evaluación de RNAm, RBS y modelos propuestos para medir su eficacia

Se sabe que un paso necesario en la producción de proteínas es el inicio de la traducción, que tiene una estrecha relación con la estructura y dinámica del RNAm (Hayashi et al. 2012). Como se mencionó en la introducción, el inicio de la traducción es la etapa donde puede afectar más la síntesis de manera positiva o negativa (Pon & L. 1990; Laursen et al. 2005; Hayashi et al. 2012). Tras el descubrimiento de la secuencia SD (Dalgarno 1975), varios estudios se enfocaron a investigar cómo los componentes no codificantes de un RNAm (particularmente la denominada RIT) interaccionaban con los ribosomas para dar lugar al inicio de la traducción, especialmente la relación entre complementariedad de SD con el extremo 3' del 16S, la estructura secundaria del RNAm y la dinámica molecular de unión de los componentes (Ringquist et al. 1992; de Smit & van Duin 1994; Carrier & Keasling 1997; Carrier & Keasling 1999; Shultzaberger et al. 2001; J. Starmer, A. Stomp, M. Vouk 2006; Salis et al. 2009; Hayashi et al. 2012).

3.2.1 Diferentes RBS y RNAm evaluados

A diferencia de los estudios realizados con los promotores, no es común generar bibliotecas con sitios de unión a ribosoma, sino generar candidatas por mutagénesis dirigida. Estudios realizados por Ringquist *et al.* (1992) revelaron que el aumento en la complementariedad de la secuencia SD con el 16S del RNAr tiene un efecto positivo en la tasa de traducción. La secuencia corta que usaron era **5'-AAGGA-3'** y la cambiaron por **5'-UAAGGAGG-3'**, sin embargo, si esta secuencia no estaba ubicada entre las 10 y 3 bases antes del codón de inicio de traducción, los niveles de síntesis disminuyeron. Este resultado se repitió con la secuencia corta, lo que indica que sólo una parte de la tasa de

traducción depende de la complementariedad entre SD-16S. Este experimento lo realizaron en *E. coli* usando al gen *lacZ* como reportero.

Siguiendo la misma línea, se han reportado otros trabajos (de Smit & van Duin 1994; Carrier & Keasling 1997) en los cuales dirigen las mutaciones al extremo **5'-no codificante (5'-NC)** para tener diferentes estructuras en el RNAm y diferentes grados de complementariedad con el 16S. En estos estudios no sólo se reafirma que la iniciación sólo depende de la hibridación sino que también encontraron que las diferentes estructuras secundarias del RNAm tienen influencia en la degradación del mismo, así como en la velocidad con la que se forma el complejo ribosomal. Sin embargo, todavía no hay suficiente información que permita discernir con sólo las estructuras entre los mensajeros con más o menos tasa de inicio de traducción (Carrier & Keasling 1999).

3.2.2 Modelos propuestos para la traducción

Debido a que la complementariedad de las bases (SD-16S) no tenía relación directa con el nivel de eficacia de traducción, se analizaron a la par secuencias existentes y secuencias sintéticas. Se analizaron varios parámetros como su estructura secundaria o sus posibles interacciones con el ribosoma. Con esta nueva información se han generado algunos modelos para la predicción del inicio de la traducción (Shultzaberger et al. 2001; J. Starmer, A. Stomp, M. Vouk 2006; Salis et al. 2009).

Shultzaberger *et al.* (2001) alinearon secuencias en la región SD, y analizaron la relación tasa de traducción y secuencia por medio de la teoría de la información. Los resultados del trabajo mostraron una importante relación entre la ubicación de SD y los niveles de traducción, localizaron a las mejores candidatas cuando tenían un espacio de separación entre 4 y 9 bases de diferencia con respecto al codón de inicio, incluso reportaron que el espacio de separación es más relevante para los procesos de traducción que la misma secuencia SD. El propósito de este trabajo fue diseñar un modelo capaz de catalogar cualitativamente un mensajero con base a la localización de SD y el codón de inicio de traducción. Sin embargo, este modelo aun carece de precisión porque no toma en cuenta otras características del RNAm.

Continuando con el trabajo de identificar las mejores secuencias de RBS en el RNAm, Starmer *et al.* (2006) recolectaron datos reportados en bancos de información de diferentes procariontes. Este trabajo concluyó que la eficiencia de los mensajeros puede deberse a la ubicación del sitio de inicio de traducción (SD y codón de inicio) y el grado de hibridación junto con la estabilidad de la estructura secundaria del RNAm. En este trabajo se describen modelos que ayudan a determinar la estabilidad

de la estructura secundaria, y en función de esto calculan la tasa de hibridación con el ribosoma. La probabilidad de hibridación la calcularon mediante la ΔG entre las interacciones del mensajero y el ribosoma. Este trabajo ilustra una forma de encontrar el sitio de inicio de traducción en secuencias nuevas y las secuencias que cualitativamente tienen más potencial para generar más producto.

También hay modelos propuestos para predecir cuantitativamente los niveles de traducción (Salis et al. 2009; Kosuri et al. 2013). El trabajo de Salis *et al.* (2009) está basado en la termodinámica de los procesos del inicio de la traducción y a diferencia de los modelos anteriormente descritos, este no sólo calcula la energía requerida para hibridar a SD con la subunidad 16S, también calcula las interacciones anteriores así como el grado de complementariedad de SD y la unión del primer RNA de transferencia (RNAt).

El cálculo de estas interacciones omite las posibles preferencias que podría tener la proteína S1 por alguna clase de mensajeros, lo que sí considera es la energía que se libera cuando se unen SD-16S ($\Delta G_{\text{RNAm:RNAr}}$), la energía liberada por la unión del codón de inicio con el RNAt (ΔG_{Start}), la energía requerida por no cumplir con el espacio ideal entre SD y el codón de inicio ($\Delta G_{\text{spacing}}$)-esta es una penalización generada por deformación del ribosoma al tener una distancia demasiada o muy poca separación entre SD y el codón de inicio-, la energía requerida para estabilizar la estructura secundaria del RNAm (ΔG_{RNAm})-para este parámetro entre más energía se necesite es más sencillo linearizar la secuencia resultando en un menor gasto energético para la hibridación- y por último, la energía necesaria para quitar cualquier estructura secundaria en el mensajero después del apareamiento entre SD-16S ($\Delta G_{\text{standby}}$), todos los parámetros son procesados por la fórmula:

$$G_{\text{total}} = \Delta G_{\text{RNAm:RNAr}} + \Delta G_{\text{Start}} + \Delta G_{\text{spacing}} - \Delta G_{\text{standby}} - \Delta G_{\text{RNAm}}$$

A partir del programa de NUPACK (Dirks et al. 2007) se usaron algoritmos junto con los parámetros de energía con el diseño de estructura secundaria del programa Mfold 3.0 (Mathews et al. 1999). El modelo se probó con secuencias existentes (tomadas de un catálogo de Biobricks) así como con secuencias sintéticas (producidas por ellos), se midió la concentración de proteínas fluorescentes con citómetro de flujo. Los resultados demostraron que el modelo de Salis *et al.* (2009) predice con alto grado de exactitud la tasa de traducción, tanto en secuencias sintéticas como en las nativas.

3.3 La combinación de promotores y RBS en una biblioteca

En el trabajo de Kosuri et al. (2013) usaron bibliotecas independientes de promotores y de RBS, generaron permutaciones con ambas regiones y midieron los efectos con los que cada una contribuía al producto final. Con un vector de mediano número de copias (pGERC) y con proteínas fluorescentes (*gfp*, mCherry), los autores midieron la cantidad de proteína producida, copias de mensajeros y del gen estructural, usando técnicas de citometría de flujo, RNAseq y DNAseq respectivamente. Los resultados mostraron que el 92% del RNAm era originado por la fuerza del promotor mientras el 8% restante le dieron atribución a una parte (4%) al RBS y la estabilidad que confiere al mensajero aumentando su vida media, y la otra a causas desconocidas; en este mismo trabajo también proponen un modelo que predice la cantidad de proteína si se conocen los niveles de transcripción y la eficiencia de traducción, este modelo también podría usarse para conocer estos dos parámetros, si se sabe la cantidad total de producto y alguno de los parámetros antes mencionados.

IV. Justificación

Un tema trascendental en Biología es el funcionamiento fisiológico de los organismos y su regulación, para develar estos procesos se han usado diversas herramientas (Arpino et al. 2013). A partir de secuencias que intervienen en el control de la expresión genética (Pribnow 1975; Rosenberg & Court 1979; Amann, Egon. Ochs, Birgit. and Abel 1988) y con una metodología similar a Sondek y Shortle (1992) en este proyecto planteamos la generación y caracterización de una biblioteca de promotores y sitios de unión a ribosoma en *E. coli* que nos permita posteriormente manipular la síntesis de proteínas con un enfoque a la ingeniería de vías metabólicas.

Este acervo en comparación con las técnicas clásicas, ofrece la posibilidad de modular la expresión sin necesidad de un plásmido que represente un gasto energético extra, entonces, estas secuencias pueden reemplazar secuencias nativas del cromosoma bacteriano.

V. Hipótesis

Las diferentes mutaciones sobre una secuencia consenso generará una gama diversa de secuencias candidatas con fuerza de expresión diferente. Cualquier mutación que altere la secuencia consenso del promotor o que agregue bases entre SD y el codón de inicio disminuirá el nivel de expresión.

VI. Objetivos

6.1 Objetivo general

Generar y caracterizar una biblioteca de secuencias promotoras y sitios de unión a ribosoma por medio de un nuevo método donde se generaron inserciones aleatorias de bases.

6.2 Objetivos particulares

- Diseñar una secuencia molde que contenga un promotor y un RBS
- Generar variantes de la secuencia molde por medio de síntesis química de oligonucleótidos, basado en mutaciones por inserciones
- Clonar las secuencias de doble hebra en el plásmido *Superglogfp*, para generar una biblioteca de secuencias promotoras y RBS
- Transformar *Escherichia coli* W3110 con la biblioteca generada

- Elegir las cepas que presenten fluorescencia y los patrones de digestión correctos del plásmido
- Caracterizar la velocidad de crecimiento y fluorescencia de cada candidata
- Secuenciar el promotor y RBS de las cepas candidatas
- Analizar los resultados y las secuencias con al menos un modelo propuesto anteriormente

VII. Materiales y métodos

7.1 Cepa, plásmido y medios

La cepa utilizada en todos los pasos de este trabajo fue *Escherichia coli* W3110. Esta cepa se caracteriza por tener un tiempo corto de generación y es incapaz de intercambiar información genética de manera horizontal (Hayashi et al. 2006).

El plásmido donde se clonó la biblioteca fue el vector conocido como pT4GFPMlu, se caracteriza por tener resistencia a kanamicina (kn), sitio de replicación pUC (alto número de copias) y un gen reportero *sgGFP* que es una proteína diferente de la GFP nativa porque tuvo una selección para ser más soluble, plegarse a 37°C, y emitir y excitar a 509_{nm} y 474_{nm}, respectivamente (Flores-Ramirez et al. 2007).

A continuación describo los medios de cultivo empleados:

- Medio bajo en sales YENB: 7.5 g/L de extracto de levadura y 8 g/L caldo nutritivo
- Medio Luria-Bertani (LB): 10 g/L de triptona, 5 g/L de extracto de levadura y 5 g/L de NaCl

7.2 Diseño del oligonucleótido molde

Las secuencias de promotor y de unión a ribosoma se eligieron con base a lo reportado en otros trabajos.

La secuencia del promotor fue el par de hexámeros del promotor consenso **5'-TTGACA-3'** y **5'-TATAAT-3'** con 17 bases de separación, para ser más específico es la misma secuencia reportada en Rossi *et al.* (1983)

5'-TTG ACA ATT AGT TAA CTA TTT GTT ATA AT-3'

Para el sitio de unión a ribosoma se usó la secuencia del gen *lacZ* que ya había sido usada en otros sistemas de expresión sintéticos (Amann et al. 1988), la diferencia es que las bases río arriba del RBS

fueron cambiadas por un programa mediante la plataforma de <http://faculty.ucr.edu/~hanneman/nettext/>, debido a que esa secuencia (entre -10 y RBS) corresponde a una secuencia operadora de *lacI*. Además del RBS también se respetaron las bases de separación con el codón de inicio

5'-GTA TTC CCT TAC CGG CCC TTT CCA TGC GCG CAC AGG AAA CAG A-3'

La secuencia molde total es de 72 bases que se representan a continuación:

**5'-TTG ACA ATT AGT TAA CTA TTT GTT ATA ATG TAT TCC CTT ACC GGC CCT TTC CAT GCG CGC ACA GGA AAC
AGA-3'**

(Los sitios subrayados corresponden a -10, -35 y RBS, respectivamente).

Con la secuencia molde establecida se sintetizaron de manera química una serie de variantes que constituyen la biblioteca de inserciones aleatorias, añadiéndole 12 y 20 bases en cada extremo para poner los sitios de restricción correspondientes a *BsaHI* y *NcoI*, que sirvieron para ligar al plásmido, la secuencia final sobre la que se hicieron las mutaciones es la siguiente:

**5'-TCG TCG GAC GCC TTG ACA ATT AGT TAA CTA TTT GTT ATA ATG TAT TCC CTT ACC GGC CCT TTC CAT GCG
CGC ACA GGA AAC AGA CCA TGG TCG TCG TCG TC-3'**

Esta secuencia cuenta con 104 bases, las subrayadas corresponden a los sitios *BsaHI* y *NcoI* respectivamente

7.3 Generación de la biblioteca

La estrategia que se siguió para generar la biblioteca fue polimerizar de manera clásica cada nucleótido descrito del oligonucleótido molde y los extremos añadidos pero, después de cada base del oligonucleótido molde se agregó una mezcla contaminante de las 4 bases en condiciones equimolares y diluidas (0.25 μ M de cada base, 1 μ M el total de la solución contaminante). Este método está basado en los trabajos realizados en Hermes et al. (1989) y en Sondek & Shortle (1992) donde también insertan bases de manera aleatoria. En la **Figura 3** se muestra un ejemplo de cómo se construyeron las variantes de la biblioteca

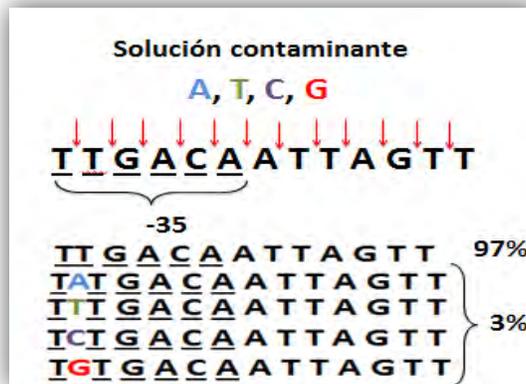


Figura 3 Generación de inserciones en la biblioteca. Muestra un ejemplo de cómo se generaron las variantes de la secuencia molde. La solución contaminante fue una mezcla hecha por la Unidad de Síntesis del Instituto de Biotecnología, UNAM. Las 4 bases estaban en las mismas concentraciones por lo que la probabilidad de inserción es la misma para cualquier base.

7.4 Clonación de la biblioteca en el plásmido y transformación

7.4.1 Complementación de la hebra sencilla de oligonucleótidos por PCR

El resultado del paso anterior fue una serie de oligonucleótidos de una sola hebra, por este motivo se complementaron las hebras por una reacción de PCR. Para realizar esta reacción se usó el siguiente oligonucleótido:



Este primer tiene 20 bases y una Tm de 69.8°C. Las condiciones para realizar la PCR fueron las siguientes:

Componente	Volumen por reacción de 100 µL	Concentración final
H ₂ O	76.21 µL	
Buffer 10x (NH ₄) ₂ SO ₄	10.00 µL	1x
MgCl ₂	6.00 µL	1.5mM
dNTPs	2.00 µL	0.2 mM por base
Oligonucleótido mutagénico*	4.16 µL	100 pmol
Oligonucleótido complementario	1.13 µL	100 pmol
Taq recombinasa	0.50 µL	2.5 unidades

*El oligonucleótido mutagénico es el producto entregado por la unidad de síntesis, es decir, la mezcla de oligonucleótidos derivados del oligonucleótido molde

Para realizar la PCR y los pasos posteriores, se usaron 3 tubos preparativos con 100 μL (con los componentes mencionados arriba). Antes de meter la reacción al termociclador, los tubos de 100 μL se dividieron a la mitad, dejando en total 6 tubos con 50 μL .

El proceso de amplificación se realizó en un termociclador marca Bio-Rad modelo "C1000 thermocycler", la reacción se dividió en 3 etapas:

1. Un ciclo de-desnaturalización a 95°C por 2 minutos
2. 10 ciclos de amplificación, cada ciclo cuenta con un paso de desnaturalización (95°C por 1 minuto), uno de alineamiento (55°C por 1 minuto) y uno de polimerización (72°C por 1 minuto)
3. Un ciclo de extensión final (72°C por 10 minutos)

7.4.2 Digestión del DNA de las candidatas y el plásmido receptor

El producto de PCR se purificó con el kit de purificación "High Pure PCR Product Purification" de la marca Roche, la purificación se realizó con los 300 μL de la reacción. La concentración final de DNA quedó en 63.1 ng/ μL , en un volumen final de 100 μL .

A estas candidatas se les hizo una doble digestión en los extremos correspondientes a *Bsa*HI y *Nco*I. La primera digestión tuvo las siguientes particularidades:

Componente	Volumen por reacción de 70 μL	Concentración final
H ₂ O	24 μL	
Buffer 4 10x	7 μL	1x
Albúmina 10x	7 μL	1x
Enz. <i>Bsa</i> HI	2 μL	20 unidades
DNA (secuencia candidata)	30 μL	2 μg

La enzima y los buffers se usaron de la marca Biolabs. La reacción duró 2 horas a 37°C.

La segunda digestión:

Componente	Volumen por reacción de 150 μL	Concentración final
H ₂ O	66.0 μL	
Buffer 10x Tango	11.0 μL	1x
Enz. <i>Nco</i> I	2.50 μL	25 unidades
DNA (secuencia candidata)	70.0 μL	2 μg

El buffer y la enzima usados eran de la marca Thermo Scientific Fermentas. La reacción duró 2 horas a 37°C.

Se volvió a purificar esta biblioteca pero ahora con el kit de Bio Basic Inc “BS 363 EZ-10 Spin Column PCR Purification kit”.

Por otro lado, se extrajo el plásmido supergloGFP por la técnica de miniprep de lisis alcalina. Se comprobó la identidad del plásmido por una digestión con *BsaHI* (que lineariza el plásmido por ser un sitio único de corte, en la **Figura 4** se muestra un mapa del plásmido) y posteriormente se hizo otra digestión en el sitio *NcoI*. Para la primera digestión estas son las especificaciones

Componente	Volumen por reacción de 70 µL	Concentración final
H ₂ O	39 µL	
4 Buffer 10x	7 µL	1x
Albumina 10x	7 µL	1x
Enz. <i>BsaHI</i>	2 µL	20 unidades
DNA (supergloGFP)	15 µL	2 µg

La enzima y los buffers se usaron de la marca Biolabs. La reacción duró 2 horas a 37°C.

Para la segunda reacción:

Componente	Volumen por reacción de 150 µL	Concentración final
H ₂ O	66.50 µL	
Buffer 10x Tango	11.00 µL	1x
Enz. <i>NcoI</i>	2.50 µL	25 unidades
DNA (supergloGFP)	70.00 µL	2 µg

El buffer y la enzima usados eran de la marca Thermo Scientific Fermentas. La reacción duró 2 horas a 37°C.

A este plásmido doblemente digerido se le hizo una purificación de banda con el kit de Roche “High Pure PCR Product Purification”

7.4.3 Ligación de la biblioteca al plásmido

Con las secuencias de candidatas purificadas y el vector linearizado, se realizó la ligación en las siguientes condiciones:

Componente	Volumen por reacción de 150 µL	Concentración final
H ₂ O	1 µL	
Buffer 10x T4 ligasa	3 µL	1x
T4 ligasa	1 µL	1 unidad
DNA (supergloGFP)	25 µL	200 ng
DNA (secuencia candidata)	1 µL	15.1 ng

El buffer y la enzima usados eran de la marca Thermo Scientific Fermentas. La reacción duró 38 minutos a 25°C.

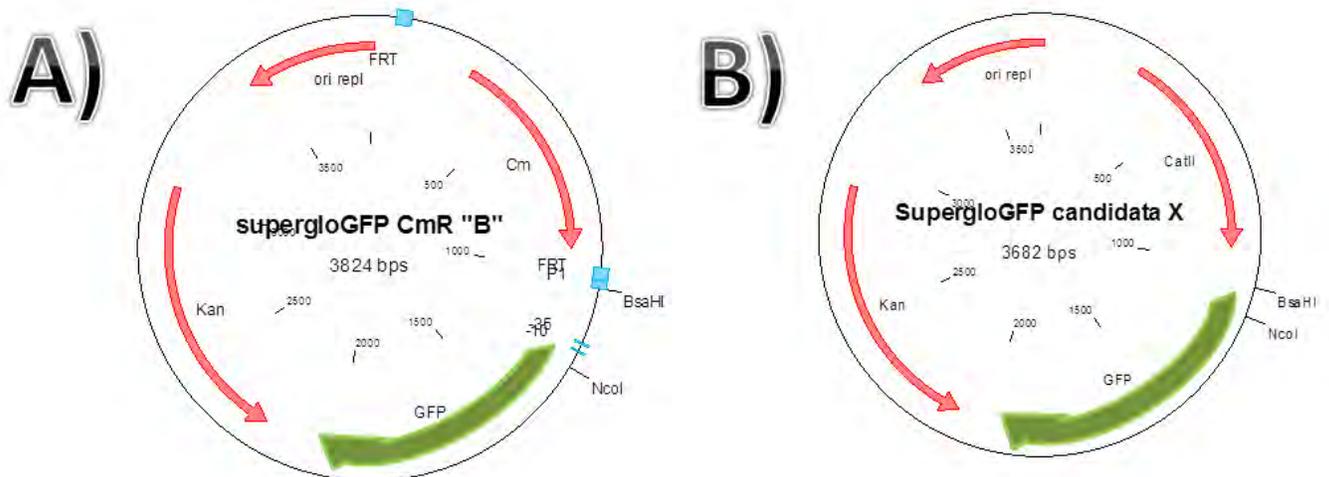


Figura 4 Mapas de los plásmidos usados. **A)** Es el plásmido original que se usó para generar la biblioteca. Los sitios únicos *NcoI* y *BsaHI* marcados fueron usados para reemplazar por la biblioteca de secuencias. **B)** Un ejemplo del tamaño aproximado en el que quedaron las secuencias.

7.4.4 Transformación

El plásmido completo (con el inserto de candidata) se transformaron en *E. coli* por electroporación, con un pulso eléctrico (2,500 volts) en celdas de la marca Eppendorf, permitió la entrada del plásmido y se recuperaron en 1 mL de medio YENB a 37°C por 1 hora. Para esta electroporación se usaron células electrocompetentes de *E. coli* W3110 preparadas previamente en el mismo medio YENB; se ocupó un volumen de 100 μ L de estas células y 4 μ L del producto de ligación.

7.5 Selección de candidatas

Se realizó un tamizaje por cultivo en placa, en medio LB sólido con los antibióticos Cm (20 μ g/mL) y Kn (30 μ g/ μ L), ahí se sembraron diferentes volúmenes (200, 100, 50 μ L) de las células transformadas. De las colonias que crecieron en caja, sólo se eligieron las que en presencia de luz oscura presentaban fluorescencia, porque esto indicaba que la secuencia candidata insertada en el plásmido era funcional, es decir, tanto el promotor como el RBS intervenían de manera positiva en la generación de productos.

7.6 Caracterización de la biblioteca

7.6.1 Secuenciación de las candidatas

A las colonias seleccionadas en placa se les hizo miniprep por lisis alcalina, del plásmido resultante se secuenciaron los sitios promotores, RBS y el gen estructural de *gfp*, descartando a las secuencias repetidas de candidatas y a las que tenían un margen de error elevado en el cromatograma, es decir, los picos de lectura denotaban

7.6.2 Determinación del crecimiento y fluorescencia

Luego de haber seleccionado a las candidatas que presentaban fluorescencia y a la par de la determinación de la secuenciación, se realizaron cinéticas de fluorescencia y crecimiento. Durante 16 horas se tomó muestra de cada cultivo (cada hora en las primeras 12 horas y cada 2 horas después), el cultivo se realizó en matraces *baffleados* con capacidad de 125 mL con 25 mL de medio LB por cada cultivo junto con los antibióticos de selección (Kanamicina 30 $\mu\text{g}/\mu\text{L}$ y Cloramfenicol 20 $\mu\text{g}/\text{mL}$), se incubaron con agitación 300 rpm a 37°C. Los cultivos se iniciaron con un primer inóculo de 0.05 $\text{DO}_{600\text{nm}}$, la procedencia de este inóculo es de un cultivo previo 14 horas en 3 mL de medio LB con y Kn (30 $\mu\text{g}/\mu\text{L}$) Cm (20 $\mu\text{g}/\text{mL}$), en tubos de ensayo con agitación (300 rpm) a 37°C.

Durante la etapa de crecimiento se tomaron muestras para registrar la densidad óptica y la fluorescencia (unidades relativas de fluorescencia, URF). Se cuantificó el crecimiento bacteriano midiendo la absorbancia de cada cultivo a 600 nm con un espectrofotómetro (Beckman DU-70) y un lector Klett (Scienceware Klett colorimeter), se tomó 1 mL de cultivo cuando la lectura era menor a 0.7 $\text{DO}_{600\text{nm}}$ después se hicieron diluciones 1:10. Por otro lado la fluorescencia se midió en un espectrofluorímetro (Perkin Elmer LS 50B) a 474 nm de excitación y 504 nm de emisión con una ventana de apertura de 5 nm; en cada tiempo de muestra se recolectó 1 mL, el cual se centrifugó por 5 minutos a 13,200 rpm para limpiar las células del medio LB (debido a que este medio tiene fluorescencia *per se* (De Mey et al. 2007)) a continuación se resuspendió la pastilla sobrante en 1 mL de solución PBS (8 g/L de NaCl, 0.2 g/L de KCl, 0.24 g/L de Na_2HPO_4 , todo aforado en H_2O a 1 L en pH 7.4), se volvió a centrifugar pero esta vez sólo por 1 minuto a 13,200 rpm y al final sólo se agitó en Vortex sin tirar el medio, estas muestras fueron almacenadas a 4°C por 1 día (este día fue sólo por el tiempo de apartado del equipo para medir la fluorescencia), este proceso no altera la estructura de la proteína verde fluorescente.

Otro control fue medir el crecimiento y fluorescencia de la misma cepa con el mismo plásmido pero que en lugar de tener las variantes de promotor y RBS del resto de la biblioteca, se le insertó una región del gen estructural *lacZ*, de la región centro del gen 250 bases río arriba y 250 río abajo, la secuencia agregada es la siguiente:

5'- GAT GAT CCG CGC TGG CTA CCG GCG ATG AGC GAA CGC GTA ACG CGA ATG GTG CAG CGC GAT CGT AAT CAC CCG AGT GTG ATC ATC TGG TCG CTG GGG AAT GAA TCA GGC CAC GGC GCT AAT CAC GAC GCG CTG TAT CGC TGG ATC AAA TCT GTC GAT CCT TCC CGC CCG GTG CAG TAT GAA GGC GGC GGA GCC GAC ACC ACG GCC ACC GAT ATT ATT TGC CCG ATG TAC GCG CGC GTG GAT GAA GAC CAG CCC TTC CCG GCT GTG CCG AAA TGG TCC ATC AAA AAA TGG CTT TCG CTA CCT GGA GAG ACG CGC CCG CTG ATC CTT TGC GAA TAC GCC CAC GCG ATG GGT AAC AGT CTT GGC GGT TTC GCT AAA TAC TGG CAG GCG TTT CGT CAG TAT CCC CGT TTA CAG GGC GGC TTC GTC TGG GAC TGG GTG GAT CAG TCG CTG ATT AAA TAT GAT GAA AAC GGC AAC CCG TGG TCG GCT TAC GGC GGT GAT TTT GGC GAT ACG -3'

Para la secuenciación se usaron 2 oligonucleótidos diferentes para flanquear la región insertada, estos cebadores tenían el objetivo de monitorear que no tuvieran mutaciones en el gen estructural, la secuencia de dichos oligonucleótidos es la siguiente:

Sentido forward **5'- TTA TACGCA AGG CGA CAA GG-3'**

Sentido reverse **5'- GCC GCA CTA GTC TCG ACT TA-3'**

7.7 Análisis de resultados

Una vez que se registró la eficiencia de cada candidata y con base en los modelos propuestos por Brewster *et al.* (2012) y Salis *et al.* (2009), se analizaron las secuencias promotoras y RBS respectivamente. Se escogieron dichos modelos porque ambos usan un parámetro objetivo, es decir, que no involucra el número de muestras de la biblioteca ni la clasificación subjetiva que el espectador le pueda dar, ellos usan modelos termodinámicos que reportan un alto nivel de certeza en la predicción de interacciones, esto aunado a que hasta ahora son los modelos más recientes para cada tipo de secuencia.

Para el análisis de las secuencias promotoras, se empleó una matriz que calcula la energía que requiere una RNAPol para unirse a una secuencia de promotor; la matriz reportada en Brewster *et al.* (2012) se usó por primera vez en Kinney *et al.* (2010). La matriz se originó de los resultados de una biblioteca de promotores y su análisis computacional (Kinney *et al.* 2010), se comprobó la eficiencia predictiva generando secuencias según la matriz (Brewster *et al.* 2012). Si bien este modelo no indica

cuantitativamente la actividad de un promotor, sí predice cualitativamente cuál será mejor promotor. La matriz original tiene una separación de 18 bases entre la región hexamérica pero, debido a que nuestra secuencia molde de biblioteca sólo tiene una separación de 17 bases, omitiremos la -30 de la matriz original (como se muestra en la **tabla 1**) porque esta matriz ha demostrado con varios ejemplos que a pesar de no usar una o dos bases conserva sus propiedades; un ejemplo es el promotor -10 extendido, que conserva el bajo gasto energético en esa zona, además, para nuestra secuencia molde resulta más parsimonioso omitir la base de aquel sitio, ya que si no tomamos en cuenta la base del otro extremo el gasto energético sería de $1.6 k_B T$ y omitiendo el que sugerimos el gasto sería de $-0.6 k_B T$ ($1 \text{ Kcal/mol} = 1.62 k_B T$).

	-10										-35																	
	-12	-11	-10	-9	-8	-7	-12	-11	-10	-9	-8	-7	-12	-11	-10	-9	-8	-7	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3
A	0.4	0.2	0.2	0.2	0.2	0.2	0.4	0.2	0.2	0.2	0.2	0.2	0.4	0.2	0.2	0.2	0.2	0.2	0.4	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
C	0.8	0.4	1.4	1	0	0.6	0.8	1	0.2	0.4	0.2	1	1.4	1	0.2	0.2	0.2	0.2	1.4	1	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
G	0.4	0.2	-0.2	0.2	0.2	0.4	1	0.8	0.4	0.6	0.6	0.8	-0.2	0.2	0.2	0.2	0.2	0.2	-0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0
T	-0.2	-0.2	0.6	0.6	-0.2	0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2

Tabla 1 Matriz de energía requerida por la RNAPol de acuerdo a los nucleótidos y su posición en la secuencia. En la parte superior se indican los sitios de la secuencia con respecto al sitio de inicio de transcripción, en la parte lateral del lado izquierdo están las bases que podrían ocupar ese lugar.

En contraste con el modelo de promotores, el modelo de RBS sí tiene un sistema de predicción que permite clasificar cuantitativamente las secuencias. El modelo se fundamenta en la estructura secundaria del RNAm y la dinámica de interacción que tiene con el ribosoma (Mathews et al. 1999; Dirks et al. 2007). Los parámetros que se consideran en el modelo incluyen las siguientes suposiciones: El RBS no se sobrepone con el codón de inicio, sólo hay un marco abierto de lectura por cada secuencia codificante, los estados intermedios de la subunidad 30S no son considerados relevantes para las interacciones y no se considera ningún aporte de preferencia que pudiera tener la proteína S1 por algún mensajero; asimismo con referencias previas sobre el plegamiento secundario e interacción de RNAm con el RNAr (programa Mfold 3.0 con sus parámetros a 37°C (Mathews et al. 1999) y el algoritmo NUPACK (Dirks et al. 2007) se elaboró una plataforma capaz de valorar los estados energéticos de la secuencia codificante y la subunidad pequeña del ribosoma, particularmente de:

- $\Delta G_{\text{RNAm:RNAr}}$: La energía que se libera cuando se unen SD-16S
- ΔG_{Start} : La energía liberada por la unión del codón de inicio con el RNAr
- $\Delta G_{\text{spacing}}$: La energía requerida por no cumplir con el espacio ideal entre SD y el codón de inicio, esta penalización es ocasionada por alterar esta distancia, podría tener efecto en la deformación del ribosoma
- ΔG_{RNAm} : La energía requerida para estabilizar la estructura secundaria del RNAm, entre más energía se necesite es más sencillo linearizar la secuencia, resultando en un menor gasto energético para la hibridación, y por último
- $\Delta G_{\text{standby}}$: La energía necesaria para quitar cualquier estructura secundaria en el mensajero después del apareamiento entre SD-16S

Todos los parámetros son procesados en la siguiente fórmula,

$$G_{\text{total}} = \Delta G_{\text{RNAm:RNAr}} + \Delta G_{\text{Start}} + \Delta G_{\text{spacing}} - \Delta G_{\text{standby}} - \Delta G_{\text{RNAm}}$$

La plataforma para caracterizar RBS tiene por dirección web <https://www.denovodna.com/software/reverse> (versión 1.1). Se ajustaron las variables del programa en: secuencia del 16S RNAr de *E. coli* str. K-12 subtr W3110 (ACCTCCTTA) y 35 bases antes del codón de inicio y 35 bases después del codón de inicio. Los resultados de este programa están medidos en

“au” que significa el promedio de proteínas fluorescentes por el origen de cómo se construyó el programa (Salis et al. 2009).

VIII. Resultados y discusión

Se logró generar una biblioteca de promotores y RBS a partir de una secuencia molde inicial, mediante la inserción de bases aleatoria. Se descartaron las candidatas que presentaban inconsistencias en la secuenciación (con picos poco definidos del cromatograma (datos no incluidos)) o las que tenían desviaciones más grandes a 40% en las cinéticas de fluorescencia. Al final de los tamizajes de selección, conservamos 14 sujetos que representan diversos niveles de expresión.

8.1 Generación de la biblioteca

El método que usamos para generar la biblioteca se basa en un proceso de variación que incluye las secuencias del promotor y la del RBS, esta diversidad se da a partir de la secuencia molde, a esta se le hicieron inserciones aleatorias. La secuencia molde escogida tiene el promotor consenso tipo σ^{70} de *E. coli* y el RBS de *lacZ*:

5'-TTG ACA ATT AGT TAA CTA TTT GTT ATA ATG TAT TCC CTT ACC GGC CCT TTC CAT GCG CGC ACA GGA AAC
AGA-3'

Los sitios subrayados corresponden al promotor (-35y -10 respectivamente) y RBS.

Un estudio reciente que agrupa ambos sitios en una biblioteca (Kosuri et al. 2013), es un trabajo que elige por separado ambas secuencias de bibliotecas descritas, hacen permutaciones para medir la influencia de cada secuencia tanto en los procesos de transcripción como de traducción. Contrastando nuestro acervo con los anteriores reportados, encontramos que los otros realizan sus variantes por la técnica de PCR propenso a errores (Mijakovic et al. 2005; De Mey et al. 2007), mientras que de los pocos reportes que hay de las bibliotecas de RBS, usan la estrategia de mutagénesis dirigida con un enfoque de cambio conformacional en la estructura secundaria del RNAm (Carrier & Keasling 1999).

Después de la síntesis de los oligonucleótidos de la biblioteca, se realizó una reacción de PCR para complementar la cadena sencilla. Como se muestra en la **Figura 5** se corrió un gel de acrilamida comprobando la diferencia entre los oligonucleótidos de hebra sencilla y los de doble cadena. Este

gel ayuda a diferenciar los tipos de secuencia por el tamaño final de corrimiento, el producto de DNA con una única cadena tiene menos peso y corre más que el de doble cadena.



Figura 5 Gel de comprobación del DNA de doble hebra de las candidatas de la biblioteca. Se muestra un gel de acrilamida nativo al 10% donde se corrió el oligonucleótido de hebra sencilla puro (recién salido de la síntesis química) (carril 1) y el producto de PCR con el que se complementó esa hebra sencilla por una de doble cadena (carril 2).

Con el producto de doble cadena se hicieron ligaciones al plásmido supergloGFP. La ligación se realizó por medio de los sitios de restricción *Bsa*HI y *Nco*I. Después de la ligación, se volvieron a hacer digestiones para ver que cada candidata conservaba aún los sitios de corte, como se muestra en la **Figura 6**, que es una muestra de un plásmido que se digirió en ambos sitios y de manera individual. Estas pruebas se hicieron para confirmar que la ligación fue exitosa con el peso molecular esperado (3500 pb). La digestión con un sitio lineariza el plásmido y con 2 pierde un fragmento (el que incluye al promotor y al RBS) de 100 pb aproximadamente.

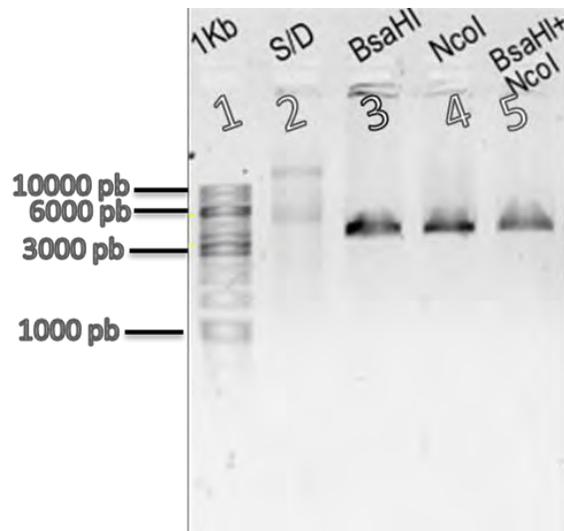


Figura 6 Gel de comprobación de la ligación de las secuencias candidatas al plásmido. Ilustra un gel de agarosa con los productos de la digestión del plásmido supergloGFP por *BsaHI* (carril 3), *NcoI* (carril 4) y *BsaHI-NcoI* (carril 5).

Después de la comprobación de la ligación, se realizó la purificación de gel y se transformó en la cepa *E. coli* W3110 por electroporación; sólo se seleccionaron las cepas que presentaron fluorescencia en el cultivo de placa, como se muestra en la **figura 7**. En las cajas se ven las colonias que crecieron después de 24 horas de incubación, hay diversas candidatas, desde aquellas que parece que perdieron actividad hasta las muy brillantes, indicando cambios positivos en la secuencia para la producción de proteínas. Este criterio de selección filtró la selección de mutantes a sólo las que conservaran su función de manera evidente, en consecuencia, la secuenciación mostró los lugares donde más variabilidad pueden tener estas dos regiones.

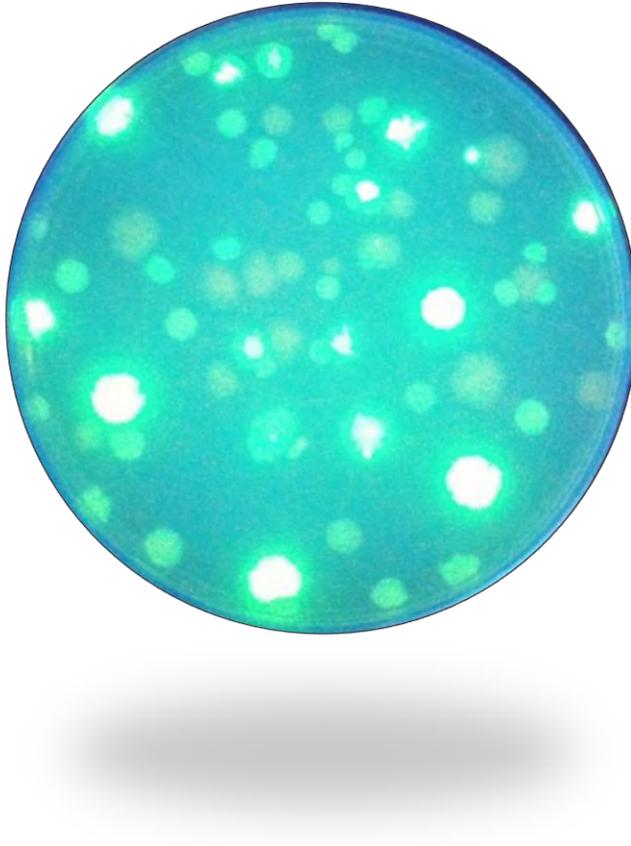


Figura 7 La recuperación de las células después de la transformación. Colonias vistas con luz UV, esta foto tiene un propósito ilustrativo.

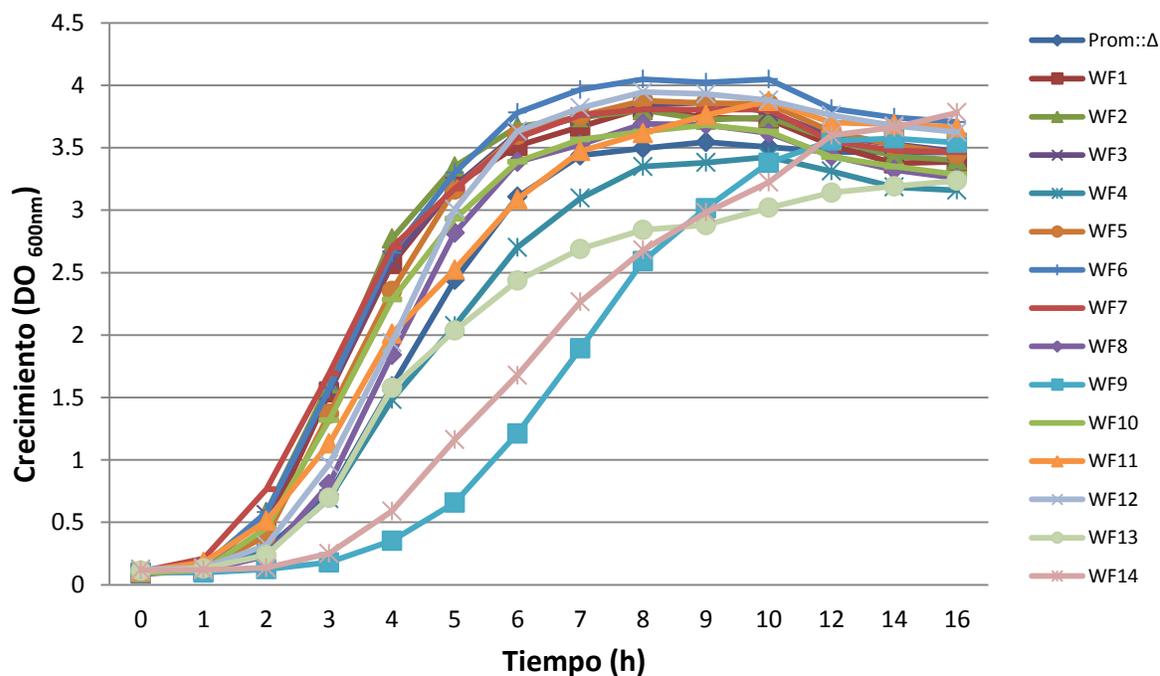
8.2 Caracterización de la biblioteca

8.2.1 Cinéticas de crecimiento y fluorescencia

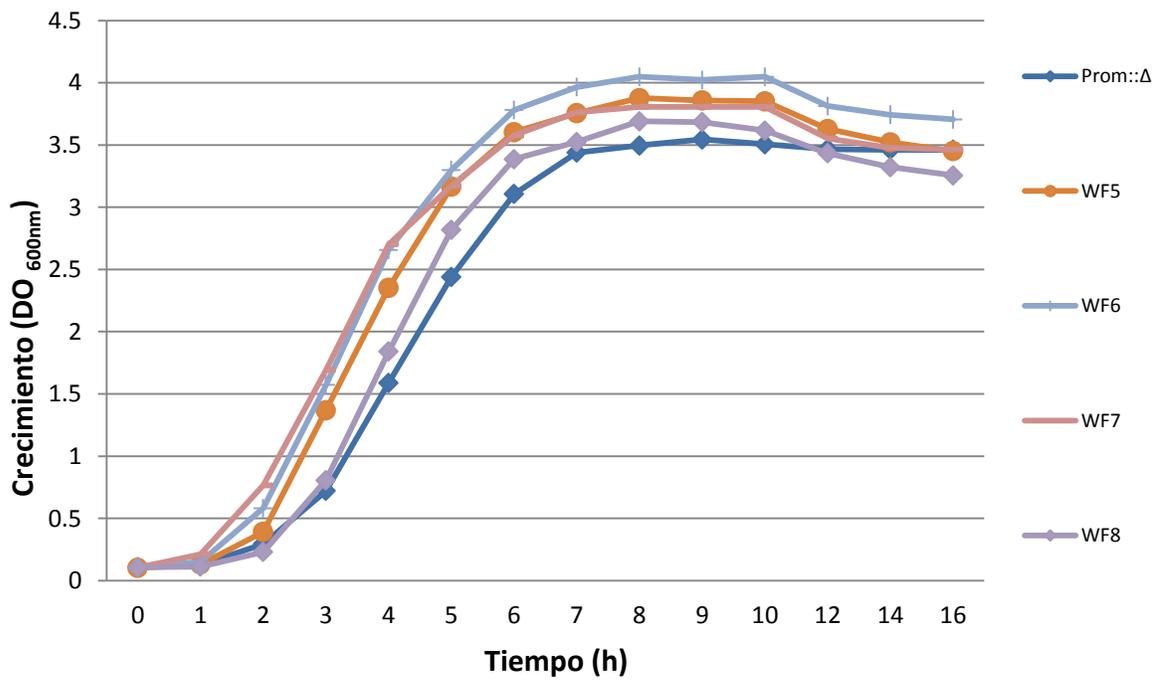
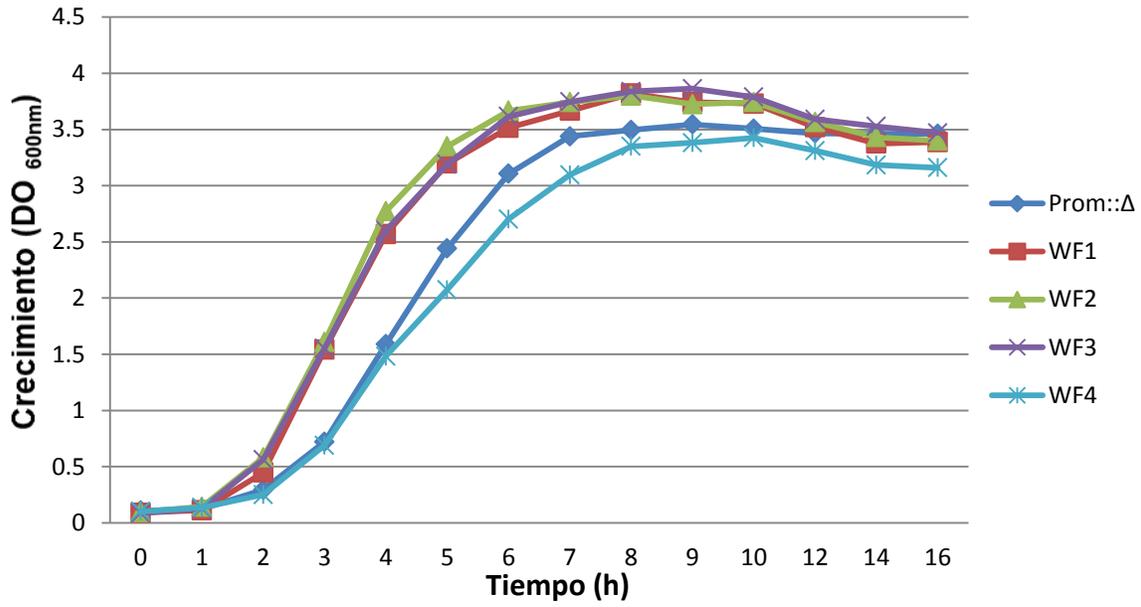
Como se describió en los métodos, se determinó crecimiento y fluorescencia durante 16 horas. A pesar del gasto energético que representa tener un plásmido de alto número de copias y la generación exacerbada de un producto como es la proteína verde fluorescente, no se encontró correlación entre el crecimiento celular y la fluorescencia, como se muestra en las **figuras 8 y 9**. La numeración de las candidatas tiene relación con su posición de producción de GFP, es decir, se organizaron de la menos fluorescente (WF1) a la más fluorescente (WF14), la candidata nombrada Prom:: Δ corresponde a la candidata que suponemos no tiene capacidad para producir GFP. La candidata WF8 corresponde a la secuencia molde con la que se generó la biblioteca.

Con nuestro plásmido Prom:: Δ (que tiene parte de una secuencia de gen estructural del gen *lacZ* y por tanto suponemos que no tiene funcionalidad en las regiones de promotor y de RBS) se demostró que no hubo interferencia significativa en el crecimiento de los cultivos a las 16 horas. Por otro lado, en horas previas se observa un retraso de crecimiento en algunas candidatas, este efecto podría atribuirse a la producción elevada de la proteína reportera, aunque en contraste hubo otras candidatas con una fluorescencia mayor o muy parecida y que sin embargo tuvieron una velocidad de crecimiento similar al de la cepa control. Estos resultados sugieren hacer análisis en la cantidad de ARNm durante la cinética, porque una razón que puede estar afectando la velocidad de crecimiento de las candidatas, es la sobreproducción de ARNm que no todo se puede ver reflejado en la proteína reportera, porque se puede degradar antes.

A)



B)



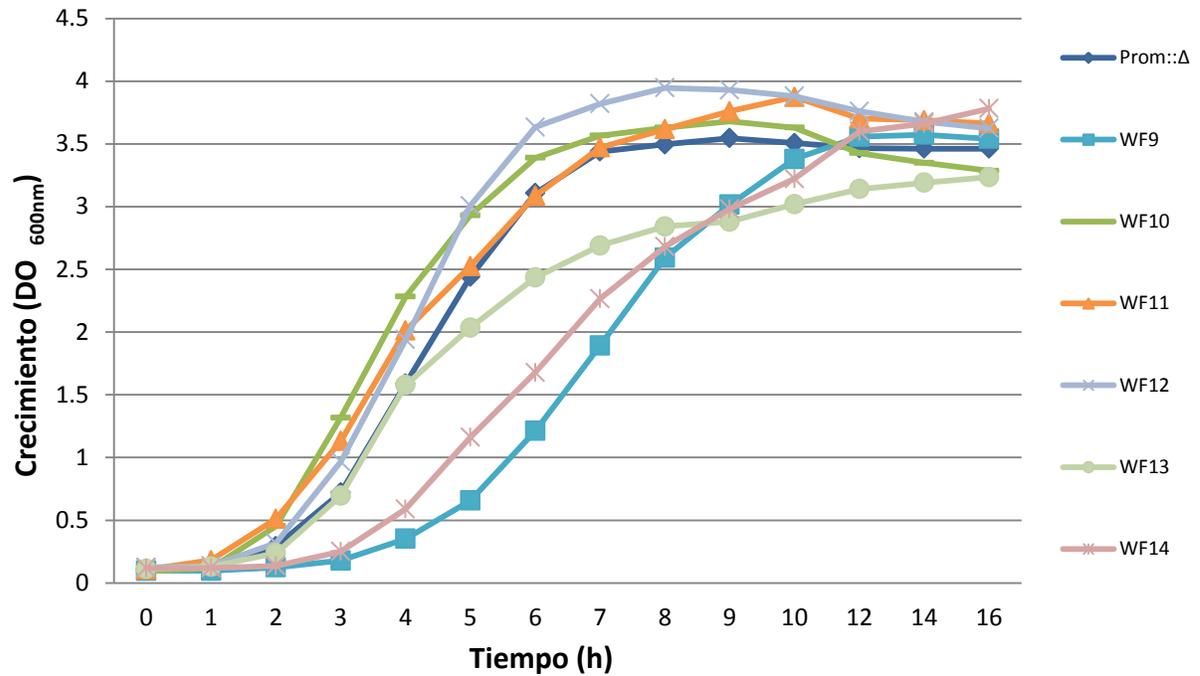
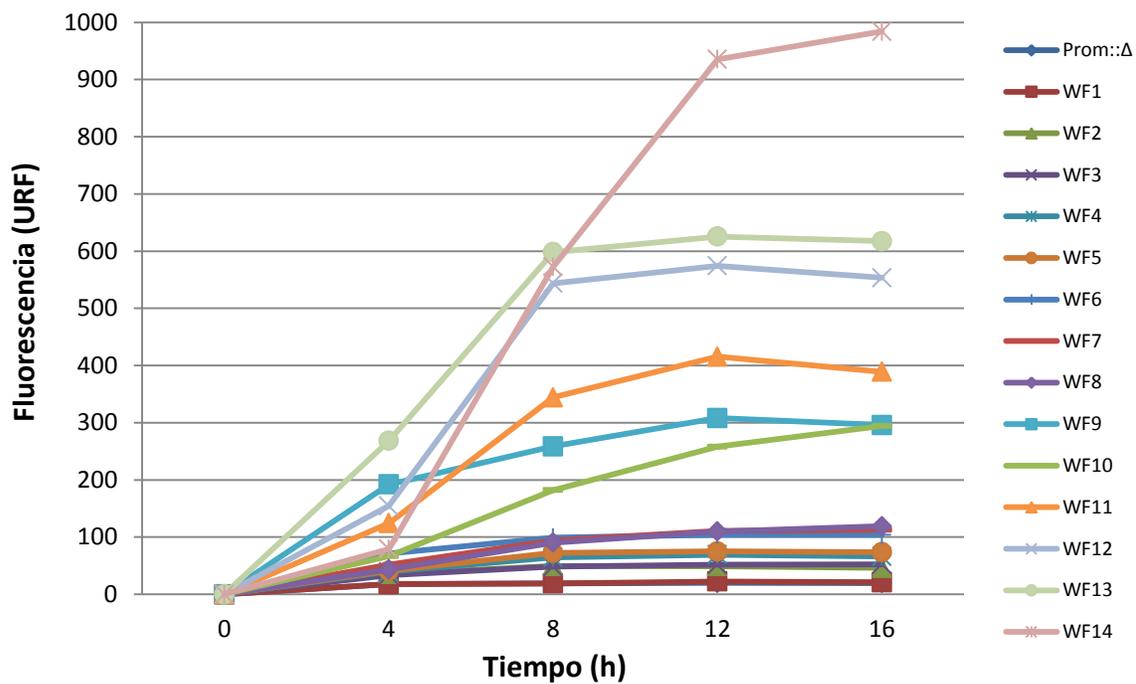
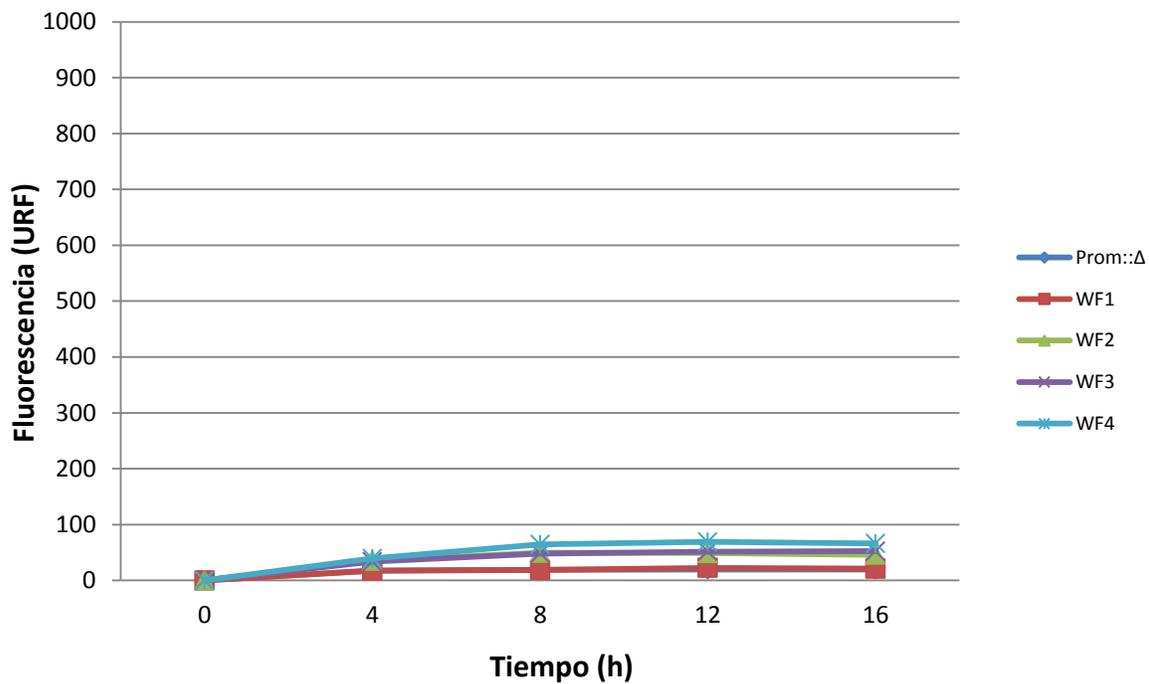


Figura 8 Cinéticas de crecimiento. Se presentan los datos que corresponden a la cinética de crecimiento por 16 horas (en medio LB). **A)** Muestra una gráfica que unifica todas las candidatas y se nota un retraso en el crecimiento de las candidatas WF14, WF13 y WF9, **B)** Los resultados de las cinéticas desglosadas por grupos de 4 más la candidata Prom::Δ.

A)



B)



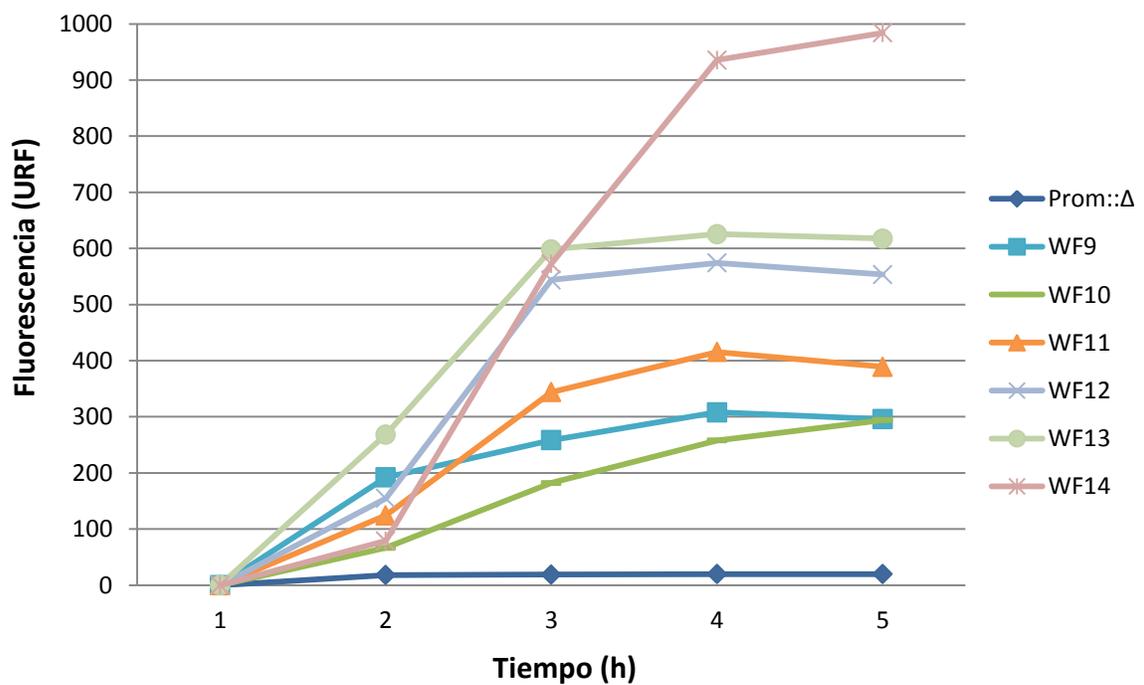
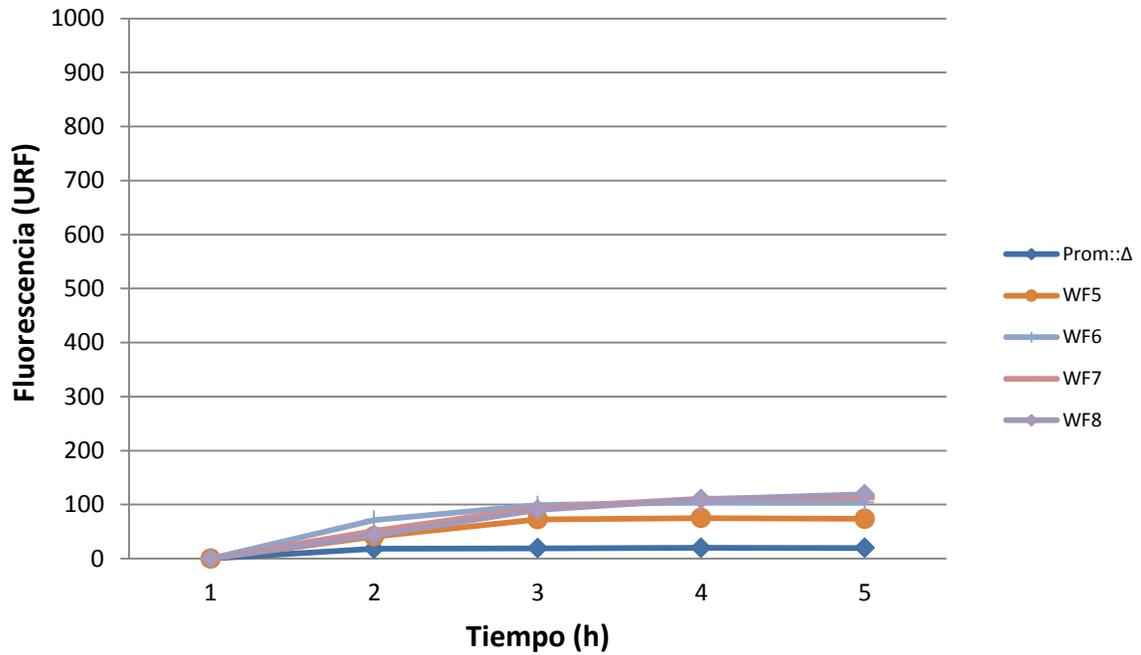
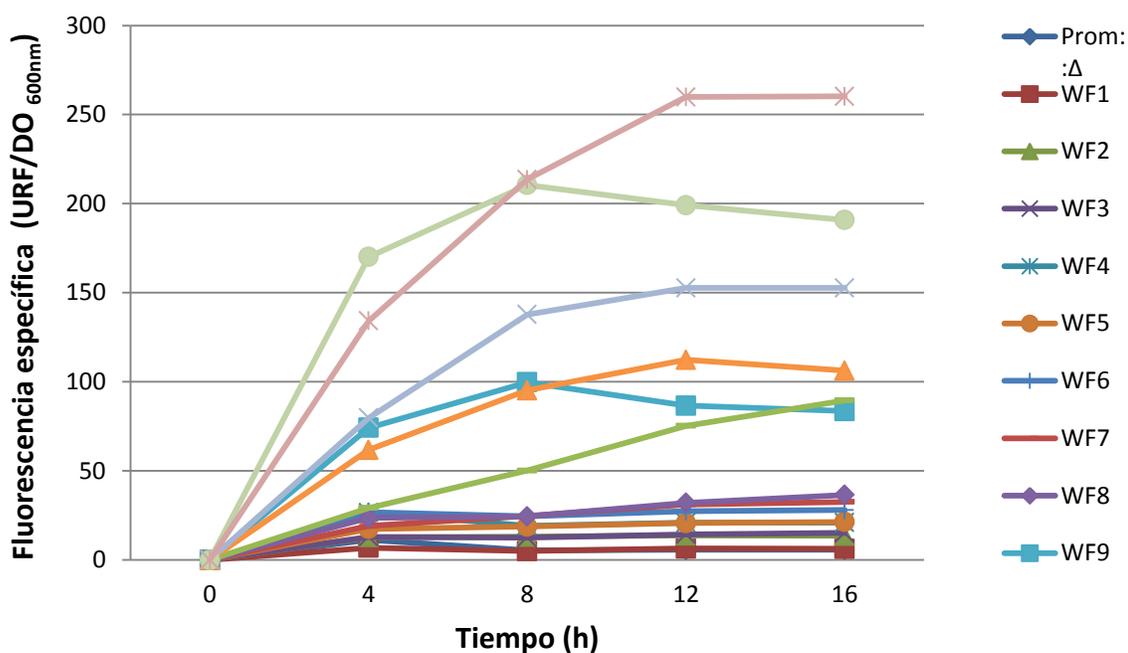


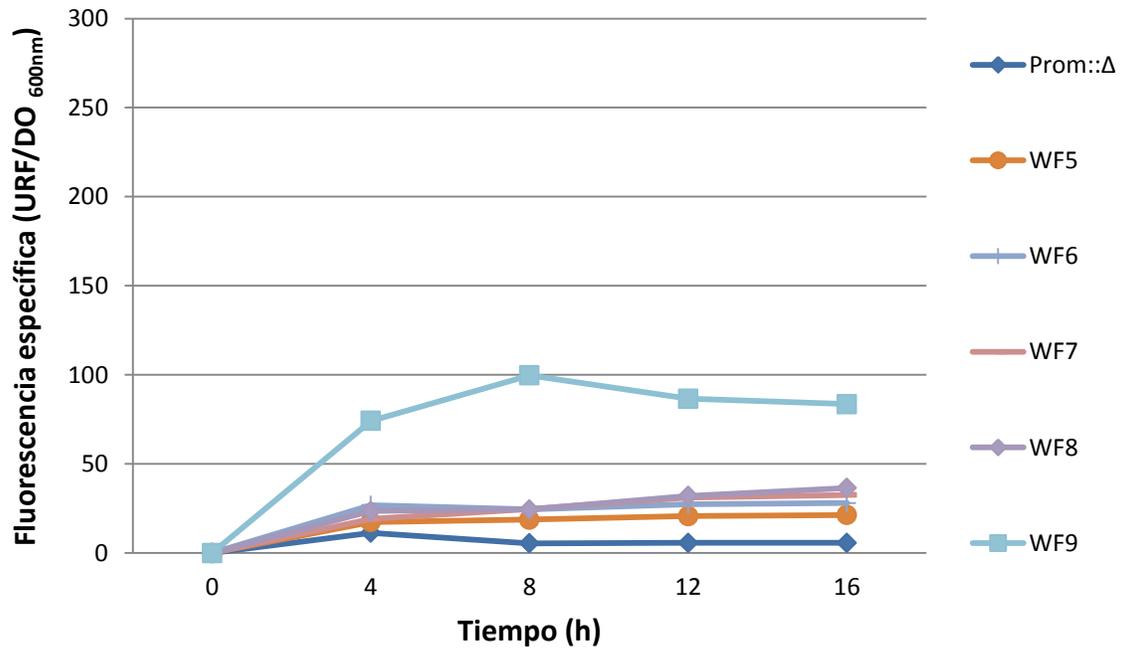
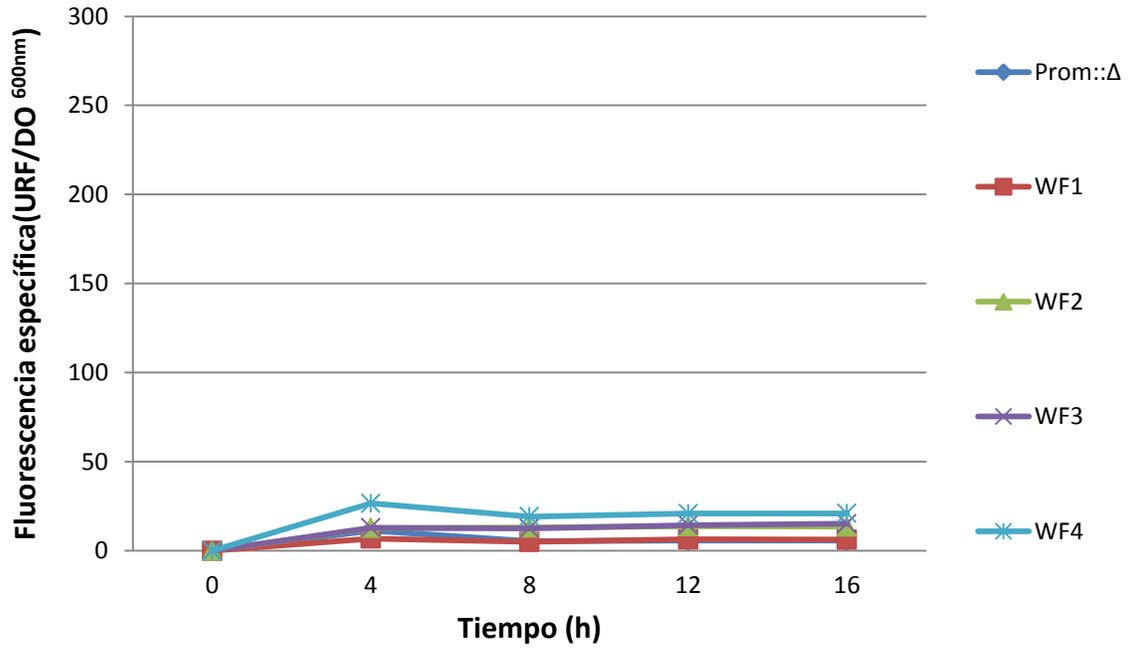
Figura 9 Cinéticas de fluorescencia. Se presentan las cinéticas de fluorescencia durante 16 horas (en medio LB), las unidades de fluorescencia relativa (UFR) calculadas por el espectrofluorímetro. **A)** Muestra una gráfica de todas las candidatas juntas, **B)** Las gráficas desglosadas por grupos

Como se mencionó anteriormente, las candidatas tienen una numeración que hace referencia al nivel de fluorescencia registrado, es decir, la candidata con menor fluorescencia ocupa el lugar 1 y se llama WF1, la más alta es WF14 y la que tiene la secuencia molde quedó como WF8. La fluorescencia (URF) se dividió entre la densidad óptica (D.O. $_{600\text{ nm}}$) para evitar los sesgos que pueden ocasionar la concentración de células como se recomienda en De Mey *et al.* (2007). Se reajustaron las gráficas como se muestra en las **figuras 10 y 11**, en dichas imágenes se reafirma el orden de numeración para cada candidata. En algunas candidatas se encontraron variaciones grandes entre un experimento y otro, esto puede explicarse por el alto número de copias del plásmido, debido a que la cantidad de plásmidos puede variar de un cultivo a otro (Hornung *et al.* 2012).

A)



B)



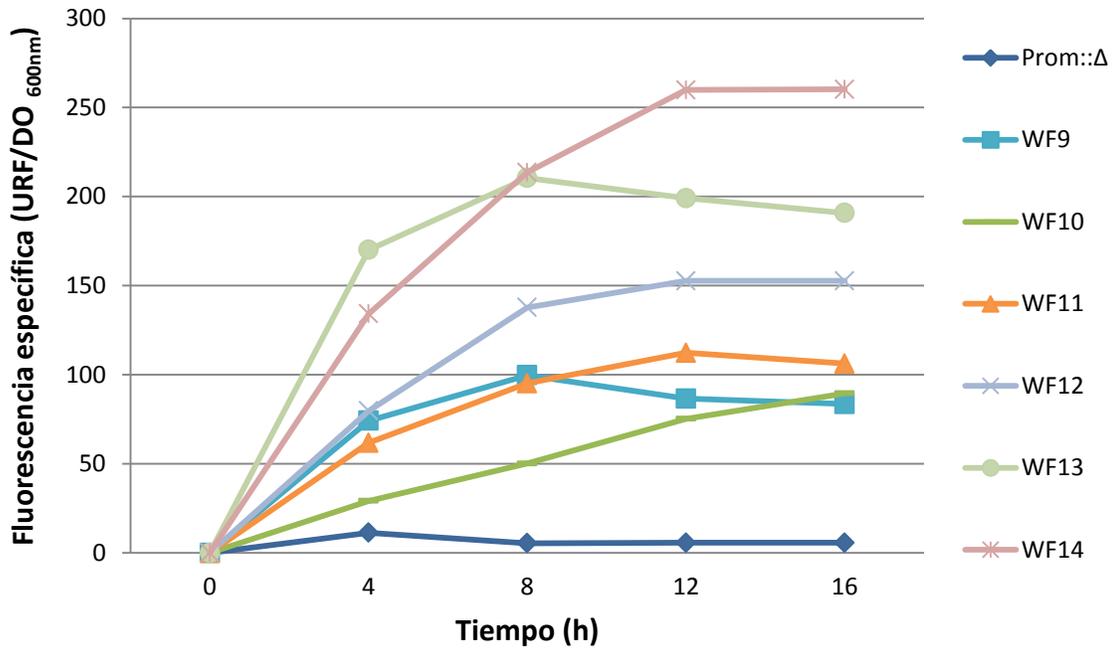
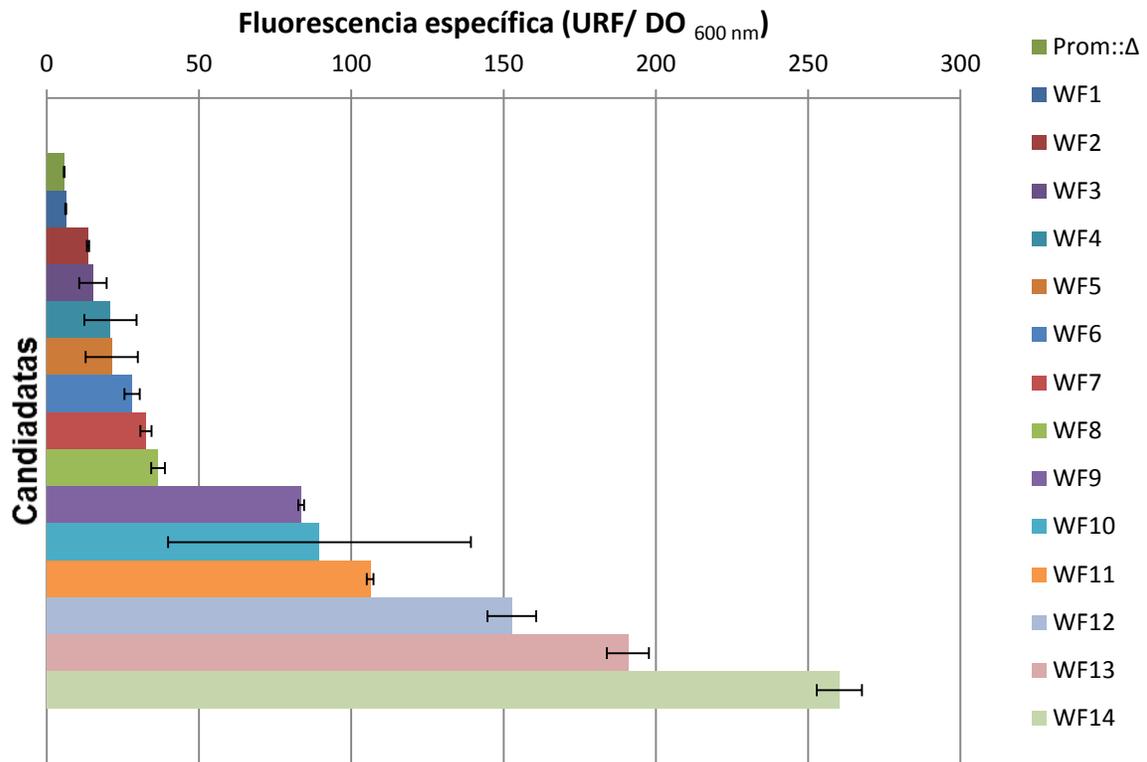


Figura 10 Cinéticas de fluorescencia específica. Las tendencias de fluorescencia específica (URF/DO_{600 nm}) relacionan los resultados de las cinéticas de crecimiento y fluorescencia para corregir posibles errores de cuantificación causados por la variación de cantidad de plásmido de cada candidata. **A)** Muestra todo el conjunto de candidatas, reafirmando las tendencias observadas en las gráficas de la **figura 9**. **B)** Son las gráficas desglosadas por grupos para simplificar su visualización



Sujetos	Prom::Δ	WF1	WF2	WF3	WF4	WF5	WF6	WF7	WF8	WF9	WF10	WF11	WF12	WF13	WF14
URF/DO	5.7	6.2	13	15	21	22	28	32	36	83.6	89.5	106	153	191	260

Figura 11 Comparación de niveles de fluorescencia específica a la hora 16. Esta gráfica de barras muestra la diversidad de fluorescencia de cada secuencia. Todos los experimentos contaron con al menos dos repeticiones, en el caso de Prom::Δ y WF8 hasta 7 repeticiones. Se excluyeron las candidatas con más de 40% de desviación.

8.2.2 Análisis de las secuencias

Al mismo tiempo de realizar cinéticas, se secuenciaron varias candidatas para identificar mutaciones tanto en la región promotora como en la RBS, estos resultados ayudaron a descartar las candidatas que tenían mutaciones en el gen estructural de GFP o las que estaban repetidas. Estas reacciones se realizaron con dos oligonucleótidos diferentes, uno que hibrida a la mitad del cassette de resistencia de cloranfenicol (con dirección hacia la región insertada) y otro que hibrida a la mitad del gen estructural de GFP (con dirección hacia la región insertada).

En la **figura 12** se aprecia que la mayor parte de las inserciones a la secuencia molde, ocurre entre la última base del -10 y antes de llegar al RBS, este resultado podría indicar que los sitios del promotor y RBS son sensibles a las inserciones.

A)

WF1	GGCGCC <i>TTGACA</i> ATTAGTAACTATTTGT <i>TATAAT</i> GTATTCCCTTACCGGCCCTTTCCATGCGCGCACAGGAAACAGAA <i>A</i> CCATGG
WF2	GGCGCC <i>TTGACA</i> ATTAGTAACTATTTGT <i>TATAAT</i> GTATTCC <i>T</i> CTTACCGGCCCTTTCCATGCGCGCACAGGAAACAG <i>T</i> ACCATGG
WF3	GCGCC <i>TTGACA</i> ATTAGTAACTATTT <i>G</i> TT <i>TATAAT</i> GTATTCCCTTACCGGCCCTTTCCATGCGCGCAC <i>C</i> AGGAAACAGACCATGG
WF4	GGCGCC <i>TTGACA</i> ATTAGTAACTATTTGT <i>TATAAT</i> GTATTCCCTTACCGGCCCTTTCCA <i>T</i> GCGCGCACAGGAAACAGACCATGG
WF5	GGCGCC <i>TTGACA</i> ATTAGTAACTATTTGT <i>AATAAT</i> GTATTCCCTTACCGG <i>G</i> CCCTTTCCATG <i>T</i> CGCGCA <i>TTAGG</i> AAACAGACCATGG
WF6	GGCGCC <i>TTGACA</i> ATTAGTAACTATTTGT <i>TATAAT</i> G <i>A</i> ATTCCCTTACCGGCCCTT <i>A</i> TCCATGCGCGCACAGGAAACAGACCATGG
WF7	GGCGCC <i>TTGACA</i> ATTAGTAACTATTTGT <i>TATAAT</i> GTATTCCCTTAC <i>TCA</i> GGCCCTTTCCATGCGCGCAC <i>G</i> AGGAAACAGACCATGG
WF8	GGCGCC <i>TTGACA</i> ATTAGTAACTATTTGT <i>TATAAT</i> GTATTCCCTTACCGGCCCTTTCCATGCGCGCACAGGAAACAGACCATGG
WF9	GGCGCC <i>TTGACA</i> ATTAGTAACTATTTGT <i>TATAAT</i> GTATTCCCTTACCGGCCCTTTCCAT <i>T</i> CGCGCACAGGAAACAG <i>TA</i> ACCATGG
WF10	GGCGCC <i>TTGACA</i> ATTAGTAACTATTTGT <i>TATAAT</i> GTATTCCCTTACCGGCCCTTT <i>A</i> CCATGCGCGCACAGGAAACAGACCATGG
WF11	GGCGCC <i>TTGACA</i> ATTAGTAACTATTTGT <i>TATAAT</i> GTATTCCCTTACCGGCCCTTTCCATG <i>C</i> GCGCACAGG <i>A</i> AACAGACCATGG
WF12	GGCGCC <i>TTGACA</i> ATTAGTAACTATTTGT <i>TATAAT</i> GTATTCCCTTACCGGCCCTTT <i>A</i> CCATGCGCGCACAGG <i>A</i> GAAC <i>A</i> AGACCATGG
WF13	GGCGCC <i>TTGACA</i> ATTAGTAACTATTTGT <i>TATAAT</i> GTATTCCCTTACCGGCCCTTTCCATGCGCGCACAGG <i>A</i> GAAC <i>A</i> AGACCATGG
WF14	GGCGCC <i>TTGACA</i> ATTAGTAACTATTTGT <i>TATAAT</i> GTATTCCCTTACCGGCCCTTT <i>A</i> CCATGCGCGCACAGGAAACAGACCATGG
Molde	NNNNNN <i>TTGACA</i> NNNNNNNNNNNNNNNNNN <i>TATAAT</i> NN <i>NAGGAAACAG</i> NNNNNN

B)



Figura 12 Inserciones en las secuencias de los sujetos evaluados. **A)** Son las secuencias individuales de los sujetos estudiados. Remarcadas en negro y escritas en cursiva se encuentran los sitios -35, -10 y RBS, respectivamente, subrayadas y remarcadas están las bases insertadas. **B)** Con ayuda del sitio web <http://weblogo.berkeley.edu/logo.cgi> se generó un esquema de la secuencia con los nucleótidos más frecuentes en cada posición. Esta imagen permite visualizar los sitios con más variabilidad de la biblioteca, correspondientes a la parte intermedia de SD y -10.

8.3 Análisis de la biblioteca

Por medio de una matriz de datos para analizar los promotores (Brewster et al. 2012) se calculó la cantidad de energía requerida por una RNAPol para unirse al promotor y con ayuda de los programas para los RBS (Salis et al. 2009) se analizaron la eficiencia de traducción de las secuencias de la biblioteca. De la región promotora sólo se analizaron 3 secuencias diferentes, ya que la biblioteca sólo tuvo esas 3 sujetos con variaciones en esta zona como se muestra en la **Tabla 2**, los cambios en el promotor pudieron influenciar el aumento o disminución de producto.

El resultado de nuestro análisis de promotores no se ajusta al modelo Brewster *et al.* (2012) porque, en algunos casos las secuencias que en teoría requerían menor energía para el pegado de la polimerasa quedan por debajo de otras que aparentemente no tienen tan buena secuencia promotora, aun cuando en ambas candidatas tienen secuencia de RBS muy parecidas y casi con la misma tasa de inicio de traducción. Las candidatas con mutaciones fueron WF3, WF5 y WF6; su análisis por la matriz resultó en $-0.6 k_B T$, $1.8 k_B T$ y $5.8 k_B T$ respectivamente ($1 \text{ kcal/mol} = 1.62 k_B T$). El resultado de la secuencia molde fue de $-0.6 k_B T$, misma energía que requiere la polimerasa para unirse al promotor de la candidata WF3 y más del triple que requieren las candidatas WF5 y WF6. Si sólo nos basáramos en estos resultados, diríamos que la candidata WF6 es de las menos eficientes y WF3 de las más eficientes en el pegado de la RNAPol, no obstante, no se cumple la predicción. Los resultados de predicción sólo nos informa sobre la fuerza que necesita la RNAPol para pegarse al promotor, sin embargo, el pegado de la polimerasa no implica necesariamente el inicio de la transcripción (Brewster et al. 2012), y esta puede ser una razón por la cual, a pesar de tener un promotor de poca eficiencia de pegado con la polimerasa, se pueda encontrar mucho producto, porque puede estar aumentando la tasa de inicio de transcripción.

Secuencias de promotor de WF1, WF2, WF4, WF7, WF8, WF9, WF10, WF11, WF12, WF13 y WF14

	-36	-35	-34	-33	-32	-31	-30	-29	-28	-27	-26	-25	-24	-23	-22	-21	-20	-19	-18	-17	-16	-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3
A	0.4	0.2	1.4	-0.2	1	-0.2	-0.2	0.2	0	0.2	-0.2	0.2	-0.2	0.2	0.2	0.2	-0.2	0.2	0.2	0.2	0.2	0.2	0.4	0.2	0.4	-0.2	0.2	-0.2	-0.2	0.6	-0.2	0	-0.2	0
C	0.8	0.4	1.4	1	0	0.6	0.2	0.2	0.2	0.2	0.2	0.2	0.2	-0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.8	0.2	0.8	1	0.2	0.4	0.2	1	0.2	0.2	0.2	0.2
G	0.4	0.2	-0.2	0.2	1	0.4	0.2	-0.2	-0.2	0.2	0.2	0.2	0	0.2	0	0.2	0	0.2	0.2	0.2	0	-0.2	-0.2	1	0.8	0.4	0.6	0.6	0.8	0.2	0.2	0.2	0	
T	-0.2	-0.2	0.6	0.6	-0.2	0.2	0	0.2	0.2	-0.2	0.2	-0.2	0.4	0.2	-0.2	-0.2	0.2	0	-0.2	0	0.4	0.2	0.4	0.2	-0.2	1	-0.2	0.6	0.6	-0.2	0.2	-0.2	0.2	-0.2

Secuencias de promotor de WF3

	-36	-35	-34	-33	-32	-31	-30	-29	-28	-27	-26	-25	-24	-23	-22	-21	-20	-19	-18	-17	-16	-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3
A	0.4	0.2	1.4	-0.2	1	-0.2	-0.2	0.2	0	0.2	-0.2	0.2	-0.2	0.2	0.2	0.2	-0.2	0.2	0	0.2	-0.2	0.2	0.4	0.2	0.4	-0.2	0.2	-0.2	-0.2	0.6	-0.2	0	-0.2	0
C	0.8	0.4	1.4	1	0	0.6	0.2	0.2	0.2	0.2	0.2	0.2	0.2	-0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.8	0.2	0.8	1	0.2	0.4	0.2	1	0.2	0.2	0.2	0.2
G	0.4	0.2	-0.2	0.2	1	0.4	0.2	-0.2	-0.2	0.2	0.2	0.2	0.2	0	0.2	0	0.2	0	0.2	0.2	0.2	0	-0.2	-0.2	1	0.8	0.4	0.6	0.6	0.8	0.2	0.2	0.2	0
T	-0.2	-0.2	0.6	0.6	-0.2	0.2	0	0.2	0.2	-0.2	0.2	-0.2	0.4	0.2	-0.2	-0.2	0.2	0	-0.2	0	0.4	0.2	0.4	0.2	-0.2	1	-0.2	0.6	0.6	-0.2	0.2	-0.2	0.2	-0.2

Secuencias de promotor de WF5

	-36	-35	-34	-33	-32	-31	-30	-29	-28	-27	-26	-25	-24	-23	-22	-21	-20	-19	-18	-17	-16	-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3
A	0.4	0.2	1.4	-0.2	1	-0.2	-0.2	0.2	0	0.2	-0.2	0.2	-0.2	0.2	0.2	0.2	-0.2	0.2	0	0.2	-0.2	0.2	0.4	0.2	0.4	-0.2	0.2	-0.2	-0.2	0.6	-0.2	0	-0.2	0
C	0.8	0.4	1.4	1	0	0.6	0.2	0.2	0.2	0.2	0.2	0.2	0.2	-0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.8	0.2	0.8	1	0.2	0.4	0.2	1	0.2	0.2	0.2	0.2
G	0.4	0.2	-0.2	0.2	1	0.4	0.2	-0.2	-0.2	0.2	0.2	0.2	0.2	0	0.2	0	0.2	0	0.2	0.2	0.2	0	-0.2	-0.2	1	0.8	0.4	0.6	0.6	0.8	0.2	0.2	0.2	0
T	-0.2	-0.2	0.6	0.6	-0.2	0.2	0	0.2	0.2	-0.2	0.2	-0.2	0.4	0.2	-0.2	-0.2	0.2	0	-0.2	0	0.4	0.2	0.4	0.2	-0.2	1	-0.2	0.6	0.6	-0.2	0.2	-0.2	0.2	-0.2

Secuencia de promotor de WF6

	-36	-35	-34	-33	-32	-31	-30	-29	-28	-27	-26	-25	-24	-23	-22	-21	-20	-19	-18	-17	-16	-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3
A	0.4	0.2	1.4	-0.2	1	-0.2	-0.2	0.2	0	0.2	-0.2	0.2	-0.2	0.2	0.2	0.2	-0.2	0.2	0	0.2	-0.2	0.2	0.4	0.2	0.4	-0.2	0.2	-0.2	-0.2	0.6	-0.2	0	-0.2	0
C	0.8	0.4	1.4	1	0	0.6	0.2	0.2	0.2	0.2	0.2	0.2	0.2	-0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.8	0.2	0.8	1	0.2	0.4	0.2	1	0.2	0.2	0.2	0.2
G	0.4	0.2	-0.2	0.2	1	0.4	0.2	-0.2	-0.2	0.2	0.2	0.2	0.2	0	0.2	0	0.2	0	0.2	0.2	0.2	0	-0.2	-0.2	1	0.8	0.4	0.6	0.6	0.8	0.2	0.2	0.2	0
T	-0.2	-0.2	0.6	0.6	-0.2	0.2	0	0.2	0.2	-0.2	0.2	-0.2	0.4	0.2	-0.2	-0.2	0.2	0	-0.2	0	0.4	0.2	0.4	0.2	-0.2	1	-0.2	0.6	0.6	-0.2	0.2	-0.2	0.2	-0.2

Tabla 2 Matrices de cálculo en la unión de la polimerasa. Estas matrices muestran en cada casilla la cifra que representa la energía requerida por la polimerasa para unirse al promotor, las unidades expresadas en $k_B T$, que por cada $1.62 k_B T$ hay 1 kcal/mol . Se coloreó cada casilla de cada base correspondiente en la posición que arriba se marca (la numeración de la secuencia está con relación al sitio de inicio de transcripción), en colores están las bases que conforman la secuencia molde, en azul los hexámeros (-35, -10), en rojo la región intermedia, en naranja las bases que están fuera del sitio de unión de la RNAPol pero que siguen influenciando el pegado de la enzima y en morado las inserciones que se realizaron.

Por otro lado, el análisis de RBS dio un resultado más parecido a lo visto en los experimentos, como la **tabla 3** revela, si bien estos análisis no son del todo exactos, este programa al menos agrupa tres de las secuencias más fuertes (WF11, WF12 y WF13) en los sitios con más tasa de traducción, dos de las secuencias intermedias cercanas a su rango observado (WF9 y WF7) y dos de las más bajas en las tasas más bajas (WF2 y WF1). La razón por la que a partir del modelo de Salis *et al.* (2009) no haya predicho el resultado que nosotros obtuvimos (omitiendo WF3, WF5 y WF6 que tuvieron

modificaciones en la región promotora) puede deberse a que no se incluyen factores de la dinámica de traducción como la actividad de la proteína S1 y la vida media del RNA mensajero.

Hasta el día de hoy no existen modelos para la predicción de la expresión conjuntando sólo la secuencia de promotores y RBS, pero sí hay algunos modelos para saber cuánto aportará cada región si se conocen sus productos de transcripción y eficiencia de traducción (Kosuri et al. 2013).

A)

	URF/DO	ΔG_{total}	$\Delta G_{mRNA:rRNA}$	$\Delta G_{spacing}$	ΔG_{stanby}	ΔG_{start}	ΔG_{mRNA}	Tasa de inicio de traducción (au)
WF1	6.22	7.11	-3.9	0.29	5.48	-1.19	-17.4	101.79
WF2	13.54	8.91	-3.9	0.29	5.48	-1.19	-19.2	45.28
WF3	15.15	5.73	-3.9	0	5.78	-1.19	-16.6	190.1
WF4	20.92	7.73	-4.2	0	5.18	-1.19	-18.3	77.28
WF5	21.30	2.03	-4.3	0	5.78	-1.19	-13.3	1004.96
WF6	28.05	6.93	4.2	0	4.78	-1.19	-17.1	110.77
WF7	32.55	4.42	-5.2	0	5.38	-1.19	-16.2	341.24
WF8	36.53	7.03	-3.9	0	5.48	-1.19	-17.6	105.9
WF9	83.55	2.7	-3.6	0.67	5.78	-1.19	-12.6	742.69
WF10	89.52	6.52	-4.2	0	4.78	-1.19	-16.7	132.62
WF11	106.21	-2.09	-9.7	0.29	5.78	-1.19	-14.3	6395.12
WF12	152.67	0.6	-6.98	0.67	4.78	-1.19	-12.7	1910.98
WF13	190.84	2	-6.5	0.67	5.48	-1.19	-14.5	1017.71
WF14	260.23	6.02	-4.2	0	5.28	-1.19	-16.7	166.08

B)

	URF/DO	ΔG_{total}	$\Delta G_{mRNA:rRNA}$	$\Delta G_{spacing}$	ΔG_{stanby}	ΔG_{start}	ΔG_{mRNA}	Tasa de inicio de traducción (au)
WF2	13.54	8.91	-3.9	0.29	5.48	-1.19	-19.2	45.28
WF4	20.92	7.73	-4.2	0	5.18	-1.19	-18.3	77.28
WF1	6.22	7.11	-3.9	0.29	5.48	-1.19	-17.4	101.79
WF8	36.53	7.03	-3.9	0	5.48	-1.19	-17.6	105.9
WF6	28.05	6.93	4.2	0	4.78	-1.19	-17.1	110.77
WF10	89.52	6.52	-4.2	0	4.78	-1.19	-16.7	132.62
WF14	260.23	6.02	-4.2	0	5.28	-1.19	-16.7	166.08
WF3	15.15	5.73	-3.9	0	5.78	-1.19	-16.6	190.1
WF7	32.55	4.42	-5.2	0	5.38	-1.19	-16.2	341.24
WF9	83.55	2.7	-3.6	0.67	5.78	-1.19	-12.6	742.69
WF5	21.30	2.03	-4.3	0	5.78	-1.19	-13.3	1004.96
WF13	190.84	2	-6.5	0.67	5.48	-1.19	-14.5	1017.71
WF12	152.67	0.6	-6.98	0.67	4.78	-1.19	-12.7	1910.98
WF11	106.21	-2.09	-9.7	0.29	5.78	-1.19	-14.3	6395.12

Tabla 3 La predicción del modelo de Salis *et al.* (2009) contra lo observado. Se muestran los resultados generados a partir del programa <https://salis.psu.edu/software/> usando 35 bases río arriba y 35 bases río debajo de cada secuencia. Los términos en los que se midió la eficiencia de cada secuencia están descritos en los métodos. **A)** Se muestran las secuencias en orden de fluorescencia que nosotros registramos **B)** Están ordenadas de manera ascendente de acuerdo a su tasa de inicio de transcripción para contrastar con la parte **A)**, las diferencias de lo observado y lo esperado.

IX. Conclusiones

Nuestro objetivo se logró con la generación y caracterización de un acervo de secuencias sintéticas de la región promotora y RBS. Esta herramienta podría usarse para modificar la expresión de manera precisa de una amplia diversidad de genes.

De las candidatas seleccionadas, hubo 7 sujetos por debajo de la expresión de la secuencia molde (WF8) con la que se generó la biblioteca y 6 sujetos con una expresión mayor.

Los sujetos que tuvieron cambios en la región promotora sólo fueron WF3 y WF5, ambas tienen una expresión menor que la secuencia WF8 por 0.41 y 0.58 veces menor. Por otro lado, los otros sujetos con menor expresión que la secuencia molde no presentan una tendencia evidente sobre sus cambios. Estos cambios incluyen inserciones junto al RBS o en la región intermedia de -10 y RBS. La candidata que más se acerca a la expresión de WF8 es WF7 con 0.89 veces menos de expresión.

Las candidatas que sobrepasaron la expresión de WF8 van desde 2.29 (WF9) a 7.12 (WF14) veces más de expresión. En este grupo tampoco se encontraron patrones, sin embargo, cabe resaltar que la candidata WF14 sólo tuvo una inserción intermedia entre el -10 y RBS.

Los modelos de predicción usados en este trabajo fueron propuestos en años recientes (Brewster et al. 2012; Salis et al. 2009), a pesar de demostrar su acertado funcionamiento, para el análisis de nuestras secuencias no se ajustan los resultados con toda precisión. Este motivo puede deberse a que la biblioteca aquí presentada tiene cambio en ambas regiones y los modelos sólo calculan de manera individual un paso de la expresión (la transcripción o la traducción), además que de estos pasos únicamente usan los parámetros más representativos.

X. Perspectivas

Esta herramienta generada da lugar a nuevas estrategias experimentales para modificar la expresión genética, que pudieran ser usados para la investigación de vías metabólicas, particularmente podrían ser usados en reemplazos cromosomales, los cuales ayudarían a caracterizar de mejor manera la biblioteca, además de estudiar los efectos fisiológicos que tendría la modificación de expresión de

genes relacionados con vías metabólicas; asimismo podrían hacerse estudios específicos de la transcripción y traducción que permitieran proponer un nuevo modelo que incluya las secuencias de promotor y RBS.

XI. Bibliografía consultada

Almagor, H., 1983. Analysis of DNA sequences. *Journal of Theoretical Biology*, 104: 633–645.

Amann, Egon. Ochs, Birgit. and Abel, K.-J., 1988. Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. *Gene*, 69:301–315.

Arpino, J. a J. et al., 2013. Tuning the dials of synthetic biology. *Microbiology (Reading, England)*, 159:1236–53.

Bintu, L. et al., 2005. Transcriptional regulation by the numbers: models. *Current Opinion in Genetics & Development*, 15:116–24.

Brewster, R.C., Jones, D.L. & Phillips, R., 2012. Tuning promoter strength through RNA polymerase binding site design in *Escherichia coli*. *PLoS Computational Biology*, 8: 1-10

Busby, D.F.B. and S.J.W., 2004. The regulation of bacterial transcription initiation, 2:1 - 9.

Carrier, T. a & Keasling, J.D., 1997. Controlling messenger RNA stability in bacteria: strategies for engineering gene expression. *Biotechnology progress*, 13: 699–708.

Carrier, T. a & Keasling, J.D., 1999. Library of synthetic 5' secondary structures to manipulate mRNA stability in *Escherichia coli*. *Biotechnology Progress*, 15: 58–64.

Cox, David L. Nelson, M.M., 2009. *Lehninger principios de bioquímica*, Quinta Edición, Barcelona.

Dalgarno, J.S.L., 1975. Determinant of cistron specificity in bacterial ribosomes. *Nature*, 254: 34 – 38.

Delvillani, F. et al., 2011. S1 ribosomal protein and the interplay between translation and mRNA decay. *Nucleic Acids Research*, 39: 7702–7715.

Dirks, R.M. et al., 2007. Thermodynamic analysis of interacting nucleic acid strands. *SIAM Review*, 49: 65–88.

Flores-Ramirez, Gabriela et al., 2007. The effect of amino acid deletions and substitutions in the longest. *BioMedCentral Chemical Biology*, 15: 1116 -1124.

Gourse, R.L., Ross, W. & Gaal, T., 2000. UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition. *Molecular microbiology*, 37: 687–695.

- Grigorova, I.L. et al., 2006. Insights into transcriptional regulation and sigma competition from an equilibrium model of RNA polymerase binding to DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 103: 5332–5337.
- H. Alper, C. Fischer, N. Elke, G.S., 2006. Tuning genetic control through promoter engineering, 103: 12678-12683
- Hawley, D.K., McClure, W.R. & Limited, I.R.L.P., 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucleic Acid Research*, 11: 2237-2255
- Hayashi, G. et al., 2012. Activation of prokaryotic translation by antisense oligonucleotides binding to coding region of mRNA. *Biochemical and Biophysical Research Communications*, 429: 105–110.
- Hayashi, K. et al., 2006. Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. *Molecular Systems Biology*, 2: 2006-0007
- Hermes, Jeffrey, Prekh, Shirish, Koster, Hubert, Knowles, J., 1989. A reliable method for random mutagenesis: deoxyribonucleotide primers. *Gene*, 84: 143–151.
- Hershey, J.W.B., Sonenberg, N. & Mathews, M.B., 2012. Principles of translational control: an overview. *Cold Spring Harbor Perspectives in Biology*, 4: 1-10
- Hornung, G. et al., 2012. Noise-mean relationship in mutated promoters. *Genome Research*, 22: 2409–2417.
- Huerta, A.M. & Collado-Vides, J., 2003. Sigma⁷⁰ promoters in *Escherichia coli*: Specific transcription in dense regions of overlapping promoter-like signals. *Journal of Molecular Biology*, 333: 261–278.
- J. Starmer, A.Stomp, M. Vouk, D.B., 2006. Predicting Shine–Dalgarno sequence locations. *Computational Biology*, 2: 454–466.
- Kaczanowska, M. & Rydén-Aulin, M., 2007. Ribosome biogenesis and the translation process in *Escherichia coli*. *Microbiology and Molecular Biology Reviews*, 7: 477–494.
- Kinney, J.B. et al., 2010. Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. *Proceedings of the National Academy of Sciences of the United States of America*, 107: 9158–9163.
- Korzheva, N., 2000. A structural model of transcription elongation. *Science*, 289: 619–625.
- Kosuri, S. et al., 2013. Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, 110: 14024–1409.
- Kyle Jensen, Hal Alper, Curt Fischer, and G.S., 2006. Identifying functionally important mutations from phenotypically diverse sequence data. *Applied and Environmental Microbiology*, 72: 3696 – 3701.

- Laursen, B.S. et al., 2005. Initiation of protein synthesis in bacteria. *Microbiology and Molecular Reviews*, 69: 101–123.
- Lu, C., Bentley, W.E. & Rao, G., 2004. A high-throughput approach to promoter study using green fluorescent protein. *Biotechnology Progress*, 20: 1634–1640.
- M. François Jacob, M.A.U. and J.M., 1964. Le promoteur, élément génétique nécessaire à l'expression d'un opéron. *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences*, 16: 3125 – 3128.
- Mathews, D.H. et al., 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *Journal of Molecular Biology*, 288:911–940.
- De Mey, M. et al., 2007. Construction and model-based analysis of a promoter library for *E. coli*: an indispensable tool for metabolic engineering. *BMC Biotechnology*, 7:34.
- De Mey, M. et al., 2010. Promoter knock-in: a novel rational method for the fine tuning of genes. *BMC Biotechnology*, 10: 26
- Mijakovic, I., Petranovic, D. & Jensen, P.R., 2005. Tunable promoters in systems biology. *Current Opinion in Biotechnology*, 16: 329–35.
- Peter Ruhdal Jensen, and K.H., 1998. The Sequence of spacers between the consensus sequences. *Applied and Environmental Microbiology*, 64: 82 – 87.
- Pon, C.O.G. & L., and C., 1990. Initiation of mRNA translation in prokaryotes? *American Chemical Society*, 29: 5881 – 5889.
- Pribnow, D., 1975. Bacteriophage T7 early promoters: nucleotide sequences of two RNA polymerase binding sites. *Journal of Molecular Biology*, 99: 419–43.
- Qu, X. et al., 2012a. Ribosomal protein S1 unwinds double-stranded RNA in multiple steps. *Proceedings of the National Academy of Sciences of the United States of America*, 109: 14458-14463.
- Recknagel, K.E. & Rolf-Dieter, A., 1994. Promoter strength prediction based on occurrence frequencies of consensus patterns. *Journal of Theoretical Biology*, 171: 355–359.
- Reznikoff William, L.G., 1986. *Maximizing gene expression*, Boston: Butterworth Publishers, Stoneham, MA.
- Rhodium, V. a et al., 2006. Conserved and variable functions of the sigma^E stress response in related genomes. *PLoS Biology*, 4: 43-59
- Rhodium, V. a & Mutalik, V.K., 2010. Predicting strength and function for promoters of the *Escherichia coli* alternative sigma factor, sigma^E. *Proceedings of the National Academy of Sciences of the United States of America*, 107: 2854–2859.

- Ringquist, S. et al., 1992. Translation initiation in *Escherichia coli* : sequences within the ribosome-binding site. *Molecular Microbiology*, 6: 1219 – 1229.
- Rosenberg, M. & Court, D., 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. *Annual Review of Genetics*, 13: 319–353.
- Rossi, J.J. et al., 1983. Biological expression of an *Escherichia coli* consensus sequence promoter and some mutant derivatives *Biochemistry. Proceedings of the National Academy of Sciences of the United States of America*, 80: 3203–3207.
- Salis, H.M., Mirsky, E. a & Voigt, C. a, 2009. Automated design of synthetic ribosome binding sites to control protein expression. *Nature Biotechnology*, 27: 946–950
- Santos, C.N.S. & Stephanopoulos, G., 2008. Combinatorial engineering of microbes for optimizing cellular phenotype. *Current Opinion in Chemical Biology*, 12: 168 –176.
- Santos, C.N.S. & Stephanopoulos, G., 2008b. Combinatorial engineering of microbes for optimizing cellular phenotype. *Current Opinion in Chemical Biology*, 12: 168 –176.
- Shultzaberger, R.K. et al., 2001. Anatomy of *Escherichia coli* ribosome binding sites. *Journal of Molecular Biology*, 313: 215–228
- De Smit, M.H. & van Duin, J., 1994. Control of translation by mRNA secondary structure in *Escherichia coli*. A quantitative analysis of literature data. *Journal of Molecular Biology*, 244: 144–150.
- Sofia Österberg, T. del P.-S. and V.S., 2011. Regulation of alternative sigma factor use. *Annual Review of Microbiology*, 65: 37–57.
- Sondek, J. & Shortle, D., 1992. A general strategy for random insertion and substitution mutagenesis: substoichiometric coupling of trinucleotide phosphoramidites. *Proceedings of the National Academy of Sciences*, 89: 3581–3585.
- Tumell, B. & Harrison, D., 1990. Consensus patterns in DNA. *Methods in Enzymology* , 183: 211-221
- Zaccolo, M. & Gherardi, E., 1999. The effect of high-frequency random mutagenesis on in vitro protein evolution: a study on TEM-1 beta-lactamase. *Journal of molecular biology*, 285:775–783.
- Zehua Chen, K.A.L. and T.D.S.R.K.S., 2007. Anatomy of *Escherichia coli* sigma⁷⁰. *Nucleic Acids Research*, 35: 771 – 788.