



UNIVERSIDAD NACIONAL AUTÓNOMA DE MEXICO
DOCTORADO EN CIENCIAS BIOMÉDICAS
CENTRO DE CIENCIAS GENÓMICAS

ANÁLISIS DEL EFECTO DE ESTRUCTURAS SECUNDARIAS ALREDEDOR
DEL SITIO DE UNIÓN A RIBOSOMA SOBRE LA TRADUCCIÓN DE
PROTEÍNAS. GFP COMO CASO DE ESTUDIO.

QUE PARA OPTAR POR EL GRADO DE:

DOCTOR EN CIENCIAS

PRESENTA:

M. EN B. JOSÉ LUIS RODRÍGUEZ MEJÍA

TUTOR PRINCIPAL

DR. ENRIQUE MERINO PÉREZ
INSTITUTO DE BIOTECNOLOGÍA

MIEMBROS DEL COMITÉ TUTOR

DR. GUILLERMO GOSSET LAGARDA
INSTITUTO DE BIOTECNOLOGÍA

DR. ALEJANDRO GARCÍA DE LOS SANTOS
CENTRO DE CIENCIAS GENÓMICAS

CUERNAVACA, MORELOS, MAYO 2017



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.

Índice	
Índice de tablas y figuras	1
Abreviaturas	3
Agradecimiento	4
Resumen	5
Summary	6
1. Introducción	7
1.1 Estructuras secundarias y su relevancia en la regulación traduccional	7
1.2 GFP como modelo de expresión	14
1.3 GFP y la mejora en la expresión de proteínas	17
2. Antecedentes	20
3. Hipótesis	25
4. Objetivo	26
5. Objetivos particulares	26
6. Estrategia experimental	27
7. Materiales y Métodos	28
7.1 Plásmidos y reactivos	28
7.1.2 Amplificación de las variantes del gen EGFP mediante PCR	28
7.1.3 Restricción enzimática	30
7.1.4 Cepas utilizadas en este estudio	30

7.1.5 Ligación de productos de PCR en plásmido	31
7.1.6 Transformación mediante el método de electroporación de las ligaciones obtenidas	31
7.1.7 Electroforesis de ácidos nucleicos y proteínas	32
7.1.8 Purificación de DNA plasmídico, productos de PCR y limpieza de reacciones enzimáticas	32
7.1.9 Análisis de proteínas	33
7.1.10 Análisis de crecimiento y fluorescencia	34
7.1.11 Análisis de estructura secundaria en el mRNA	35
7.1.12 Oligonucleótidos utilizados	36
8. Resultados	37
8.1 Obtención de las variantes moleculares del gen EGFP	37
8.2 Análisis de transformantes y su fluorescencia	43
8.3 Análisis de proteína por Western-blot	45
8.4 Análisis de las construcciones en una cinética de crecimiento y fluorescencia	50
8.5 Análisis del efecto producido por la remoción del codón que codifica para la Gly4 en GFP y el estructuramiento del mRNA de sus correspondientes variantes	56
9. Discusión	70
10. Conclusiones	73
11. Perspectivas	74

12. Manuscrito: A Codon Deletion at the Beginning of Green Fluorescent Protein Genes Enhances Protein Expression	75
13. Tabla suplementaria del manuscrito	86
14. Bibliografía	87

Índice de tablas y figuras

Tabla 1. Componentes y volúmenes utilizados en las reacciones de PCR	29
Tabla 2. Programa general utilizado para las reacciones de PCR	29
Tabla 3. Componentes para las reacciones enzimáticas	30
Tabla 4. Oligonucleótidos utilizados para la obtención de las variantes del gen EGFP	36
Tabla 5. Análisis de densitometría Western-blot figura 13. Los porcentajes obtenidos fueron evaluados respecto al área analizada	47
Tabla 6. Análisis de densitometría Western-blot figura 14. Los porcentajes obtenidos fueron evaluados respecto al área analizada	49
Tabla 7. Análisis de las variables de crecimiento en las ocho variantes moleculares analizadas	52
Figura 1. Factores que controlan y afectan el proceso de traducción	11
Figura 2. Sesgo estadístico de los valores de energía libre de desestabilización de los mRNA de <i>E. coli</i>	13
Figura 3. Estructura 3D de la proteína verde fluorescente GFP de <i>Aequorea victoria</i>	15
Figura 4. Esquemización de las diferentes mutaciones en cinco variedades de GFP	16
Figura 5. Esquemización de las diferentes mutaciones en cuatro variedades de DsRed	19
Figura 6. Sobreposición 3D de EGFP y EGFP Δ G4	24
Figura 7. Diagrama de flujo estrategia experimental	27
Figura 8. Amplificación del gen <i>EGFP</i>	38
Figura 9. Amplificación del gen <i>EGFPΔG4</i>	39
Figura 10. Amplificación de los genes <i>EGFP</i> , <i>EGFPΔG4</i> , <i>rsGFP</i> y <i>rsGFPΔG4</i>	41
Figura 11. Amplificación del gen <i>rsEGFP</i>	42
Figura 12. Estriados en LB solido de las ocho variedades de GFP obtenidas	43

Figura 13. Western-blot para la detección de proteína en las variantes Trc-EGFP y Trc-rsEGFP	46
Figura 14. Análisis de extracto total de proteína en gel de acrilamida desnaturizante SDS-PAGE.	48
Figura 15. Cinética fluorescencia de las cepas de <i>E. coli</i> BL21-Gold (DE3) y MC1061 transformadas con cada una de nuestras construcciones de GFP	51
Figura 16. Cinética fluorescencia de las cepas de <i>E. coli</i> BL21-Gold (DE3) y MC1061 transformadas con T7-EGFP, Trc-EGFP, Trc-rsEGFP y Trc-rsGFP.	53
Figura 17. Cinética fluorescencia de las cepas de <i>E. coli</i> BL21-Gold (DE3) y MC1061 transformadas con cada una de nuestras construcciones de GFP con la remoción del codón Gly4	55
Figura 18. Esquema de las secuencias nucleotídicas de los promotores Trc y T7	57
Figura 19. Secuencias nucleotídicas de la región líder e inicio del gen codificante de las diversas construcciones empleadas en el estudio	58
Figura 20. Estructuras secundarias de la variante T7-EGFP y T7-EGFP Δ G4	62
Figura 21. Estructuras secundarias de la variante Trc-EGFP y Trc-EGFP Δ G4	64
Figura 22. Estructuras secundarias de la variante Trc-rsEGFP y Trc-rsEGFP Δ G4	66
Figura 23. Estructuras secundarias de la variante Trc-rsGFP y Trc-rsGFP Δ G4	68
Figura 24. Gráficas de entropía posicional para las ocho variantes de GFP	69

Abreviaturas

DNA: Ácido de desoxirribonucleico.

mRNA: RNA mensajero, molécula que se produce de la transcripción del ADN.

GFP: por sus siglas en inglés Green Fluorescent Protein.

avGFP: Versión parental de la GFP, obtenida de la medusa *Aequorea victoria*.

EGFP: por sus siglas en inglés Enhanced Green Fluorescent Protein, versión mejorada de la avGFP.

rsGFP: por sus siglas en inglés Red Shift Green Fluorescent Protein, versión comercial en el plásmido pQBI25.

rsEGFP: por sus siglas en inglés Red Shift Green Fluorescent Protein, versión obtenida por mutación sitio específica por el equipo del Dr. Paul Gaytán IBT-UNAM.

G4 y Gly4: abreviaciones que refieren a la posición del aminoácido glicina en la posición cuatro de las diferentes versiones de GFP, en este trabajo.

ΔG4: abreviación que refiere la remoción del aminoácido glicina en la posición cuatro de las diferentes versiones de GFP, en este trabajo.

T7: abreviación que refiere al promotor sintético placT7, inscrito en el plásmido pET-28^a.

Trc: abreviación que refiere al promotor Trc del plásmido pJOQ, tomado del plásmido pTrc97.

UTR: por sus siglas en inglés Untranslated Region, denota la región que no se traduce en el mRNA.

RBS: Ribosomal Binding Site por sus siglas en inglés, es el segmento de mRNA que consiste en +/- 18 nucleótidos de cada lado del ATG y que en su región 5' contiene la secuencia Shine-Dalgarno (Steitz & Jakes 1975).

SD: Siglas de Shine-Dalgarno, es una secuencia nucleotídica propia de mRNAs procariontes situada a 6 o 7 nucleótidos antes del ATG, posee una secuencia consenso AGGAGG y forma parte del RBS.

REDOX: Reducción Oxidación, denota la transferencia de electrones en una reacción.

in vivo: del latín, dentro de lo vivo, experimento realizado en un ambiente controlado.

in silico: no proviene del latín y significa simulación computacional.

LED: por sus siglas en inglés de Light Emitting Diode.

mM: concentración molar expresada en 10⁻³ mol/L.

nm: unidad de medición de radiación ultravioleta, radiación infrarroja y luz visible.

PCR: siglas en inglés que refieren a la reacción en cadena de la polimerasa

IPTG: abreviación para isopropil-β-D-1-tiogalactopiranosido, carbohidrato utilizado para la inducción de actividad β-galactosidasa.

mRFP1: monómero funcional obtenido a partir de Red Fluorescent Protein de *Discomia sp.*

DsRed: acrónimo para *Discomia sp.* Fluorescent Protein.

mCherry: monómero funcional obtenido a partir de mRFP1.

MVSKGEE: siglas de los aminoácidos Metionina, Valina, Serina, Lisina, Glicina y Ácido glutámico, que conforman el péptido reportado por (Shaner et al. 2004).

Quisiera agradecer al Programa de Doctorado en Ciencias Biomédicas de la UNAM, a la Dirección General de Estudios de Posgrado de la UNAM y al Consejo Nacional de Ciencia y Tecnología (CONACYT) por la beca proporcionada con No. 176337 para realizar este proyecto de doctorado. Al igual que los fondos proporcionados para asistir a congresos que me otorgaron ambas instituciones.

Agradezco también al comité revisor

Dr. Norma Adriana Valdez Cruz

Dr. Rosa Estela Navarro González

Dr. Nuria Victoria Sánchez Puig

Dr. José Adelfo Escalante Lozada

Dr. Enrique Merino Pérez

Esta tesis se realizó en el Departamento de Microbiología Molecular, en el Instituto de Biotecnología de la Universidad Nacional Autónoma de México

RESUMEN

A lo largo de más de cuatro décadas los esfuerzos por obtener mejoras en la expresión de genes o sistemas de importancia biotecnológica han sido objetivo principal de muchos grupos de investigación. Trabajos como el de Laila Berg y colaboradores en 2012 exploran la región 5'UTR del RNA como posible blanco en la mejora de la expresión de un gen recombinante (Berg et al. 2012). En 2013 Levin-Karp y colaboradores desarrollan un análisis cuantitativo para la mejora de la expresión de genes a sus productos en un operón sintético (Levin-Karp et al. 2013). Por último, en 2014 Woo Seo y colaboradores implementan una metodología *in silico* acoplada a un sistema de expresión *in vivo*, que revela la importancia de un marco nucleotídico alrededor del codón de inicio de sus genes reporteros (Seo et al. 2014). Si bien se ha explorado ampliamente el destino de la expresión de genes recombinantes, aun hoy en día quedan algunas variables por desvelar, en el sentido de que año tras año se incrementa la demanda en el aprovechamiento de nuevas proteínas o la mejora de la obtención de éstas.

En este trabajo se desarrolló el estudio de la delección de un codón al inicio del marco de lectura del gen que codifica para la proteína verde fluorescente y su repercusión en la fluorescencia final de ésta. Un análisis integrativo *in silico* e *in vivo* nos permitió concluir que el incremento en la fluorescencia de las diferentes versiones de GFP y EGFP se da por la reducción de estabilidad de estructuras secundarias en la región líder de sus mRNAs tal y como se demuestra en la configuración de la estructura secundaria de cada versión del gen GFP analizada y que se pudiera relacionar con una mayor tasa del inicio de su traducción y su correspondiente incremento en la síntesis final de su proteína.

SUMMARY

For more than four decades, efforts to improve gene expression or biotechnology-relevant systems have been the focus of many research groups. Works such as that of Laila Berg et al., in 2012 exploring the 5'UTR region of RNA as a possible target in enhancing the expression of a recombinant gene (Berg et al., 2012); in 2013 Levin-Karp and colleagues developed a quantitative analysis to improve the expression of genes to their products using a synthetic operon (Levin-Karp et al., 2013); finally in 2014 Woo Seo and colleagues implement an *in silico* methodology coupled with an *in vivo* expression system that reveals the importance of a nucleotide frame around the start codon of their reporter genes (Seo et al., 2014). Although the fate of expression of recombinant genes has been widely explored, there are still some variables to be unveiled today, in the sense that year after year the demand for new proteins increases or the improvement of the production of these.

In this work the study of the deletion of a codon at the beginning of the reading frame of the gene coding for the green fluorescent protein and its repercussion in the final fluorescence of this is developed. An integrative *in silico* and *in vivo* analysis allowed us to conclude that the increase in fluorescence of the different versions of GFP is due to the reduction of stability of secondary structures in the leader region of their mRNAs as demonstrated in the configuration of the secondary structure of each version of the analyzed GFP gene and that could be related to a greater translation initiation rate and with the corresponding final protein synthesis.

1. INTRODUCCIÓN

1.1 Estructuras secundarias y su relevancia en la regulación traduccional

La expresión eficiente de proteínas recombinantes no es un proceso sencillo por lo que a través de los años se han implementado innumerables estrategias que incluyen la optimización en los procesos de transcripción, traducción, plegamiento y estabilidad de las proteínas a ser sobreexpresadas. Respecto al proceso de traducción, éste se ha dividido en tres etapas consideradas cruciales en la optimización de todo el proceso: inicio, elongación y término, siendo uno de los más importantes, el del inicio de la traducción (Li 2015).

El inicio de la traducción en organismos bacterianos está determinado por la unión entre la subunidad ribosomal 30S y una región en el 5' del mRNA llamada sitio de unión a ribosomal (RBS, por sus siglas en inglés) o secuencia Shine-Dalgarno, en honor a los investigadores John Shine y Lynn Dalgarno quienes fueron los primeros en caracterizarla (Shine & Dalgarno 1975). Estudios realizados en 2003 describen que la subunidad 30S es capaz de anclar el mRNA con una estructura secundaria intrínseca y una vez que este se ha "relajado" es colocado sobre la plataforma de inicio de la traducción. Este proceso fue descrito al analizar la expresión de una proteína de cápside del bacteriófago MS2, en función de la estabilidad de la estructura en el RBS y la afinidad de la subunidad 30S para un RBS sin estructura. Este modelo ha permitido sugerir la existencia de un estado de espera a nivel ribosomal descrito como "*standby site*" (de Smit & van Duin 2003).

En el año de 2005, Marilyn Kozak publica una amplia revisión acerca de la regulación traduccional en procariontes y eucariontes en donde describe el efecto de estructuras secundarias estables en las regiones 5' del mRNA alrededor del RBS. Tanto Kozak como de Smit y van Duin describen dos posibles procesos de regulación. El primer proceso involucra la secuencia nucleotídica del SD en el mRNA (según consenso universal en procariontes AGGAGG) que interactúa con la secuencia anti-SD en el 16S ribosomal de la subunidad 30S ribosomal, que puede ser eficiente si la conservación es alta, o deficiente si la conservación es baja o alejada del consenso. El segundo proceso es mediado por el grado de estructuración que puede adoptar el mRNA al rededor del RBS, debido a la secuencia primaria, siendo aquellas más estructuradas las que impiden el reconocimiento del SD con una consecuente reducción del inicio de traducción (Kozak 2005; de Smit & van Duin 2003). En este mismo sentido, Schulz y Reznikoff al analizar la relación entre la estabilidad de las estructuras secundarias alrededor del RBS y la actividad de β -galactosidasa provenientes del gen reportero *lacZ* colocado inmediatamente río abajo, demostraron que la estructura secundaria del mRNA puede ser determinante en la eficiencia con la que la traducción se lleva a cabo (Schulz & Reznikoff 1990). Una de las principales conclusiones de este estudio fue que la eficiencia de un RBS es linealmente proporcional a la fracción de los mismos que carecen de estructuras secundarias.

La búsqueda para desvelar la importancia de las estructuras secundarias y qué papel tienen a nivel traduccional se describe por Studer y Joseph, quienes identifican la dinámica de ensamble del mRNA y la subunidad 30S del complejo

de iniciación de la traducción. Utilizando un conjunto de doce mRNAs diseñados y sintetizados *ad hoc* por tener estructuras secundarias con diferentes estabilidades y secuencias Shine-Dalgarno en diferentes posiciones, demostraron que existe una fuerte correlación entre la estabilidad de la estructura del mRNA y las constantes de asociación y velocidad de disociación. Adicionalmente observaron que para lograr la desestabilización de estructuras secundarias del mRNA se necesitan una secuencia SD, el codón de inicio, el tRNA iniciador para la N-Formilmetionina (fMet-tRNA), y la forma vinculada a GTP de factor de iniciación 2 unido a la subunidad 30S. Dado lo anterior, la desestabilización de las estructuras secundarias requieren de la interacción de la secuencia Shine-Dalgarno y la región anti-Shine-Dalgarno de la región 16S que se encuentra en la subunidad 30S, para así “fijar al mRNA y aplanarlo” dando como resultado un complejo estable entre el mRNA y la subunidad 30S. Cabe destacar que no se obtuvieron datos a nivel traduccional, pero de esta manera se entiende que los procariontes tienen formas de contender con obstáculos de este tipo, lo que pudiera interpretarse como un sistema ancestral al mecanismo poco entendido y más especializado de los eucariontes (Studer & Joseph 2006).

Las estructuras secundarias en el mRNA son una forma más de regulación en la expresión global de los genes como se detalla en la figura 1. En organismos procariontes, la ausencia o presencia de estructuras secundarias son elementos críticos ya que su presencia genera puntos de control en el tránsito hacia su traducción; pudiendo controlar la producción de proteínas que en demasía fueran tóxicas o aquellas que solo son necesarias en concentraciones pequeñas (Kozak 2005; Mahalik et al. 2014).

Un control de tránsito hacia la traducción pudiera ser un paso clave en la formación de proteínas. Estudios de visualización por crio-microscopía electrónica determinaron que el complejo traduccional y estructuras secundarias hacia el 5' de mRNAs interactúan acoplándose temporalmente en lo que la estructura secundaria es relajada, incluidos complejos secundarios como los Riboswitches, que convergen directamente en la plataforma de la subunidad 30S. A esta temporalidad de sujeción entre ambos elementos se le nombró como *standby site* (paso que solo se había especulado para el año 2003 con los trabajos de de Smit y Van Duin) que con nueva evidencia experimental se considera a las estructuras secundarias como partícipes activos en el destino de la traducción de mRNAs (Marzi et al. 2007).

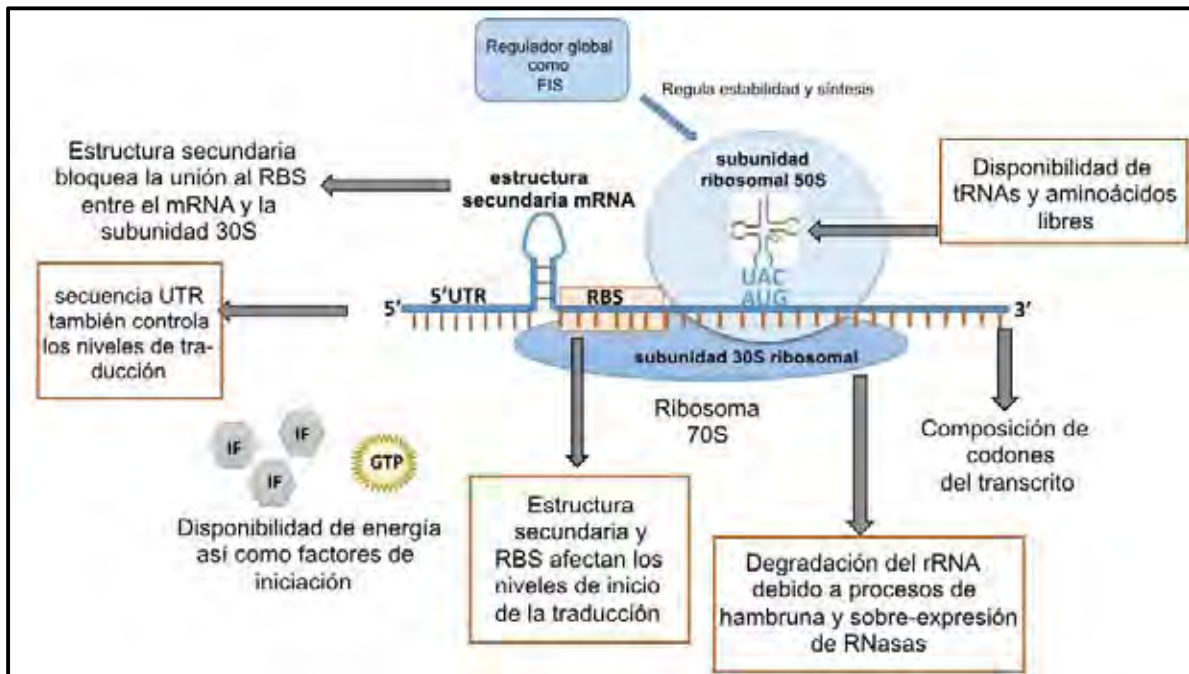


Figura 1. Factores que controlan y afectan el proceso de traducción. A nivel celular podrían estar involucrados una serie de mecanismos que afectan la traducción, estructuras secundarias en el mRNA, el sitio RBS y su unión a la región correspondiente de la subunidad 30S, la composición nucleotídica del mRNA, la región UTR de l mRNA. En adición a los anteriores elementos, el proceso de traducción puede verse influido por procesos circundantes tales como disponibilidad de factores de iniciación y energía, disponibilidad de tRNAs y aminoácidos, procesos de hambruna y la participación de algunos reguladores globales como el caso de FIS. Esquema adaptado de Mahalik *et al.*, 2014.

Una de las ideas centrales para la sobreexpresión de proteínas heterólogas es el de reducir las diferencias entre el uso de codones preferenciales del organismo huésped (por ejemplo *Escherichia coli*), y el uso de codones del gen a ser expresado. Con el propósito de evaluar el efecto de codones atípicos (comúnmente llamados codones “raros”) en la sobreexpresión de proteínas heterólogas, en 2009 Kudla y colaboradores generaron una librería de 154 genes

de GFP modificados aleatoriamente para introducir modificaciones sinónimas, de tal forma de mantener siempre la misma secuencia proteínica. Sus resultados demostraron que más del 50 % del efecto sobre los niveles de proteína no estaban relacionados por un sesgo hacia el uso preferencial de codones, sino por la presencia de estructuras secundarias estables dentro de la cercanía del sitio de unión a ribosoma. Los diferentes niveles de fluorescencia estuvieron relacionados con la cantidad total de proteína, misma que fue analizada mediante Western-blots. Los datos estadísticos sobre la posición nucleotídica de los codones sinónimos indicaron que ciertos arreglos locales de la secuencia estaban relacionados con la capacidad de formar estructuras secundarias alrededor del Shine-Dalgarno del mRNA que reducían la traducción de sus correspondientes mRNAs disminuyendo sus niveles de expresión (Kudla et al. 2009). Estos resultados fueron la base para que Gu y colaboradores en 2010 realizaran un análisis bioinformático de las tendencias generales del grado de estructuración del RNA alrededor del inicio de traducción de los genes. En su estudio se incluyó el análisis de la secuencia de 340 genomas completos de diferentes especies, incluidas bacterias, arqueas, hongos, plantas, peces, aves y mamíferos (Gu et al. 2010). Las conclusiones del estudio señalan que tanto en organismos procariontes como eucariontes existe una tendencia de encontrar regiones no estructuradas alrededor del inicio de las regiones codificantes de los genes, lo que podría facilitar el reconocimiento eficiente del codón de inicio de la traducción (Gu et al. 2010). En la figura 2 se muestran los valores promedio de energía libre de los mRNAs de *E. coli*. Como puede verse, la región cercana del codón de inicio (rango 1 a 21 nucleótidos) tiende a ser poco estructurada.

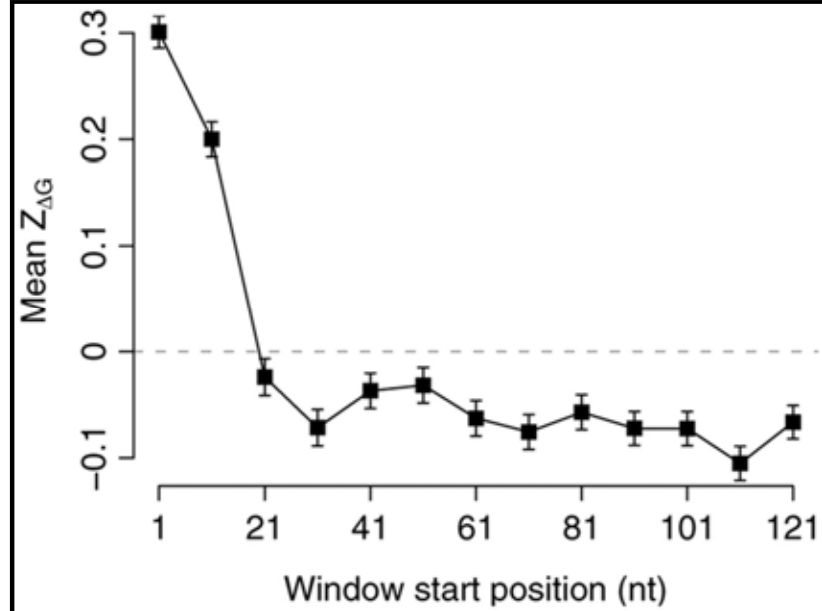


Figura 2. Sesgo estadístico de los valores de energía libre de desestabilización de los mRNA de *E. coli*. La Energía libre de plegamiento local (ΔG) a lo largo de los primeros 120 nt de la región codificante de los mRNAs de *E. coli* fue evaluada utilizando ventanas de análisis de 30 nucleótidos moviéndose desde el codón de inicio hasta el nucleótido 120 en pasos de 10 nucleótidos. Para evaluar la significancia estadística de los valores de energía dada de la secuencia real de aminoácidos de las proteínas y el uso preferencial de codones del organismo, se calculó la ΔG de genes hipotéticos producto de las permutadas aleatorias de codones sinónimos. El valor $Z_{\Delta G}$ mide el sesgo estadístico de la desviación de los valores observados respecto a los esperados al azar, valores de $Z_{\Delta G}$ positivos indican poco estructuramiento, valores negativos lo contrario. Figura tomada de (Gu et al. 2010).

1.2 GFP como modelo de expresión

Muchos han sido los genes utilizados como reporteros, cuyas cualidades han sido aprovechadas desde distintas aplicaciones de manera que permitan entender los diferentes factores que afectan la producción y expresión de proteínas recombinantes, siendo una de las más utilizadas la proteína verde fluorescente. La proteína verde fluorescente proveniente de la medusa marina *Aequorea victoria* (avGFP o GFP, por sus siglas en inglés) ha sido ampliamente utilizada desde la década de los 80's. La formación del cromóforo de esta proteína es espontánea, activado de manera post-traducciona l y requiriendo únicamente oxígeno y un buen plegamiento que la hace ser funcional en un amplio rango de organismos (Zimmer 2002). La proteína GFP está compuesta tan solo por 238 aminoácidos, desplegando una forma muy particular en forma de barril β conformado por 11 láminas β y al centro un alfa hélice que contiene las uniones covalentes del cromóforo, tal y como se muestra en la figura 3 (Tsien 1998).

Desde 1994 la búsqueda de variantes mejoradas de GFP ha sido el trabajo de muchos grupos de investigación, obteniendo así, el incremento en el rendimiento cuántico (*quantum yield*) de fluorescencia (Cormack et al. 1996), foto estabilidad, solubilidad, termotolerancia (Siemering et al. 1996; Zimmer 2002), color o sensibilidad a diferentes parámetros bioquímicos como pH o estados REDOX (Shaner et al. 2007), e incluso ha sido fusionada a polipéptidos con deficiencia en su plegado para incrementarlo (Pédelacq et al. 2006).

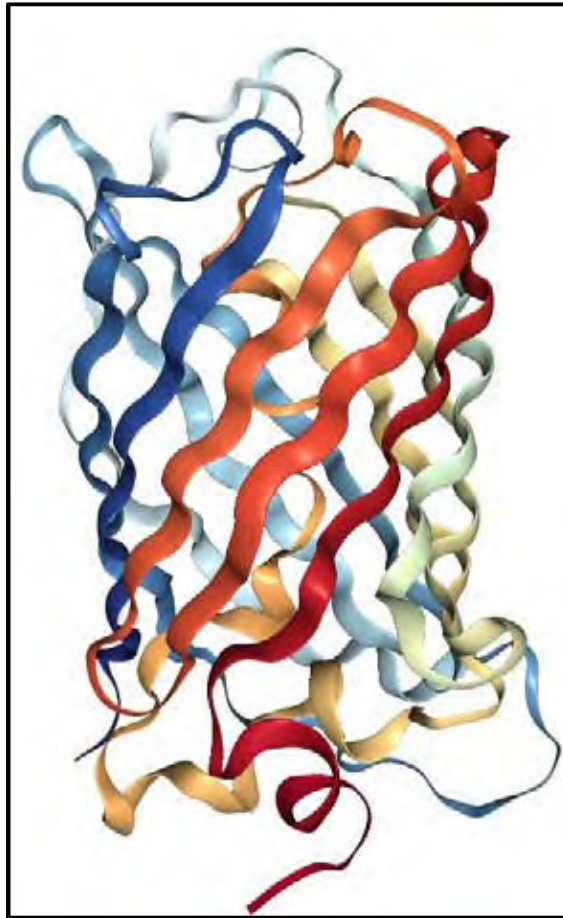


Figura 3. Estructura 3D de la proteína verde fluorescente GFP de *Aequorea victoria*. En la figura se muestran las 11 láminas β formando un cilindro a través del cual, está una alfa hélice que constituye el cromóforo. Adaptado de (Yang et al. 1996). Estructura obtenida de (Rose et al. 2016).

Muchas cualidades han sido reportadas y desarrolladas por diferentes grupos de investigación, incluidas modificaciones que permiten contar con cambios en la emisión de color, desde el verde, cian, amarillo, azul, hasta los colores aumentados como el caso de las “Enhanced GFP’s” (Rekas et al. 2002). Así también, se ha aprovechado la flexibilidad de sus secuencias terminales (amino-carboxilo) para mejorar y ampliar la paleta de colores de proteínas

fluorescentes derivadas de otros cnidarios y de la propia GFP de *Aequorea victoria*, figuras 4 y 5 (Shaner et al. 2004; Shaner et al. 2007).

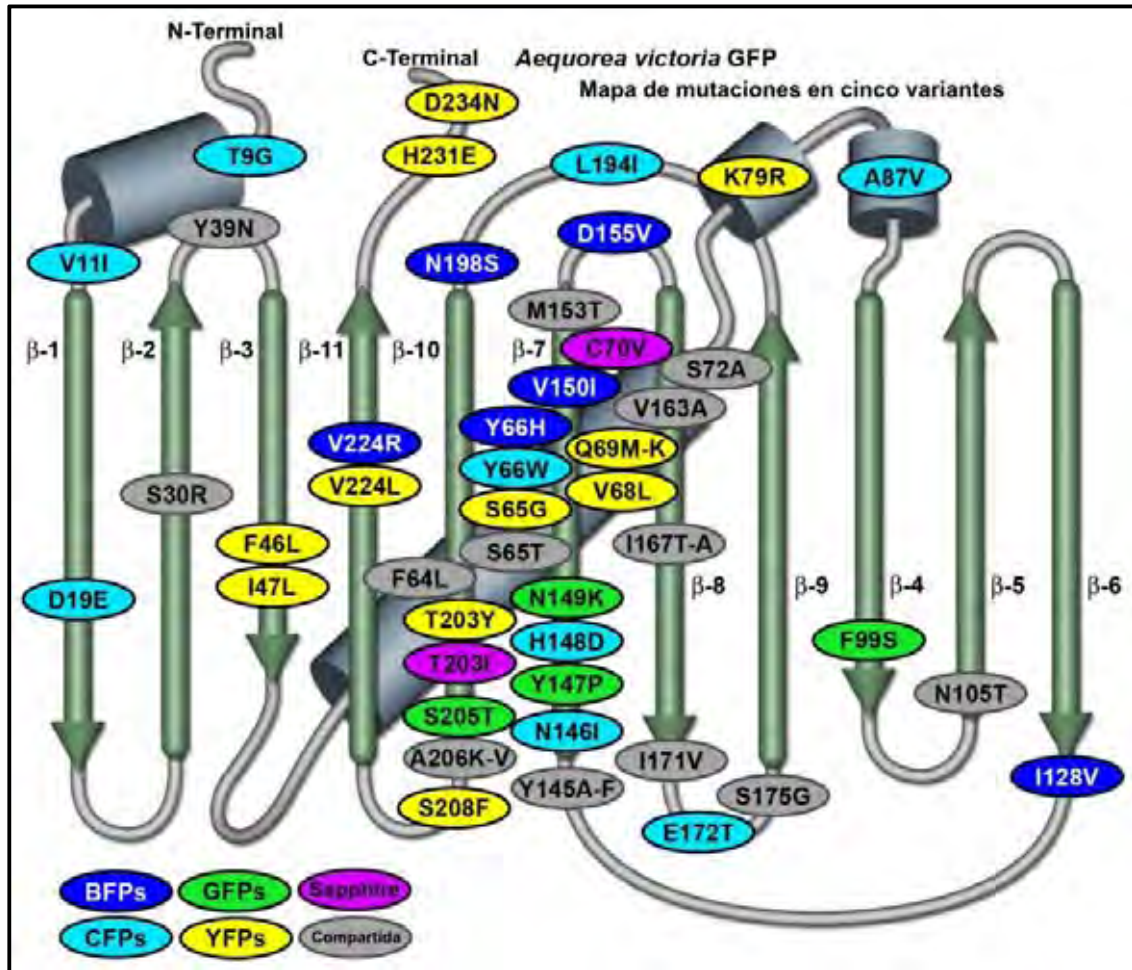


Figura 4. Esquemización de las diferentes mutaciones en cinco variedades de GFP. Las mutaciones que determinan el color para cada variedad está resaltado en el mismo. Amarillo: YFP's, Azul: BFPs, Cian: CFPs, Fucsia: Sapphire y el Gris denota las mutaciones que comparten entre estas variantes. Imagen adaptada de Shaner *et al.*, 2007.

1.3 GFP y la mejora en la expresión de proteínas

Existen diferentes aproximaciones metodológicas que han sido utilizadas para la sobreexpresión y modificación de las propiedades de proteínas fluorescentes, siendo uno de ellos la "Evolución dirigida". Este enfoque metodológico consiste básicamente en la generación artificial de un gran número de variantes de proteínas seguidas de un proceso de selección. Un primer ejemplo del uso de Evolución dirigida aplicado a la obtención de proteínas fluorescentes con propiedades mejoradas lo constituye la obtención de la proteína monomérica denominada mRFP1 (monomeric red fluorescent protein) (Campbell et al. 2002). Como es sabido, las proteínas fluorescentes silvestres conocidas actualmente son proteínas multiméricas. Con el objeto de obtener una versión de proteína fluorescente monomérica para su uso como proteína de fusión en estudios de localización celular, Robert E. Campbell y colaboradores sometieron al gen que codifica para la proteína fluorescente "DsRed" de *Discosoma sp.* a evolución dirigida para aumentar su velocidad de maduración, romper las interfaces de interacción entre sus subunidades, y finalmente restauración su fluorescencia ya como proteína monomérica.

Pese a la gran ventaja de ser monomérica, la proteína mRFP1 presentaba poca capacidad para tolerar fusiones en sus extremos amino y carboxilo terminal, por tal motivo, Shaner y colaboradores sometieron al gen mRFP1 a nuevos procesos de evolución dirigida mRFP1. Para evitar obtener mutaciones en el gen de mRFP1 que reestablecieran la agregación entre los monómeros de esta proteína y al mismo tiempo se tolerara fusiones en sus extremos amino y

carboxilo, se utilizaron los primeros siete aminoácidos del amino terminal y los últimos siete del carboxilo terminal de GFP de *Aequorea victoria*, que de manera natural es apta para fusiones traduccionales en sus extremos. Las diferentes variedades obtenidas presentaron emisiones distinguibles entre la gama de amarillos a los rojos (Fig. 5). Además las variedades obtenidas presentaron mejoras considerables, desde la maduración de sus cromóforos, tolerantes a fusiones en sus extremos terminales y mejoras en su fotoestabilidad (Shaner et al. 2004).

En 2006 Pédelacq y colaboradores buscaron proteínas GFP mejor plegadas e incorporaron una estrategia de evolución molecular por recombinación *in vitro*, la cual consiste en la digestión parcial enzimática de los genes blanco y su recombinación en una PCR sin oligonucleótidos, en cada ciclo cada uno de los fragmentos actúa como oligonucleótido dada su homología, la recombinación ocurre cuando cada fragmento generado sirve como oligonucleótido para otra copia. El método desarrollado por los autores fue denominado “método de interferencia de plegado” y consistía en la fusión de la proteína GFP con proteínas con plegado deficiente (dominios de interferencia). Inicialmente este tipo de fusión resultaba en proteínas híbridas con fluorescencia muy pobre o sin fluorescencia. Posterior a varios ciclos de mutaciones, se seleccionó aquella proteína que recuperó la fluorescencia. A la sección proteína GFP con mejor fluorescencia se le denominó “superfolder green fluorescent protein” (Pédelacq et al. 2006).

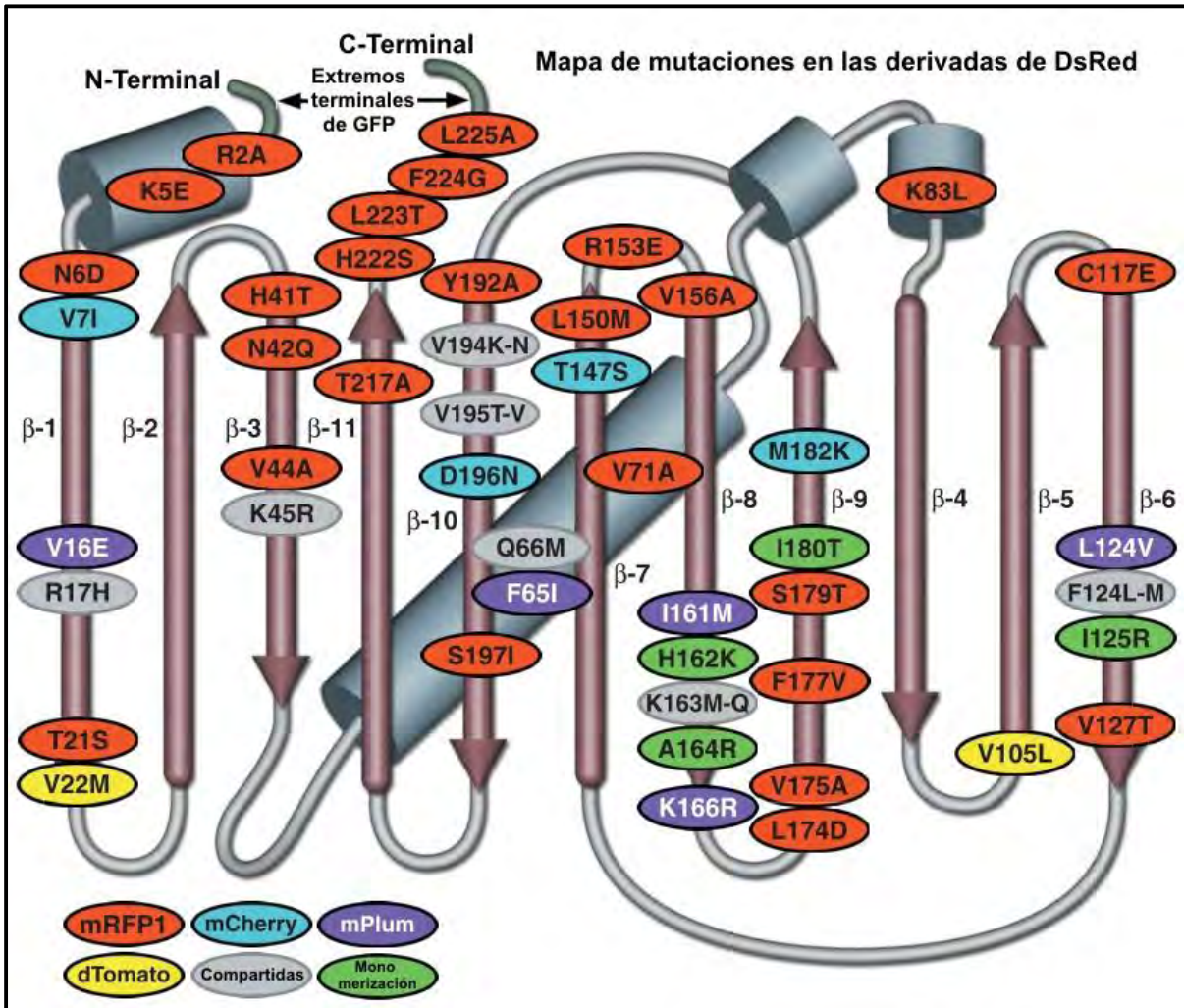


Figura 5. Esquematación de las diferentes mutaciones en cuatro variedades de proteína DsRed. Las mutaciones que determinan el color para cada variedad están resaltadas en la misma figura. Naranja: mRFP1, Cian: mCherry, Morado: mPlum, Amarillo: dTomato, Gris denota las mutaciones que comparten entre estas variantes y Verde: mutaciones para la monomerización. Nótese el uso de los extremos terminales para la obtención de estas derivadas. Imagen adaptada de Shaner *et al.*, 2007.

2. Antecedentes

Una de las metas principales de la Ingeniería de proteínas es lograr la sobreproducción de proteínas o la generación de variantes proteicas con actividades mejoradas. Con tal fin, en los últimos años, la Ingeniería de proteínas ha implementado diferentes estrategias, una de las cuales ampliamente utilizada es la mutagénesis al azar. La sustitución al azar de aminoácidos es fácil de implementar mediante procesos de barajeo de DNA o por PCR errónea (controlada por el uso de polimerasas de baja fidelidad). Por el contrario, la delección azarosa de aminoácidos es más difícil de desarrollar ya que requiere el uso de librerías especiales de oligonucleótidos o el uso de transposones o pasos de modificación multi-enzimático (Jones 2005; Murakami et al. 2002). Contrario a estas metodologías, la mutagénesis por deleciones, utilizando oligonucleótidos, genera remoción de codones de manera limpia y específica, a diferencia de las sustituciones enzimáticas que generan cambios no deseados alrededor de las deleciones que se pretenden hacer. Teniendo esto en cuenta, el Dr. Paul Gaytán del Instituto de Biotecnología-UNAM y colaboradores en el año 2004 desarrollaron una estrategia de armado de librerías por oligonucleótidos que fue nombrada COBARDE (por sus siglas en inglés, COdon-BAsed Random Deletion) y que fue utilizada para la obtención de mutantes en el gen de la proteína β -lactamasa alrededor del lazo omega sin perder propiedades catalíticas, como lo demostraron por los ensayos de resistencia bacteriana al antibiótico (Osuna et al. 2004). La aplicación de COBARDE demostró que este método puede generar la mezcla necesaria de modificaciones a lo largo de casi cualquier secuencia nucleotídica,

generando cambios aleatorios en todo el marco de lectura o mediante el cambio específico de regiones codificantes de dominios (Osuna et al. 2004).

En 2007 la estrategia COBARDE fue utilizada para generar modificaciones en el lazo más grande de la proteína sgGFP, correspondiente a los residuos 129 al 142 y que en previos reportes no se habían obtenido con éxito. 16,384 variantes fueron obtenidas y solo dos de ellas tuvieron actividad fluorescente: sgGFP- Δ I 129 y sgGFP- Δ D130. Ambas mutantes presentaron fluorescencia mayor a la parental a una temperatura de 30 °C, no así a una temperatura de 37 °C. COBARDE tuvo una cobertura de mutagénesis del 50%, demostrando ser una estrategia viable para la introducción de cambios y deleciones específicas en otros modelos de estudio. Ambas variantes presentaron datos interesantes acerca de la tolerancia a mutaciones de este lazo en el que se presentan desde aminoácidos espaciadores hasta aminoácidos que permiten un mejor plegamiento estructural, así como, que las deleciones en esta región sí tienen un papel en la configuración exitosa en la proteína completa, afectando directamente su estabilidad (Flores-Ramírez et al. 2007).

Alentados por la búsqueda de mejores variantes de la ya comercializada EGFP, recientemente Arpino y colaboradores en 2014 reportaron la obtención de una proteína EGFP cuya recombinación en el residuo Gly4 (Δ Gly4) incrementa la fluorescencia considerablemente a diferencia de la EGFP parental, ambas cuando son transcritas a partir del promotor T7, Uno de los objetivos desarrollados por Arpino y colaboradores fue la de cubrir un análisis de evolución por deleción de codones a lo largo de los 720 pb que corresponden al marco de lectura de la

proteína. El banco de mutantes se obtuvo utilizando mutagénesis por transposon y se obtuvieron 87 eventos únicos de delección de codones, 42 de estos sugieren ser mejor tolerados contrario a lo que se había reportado por Flores Ramírez y colaboradores en 2007 (Arpino et al. 2014). De acuerdo a sus resultados, una de las mutaciones se ubica en el amino terminal de la EGFP antes del inicio de la hélice 1, en el aminoácido cuatro que corresponde a una Glicina. Este cambio presentó una fluorescencia mayor a la parental. Los datos de fluorescencia se obtuvieron utilizando cepas de *E. coli* crecidas a 25 °C y 37 °C y proponiendo que la delección de la Gly4 habría cambiado la forma en la que la EGFP se pliega, dando como resultado el incremento en fluorescencia total. De manera muy especulativa se afirmó que esta pequeña delección creaba un cambio enorme, ya que utilizando una representación gráfica de “Wheel” en la que solo se incluyen seis residuos de la región amino terminal y no los 11 que la conforman, aprecia una variación o desplazamiento de aminoácidos ácidos por hidrofóbicos. Este cambio se analizó realizando un análisis de área de la superficie accesible a solvente (SASA, por sus siglas en inglés) que es utilizada para determinar el área de la superficie de una biomolécula que es accesible a un solvente, ya que al desplazar un aminoácido hidrofóbico a una región estructuralmente expuesta a solvente, se quería saber si los valores de accesibilidad cambiaban. El análisis solo incluyó la región entre los residuos K3-L231 (parte de la tapa de uno de los extremos del cilindro de EGFP) y se determinó que el haber quitado la Gly4 generó una constricción estructural que hacía menos accesible a solvente esta área. Para confirmar este cambio se incluyó un sobre-posicionamiento de

estructuras 3D en el que se denota la posición de la Gly4 en la estructura (Fig. 6) (Arpino et al. 2014).

El reporte de esta nueva remoción Gly4 propuesta por Arpino y colaboradores es importante porque la remoción del residuo Gly4 que se genera en la región amino-terminal de la proteína no había sido caracterizada ni reportada, ya que los residuos de esta región terminal no son comúnmente considerados como relevantes para la fluorescencia. De hecho, reportes experimentales indican que la delección de toda la hélice 1 (correspondiente a los residuos 3 al 9) de la GFP no afecta la fluorescencia de la proteína y que solo los residuos que se encuentran en las hélices internas son esenciales, ya que la destrucción de éstas elimina completamente la actividad de la GFP (Li et al. 1997).

En nuestro laboratorio hemos clonado las versiones del gen *EGFP* reportadas por Arpino y colaboradores, utilizando el vector de expresión pJOQ, sin embargo, los resultados son opuestos a los propuestos por Arpino, la versión del gen *EGFP* produce más fluorescencia que la versión del gen *EGFP Δ G4*. Este resultado indicaría que la remoción de la Gly4 no tiene efecto sobre el plegamiento de la proteína, sino un efecto en la expresión del gen que se ve reflejada en la fluorescencia observada.

Esta nueva variante de EGFP abre la posibilidad de realizar un estudio con una estrategia integrativa que permita elucidar el efecto que tiene la remoción del residuo Gly4 sobre la expresión de EGFP, al haber una mejora intrínseca en esta

modificación, su comprensión permitiría extrapolar su aplicación en otros genes recombinantes de interés biotecnológico.

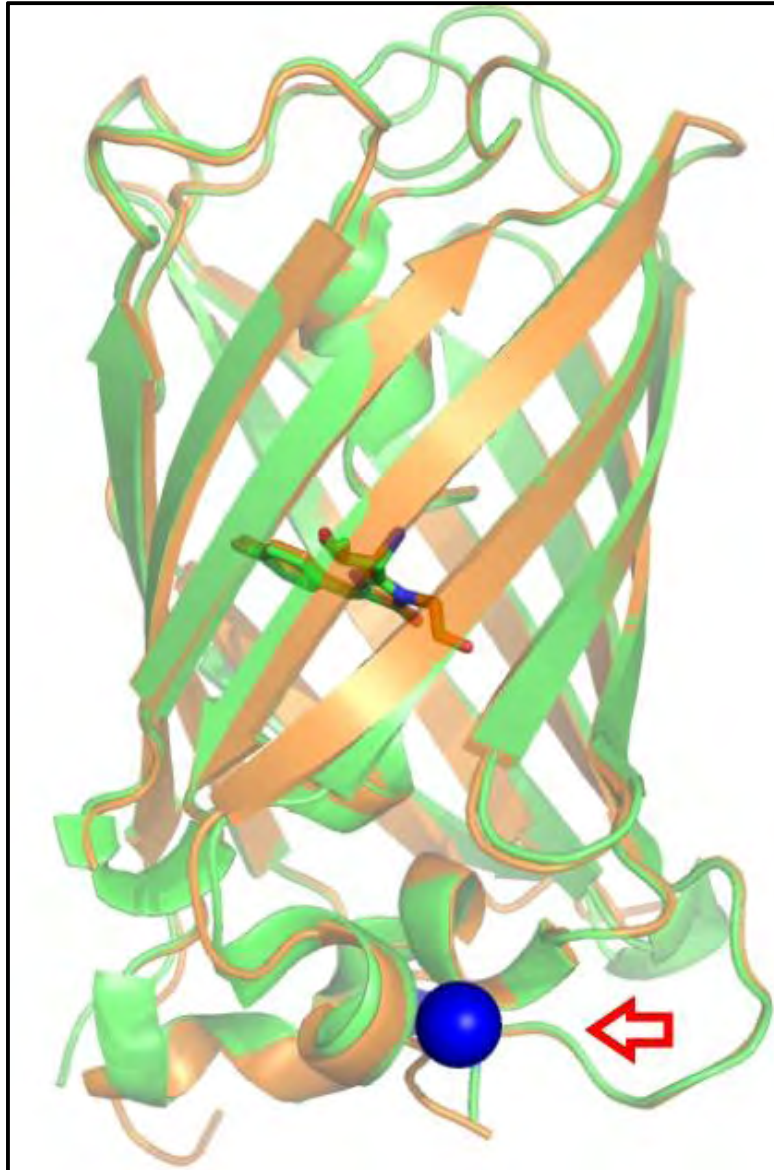


Figura 6. Sobreposición de EGFP (en verde) y EGFP Δ G4 (en naranja). La sobre posición de la estructura se utilizó para mostrar el área en la que se encuentra la Gly4 (flecha roja). Adaptado de Arpino *et al.*, 2014.

3. Hipótesis

La remoción del codón de glicina en la cuarta posición del gen que codifica para la proteína *GFP* afecta el proceso del inicio de la traducción de su correspondiente mRNA como consecuencia del cambio en la estabilidad de estructuras secundarias que pueden formarse alrededor del sitio de unión a ribosoma.

4. Objetivo

Analizar el efecto en la expresión del gen *GFP* producido por la remoción del codón de glicina en la cuarta posición de su correspondiente gen.

5. Objetivos particulares

- Evaluar la expresión de variantes moleculares del gen *GFP* con y sin el codón que codifica para la glicina cuatro.
- Análisis *in silico* de las estructuras secundarias de RNA de las regiones 5' de cada variante molecular de *GFP*.
- Analizar los resultados obtenidos *in silico* e *in vivo* para interpretar la relación existente entre estructura secundaria del RNA y expresión.

6. Estrategia experimental



Figura 7. Diagrama de flujo estrategia experimental. Para cumplir con el objetivo general de esta tesis se diseñó una estrategia que permite realizar comprobaciones *in vivo* retroalimentadas por los análisis bioinformáticos. Las variantes de la proteína EGFP se analizarán *in vivo* y su expresión final es analizada por estudios de Western-blot. Estos datos junto con el análisis *in silico* son usados para conocer el efecto de la delección de la glicina cuatro en la región RBS y su posible efecto sobre la traducción de cada mRNA.

7. Materiales y Métodos

7.1 Plásmidos y reactivos

Todos los reactivos y enzimas utilizados para la amplificación por PCR y clonación se compraron a New England BioLabs (NEB), incluyendo la enzima Vent DNA polimerasa. Los kits para la purificación de productos de PCR, extracción con gel de agarosa y plásmidos se compraron a Bio Basic, Inc. Para el análisis de proteínas los buffer de lisis y solubilización se compraron a Thermo Scientific.

Los genes expresados bajo el control del promotor T7 se clonaron como insertos *NcoI-XhoI* en el plásmido pET-28a (Novagen), mientras que los genes expresados bajo el control del promotor Trc se clonaron como insertos *NdeI-XhoI* en el plásmido no comercial denominado pJOQ. Este plásmido multicopia de 2001 bp fue construido por el Dr. Joel Osuna-Quintero usando una porción del plásmido comercial pET-28a y una porción del plásmido pTRC99a que posee el promotor Trc (Amann et al. 1988). pJOQ contiene el gen de resistencia al antibiótico kanamicina y el origen de replicación del pBR322. El plásmido carece del gen *lacI*, por lo que los genes clonados se expresan constitutivamente, en cepas de *E. coli* que carecen del gen *lacI* como la cepa MC1061.

7.1.2 Amplificación de las variantes del gen EGFP mediante PCR

Las variantes de *EGFP* se obtuvieron utilizando la técnica de reacción en cadena de la polimeras (PCR, por sus siglas en ingles), las reacciones de rutina (amplificación de un solo gen) se realizaron utilizando Vent polimerasa de New

England Biolabs (NEB). La amplificación por PCR se realizó utilizando las condiciones y recomendaciones por el fabricante, aunque en un volumen final de 100 μ L. A continuación se describe la reacción:

Tabla 1. Componentes y volúmenes utilizados en las reacciones de PCR

Reactivo	Volumen (μ L)
Templado (2-100 ng/ μ L)	1
dNTPs 10X	10
Buffer 10X	10
MgSo4 100mM	4
Enzima Vent (2U/ μ L)	1
Oligonucleótido Forward 10 μ M	5
Oligonucleótido Reverse 10 μ M	5
Agua desionizada	Llevar a 100 μ L

Las reacciones de PCR se realizaron en un termociclador C1000 Touch™

Thermal Cycler de Bio Rad, con el siguiente programa:

Tabla 2. Programa general utilizado para las reacciones de PCR

Paso	Temperatura °C	Tiempo (min.)	Ciclos
1	94 °C	3	1
2	94 °C	1	25
	58 °C	1	
	72 °C	1	
3	72 °C	5	1

Para la obtención de la variante rsEGFP se utilizó la técnica de mutagénesis por PCR sobrelapados (Ho et al. 1989). Esta técnica consiste en generar productos de PCR que complementan por homología y que contengan la mutación requerida, esta generada directamente en el oligonucleótido. Finalmente los productos, que pueden ser varios dependiendo el tamaño del gen a mutar, son utilizados en una PCR final para amplificar un solo producto que contiene el gen completo con los cambios requeridos.

7.1.3 Restricción enzimática

Las enzimas de restricción utilizadas son *NcoI*, *XhoI* y *NdeI*, de la marca New England Bio Labs (NEB). Las reacciones enzimáticas se realizaron en un volumen final de 80 μL a una temperatura de 37 $^{\circ}\text{C}$ durante 18 h, como se describe en la siguiente tabla:

Tabla 3. Componentes para las reacciones enzimáticas

Reactivo	Cantidad (μL)
Producto de PCR o plásmido (50-100 ng/ μL)	68 μL
Enzima (<i>NdeI</i> o <i>NcoI</i>) 10 U/ μL	1.5 μL
Enzima <i>XhoI</i> 10 U/ μL	1.5 μL
Buffer 3.1, <i>NdeI</i> , <i>XhoI</i> y <i>NcoI</i> presentan 100% actividad	8 μL
Agua desionizada	Llevar a 80 μL

7.1.4 Cepas utilizadas en este estudio

Se utilizaron dos cepas de *Escherichia coli*:

BL21 (DE3) genotipo: *E. coli* str. B F⁻ *ompT gal dcm lon hsdS_B(r_B⁻m_B⁻) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB⁺]_{K-12}(λ^S).*

MC1061 genotipo: *E. coli* str. K-12 F⁻ λ⁻ Δ(*ara-leu*)7697 [*araD139*]B/r Δ(*codB-lacI*)3 *galK16 galE15 e14⁻ mcrA0 relA1 rpsL150(Str^R) spoT1 mcrB1 hsdR2(r⁻m⁺)*.

La cepa de *E. coli* BL21 (DE3) se utilizó para la expresión de la proteína EGFP bajo la regulación del promotor T7. *E. coli* MC1061 se utilizó para la expresión de la proteína EGFP bajo la regulación del promotor Trc.

7.1.5 Ligación de productos de PCR en plásmido

Para la reacción de ligación, se mezclaron 200 ng de plásmido digerido, 200 ng de producto de PCR digerido, 1,5 unidades de T4 ADN ligasa (The Mo Bio Laboratories), 2 μL de buffer de ligasa y agua Milli Q estéril hasta un volumen final de 20 μL. La reacción se incubó durante 16 a 18 h a 22°C.

7.1.6 Transformación mediante el método de electroporación de las ligaciones obtenidas

De cada reacción de ligación se utilizaron únicamente 2 μL del volumen total para transformar 100 μL de células electrocompetentes. Las células competentes se obtuvieron por el método reportado por (Sharma & Schimke 1996). Cada alícuota transformada se sembró en placas de LB sólido suplementadas con 25 μg/mL de kanamicina más IPTG a una concentración final de 0.1 mM únicamente en la cepa BL21 (DE3), (para las construcciones bajo la regulación del promotor Trc en la cepa MC1061, no fue necesaria la adición de IPTG como inductor). Las placas se incubaron durante 18 h a 37 °C. Se eligieron

tres colonias fluorescentes, de las cuales se aislaron los plásmidos y se secuenciaron para confirmar la correcta amplificación y clonación del gen.

7.1.7 Electroforesis de ácidos nucleicos y proteínas

La electroforesis de ácidos nucleicos se realizó en gel de agarosa al 1%, los geles utilizados fueron de 8 cm X 6 cm de dimensión, se utilizó buffer TBE 1X como buffer de corrida. La corrida se realizó a 6.6 Volts/cm durante 40 min. Se utilizó como intercalante bromuro de etidio. Marcador molecular Thermo Scientific GeneRuler 100 bp DNA Ladder.

La electroforesis de proteínas se realizó en gel de acrilamida al 11.2% como gel separador y al 2% en una sección concentradora, los geles utilizados fueron de 12 cm X 10 cm de dimensión, se utilizó buffer Tris 1% Glicina 1% y SDS 0.1% como buffer de corrida. La corrida se realizó durante 1 h aproximadamente a 110 volts. El gel se tiñó en una solución de azul de Coomassie al 0.01% en una solución de metanol al 50% y ácido acético al 10% por 20 min. Se destiñeron en una solución de etanol al 20% y ácido acético al 5% por una hora. Marcador molecular PageRuler™ Prestained Protein Ladder, 10 to 180 kDa.

Los geles de agarosa y acrilamida se visualizaron en el equipo ESSENTIAL V2 Chrom CV-415.LS de UV ITEC® y se fotodocumentaron utilizando el software Essential V2.

7.1.8 Purificación de DNA plasmídico, productos de PCR y limpieza de reacciones enzimáticas

Todas las extracciones de plásmido se realizaron utilizando el kit EZ-10 Spin Column Plasmid DNA Minipreps Kit Bio Basic Inc, La purificación de los

productos de PCR obtenidos y la limpieza de las reacciones enzimáticas realizadas en este trabajo se hicieron utilizando el kit EZ-10 Spin Column PCR Products Purification Kit Bio Basic Inc. La extracción de DNA de bandas de gel y su purificación se realizó utilizando el kit EZ-10 Spin Column DNA Gel Extraction Kit Bio Basic Inc. El uso de estos kits se llevó a cabo con las recomendaciones y protocolos del fabricante.

7.1.9 Análisis de proteínas

Las células MC1061 de *E. coli* que se transformaron con las construcciones basadas en Trc se cultivaron a 37 °C durante 18 h en LB líquido suplementado con kanamicina, mientras que las células *E. coli* BL21-Gold (DE3) que se transformaron con construcciones basadas en T7 se cultivadas de manera similar, con adición de IPTG cuando los cultivos alcanzaron una densidad celular de 0,6 DO. Después de que los cultivos se incubaron, se centrifugaron 200 µl de cultivo a 13,000 rpm durante 5 min y la pastilla celular se resuspendió y se hirvió en 20 µl de buffer de carga desnaturante durante 5 min. Para el análisis del extracto soluble, al resto del cultivo se le adicionaron 250 µL de buffer de lisis B-PER Thermo Scientific. El lisado se centrifugó a 13,000 rpm durante 5 min, del sobrenadante se tomaron 10 µL y se agregó buffer de carga desnaturante y se hirvió durante 5 min. Para el análisis del extracto insoluble la pastilla generada del lisado fue resuspendida en 250 µL de la solución Inclusion Body Solubilization Thermo Scientific, a 10 µL de este resuspendido se agregó buffer de carga y se hirvió por 5 min. Todos los extractos fueron analizados en geles duplicados mediante SDS-PAGE en condiciones desnaturantes.

Un gel fue analizado mediante tinción con azul de Coomassie, mientras que el otro se analizó por Western-blot, el gel se desmontó del cassette y fue colocado sobre una membrana de nitrocelulosa sobre cuatro papeles filtro (previamente remojados en buffer pH 8.2 compuesto de Tris 25 mM, Gly 192 mM, SDS 0.1%, metanol 20%), cuatro papeles filtro más sobre el gel. La transferencia se realizó en una cámara semi húmeda marca Hoefer, durante 1 h a 55 V. Posteriormente en buffer PBS 1X se agregó el conjugado anti-GFP con fosfatasa alcalina para la detección específica de las bandas de GFP, después de una 1 h aproximadamente se revelaron mediante la adición de 2 mL del sistema de sustrato líquido BCIP®/NBT-Blue Liquid Substrate System for Membranes SIGMA Aldrich.

7.1.10 Análisis de crecimiento y fluorescencia

El análisis de fluorescencia se realizó utilizando las cepas de *E. coli* transformadas con las construcciones moleculares obtenidas. Se crecieron pre-inoculos durante 18 h de crecimiento a 37 °C en medio LB líquido suplementado con kanamicina. Al día siguiente cada pre-inoculo se diluyó 1: 100 en medio de LB líquido fresco. Estos cultivos se hicieron crecer hasta 0.5 de DO a 600 nm y se enfriaron inmediatamente sobre hielo, de cada cultivo se obtuvo la pastilla celular centrifugando a 6,000 rpm durante 18 min, esta fue lavada dos veces con 10 mL de solución de sulfato de magnesio 10 mM fría. Al final del segundo lavado la pastilla celular se resuspendió en 3 mL de solución de sulfato de magnesio 10 mM fría. Para cada muestra resuspendida se midió la DO final a 600 nm. A continuación se preparó una micro placa de 96 pozos Corning® negra con tapa,

fondo plano transparente y capacidad volumétrica de 350 μ L, a cada pozo se agregó el volumen necesario de LB líquido, IPTG a una concentración final de 0.1 mM de ser necesario y un volumen de células resuspendidas para obtener una concentración final de 0.05 DO en un volumen final de ensayo de 200 μ l.

El ensayo en micro-placa fue monitoreado utilizando un lector de placas Synergy 2 Multi-Mode Reader de Bio Tek. Se registraron niveles de fluorescencia fijando la excitación a 474 nm y registrando la emisión a 508 nm. La estimación indirecta de crecimiento se monitoreo en paralelo midiendo la absorbancia a 600 nm. El ensayo se monitoreo durante 24 h a una temperatura de 37 °C, realizando lecturas cada 25 min, los datos se recopilaron con el software Gen5™ Reader Control and Data Analysis Software.

7.1.11 Análisis de estructura secundaria en el mRNA

La predicción y el análisis de la estructura secundaria se realizaron con el servidor RNA fold Server (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>) Instituto de Química Teórica, Universidad de Viena, incluido en los servicios de ViennaRNA Web Services (Gruber et al. 2008). Los análisis se realizaron utilizando los parámetros de valor predeterminado. El análisis de la estructura secundaria consideró las secuencias líder de los promotores T7 o Trc y 60 pb dentro del ORF de las diferentes construcciones.

7.1.12 Oligonucleótidos utilizados

Tabla 4. Oligonucleótidos utilizados para la obtención de las variantes del gen EGFP

Construcción	Templado	Oligonucleótidos utilizados	fragmento generado	Vector y sitio de clonación
T7-EGFP	pEGFP-N3	Pr1_Fw: aacag <u>cccatg</u> gtgagcaagggcgaggagctg Pr2_Rv: ctctctc <u>cgag</u> ttactgtacagctcgtccatgccgag	740 bp	pET28a (NcoI/XhoI)
T7-EGFP Δ G4	pEGFP-N3	Pr3_Fw: aacagcc <u>catg</u> gtgagcaagggagagctgttcaccggg Pr4_Rv: ctctctc <u>cgag</u> ttactgtacagctcgtccatgccgag	737 bp	pET28a (NcoI/XhoI)
Trc-EGFP	pEGFP-N3	Pr5_Fw: cacaggaacagc <u>catatg</u> gtgagcaagggcgaggagctg Pr6_Rv: ctctctc <u>cgag</u> ttactgtacagctcgtccatgccgag	747 bp	pJOQ (NdeI/XhoI)
Trc-EGFP Δ G4	pEGFP-N3	Pr7_Fw: cacaggaacagc <u>catatg</u> gtgagcaagggagagctgttcaccggg Pr8_Rv: ctctctc <u>cgag</u> ttactgtacagctcgtccatgccgag	744 bp	pJOQ (NdeI/XhoI)
Trc-rsGFP	pQBI25-f	Pr9_Fw: cacaggaacagc <u>catatg</u> gttagcaagggagagaagaactc Pr10_Rv: cgggctc <u>cgag</u> ttatcagttgtacagttcatccatgccatg	750 bp	pJOQ (NdeI/XhoI)
Trc-rsGFP Δ G4	pQBI25-f	Pr11_Fw: cacaggaacagc <u>catatg</u> gttagcaagaagaactctcactggagttgtc Pr12_Rv: cgggctc <u>cgag</u> ttatcagttgtacagttcatccatgccatg	747 bp	pJOQ (NdeI/XhoI)
Trc-rsEGFP	pQBI25-f	Pr13_Fw: ccttccgctg <u>catatg</u> gtgagcaagggagagaagaactcttc Pr14_Rv: gcattgaacacataggtcagagtagtgactag Pr15_Fw: ctagtactactctgacctatggtgtcaatgc Pr16_Rv: ttcaatggtggcgaatcttgaagttcacttt Pr17_Fw: aaagtgaactcaagattgccacacacattgaa Pr18_Rv: cgggctc <u>cgag</u> ttatcattgtacagttcatccatgccagtgtaatccagcagctgttac	(750 bp)	pJOQ (NdeI/XhoI)
Trc-rsEGFP Δ G4	Trc-rsEGFP	Pr19_Fw: ctcaggacacagc <u>catatg</u> gtgagcaagaagaagaactctcactggagttgtc Pr20_Rv: cgggctc <u>cgag</u> ttatcattgtacagttcatccatgccagtgtaatccagcagctgttac	747 bp	pJOQ (NdeI/XhoI)

Se utiliza diferente código para los siguientes oligonucleótidos, aunque son los mismos: Pr2_Rv = Pr4_Rv = Pr6_Rv = Pr8_Rv; Pr10_Rv = Pr12_Rv; Pr18_Rv = Pr20_Rv. Todos los oligonucleótidos enlistados se obtuvieron de la Unidad de Síntesis y Secuenciación de DNA (USSDNA) del Instituto de Biotecnología UNAM campus Morelos. Los nucleótidos subrayados indican los sitios de restricción, NcoI, NdeI y XhoI, según correspondan.

8. Resultados

8.1 Obtención de las variantes moleculares del gen *EGFP*

El reporte de Arpino y colaboradores en 2014 produjo una variante molecular de la proteína EGFP con mayor fluorescencia, la construcción molecular se reprodujo en nuestro laboratorio como a continuación se describe.

Se utilizó el vector pET-28a de Novagen[®] como vehículo de expresión para la variante EGFP conteniendo el promotor T7, utilizando los sitios de corte *Nco*I y *Xho*I fueron clonados los genes *EGFP* y *EGFP*ΔG4 tal y como es reportado en el artículo de Arpino y colaboradores (Arpino et al . 2014). Incluimos la construcción molecular de las versiones reportadas por Arpino y colaboradores en el vector pJOQ, lo que permitiría molecularmente identificar variaciones de expresión bajo otro sistema de promoción como el promotor Trc. Adicionalmente, para analizar el efecto que produce la remoción del codón que codifica para la glicina 4 en otras variantes del gen *EGFP* se incluyeron las construcciones de los genes *rsGFP*, *rsGFP*ΔG4 y *rsEGFP*, *rsEGFP*ΔG4 clonadas en el vector pJOQ.

El gen *EGFP* fue amplificado mediante PCR utilizando como templado el vector pEGFP-N3 y los oligonucleótidos Pr1_Fw y Pr2_Rv (sección Materiales y Métodos). El producto de PCR incluye los sitios de restricción *Nco*I y *Xho*I utilizados para su clonación sitio dirigida en pET-28a (Fig. 8). Se realizó la ligación correspondiente y una alícuota de ésta fue utilizada para transformar células electrocompetentes de *E. coli* BL21-Gold (DE3). La construcción resultante fue nombrada como T7-EGFP.

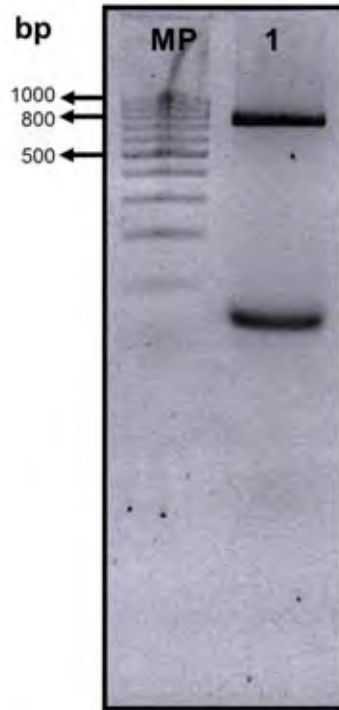


Figura 8. Amplificación del gen *EGFP*. La figura muestra la electroforesis en gel de agarosa al 1%. 1: Producto de PCR para la amplificación del gen *EGFP* utilizando como templado el plásmido pEGFP-N3. MP: GeneRuler 100 bp DNA Ladder, Thermo Scientific.

Para la amplificación por PCR del gen *EGFP Δ G4* se utilizaron los oligonucleótidos Pr3_Fw y Pr4_Rv y el templado pEGFP-N3 (sección Materiales y Métodos). En este caso, el oligonucleótido Pr3_Fw carece de 1 codón ggc que codifica para el aminoácido glicina en la posición cuatro (Gly4) en el gen *gfp* original. Se realizó el mismo procedimiento de clonación sitio dirigido *NcoI/XhoI* como en T7-EGFP y a esta construcción se le asignó el nombre de T7-EGFP Δ G4 (Fig. 9).

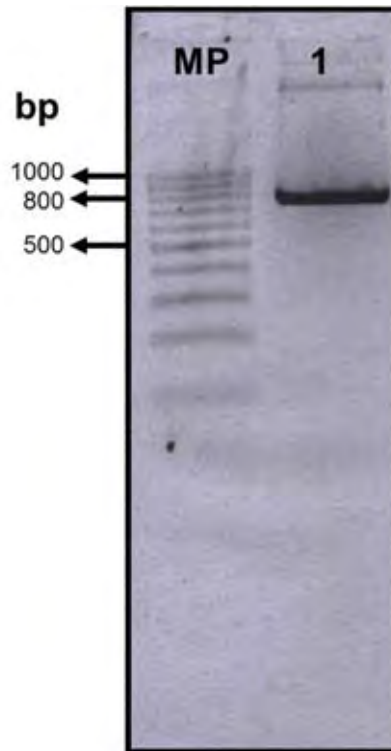


Figura 9. Amplificación del gen *EGFP Δ G4*. La figura muestra la electroforesis en gel de agarosa al 1%. 1: Producto de PCR para la amplificación del gen *EGFP Δ G4* utilizando como templado el plásmido pEGFP-N3. MP: GeneRuler 100 bp DNA Ladder, Thermo Scientific.

Los genes *EGFP*, *EGFP Δ G4*, *rsGFP*, *rsGFP Δ G4* y *rsEGFP*, *rsEGFP Δ G4*, se amplificaron y clonaron en el plásmido pJOQ para estar bajo la regulación del promotor Trc.

El gen *EGFP* fue amplificado mediante PCR utilizando como templado el vector pEGFP-N3 y los oligonucleótidos Pr5_Fw y Pr5_Rv (sección Materiales y Métodos). El producto de PCR incluye los sitios de restricción *Nde*I y *Xho*I utilizados para su clonación sitio dirigida en pJOQ (Fig. 10). Se realizó la ligación correspondiente y una alícuota de esta fue utilizada para transformar células electrocompetentes de *E. coli* MC1061. Esta construcción fue nombrada Tr-

EGFP. El mismo procedimiento fue realizado para la obtención del gen *EGFPΔG4*, utilizando como templado pEGFP-N3 y los oligonucleótidos Pr7_Fw y Pr8_Rv. El oligonucleótido Pr7_Fw no contiene el codón que codifica para la Gly4. Se llevó a cabo la restricción enzimática y la ligación necesaria, una alícuota de esta fue utilizada para transformar células electrocompetentes de *E. coli* MC1061. La construcción obtenida fue nombrada como Trc-*EGFPΔG4*.

Para la obtención de rs *GFP* se realizó una amplificación mediante PCR utilizando como templado el plásmido pQB125-f y los oligonucleótidos Pr9_Fw y Pr10_Rv (descritos en la sección Materiales y Métodos). El producto de 750 pb fue clonado sitio dirigido en el plásmido pJOQ utilizando las enzimas *NdeI* y *XhoI*. Una alícuota de esta ligación fue utilizada para transformar células electrocompetentes de *E. coli* MC1061, esta construcción se nombró como Trc-rs*GFP*. Para la obtención de la construcción Trc-rs*GFPΔG4* se utilizó el plásmido pQB125-f como templado del gen y los oligonucleótidos Pr11_Fw y Pr12_Rv, Pr11_Fw no contiene el codón gga que codifica para la Gly4. El producto fue digerido con las enzimas *NdeI* y *XhoI* para su ligación sitio dirigida en pJOQ. En la figura 10 se muestra el producto de cada una de las construcciones anteriores.

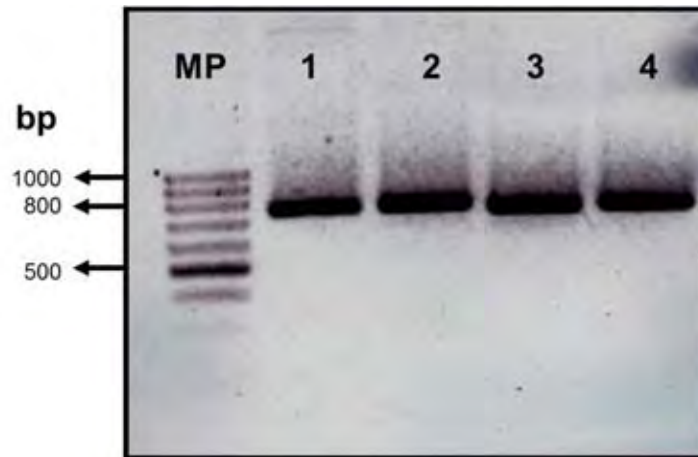


Figura 10. Amplificación de los genes *EGFP*, *EGFP* Δ *G4*, *rsGFP* y *rsGFP* Δ *G4* para su clonación en pJOQ. Electroforesis en gel de agarosa al 1%. Productos de PCR de los genes 1: *EGFP*, 2: *EGFP* Δ *G4*, 3: *rsGFP* y 4: *rsGFP* Δ *G4*. MP: GeneRuler 100 bp DNA Ladder, Thermo Scientific.

Se generó un juego adicional de genes derivados del gen *rsGFP*, pero esta vez se consideró la adición de algunos codones que permiten generar una versión parecida a la proteína EGFP, ampliamente utilizada en *E. coli* y que es parecida a la versión *avGFP*. A esta nueva versión se le llamó *rsEGFP* o versión “*ecolizada*” de *rsGFP*, que contempla el uso preferencial de codones de *E. coli* que a su vez permite contar con un control de expresión. Para la obtención de estas versiones se realizaron mutagénesis sitio dirigidas mediante PCR empalmados, en principio se amplifica el ORF del gen en tres etapas: un primer PCR1 que incluye los primeros 500 pb aproximadamente del gen. Una segunda PCR2 que amplifica 250 pb de región 3’ del gen. Ambos productos son purificados y utilizados como templado en una mezcla de 20 ng de cada uno y un set adicional de oligonucleótidos con los que se realiza la tercera amplificación consistente en un producto de 850 pb aproximadamente que forman el nuevo gen “*ecolizado*”. El

gen sintetizado es digerido con las enzimas *NdeI* y *XhoI* y ligado en el plásmido pJOQ. El juego de oligonucleótidos utilizados se describe en la sección Materiales y Métodos (Tabla 4). Una alícuota de esta ligación fue utilizada para transformar *E. coli* MC1061. La construcción resultante fue designada como Trc-*rsEGFP*, figura 11. Para la obtención de la construcción Trc-*rsEGFP* Δ G4 se utilizó como templado el plásmido Trc-*rsEGFP* y los oligonucleótidos Pr19_Fw, este no contiene el codón para la Gly4, y Pr20_Rv. El producto obtenido contiene los sitios de corte para las enzimas *NdeI* y *XhoI*, mismos que fueron utilizados para su ligación en el plásmido pJOQ.

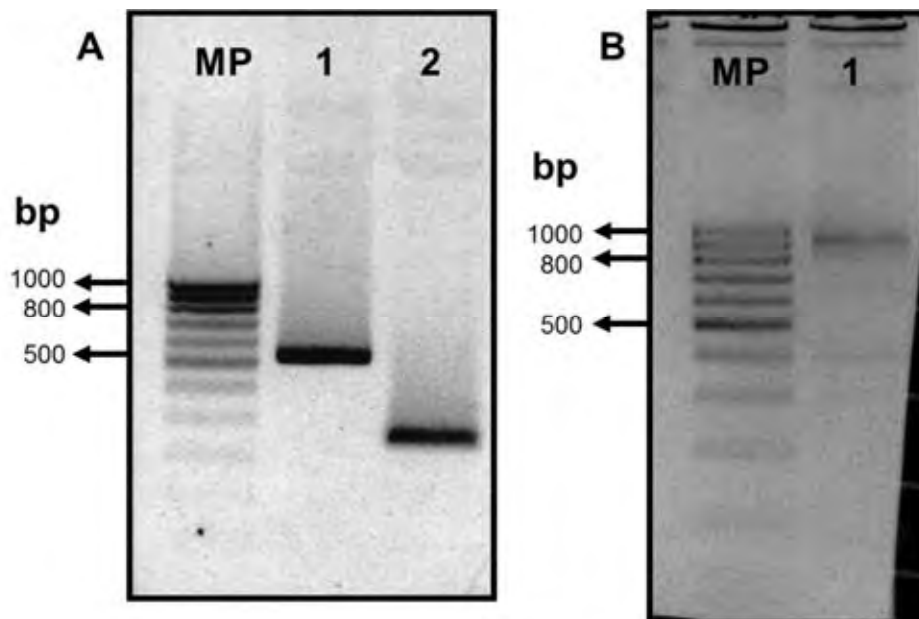


Figura 11. Amplificación del gen *rsEGFP*. Electroforesis en gel de agarosa al 1%. Panel A: productos de PCR para 1: PCR 1, 2: PCR2. Panel B: producto de PCR para la amplificación completa del gen *rsEGFP*, este producto es el producto final del empalme de PCR1 y PCR2. Para ambos paneles MP: GeneRuler 100 bp DNA Ladder, Thermo Scientific.

8.2 Análisis de transformantes y su fluorescencia

Cada construcción obtenida fue verificada visualmente irradiándola con luz azul (460 nm de emisión LED). Las clonas obtenidos fueron seleccionadas considerándolas positivas sí emitían el color verde, o negativas al no presentar coloración. Se realizaron estriados separados de cada una de las transformantes obtenidas de manera aislada (Fig. 12).

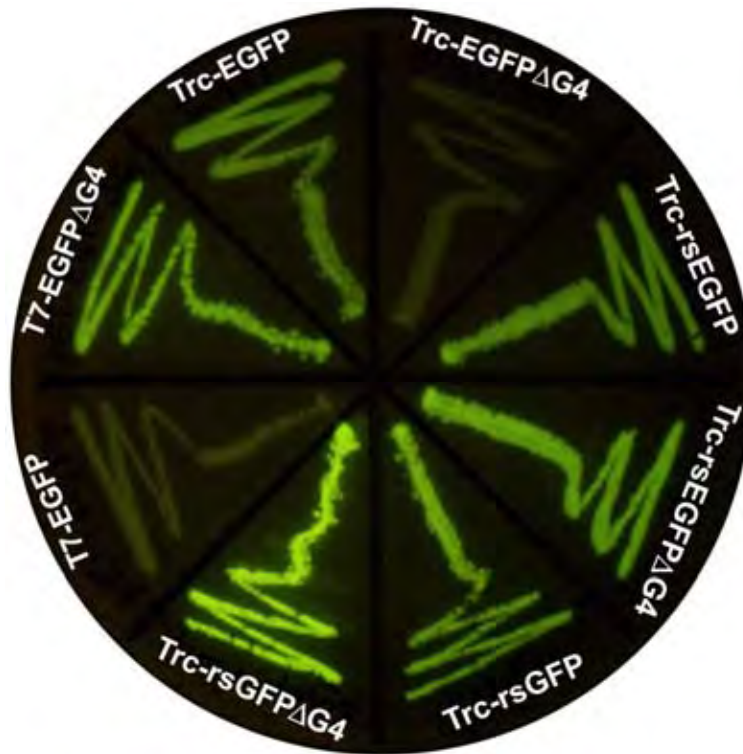


Figura 12. Estriados en LB solido de las ocho variedades de GFP obtenidas. Caja LB solido con kanamicina y 0.1 mM de IPTG como inductor. Se incubó durante 20 h a 37 °C.

Arpino y colaboradores mencionan en su reporte que la proteína EGFP Δ G4 presenta mayores niveles de fluorescencia que su versión parental

EGFP, presumiblemente por una mejora en el plegamiento de la proteína al ser removido el aminoácido Gly4. No obstante, las fluorescencias de las diferentes cepas construidas en este estudio (Fig. 12), muestran que este efecto es dependiente del tipo de construcción analizada. Así por ejemplo, nuestros resultados con las construcciones EGFP y EGFP Δ G4 son similares a los obtenidos por Arpino y colaboradores, cuando sus correspondientes genes son transcritos a partir del promotor T7, pero son completamente opuestos (menor fluorescencia con la construcción EGFP Δ G4), cuando el promotor utilizado es Trc. Nuestras variantes rsGFP y rsEGFP bajo el promotor Trc presentan una mayor fluorescencia al ser removido el residuo Gly4, tal y como se observa en el estudio (Fig. 12).

8.3 Análisis de proteína por Western-blot

Se verificó la producción de proteína para cada una de las construcciones mediante un inmunoblot usando un conjugado-anti-GFP-fosfatasa alcalina para la detección específica de GFP, en la sección Materiales y Métodos 7.1.9 se detalla el procedimiento. Este análisis se realizó para determinar si la fluorescencia observada se relaciona directamente con la producción de proteína, ya que anteriormente se ha reportado que cambios nucleotídicos afectan la eficiencia con la que se traduce el gen reportero analizado (Schulz & Reznikoff 1990; Berg et al. 2012).

Analizamos las construcciones *Trc-EGFP*, *Trc-EGFP Δ G4* que presentan un resultado opuesto al reportado por Arpino y colaboradores. También incluimos las construcciones *Trc-rsEGFP* y *Trc-rsEGFP Δ G4*, visualmente hemos observado que estas presentan mayor fluorescencia. El Western-blot nos permitió identificar que el aumento en la fluorescencia de estas variantes está relacionado con la producción de proteína. Dos geles SDS-PAGE bajo condiciones desnaturizantes fueron realizados, tal y como se describe en la sección Materiales y Métodos.

El primero de ellos se analizó por tinción con Azul de Coomassie y su par fue utilizado para el ensayo de inmunodetección (Western-blot). En la figura 13 se muestra el Western-blot con las construcciones analizadas *Trc-EGFP*, *Trc-EGFP Δ G4* y *Trc-rsEGFP*, *Trc-rsEGFP Δ G4*, cada una en una fracción proteínica: extracto total, fracción soluble y fracción insoluble.

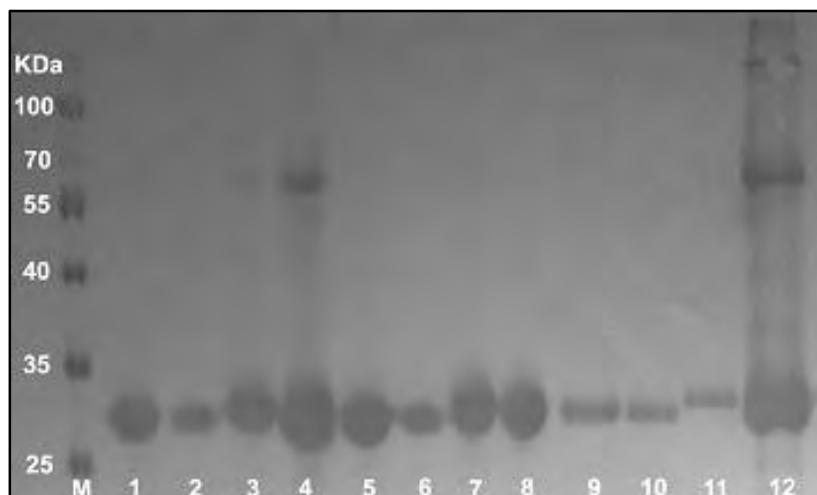


Figura 13. Western-blot para la detección de proteína en las variantes Trc-EGFP y Trc-rsEGFP. M: marcador para proteína. 1, 5 y 9: Trc-EGFP. 2, 6 y 10: Trc-EGFP Δ G4. 3, 7 y 11: Trc-rsEGFP. 4, 8 y 12: Trc-rsEGFP Δ G4. Carriles 1-4 extracto total. Carriles 5-8: fracción soluble. Carriles 9-12: fracción insoluble.

El resultado de Western-blot de este primer análisis determinó que la baja fluorescencia en la construcción Trc-EGFP Δ G4 está relacionada con la cantidad total de proteína producida, como se puede observar, esta no se encuentra en los mismos niveles de su par Trc-EGFP, ya que Trc-EGFP Δ G4 reduce la producción de proteína en promedio un 40% según análisis de densitometría, (Tabla 5). Para la construcción Trc-rsEGFP Δ G4 la remoción del codón que codifica para la Gly4 incrementa en promedio un 0.26 veces más la producción de proteína con respecto a Trc-rsEGFP. Los datos del análisis de densitometría se muestran en la Tabla 5, estos se obtuvieron utilizando el software ImageJ versión 1.50i (Rasband 2016).

Tabla 5. Análisis de densitometría del estudio Western-blot figura 13. Los porcentajes obtenidos fueron evaluados respecto al área analizada

	Variante molecular			
	Trc-EGFP	Trc-EGFPΔG4	Trc-rsEGFP	Trc-rsEGFPΔG4
Extracto total	100%	55%	100%	137%
Extracto soluble	100%	60%	100%	100%
Extracto insoluble	100%	65%	100%	141%

*El valor es calculado con el software imageJ, el porcentaje es con respecto al 100% de la variante molecular sin la delección.

El resultado del Western-blot con todas las construcciones se muestra en la figura 14, en esta se observan el par de geles de acrilamida, uno de estos teñido con Azul de Coomassie y el otro utilizado para el análisis por Western-blot.

Con el software imageJ se analizaron los productos identificados por Western-blot correspondientes a las proteínas de las ocho variantes moleculares. Para cada una de las variantes se observó que la remoción de Gly4 afecta la producción de proteína total observada, para la variante *Trc-EGFP Δ G4* se observa una disminución de hasta un 60% con respecto a la variante *Trc-EGFP*. Mientras que en el caso de la variante *T7-EGFP Δ G4*, esta incrementa un 1.77 veces más con respecto a *T7-EGFP*. Para las variantes *Trc-rsGFP*, *Trc-rsEGFP* y las variantes con la delección de Gly4, *Trc-rsGFP Δ G4* y *Trc-rsEGFP Δ G4*, se observa que el incremento ronda aproximadamente en un 0.1 veces más, de acuerdo a los datos de densitometría obtenidos (Tabla 6).

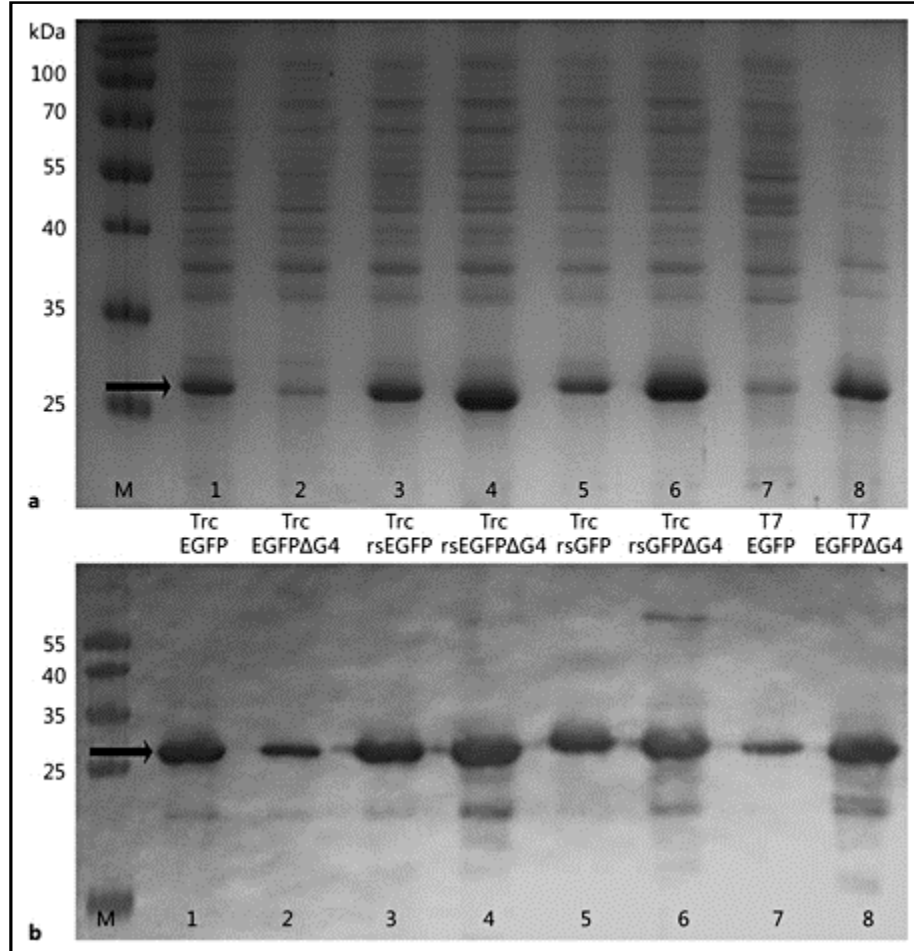


Figura 14. Análisis de extracto total de proteína en gel de acrilamida desnaturalizante SDS-PAGE. a) Gel teñido con azul de Coomassie, extracto total de las ocho variantes moleculares. **b)** Detección de proteínas GFP por Western-blot. La flecha indica las bandas correspondientes a GFP. El nombre de cada variante indicado en medio de los dos geles. La flecha negra indica la banda correspondiente a la proteína GFP en las ocho variantes. M: marcador PageRuler™ Prestained Protein Ladder.

Tabla 6. Análisis de densitometría Western-blot figura 14. Los porcentajes obtenidos fueron evaluados respecto al área analizada

	Variante molecular							
	Trc-EGFP	Trc-EGFPΔG4	Trc-rsEGFP	Trc-rsEGFPΔG4	Trc-rsGFP	Trc-rsGFPΔG4	T7-EGFP	T7-E-GFPΔG4
Porcentaje calculado*	100%	42%	100%	109%	100%	111%	100%	277%

*El valor es calculado con el software imageJ, el porcentaje es con respecto al 100% de la variante molecular sin la delección.

Los resultados obtenidos por Western-blot en las ocho variantes de GFP denotan que el incremento o disminución de fluorescencia producido por la remoción del aminoácido Gly4 es consecuencia de un efecto sobre la traducción de GFP, al menos para la variante reportada por Arpino y colaboradores en 2014, en donde es evidente un cambio dramático en la producción de proteína observada. Para las variantes Trc-rsEGFP Δ G4 y Trc-rsGFP Δ G4 la remoción de la Gly4 si refleja un efecto de mejora, se puede observar un incremento en el producto traducido (Fig. 14).

8.4 Análisis de las construcciones en una cinética de crecimiento y fluorescencia

Con el objetivo de verificar que el incremento de fluorescencia observado en algunas de las construcciones con variantes de la proteína GFP con la delección Gly4 no fuera el resultado indirecto de un mayor crecimiento celular que directamente afectaran la producción final de las proteínas, se realizó una cinética de crecimiento con su correspondiente análisis de fluorescencia. Para cada construcción se consideró un ensayo con 12 réplicas que fueron monitoreadas cada 25 min durante 24 h (Materiales y Métodos sección 7.1.10). El análisis se llevó a cabo en un lector de placas multi-modal BioTek, en el que se midió crecimiento por absorbancia a 600 nm y se monitorio la fluorescencia fijando a 474 nm la excitación y recolectando la emisión a 508 nm.

Los datos se colectaron utilizando el software Gene5™ reader Control and Data analysis Software. La figura 15 muestra los resultados para el monitoreo de crecimiento en las ocho construcciones durante 24 h.

Se calcularon las velocidades de crecimiento (μ), los valores de duplicación para cada variante molecular, así como su crecimiento total alcanzado (Tabla 7).

Las velocidades de crecimiento (μ) no presentaron un cambio significativo. Los tiempos de duplicación son muy homogéneos, incluso entre variantes moleculares (con y sin la remoción de Gly4) como se aprecia en la tabla 7, los crecimientos máximos alcanzados en 24 h para las ocho variantes fueron en promedio de 1.12 D.O.

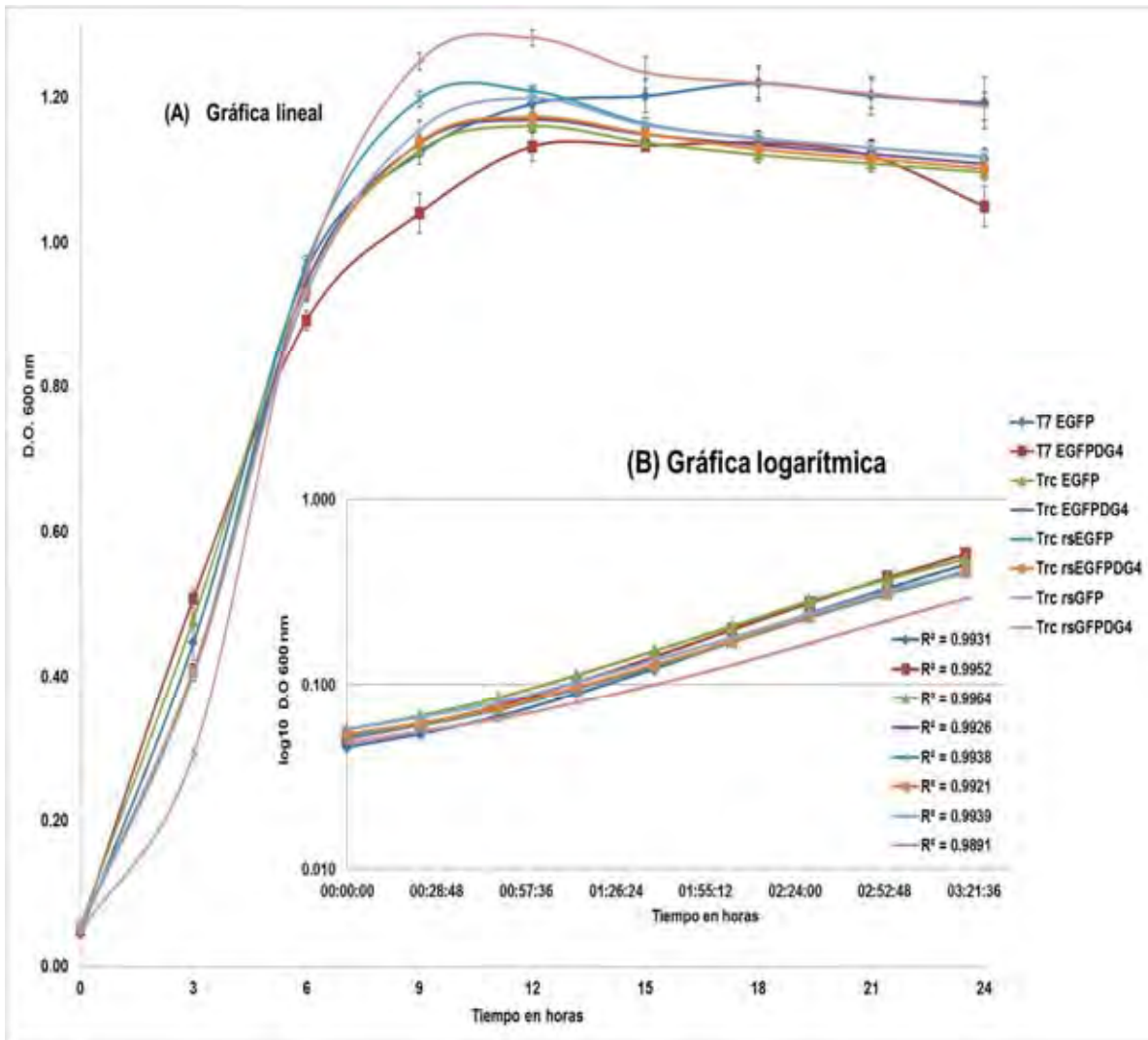


Figura 15. Cinética de crecimiento para las cepas de *E. coli* BL21-Gold (DE3) y MC1061 transformadas con cada una de nuestras ocho construcciones de GFP. (A) Gráfica en valores lineales, monitoreo durante 24 h, (B) gráfica en escala logarítmica base 10. Se muestra el comportamiento de crecimiento de 0 a 3 h. Se incluyen el valor de distribución lineal (R^2). Las versiones moleculares con las remociones de Gly4 se indican con la letra “D” sustituyendo el símbolo Δ .

Tabla 7. Análisis de las variables de crecimiento en las ocho variantes moleculares analizadas

	T7-EGFP	T7-EGFPΔG4	Trc-EGFP	Trc-EGFPΔG4	Trc-rsEGFP	Trc-rsEGFPΔG4	Trc-rsGFP	Trc-rsGFPΔG4
μ^*	0.33	0.33	0.31	0.30	0.30	0.29	0.28	0.26
G*	0.9 h	0.9 h	0.8 h	0.8 h	0.8 h	0.8 h	0.7 h	0.7 h
Crecimiento en 24 h D.O. a 600 nm	1.19	1.05	1.10	1.11	1.12	1.10	1.12	1.19

G*: tiempo duplicación de la región exponencial en h. μ^* valor de la pendiente en etapa exponencial (D.O/h⁻¹).

El tiempo de duplicación de acuerdo al estándar para *E. coli* que es de aproximadamente 21 min, se incrementó en unos 20 a 30 min para las *E. coli* que contienen nuestras variantes moleculares, debido probablemente a las condiciones de crecimiento en las que se realizó la cinética (Sección Materiales y Métodos 7.1.10). Aun con este cambio se obtuvo un crecimiento bastante optimo en 24 h de monitoreo, aun en las variantes que producen un alto nivel de proteína como: T7-EGFP Δ G4, Trc-rsEGFP Δ G4 y Trc-rsGFP Δ G4, o en las que no, como: T7-EGFP y Trc-EGFP Δ G4, de acuerdo a lo analizado en el Western-blot (Fig. 14).

En la figura 16 se muestra el resultado del monitoreo de la fluorescencia (sección Materiales y Métodos 7.1.10), en cuatro construcciones sin la remoción del codón que codifica para el aminoácido Gly4. Se observó que las construcciones con EGFP presentan niveles similares de fluorescencia, para T7-EGFP, 2222 RFUs y para Trc-EGFP, 2464 RFUs, en 24 h respectivamente. Las variantes Trc-rsGFP y Trc-rsEGFP sin remoción de Gly4 presentan una fluorescencia mayor, 3648 RFUs para Trc-rsGFP y 3399 RFUs para Trc-rsEGFP en 24 h de monitoreo. Estas versiones derivan de pQBI25-f (Quantum Biotechnologies), que es una versión más fluorescente que su contraparte EGFP.

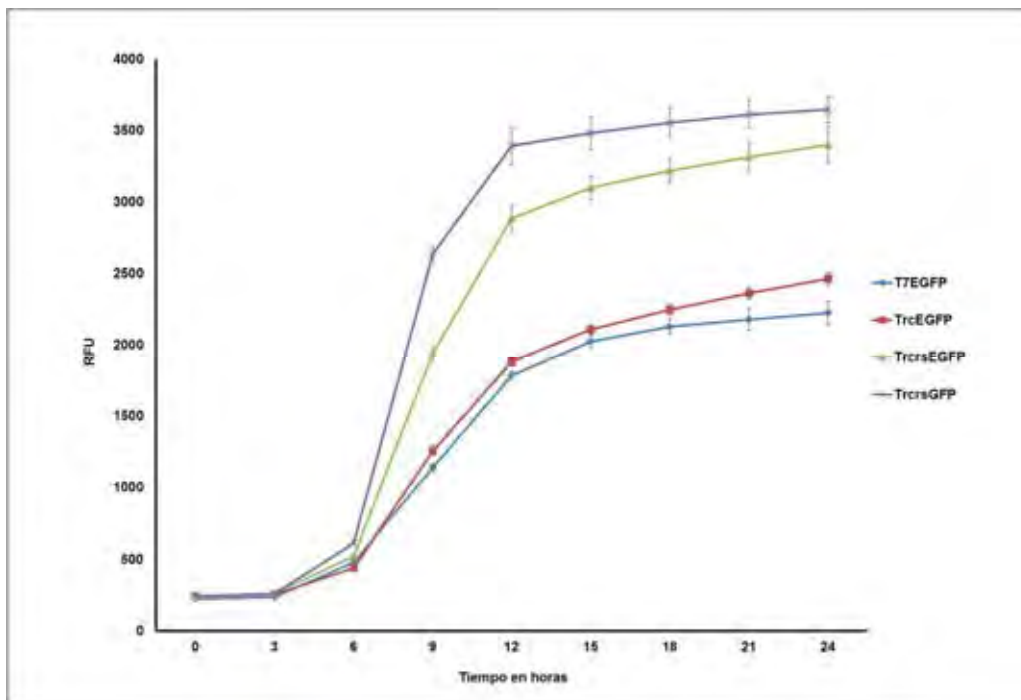


Figura 16. Cinética fluorescencia de las cepas de *E. coli* BL21-Gold (DE3) y MC1061 transformadas con T7-EGFP, Trc-EGFP, Trc-rsEGFP y Trc-rsGFP. Se muestran los datos de unidades relativas de fluorescencia, y el monitoreo en un periodo de tiempo de 24 h. Los datos se colectaron de manera simultánea a los datos de crecimiento.

En la figura 17 se muestran los datos de la cinética de fluorescencia de las construcciones con la remoción del codón que codifica para la Gly4. El efecto que se observa sobre la fluorescencia de la construcción Trc-EGFP Δ G4 es opuesto a lo que se observa en T7-EGFP Δ G4 y que como se muestra en el inmunoblot existe una reducción de hasta 58% en la detección de la proteína para Trc-EGFP Δ G4 (Tabla 6, Fig. 14). Cabe señalar que para las construcciones Trc-rsGFP Δ G4 y Trc-rsEGFP Δ G4, la remoción de la Gly4 confiere un mayor nivel de fluorescencia, cerca del 0.1 veces más que el detectado con las construcciones silvestres (Fig. 14, Tabla 6). Los incrementos en la fluorescencia son consistentes con la cantidad de proteína detectada en los ensayos de inmunodetección para las cepas correspondientes (Fig. 13, 14 y Tabla 5 y 6). En la figura 3 del artículo que se anexa en la página 79 de esta tesis, se muestran las gráficas en conjunto de los valores normalizados de fluorescencia de las ocho diferentes construcciones de nuestro estudio,

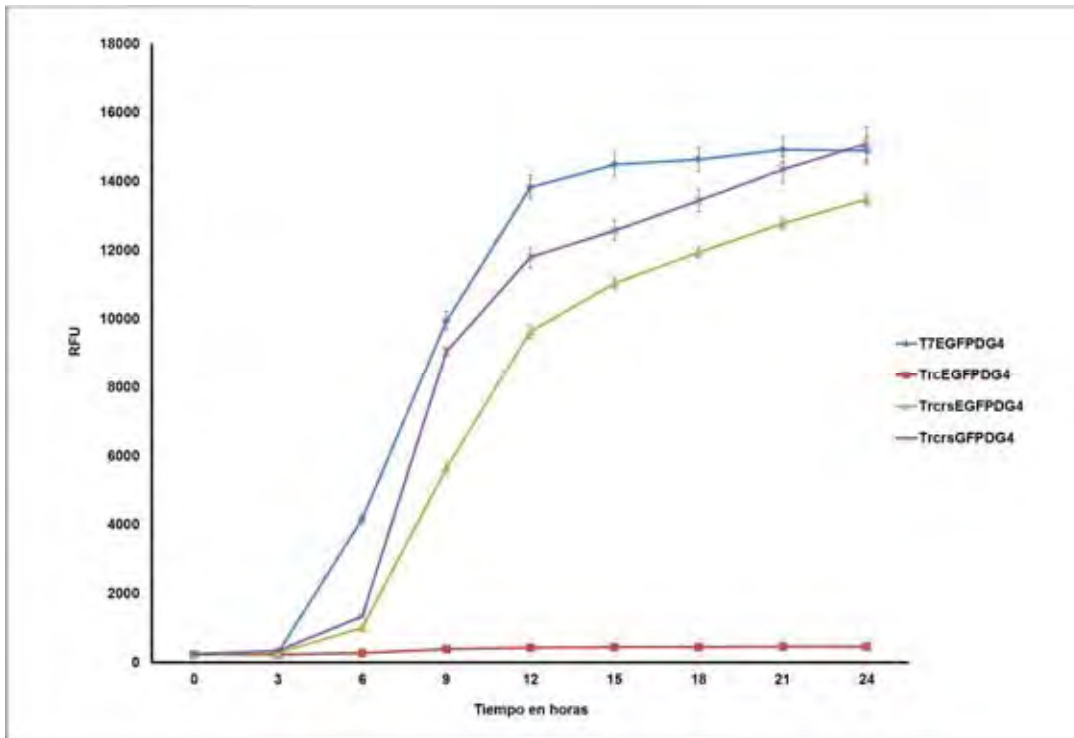


Figura 17. Cinética fluorescencia de las cepas de *E. coli* BL21-Gold (DE3) y MC1061 transformadas con cada una de nuestras construcciones de GFP con la remoción del codón Gly4. Se muestran los datos de unidades relativas de fluorescencia, y el monitoreo en un periodo de tiempo de 24 h. Los datos se colectaron de manera simultánea a los datos de crecimiento. Las versiones moleculares con las remociones de Gly4 se indican con la letra “D” sustituyendo el símbolo Δ .

8.5 Análisis del efecto producido por la remoción del codón que codifica para la Gly4 en GFP y el estructuramiento del mRNA de sus correspondientes variantes

Los resultados de nuestros análisis de fluorescencia y proteína detectada por Western-blot no son consistentes a lo propuesto (Arpino et al. 2014), referente al efecto de la remoción de Gly4 sobre un mejor plegamiento de la proteína analizada. ¿Qué otra variable no se tomó en cuenta y pudiera estar influenciando significativamente la emisión de fluorescencia? Para poder contestar esta pregunta, consideramos los resultados de las investigaciones de Kudla y colaboradores (Kudla et al. 2009) así como los de Gu y colaboradores (Gu et al. 2010). Ambos estudios señalan el efecto que pudieran tener las estructuras secundarias de RNA, estas dependientes de la secuencia nucleotídica de los primeros codones del gen en cuestión. Este efecto negativo de regiones estructuradas de RNA fue observado por Kudla y colaboradores en 2009, al encontrar que de la librería en donde se incluían todos los codones sinónimos para codificar al gen GFP, las clonas con menor actividad correspondían a codones en la región amino-terminal que generaban estructuras secundarias estables. Los resultados de análisis bioinformáticos realizados por Gu et al., (Gu et al. 2010) apuntan en el mismo sentido sobre una tendencia evidente de los genes de una gran mayoría de organismos de los tres reinos de la vida (bacteria, arquea o eucaria) para evitar estructuras secundarias de RNA que puedan afectar la eficiencia del inicio de la traducción (Fig. 2).

La idea de que una estructura secundaria, favorecida por la remoción del codón Gly4, afectara los niveles de traducibilidad de los mRNA de nuestras construcciones, se relaciona mejor con los datos observados experimentalmente y no por un cambio en el plegamiento como lo propone Arpino y colaboradores para su variante EGFP Δ G4 (Arpino et al. 2014).

Para elucidar el papel que juega la remoción de la Gly4 en nuestras variantes de GFP, realizamos los siguientes análisis:

a) Se analizaron las regiones nucleotídicas de los promotores utilizados, T7 y Trc (Fig. 18) para determinar su influencia en la formación de alguna estructura secundaria, de acuerdo a la longitud que presenta cada promotor desde el inicio de su transcripción al nucleótido +1 (ATG) en Trc y T7. Posteriormente estas secuencias fueron fusionadas a los primeros 60 nucleótidos de la región codificante de cada versión molecular de GFP, lo que permitió analizar la región 5' no traducida de cada mRNA.

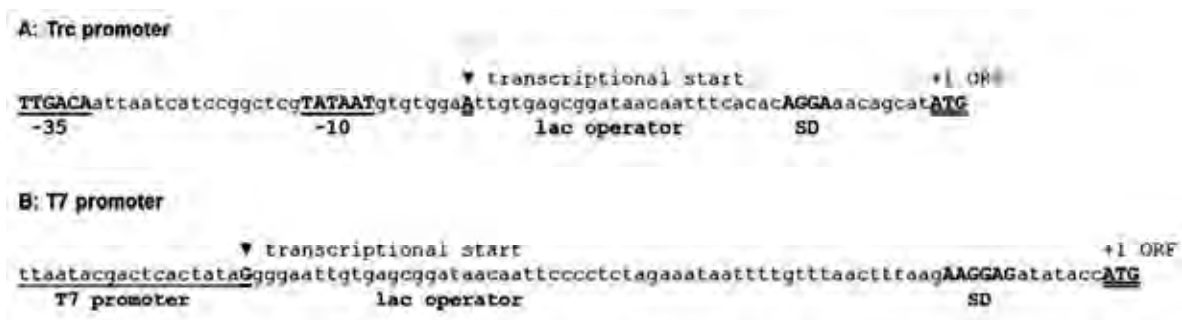


Figura 18. Esquema de las secuencias nucleotídicas de los promotores Trc y T7. El símbolo ▼ indica los inicios de transcripción para cada promotor. La secuencia Shine-Dalgarno está indicada como SD. El codón de inicio ATG está indicado como +1.

b) Se compilaron reportes literarios en los que se haya analizado la estructura secundaria del mRNA, de los cuales obtuvimos información que ayudó a seleccionar la región de este análisis. Se tomaron en cuenta los análisis de (Kudla et al. 2009; Gu et al. 2010; Katz & Burge 2003) para determinar la longitud de nuestra secuencia de análisis y comprendió desde el inicio de la transcripción hasta 60 nucleótidos río abajo del codon ATG. Es en esta región del mRNA en donde una estructura secundaria puede afectar la accesibilidad de la secuencia Shine-Dalgarno.

c) Cada construcción previamente fue secuenciada, se generaron los formatos de análisis tipo FASTA y se almacenaron como archivos de texto plano (Fig. 19).

```

>T7-EGFP
GGGGAATTGTGAGCGGATAACAATTC CCTCTAGAAATAATTTTGTTTAACTTTAAGAAG
GAGATATACCATGGTGAGCAAGGGCGAGGAGCTGTTCCACGGGGTGGTGCCCATCCTGGT
CGAGCTGGAC
>T7-EGFPΔG4
GGGGAATTGTGAGCGGATAACAATTC CCTCTAGAAATAATTTTGTTTAACTTTAAGAAG
GAGATATACCATGGTGAGCAAGGAGGAGCTGTTCCACGGGGTGGTGCCCATCCTGGTCGA
GCTGGACGGC
>Trc-EGFP
ATTGTGAGCGGATAACAATTTCCACACAGGAAACAGCATATGGTGAGCAAGGGCGAGGAGC
TGTTCCACGGGGTGGTGCCCATCCTGGTCGAGCTGGAC
>Trc-EGFPΔG4
ATTGTGAGCGGATAACAATTTCCACACAGGAAACAGCATATGGTGAGCAAGGAGGAGCTGT
TCACCGGGTGGTGCCCATCCTGGTCGAGCTGGACGGC
>Trc-rsEGFP
ATTGTGAGCGGATAACAATTTCCACACAGGAAACAGCATATGGTGAGCAAAGGAGAAGAAC
TCTTCACTGGAGTTGTCCCAATTCCTTGTGAATTAGAT
>Trc-rsEGFPΔG4
ATTGTGAGCGGATAACAATTTCCACACAGGAAACAGCATATGGTGAGCAAAGAAGAACTCT
TCACCTGGAGTTGTCCCAATTCCTTGTGAATTAGATGGT
>Trc-rsGFP
ATTGTGAGCGGATAACAATTTCCACACAGGAAACAGCATATGGCTAGCAAAGGAGAAGAAC
TCTTCACTGGAGTTGTCCCAATTCCTTGTGAATTAGAT
>Trc-rsGFPΔG4
ATTGTGAGCGGATAACAATTTCCACACAGGAAACAGCATATGGCTAGCAAAGAAGAACTCT
TCACCTGGAGTTGTCCCAATTCCTTGTGAATTAGATGGT

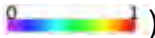
```


Figura 19. Secuencias nucleotídicas de la región líder e inicio del gen codificante de las diversas construcciones empleadas en el estudio. Se muestran las secuencias en formato FASTA para cada variante molecular obtenida. El codón de inicio ATG se indica en negritas para cada secuencia.

d) Se buscó una herramienta para el análisis de estructuras secundarias de RNA. En este sentido cabe señalar que existen una gran variedad de algoritmos tanto para análisis en Unix como para interfaces de visualización amigables (windows®). Una herramienta que integra un número amplio de variables para la predicción de estructuras secundarias y que además genera la visualización de las predicciones termodinámicas de estas, es el servidor RNAfold WebServer (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>). Este servidor es una herramienta ofrecida gratuitamente por el Instituto de Química Teórica, de la Universidad de Viena, este web server se incluye en los servicios de ViennaRNA Web Services (Gruber et al. 2008).

Los resultados obtenidos utilizando la herramienta RNAfold Web Server se interpretaron de acuerdo a la literatura descrita por Ivo L. Hofacker y Ronny Lorenz (Hofacker 2014; Hofacker & Lorenz 2014), una breve descripción del paquete que se incluye en el WebServer utilizado es hecha por Ronny Lorenz en la que se describen los alcances y limitaciones de la herramienta (Lorenz et al. 2011), por último se consultaron los detalles del Websuite, cuya descripción y manejo es detallado por Gruber y colaboradores en 2008 (Gruber et al. 2008).

La predicción de las estructuras secundarias arrojó como resultado gráficos en dos dimensiones. De acuerdo a lo sugerido por Hofacker y Lorenz el algoritmo elige la estructura que describe mejor la interpretación espacial de ésta llamándole MFE (minimum free energy, por sus siglas en inglés), en esta estructura también se incorporan datos propios de su dinámica y termodinámica, tales como la probabilidad de pares para cada base nucleotídica (la probabilidad

con la que cada base nucleotídica no forma pares y la entropía posicional de cada nucleótido). Para asignar un valor tangible a la dinámica de la molécula se describe una escala de color azul a rojo con un valor de 0 a 1 () respectivamente, siendo rojo o 1 el valor más alto obtenido para la probabilidad de pares, de lo contrario el color con tendencia a los azules adquiere un valor menor a 1 y una baja probabilidad de pares. El valor para la probabilidad de no formar pares se maneja en la misma escala de color y valor, siendo ésta interpretada como lo opuesto.

La asignación de valor termodinámico se da en una escala de color que va de los rojos a los azules (), el rojo representa el 0 y es considerado como la carencia de entropía posicional, mientras los colores hacia la escala de azules representan el incremento de la entropía posicional. Este valor describe la fuerza con la que se establece la estructura (termodinámicamente), valores dentro de la escala de los azules refieren a una estructura débil, con un nivel bajo de confianza.

Para identificar la diferencia entre los valores de dinámica y termodinámica se generaron por separado los gráficos para cada uno de estos valores sobre la misma estructura MFE. Los valores de entropía posicional para cada una de las variantes moleculares se presentan en la figura 24.

Los resultados del análisis en la variante molecular T7-EGFP predicen una estructura con cuatro brazos definidos (Fig. 20). EL primero de los brazos, el superior izquierdo, permanece bastante estable según los valores de probabilidad de pares, así como los de entropía (Fig. 24), sin embargo, la estructura

secundaria de RNA corresponde a la región del operador del regulador LacI (Fig. 18, inciso B), por lo que no influiría con el inicio de la traducción. El brazo inferior izquierdo contiene en su tallo principal la región Shine-Dalgarno (delimitada por • en la estructura), que si bien está estructurada, no es lo suficientemente estable ya que presenta niveles medios de probabilidad de pares y un alto nivel de entropía. En la figura 24, el nivel de entropía para T7-EGFP se mantiene alto en la posición 60 (región en la que se encuentra el Shine-Dalgarno). Algo muy similar pasa en el brazo inferior derecho en el que se puede observar el codon de inicio ATG (señalado como ♦ en la estructura). Este brazo también presenta una baja probabilidad de pares y un nivel intermedio de entropía, aunque para la variante T7-EGFP no se observaron regiones de RNA altamente estructuradas por lo que se obtiene un nivel de traducibilidad suficiente para ser detectado, (Fig.14 y Fig. 20).

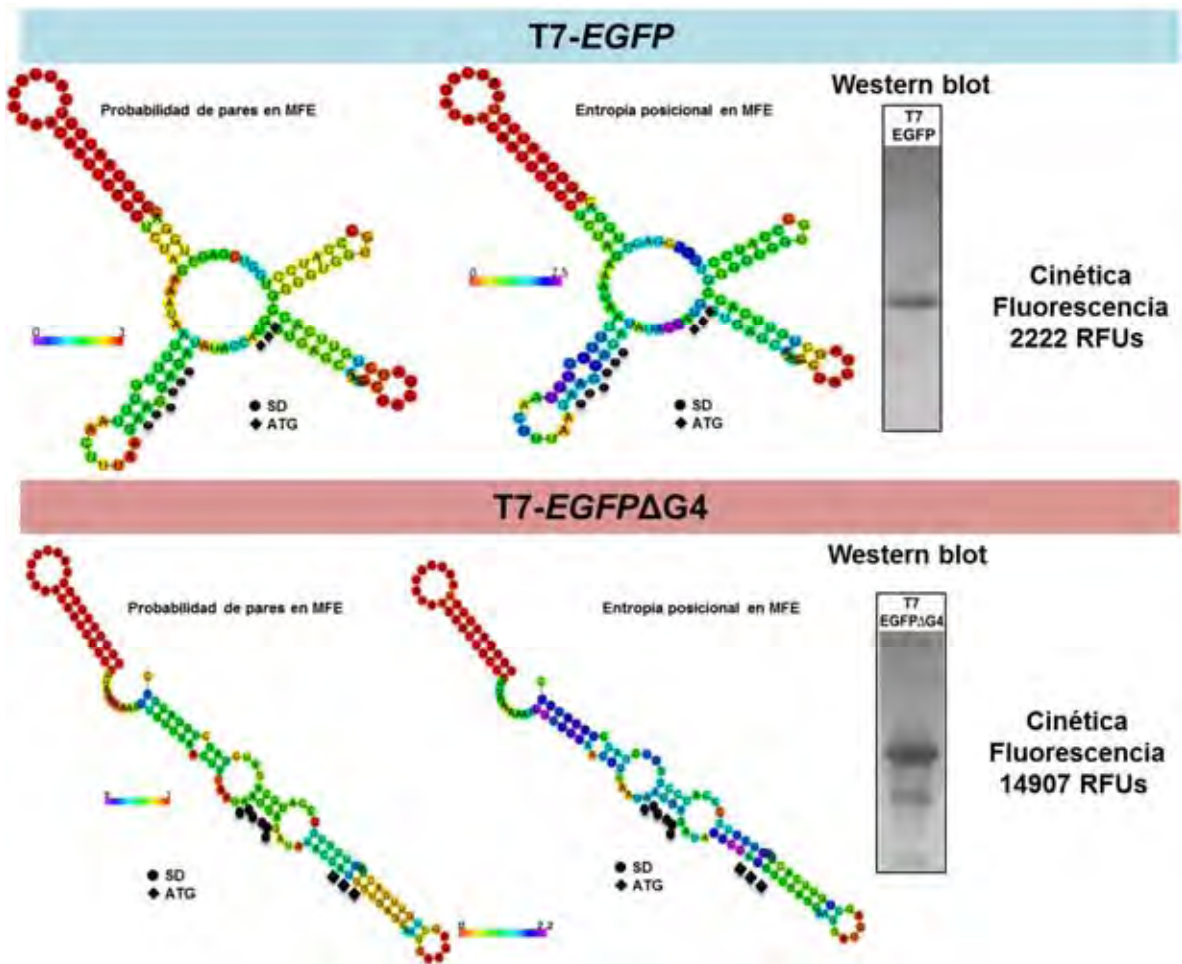


Figura 20. Estructuras secundarias de la variante T7-EGFP y T7-EGFP Δ G4. Se presentan la estructura correspondiente a MFE (m inimum free energy) y en estas se inscriben los valores de probabilidad de pares (imagen MFE izquierda) y los valores de entropía (imagen MFE derecha). La región de Shine-Dalgarno señalada con • y el codón de inicio ATG por ♦. Se incluye la detección por Western-blot de la figura 14 y el valor máximo en RFU de fluorescencia.

Para la variante molecular T7-*EGFP* Δ G4, la remoción de Gly4 genera un cambio en la conformación de la estructura (Fig. 20, T7-*EGFP* Δ G4), creando un solo tallo. Adicionalmente a lo anterior, se conserva el tallo de la región del operador *Lacl*. El sitio del Shine-Dalgarno (•) se encuentra en un tallo con una probabilidad de pares media y el sitio ATG (♦) se recorre al medio de un tallo con una probabilidad de pares más estable, sin embargo, los datos de entropía (Fig. 24) sugieren que toda la estructura de tallo pudiera ser inestable o pobremente no estructurado. El incremento en la entropía de la estructura sería la respuesta al gran incremento en los niveles de producción de proteína detectado por Western-blot, y que según datos de densitometría representan cerca de 1.77 veces de incremento (Tabla 6) y que se refleja en el incremento de fluorescencia detectada (Fig. 20).

El análisis realizado para la variante molecular *Trc-EGFP* predice una estructura en forma de tallo adornado con dos pequeños tallos en su región superior (Fig. 21, *Trc-EGFP*). El tallo pequeño de la región superior izquierda contiene el segmento correspondiente al operador *Lacl* del promotor *Trc* (Fig. 18, inciso A). La sección del tallo más largo en su región 5' contiene la secuencia Shine-Dalgarno (•) con valores medios de probabilidad de pares, hacia su extremo 3' se encuentra el ATG (♦) al inicio de un tallo con valores altos de probabilidad de pares. Aun con estos valores, los datos de entropía son lo suficiente altos para la región del tallo que contiene el Shine-Dalgarno y los tallos pequeños de la región superior de la estructura. Al igual que la estructura de la

variante T7-EGFP, la estructura de la variante Trc-EGFP permite un nivel de traducibilidad bueno, tal y como se muestra en las figuras 14 y 21.

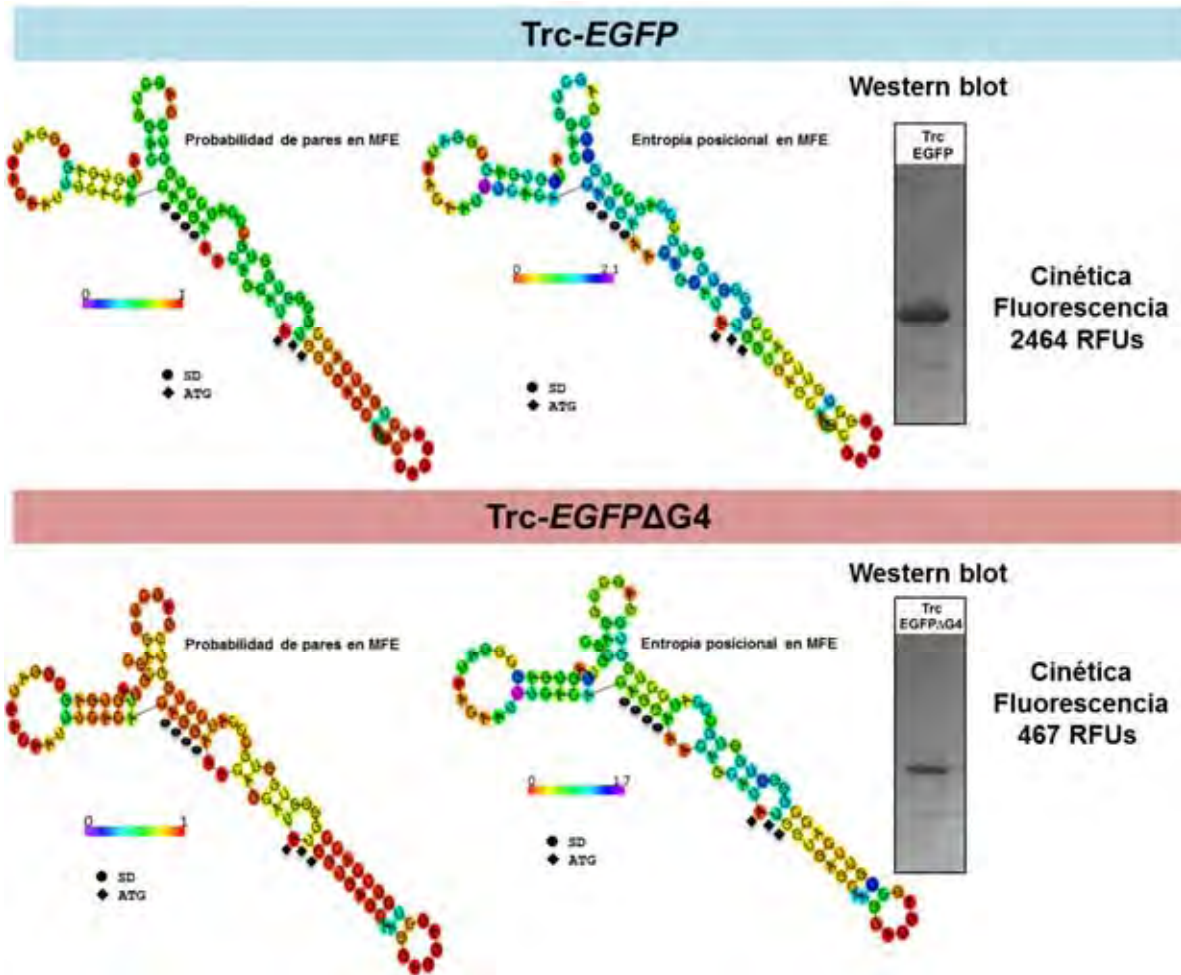


Figura 21. Estructuras secundarias de la variante Trc-EGFP y Trc-EGFP Δ G4.

Se presentan la estructura correspondiente a MFE (Minimum Free Energy) y en estas se inscriben los valores de probabilidad de pares (imagen MFE izquierda) y los valores de entropía (imagen MFE derecha). La región de Shine-Dalgarno señalada con • y el ATG por ♦. Se incluye la detección por Western-blot de la figura 14 y el valor máximo en RFU de fluorescencia.

El análisis de la variante *Trc-EGFP Δ G4* genera la misma estructura secundaria que en la variante *Trc-EGFP*, pero los datos de probabilidad de pares cambian a valores altos, es decir, el grado de estructuración corresponde a un nivel de confianza, es más real que los que se configuran para *Trc-EGFP*, *T7-EGFP* y *T7-EGFP Δ G4*, de acuerdo a la predicción del algoritmo utilizado para el análisis. Aunado a esto, los valores de entropía son bajos en el tallo que contiene la región Shine-Dalgarno (●) (Fig. 24, posición 40 a 60 nucleótidos) y así también para el tallo en el que se localiza el ATG (◆) (Hofacker 2014). Este resultado coincide con la disminución de proteína detectada por Western-blot de casi un 60% y la baja detección de fluorescencia que paso de 2464 RFUs para la variante *Trc-EGFP* a solo 467 RFUs para *Trc-EGFP Δ G4* (Fig. 14, Fig. 21).

El análisis del estructuramiento para la variante *Trc-rsEGFP* presentó valores medios para la probabilidad de pares en su MFE, conteniendo diversos colores de la escala de rojos para el tallo más distal. Este resultado nos indica que esta estructura no se genera con un nivel de confianza aceptable, además, presenta valores de entropía que van de medios a altos del nucleótido 1 al 98 (Fig. 24) a lo largo de la estructura de tallo que se predice (Fig. 22).

El tallo que se estructura contiene en su extremo 5' un pequeño tallo que contiene la región del operador *Lacl*, que no presenta niveles de probabilidad de pares altos. Hacia el inicio del tallo más grande se encuentra la secuencia Shine-Dalgarno (●) en un tallo con valores de probabilidad de pares medios y el ATG (◆) en un tallo con valores más estables de probabilidad de pares (entre colores amarillos y naranjas). Sin embargo, los valores de entropía para la región 5' del

tallo sonde se encuentra el Shine-Dalgarno son demasiado altos, y disminuyen por mitad aproximadamente hacia la región 3' que contiene el ATG. Estos valores indican que la estructura no es tan estable y permite la traducibilidad del mRNA con buena eficiencia, alcanzando valores de 3,399 RFUs.

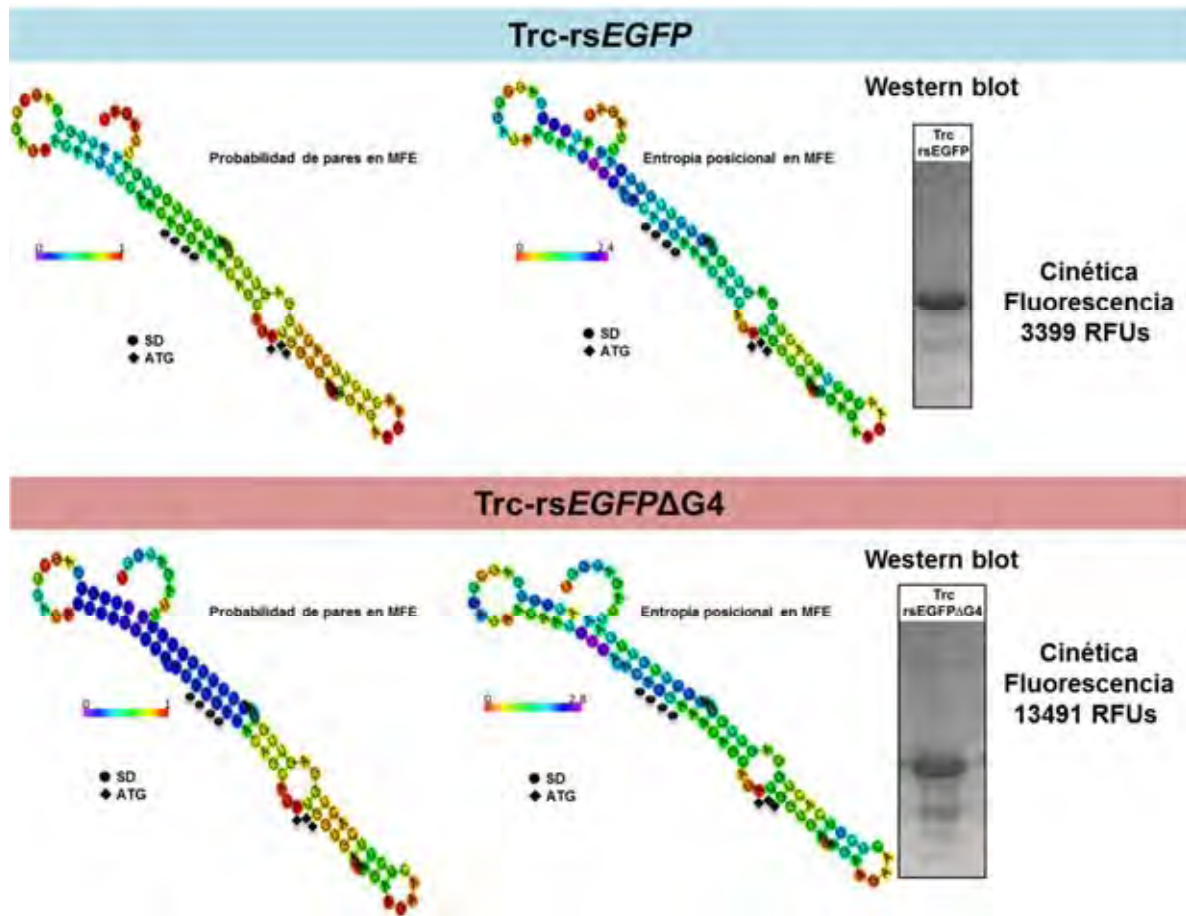


Figura 22. Estructuras secundarias de la variante *Trc-rsEGFP* y *Trc-rsEGFPΔG4*. Se presentan la estructura correspondiente a MFE (mínimum free energy) y en estas se inscriben los valores de probabilidad de pares (imagen MFE izquierda) y los valores de entropía (imagen MFE derecha). La región de Shine-Dalgarno señalada con • y el ATG por ♦. Se incluye la detección por Western-blot de la figura 14 y el valor máximo en RFU de fluorescencia.

La predicción realizada para la variante Trc-*rsEGFPΔG4* dio como resultado una estructura muy similar a la predicha en Trc-*rsEGFP*, pero resaltan los valores tan bajos de probabilidad de pares en su estructura MFE, justo en la región que contiene al sitio Shine-Dalgarno. En esta región la entropía posicional sigue siendo alta (Fig. 24). De acuerdo a la interpretación que se genera de este resultado, el estructuramiento no se encuentra en un nivel de confianza según Ivo L. Hofacker, (Hofacker 2014). La remoción de la Gly4 en esta variante generaría la pérdida de estabilidad de la estructura secundaria, por lo que no existirían límites físicos del mRNA al interactuar con la maquinaria ribosomal, generando en consecuencia un nivel alto de traducibilidad.

El resultado obtenido para el análisis de estructura de la variante Trc-*rsGFP* generó la predicción de una estructura con tres tallos, el más grande de estos contiene en su extremo 5' la región estructurada del operador Lacl, y en su región 3' se encuentra el sitio Shine-Dalgarno en uno de los tallos inferiores, el pequeño del extremo izquierdo contiene el ATG y para esta estructura la probabilidad de no formar pares es muy alta, es decir, pudiera no estarse generando esta estructura, aunque el tallo de la parte inferior derecha presenta un probabilidad de pares media. La entropía posicional para esta estructura es alta, por lo que no pudiera estarse generando esta estructura establemente (Fig. 23).

Para la versión con la remoción de Gly4, Trc-*rsGFPΔG4*, el análisis de estructura secundaria da como resultado una estructura en forma de tallo, parecido a la estructura de la versión Trc-*rsEGFPΔG4*, aunque para esta versión se observa una homogeneidad en los valores bajos de probabilidad de pares para

su estructura MFE. Incrementando los valores de entropía. Este resultado concuerda con un nivel alto de fluorescencia de hasta 15,099 RFUs, debido a la falta de estructura el acceso del mRNA para su traducción es más eficiente (de Smit & van Duin 2003).

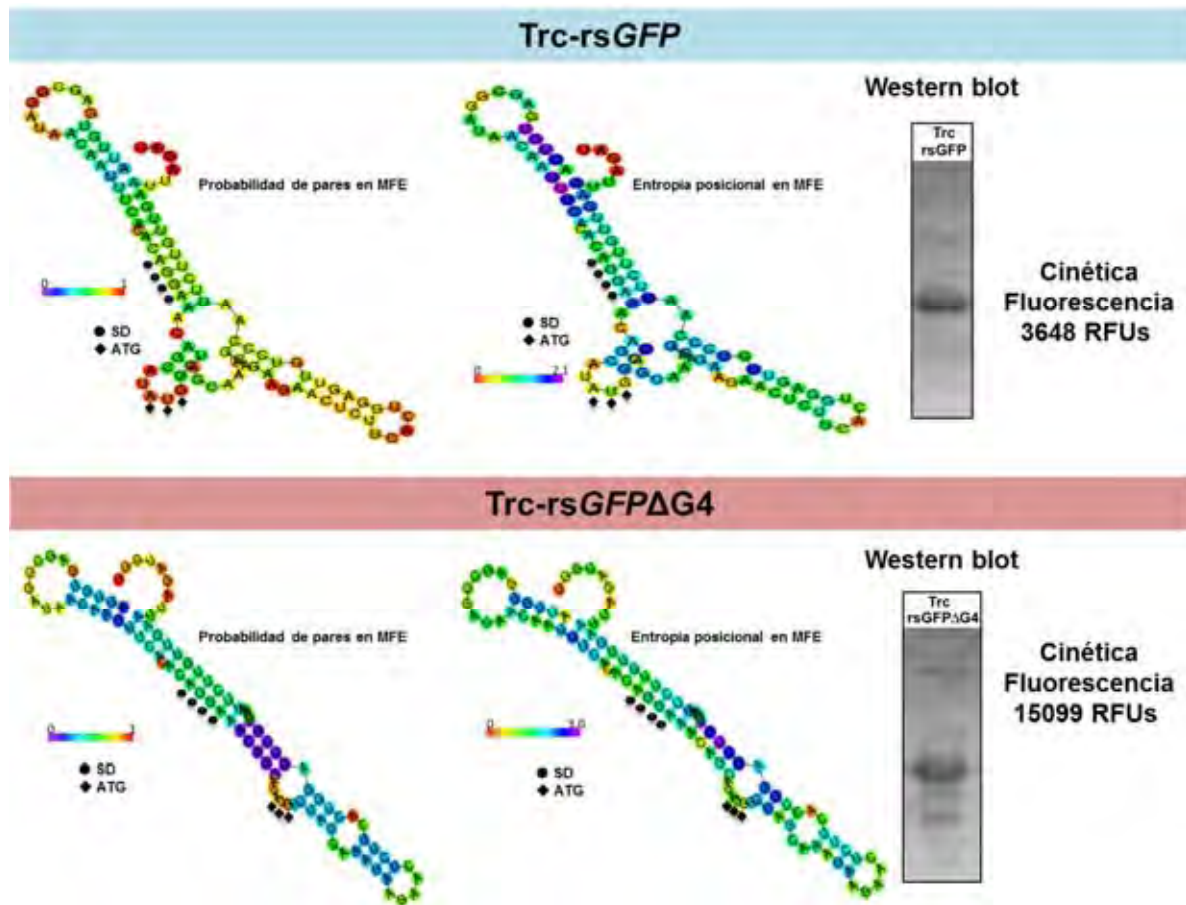


Figura 23. Estructuras secundarias de la variante Trc-rsGFP y Trc-rsGFP Δ G4. Se presentan la estructura correspondiente a MFE (minimum free energy) y en estas se inscriben los valores de probabilidad de pares (imagen MFE izquierda) y los valores de entropía (imagen MFE derecha). La región de Shine-Dalgarno señalada con \bullet y el ATG por \blacklozenge . Se incluye la detección por Western-blot de la figura 14 y el valor máximo en RFU de fluorescencia.

Gráficos entropía posicional

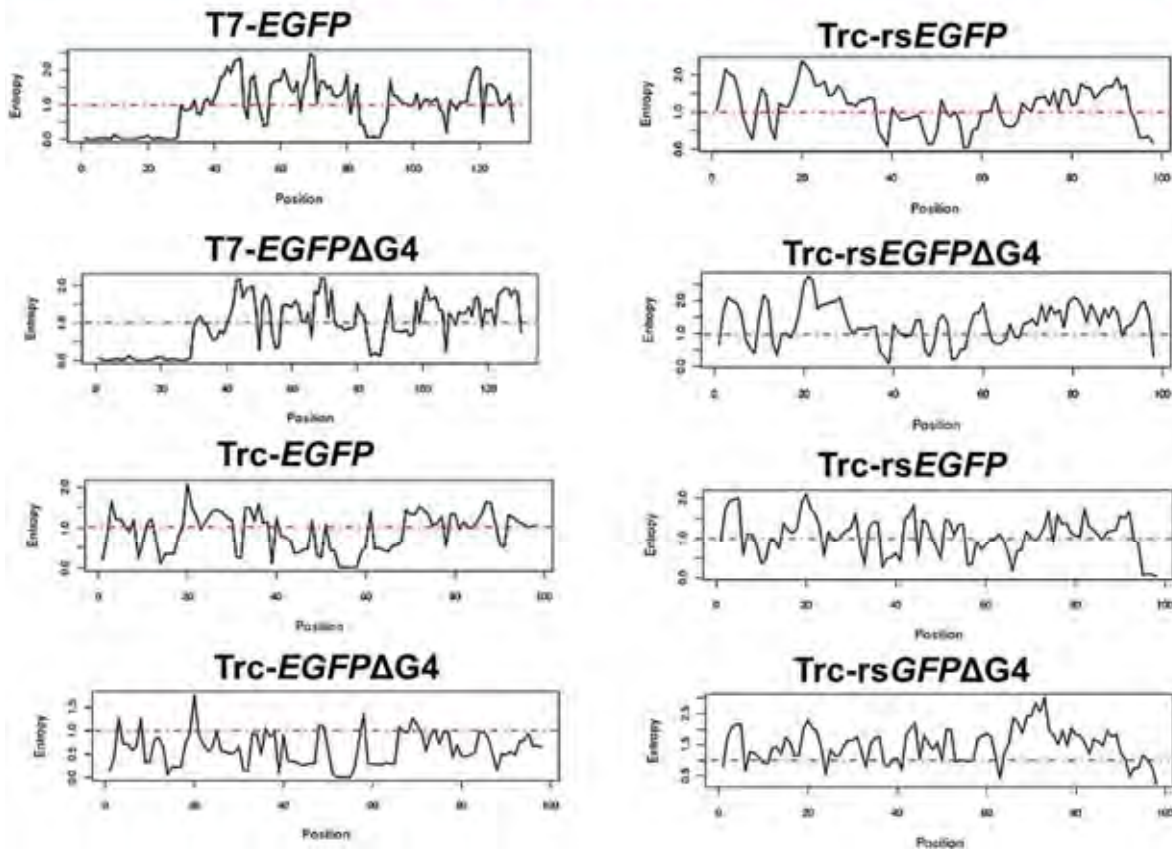


Figura 24. Gráficas de entropía posicional para las ocho variantes de GFP.

La línea roja en cada plot indica el valor medio de entropía. Valores por debajo de esta línea, determinan estabilidad de la estructura predicha. Picos por encima de la línea roja, indican un incremento en la entropía de la estructura, el incremento en número de estos picos describen la desestabilización de la estructura predicha.

9. Discusión

En el presente trabajo se obtuvieron ocho variantes moleculares de EGFP que permitieron interpretar el efecto sobre el estructuramiento del mRNA como consecuencia de la remoción del codón que codifica para la glicina en la cuarta posición del marco abierto de lectura (Gly4); cambio localizado en la región amino-terminal de la proteína EGFP. Las variantes se obtuvieron mediante una estrategia directa de mutagénesis que nos permitió de manera sencilla obtener la modificación deseada, desde la remoción de la Gly4 hasta la obtención de variantes moleculares derivadas de otras GFP.

Contrario a lo reportado por Arpino y colaboradores (Arpino et al. 2014), que proponen que la variante EGFP Δ G4 tiene un efecto sobre el plegamiento de la proteína, nuestros resultados revelaron que el efecto de la remoción de este codón está directamente relacionado con la estabilización o desestabilización de estructuras secundarias en el mRNA, que en consecuencia estarían involucradas con la disponibilidad de la región de unión a ribosoma del mRNA para su traducción, generando un incremento evidente en la producción de proteína (Fig. 13 y Fig. 14).

Los resultados del análisis *in silico* para la predicción de estructuras secundarias de las diferentes variantes moleculares, son consistentes con un efecto directo en la dinámica y termodinámica estructural de los mRNAs. El efecto de la remoción de Gly4 en la dinámica y termodinámica de la estructura secundaria de la versión T7-EGFP Δ G4, evidencia que el efecto de mejora en la fluorescencia se da por un incremento de la producción de proteína, consecuencia

de un desestructuramiento en su mRNA (Fig. 20, Fig. 24). Este resultado confirma que las estructuras secundarias mantienen un efecto de regulación a nivel traduccional, tal y como lo postuló de Smith y Van Duin en 2003, aunado a que la ausencia de estas benefician la traducibilidad de los mRNAs (Schulz & Reznikoff 1990).

Contrario al efecto observado en T7-EGFP Δ G4, Trc-EGFP Δ G4 presenta un estructuramiento más estable de acuerdo al análisis *in silico* (Fig. 21). De hecho, los picos con un nivel alto de entropía disminuyen considerablemente (Fig. 24) lo que permitió la formación de una estructura más estable que está relacionada con un fuerte bloqueo en la traducción del mRNA de esta variante y que fue reflejado en la disminución de la fluorescencia y la proteína detectada (Fig. 14, Tabla 6).

Los resultados que hemos obtenido del análisis de la mutante de GFP y sus variantes moleculares demuestran que en la mutagénesis realizada por Arpino y colaboradores (Arpino et al. 2014), no se consideró que los fenotipos observados (i.e. incremento en la fluorescencia) pudieran deberse o estar influenciados por la estructuras secundarias en el mRNA.

De acuerdo a los resultados obtenidos con relación a la producción final de proteína medidos en nuestros estudios de Western-blot, el decremento e incremento de fluorescencia, así como el desestructuramiento o estructuramiento de los mRNA analizados son consistentes con la hipótesis planteada en este trabajo sobre el potencial efecto de la remoción de Gly4 sobre la estructura secundaria en el mRNA de las GFP's analizadas y su efecto sobre los niveles con los que éstas se expresan.

La remoción de Gly4 en cada una de las versiones moleculares analizadas generan un efecto positivo en las versiones T7-*EGFP* Δ G4, Trc-rs*GFP* Δ G4 y Trc-rs*EGFP* Δ G4, incrementando la fluorescencia y producción de proteína, debido a que la remoción de este codón no genera una estructura estable que pudiera afectar la disponibilidad de la región correspondiente al Shine-Dalgarno y codón de inicio ATG. La disponibilidad de esta región para cada una de las versiones, hace que su mRNA sea susceptible a un mayor incremento en su traducción. También generamos una versión molecular en la que la remoción de Gly4, de manera contraria a las anteriores construcciones, afectó negativamente a la traducibilidad de su mRNA, en la versión Trc-*EGFP* Δ G4. En esta construcción, la remoción de Gly4 incrementó la estabilidad de la estructura secundaria en todo el mRNA, incluida la región del RBS, que al no estar disponible redujo el nivel de su traducibilidad. La remoción corre el marco de lectura incrementando el contenido de G+C en la región del RBS, favoreciendo la formación de una estructura secundaria estable.

La adecuación en el uso de codones fue considerada en la versión Trc-rs*EGFP* y Trc-rs *EGFP* Δ G4 permitió contar con un control sobre el efecto en el cambio nucleotídico a lo largo de toda la versión molecular generada. Sin embargo, al igual como lo observó Kudla y colaboradores en 2009, el cambio de codones de las versiones originales por unos sinónimos no generaron un efecto sobre los niveles de expresión observados, lo que podría indicar que el inicio de la traducción es un punto limitante para la expresión de los genes y no el proceso de elongación ya iniciada la traducción (Kudla et al. 2009).

10. Conclusiones

- La remoción del codón que codifica para el aminoácido glicina en la posición 4 de las versiones de GFP analizadas, influye directamente con la dinámica y termodinámica de una estructura secundaria en el mRNA.
- La versión molecular de EGFP Δ G4 propuesta por Arpino y colaboradores en 2014, incrementa su fluorescencia debido a un cambio en la estructura secundaria del mRNA y no por una mejora en su plegamiento.
- E GFP Δ G4 bajo la regulación de un promotor distinto a T7, como el promotor Trc, presenta un efecto contrario a lo reportado por Arpino y colaboradores en 2014. Dicha diferencia puede explicarse al cambio en el contenido de G+C de la región RBS de la sección líder del promotor y el ORF del gen que favorece la formación de estructuras estables en el mRNA.
- La remoción de Gly4 en la versión Trc-EGFP Δ G4 genera una estructura secundaria con niveles estables de dinámica y termodinámica. Este estructuramiento afecta directamente la traducción de su mRNA.
- El análisis presentado en este trabajo prueba que la eliminación de codones en la región 5' del mRNA presenta un efecto sobre la formación de estructuras secundarias y en consecuencia afecta la disponibilidad de la región RBS al interactuar con la región anti-Shine-Dalgarno del RNA 16S ribosomal, afectando directamente la traducibilidad del mensajero.

11. Perspectivas

- Implementar la estrategia de análisis desarrollada en este trabajo para la obtención de mejoras en la expresión heteróloga de proteínas.
- Utilizar la secuencia que comprende la región RBS de las variantes moleculares con mayor fluorescencia en otras proteínas y verificar el cambio en sus niveles de expresión.
- Realizar análisis de estabilidad de RNA, e identificar si un menor estructuramiento afecta la vida media del RNA.

12. Manuscrito

En esta sección se anexa la publicación generada de los resultados obtenidos del análisis de ocho variantes de la proteína verde fluorescente, así como, la influencia de la remoción de la Gly4 en el estructuramiento de sus mRNA y su efecto en la expresión de esta proteína.

El artículo lleva como título: A Codon Deletion at the Beginning of Green Fluorescent Protein Genes Enhances Protein Expression

Autores: José Luis Rodríguez Mejía, Abigail Roldán Salgado, Joel Osuna Quintero, Enrique Merino Pérez y Paul Gaytán Colin.

El trabajo se desarrolló en: Instituto de Biotecnología, UNAM Cuernavaca, Molecos. En los departamentos de Microbiología molecular e Ingeniería Celular y Biocatálisis.

A Codon Deletion at the Beginning of Green Fluorescent Protein Genes Enhances Protein Expression

José-Luis Rodríguez-Mejía^a Abigail Roldán-Salgado^b Joel Osuna^b
Enrique Merino^a Paul Gaytán^b

Departamentos de ^aMicrobiología Molecular and ^bIngeniería Celular y Biotecnología, Instituto de Biotecnología, UNAM, Cuernavaca, Mexico

Key Words

Fluorescent proteins · Codon deletions · Amino acid deletions · Heterologous protein expression · mRNA secondary structure · mRNA translation rate

Abstract

Recombinant protein expression is one of the key issues in protein engineering and biotechnology. Among the different models for assessing protein production and structure-function studies, green fluorescent protein (GFP) is one of the preferred models because of its importance as a reporter in cellular and molecular studies. In this research we analyze the effect of codon deletions near the amino terminus of different GFP proteins on fluorescence. Our study includes Gly4 deletions in the enhanced GFP (EGFP), the red-shifted GFP and the red-shifted EGFP. The Gly4 deletion mutants and their corresponding wild-type counterparts were transcribed under the control of the T7 or Trc promoters and their expression patterns were analyzed. Different fluorescent outcomes were observed depending on the type of fluorescent gene versions. In silico analysis of the RNA secondary structures near the ribosome binding site revealed a direct relationship between their minimum free energy and

GFP production. Integrative analysis of these results, including SDS-PAGE analysis, led us to conclude that the fluorescence improvement of cells expressing different versions of GFPs with Gly4 deleted is due to an enhancement of the accessibility of the ribosome binding site by reducing the stability of the RNA secondary structures at their mRNA leader regions.

© 2016 S. Karger AG, Basel

Introduction

Green fluorescent protein (GFP) is an excellent genetic molecular reporter because the formation of its chromophore is spontaneously achieved in a posttranslational mode and because this protein requires only appropriate folding and oxygen to be functionally active in several organisms [Zimmer, 2002]. This protein, which is composed of only 238 amino acids, displays a unique cylinder shape with a β -barrel structure consisting of eleven β -strands and with an α -helix running through the center of the barrel, which contains the covalently bonded chromophore [Flores-Ramírez et al., 2007] formed by the three contiguous residues Ser65, Tyr66 and Gly67 in the

KARGER

© 2016 S. Karger AG, Basel

E-Mail karger@karger.com
www.karger.com/jmb

Enrique Merino or Paul Gaytán
Instituto de Biotecnología, UNAM
Avenida Universidad 2001
Cuernavaca, Morelos CP62119 (Mexico)
E-Mail merino@ib.unam.mx or paul@ib.unam.mx

Downloaded by:
Univ. of Toronto on 08/28/16
143.166.201.117 - 143.166.201.117

original protein *avGFP*, which was cloned from the jellyfish *Aequorea victoria* [Prasher et al., 1992]. However, the properties of *avGFP* are not optimal for general use in the laboratory because this protein is poorly expressed at 37°C and exhibits two absorption peaks, a slow maturation rate and fast photobleaching.

For these reasons, *avGFP* suddenly became one of the preferred targets with which to perform protein engineering to improve its properties [Cormack et al., 1996; Cramer et al., 1996; Delagrave et al., 1995; Heim et al., 1994, 1995]. From these early studies, enhanced GFP (EGFP) was isolated [Cormack et al., 1996] and commercialized by Clontech in a humanized version of the codons. The double mutation F64L/S65T found in EGFP improved the expression of the protein at 37°C and simplified the excitation spectrum to only one peak in the blue region of the visible spectrum, producing a much better reporter than *avGFP*. Now, newly engineered versions of GFP are continuously being reported that claim to be better than the previous versions. However, all improved artificial mutants share a common feature, namely they were created by amino acid substitutions performed either in a site-directed, combinatorial or completely random mutagenesis mode, and only one example has been soundly attributed to an amino acid deletion. Arpino et al. [2014] recently reported that the mutant gene EGFPΔG4 (coding for EGFP with Gly4 deleted) produced cells with greater fluorescence than did the parental gene EGFP under control of the T7 promoter, concluding after a speculative description of the crystal structure that the Gly4 deletion improved the folding process, increasing the amount of soluble protein. However, when we cloned both genes under the control of the T7 promoter the results were opposite: cells expressing the EGFP gene produced more fluorescence than did cells expressing EGFPΔG4.

Intrigued by this result, we decided to study in depth the effect of Gly4 deletion in the other two GFP genes: red-shifted GFP (*rsGFP*) and *rsEGFP*. In comparison to the original *avGFP*, *rsGFP* contains the mutations F64L, S65C, I167T, L231H and K238N, as well as an additional amino acid numbered Ala-1b, whereas *rsEGFP*, which was obtained from multiple site-directed mutagenesis of *rsGFP*, carries the mutations F64L, S65T and Val-1b. Therefore, EGFP and *rsEGFP* code for exactly the same fluorescent proteins that are expressed differently due to the different effects of the Gly4 deletion on their corresponding SD accessibility, as will be shown in the present research.



Fig. 1. Streaks of *E. coli* expressing different GFP variants and their corresponding mutants with Gly4 deleted. The cells were streaked on solid LB kanamycin-containing plates in the presence of IPTG as a gratuitous inducer, and were incubated at 37°C for 20 h. Transcription was performed from promoters T7 or Trc. The EGFP and *rsEGFP* genes code for exactly the same GFP amino acid sequence; nevertheless, their nucleotide compositions are very different.

Results

Expression of GFP Variants

As the stability of mRNA folding near the ribosomal binding site explains the variable protein level from synonymous genes containing silent base substitutions [Kudla et al., 2009], we decided to prepare eight different constructs to evaluate the importance of the promoter sequence in the expression of genes carrying a codon deletion near the 5' end and to determine an alternative explanation for the fluorescence improvement observed in EGFPΔG4 with respect to the parental protein EGFP as previously reported [Arpino et al., 2014]. The tested promoters were T7 and Trc and the expression of the GFP variants were as outlined below.

T7-EGFP and T7-EGFPΔG4. When the plasmids containing each gene (see Experimental Procedures for details of plasmid constructions) were independently used to transform the *Escherichia coli* strain BL21 gold DE3, the resulting colonies behaved as previously described

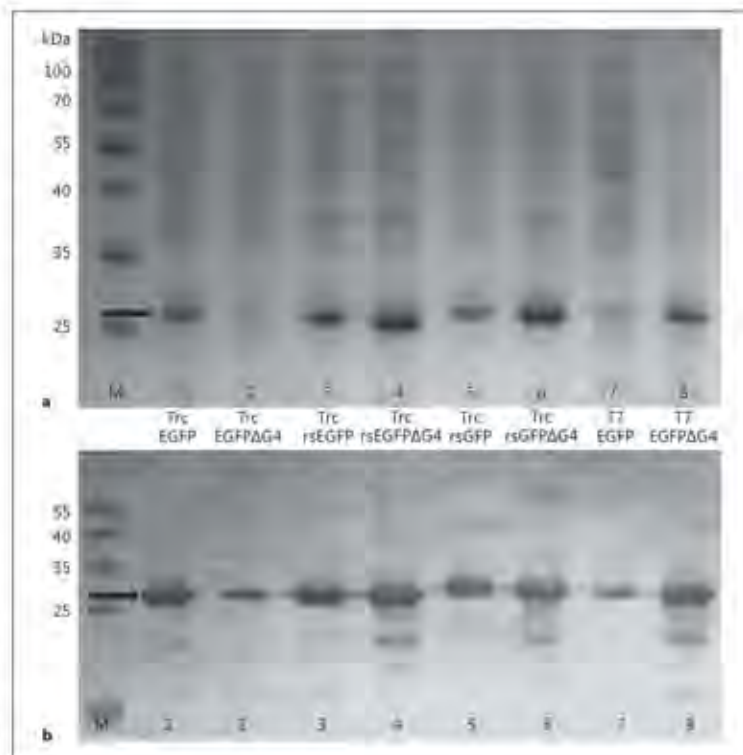


Fig. 2. SDS-PAGE analysis of total protein extracts under denaturing conditions. **a** Gel stained with Coomassie blue; **b** Proteins were detected by Western blot using alkaline phosphatase-conjugated anti-GFP. GFP bands are indicated with a bold arrow. M = Molecular protein marker.

[Arpino et al., 2014]: cells expressing the deletion mutant T7-EGFPΔG4 were more fluorescent than those expressing the parental protein, as shown in figure 1. Because both proteins exhibit similar extinction coefficients and quantum yields [Arpino et al., 2014], the reason for the difference in fluorescence was increased production of EGFPΔG4 with respect to EGFP as demonstrated by SDS-PAGE analysis of the total extracts shown in figure 2 (lanes 8 and 7, respectively).

Trc-EGFP and Trc-EGFPΔG4. *E. coli* strain MC1061 cells were separately transfected with both constructs (see Experimental Procedures for details of plasmid constructions) and after the cells were incubated, the cells expressing the deletion mutant *Trc-EGFPΔG4* were less fluorescent than were those cells expressing the parental protein, as shown in figure 1. Thus, EGFPΔG4 production was lower than EGFP production, as shown in lanes 2 and 1, respectively, in figure 2. This opposite result compared to that observed with the T7-based constructs suggested that Gly4 deletion did not influence the protein folding

process, otherwise EGFPΔG4 would have been more abundant than the parental EGFP, independent of the promoter or bacterial strain. In line with these results, Liu et al. [2015] recently reported that the cells expressing the deletion mutant GFPuvΔG4 also had lower fluorescence levels than did those cells expressing the parental protein GFPuv when transcription was performed from the *lac* promoter. To study the effect of Gly4 deletion in fluorescent proteins in greater detail, we focused our research on the fluorescent variant rsGFP, which is contained in the commercial plasmid pQB125 and is well expressed in bacterial cells. Although EGFP and rsGFP differ by only six amino acids, their genes differ by 172 bp: EGFP is composed of the preferred human codons, whereas rsGFP is more similar to the original avGFP. However, the photochemical properties of both proteins are very similar.

Trc-rsGFP and Trc-rsGFPΔG4. After the cells were transformed with these plasmids (see Experimental Procedures for details of plasmid constructions) and grown in solid LB, cells expressing the deletion mutant rsGFPΔG4

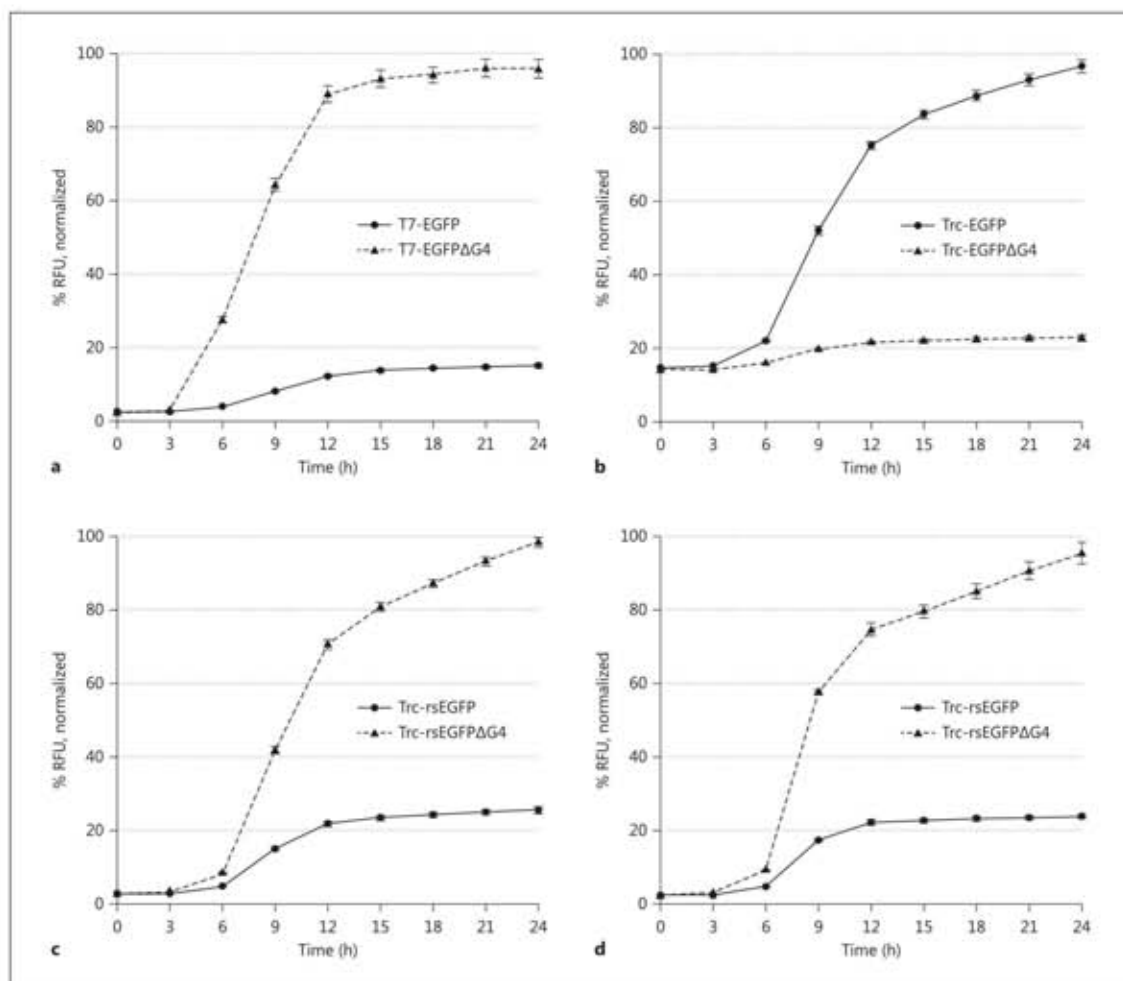


Fig. 3. Normalized kinetic fluorescence of *E. coli* cells expressing different GFP constructs. Cell fluorescence was monitored over a 24-hour period. Circles correspond to cells expressing the complete GFP genes, while triangles correspond to cells expressing the mutants lacking Gly4. **a** Cells expressing EGFP variants under the control of the T7 promoter. **b** Cells expressing EGFP variants un-

der the control of the Trc promoter. **c** Cells expressing rsEGFP variants under the control of the Trc promoter. **d** Cells expressing rsGFP variants under the control of the Trc promoter. Standard deviations of the normalized values are represented as vertical bars.

generated more fluorescent colonies than the parental protein rsGFP (fig. 1), as was observed with the pair T7-EGFP and T7-EGFPΔG4 (fig. 2, lanes 6 and 5, respectively). This result suggested that fluorescence improvement of rsGFPΔG4 depended on the combined sequence of the promoter and gene and not on the protein sequence alone. However, to confirm this hypothesis, we converted

the rsGFP gene into rsEGFP by multiple site-directed mutagenesis as detailed in Experimental Procedures. In this sense, the proteins EGFP and rsEGFP are identical in amino acid sequence, but their genes differ by 160 bp.

Trc-rsEGFP and Trc-rsEGFPΔG4. Transformation of *E. coli* MC1061 cells with both constructs (see Experimental Procedures for details of plasmid constructions)

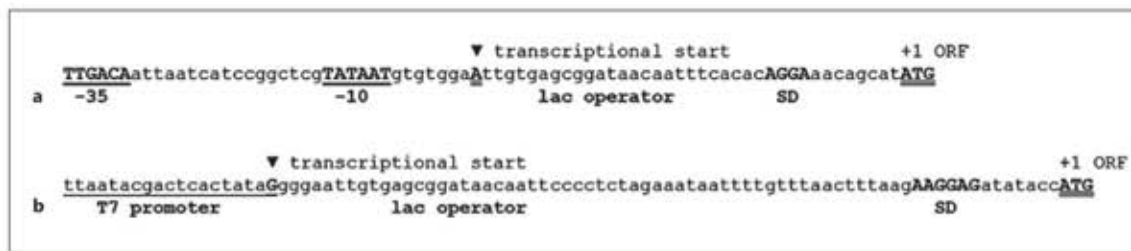


Fig. 4. Scheme of the regulatory regions used in this study. **a** Regulatory region using the Trc promoter. The -10 and -35 boxes are underlined. **b** Regulatory region using the T7 promoter (underlined). The SD sequences are shown in bold, while the transcription start sites and the ATG start codons are double underlined.

produced more fluorescent colonies with the deletion mutant than with the complete gene, a result similar to that observed with the construct pairs T7-EGFP/T7-EGFPΔG4 and Trc-rsGFP/Trc-rsGFPΔG4. In line with the observed fluorescence intensities of the colonies, production of the deletion protein rsEGFPΔG4 was higher than that of the complete protein rsEGFP, as observed in lanes 4 and 3, respectively, in figure 2. Therefore, these experiments indicated that production of the deletion mutants depended on the combined promoter-gene nucleotide sequence and not on a folding event. In agreement with these results, Li et al. [1997] demonstrated relatively early that EGFP could tolerate deletion of its last nine amino acids or the first six amino acids, including Gly4, which did not affect the fluorescence of the protein. For this reason, deletion of Gly4 (ΔG4) in EGFP is unlikely to improve the protein folding when it is not necessary for its production. Gly4 is spatially far from the chromophore, which is located in the center of the structure, whereas most of the reported mutations that improve GFP folding are located in the central helix and in β-sheet strands 7, 8 and 10 surrounding the chromophore [Shaner et al., 2007].

Fluorescence Kinetic Analysis

To characterize the in vivo fluorescence activity of the different constructs carrying ΔGly4, we performed a kinetic study of whole cells for 24 h, with measurements every 25 min. The results of this analysis are shown in figure 3 and are in good agreement with the apparent fluorescence intensities of the streaks shown in figure 1 and with the densities of the GFP bands observed in figure 2. Compared to cells expressing their parental proteins, a positive effect of ΔGly4 on fluorescence was observed for cells expressing the constructs T7-EGFPΔG4, Trc-

rsEGFPΔG4 and Trc-rsGFPΔG4, but a negative effect of ΔGly4 on fluorescence was observed for the cells expressing Trc-EGFPΔG4.

RNA Structure and Stability Bioinformatic Analysis

Prediction of RNA secondary structures and their corresponding base pair probabilities was performed using the ViennaRNA Package 2.0 [Lorenz et al., 2011] and the ViennaRNA Web Services [Gruber et al., 2008] because this software has been proven to be a powerful tool that is robust and dynamic [Andronescu et al., 2010; Mathews et al., 2004]. We compared the RNA secondary structures of the 5' UTR leader sequences and 60 bp within the different ORF constructs. The analyzed sequences depended on the different types of GFP genes (EGFP, EGFPΔG4, rsEGFP, rsEGFPΔG4, rsGFP and rsGFPΔG4) and on the leader sequence following each promoter (T7 or Trc). As shown in figure 4, these differences affected the transcription start site, the leader region size, and the ribosomal binding site or Shine-Dalgarno (SD) sequences. Notably, the SD sequence of the Trc constructs (AGGA) is less conserved than that of T7 constructs (AAGGAG) [Amann et al., 1988; Dubendorff and Studier, 1991].

Comparison of the secondary structures, base-pairing probability and positional entropy values shown in figure 5 resulted in the following observations. (1) Deletion of the Gly4 codon on T7-EGFP produces a change in the predicted secondary structure of the RNA leader sequence with a concomitant change in the secondary structures that contain the SD sequences. In T7-EGFP, the SD sequence is located in the stem and loop structure, whereas in T7-EGFPΔG4 the SD sequence is part of a short stem (4 bp long) flanked by two loops. This new location of the SD sequence in T7-EGFPΔG4 may contribute to increased ribosome accessibility with a correspond-

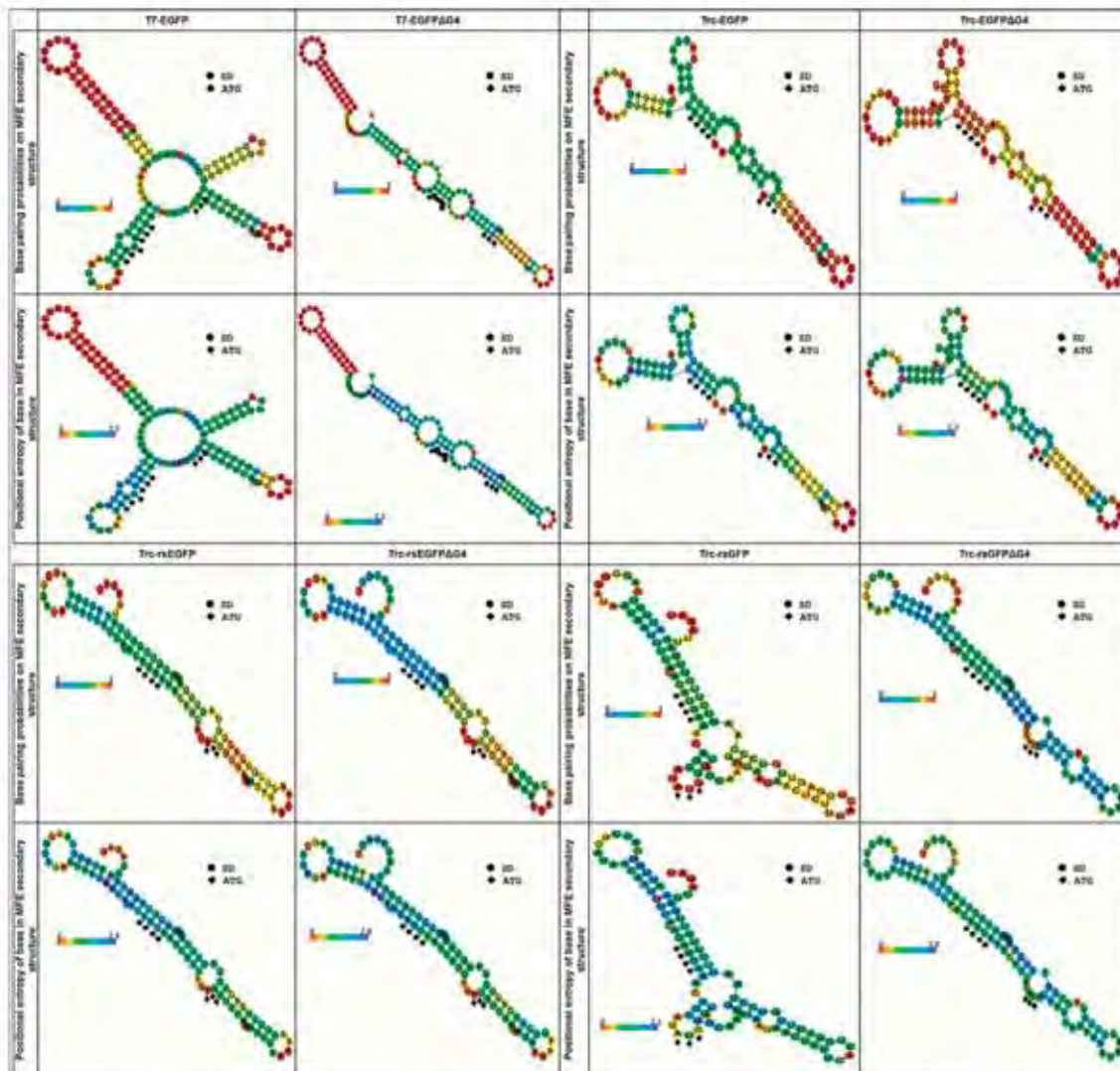


Fig. 5. Secondary structures of the leader regions and 60 bp within the ORF used in our study. Nucleotides are color coded according to their base-pairing probabilities of occurring in the optimal and suboptimal secondary structures predicted to be formed for a given RNA sequence using the algorithms implemented in the Vienna RNA Package [Gruber et al., 2008; Lorenz et al., 2011]. For unpaired regions, the color denotes the probability of being un-

paired. The probability value of each base can vary from 1 to 0, with 1 being the highest and 0 being the lowest, and are represented by red and blue, respectively [Lorenz et al., 2011]. Nucleotides are color coded according to their positional entropy, which was evaluated considering the number of different pairings of that base in a set of suboptimal structures. Well-defined regions should have low positional entropy and high pair probabilities.

ing increase in the rate of translation initiation. (2) In contrast to the previous case, deletion of the Gly4 codon in Trc-EGFP did not result in an evident change in the overall RNA secondary structure. Nevertheless, a clear increase in the base pair binding probability of the residues that form the stem and loop structure containing the SD sequence and a decrease in the positional entropy of the nucleotides around it could be observed. This increase in the base pair binding probability and the decrease in the positional entropy correspond to a better-defined secondary structure containing the SD sequence, which is in accordance with a possible inhibitory effect on EGFPΔG4 translation initiation and with the observed decreases in the GFP concentration in figure 2 and the fluorescence activity shown in figures 1 and 3. (3) Gly4 codon deletion on Trc-rsEGFP results in a clear reduction of the base pair binding probabilities of those residues that form the stem and loop structure containing the SD sequence. (4) Gly4 codon deletion on Trc-rsGFP produces a change in the predicted secondary structure of the RNA leader. The aforementioned results are in line with the premise that changes in the apparent cellular fluorescence of the different constructs are due to changes in the mRNA secondary structure and in the frequency with which ribosomes can recognize the SD sequence and start translation of the mRNA.

Discussion

Recent developments in protein engineering have proposed random amino acid deletions as a good alternative to classic amino acid substitution mutagenesis to expand protein sequence space [Osuna et al., 2004; Raghunathan et al., 2012]. Here we presented evidence that codon deletion is a feasible approach with which to enhance SD accessibility and protein expression.

Whereas random amino acid substitutions are easily performed by error-prone PCR, DNA shuffling or conventional degenerated oligonucleotides, random amino acid deletions are more difficult to perform because they require the use of either special oligonucleotide libraries or enzymatic methods that make use of transposons [Jones, 2005] or multi-enzymatic steps that are difficult to implement [Murakami et al., 2002]. Contrary to oligonucleotide-based deletions, which generate clean amino acid removal, enzymatic methods also generate many undesired amino acid substitutions near the amino acid deleted, making the study of the deletions technically difficult.

In our previous work, we used an oligonucleotide library assembled by the Codon-Based Random Deletion (COBARDE) mutagenesis approach [Osuna et al., 2004] and successfully isolated two functional deletion mutants, sgGFPΔI128 and sgGFPΔD129, which remained robustly fluorescent under direct visualization with sunlight. In fact, *E. coli* cells expressing sgGFPΔD129 at 37°C were slightly less fluorescent than were those expressing the parental protein sgGFP and were more fluorescent at 30°C. A similar result was obtained by Liu et al. [2015] for several internal deletion mutants of GFPuv, the relative fluorescence of which was improved at a lower temperature. These results are in line with the general trend observed in proteins containing amino acid deletions in which stability is reduced due to folding problems because many of the internal contacts between the amino acids surrounding the deletion are modified or lost. In contrast, RNA secondary structures near the ribosomal binding site are known to decrease the mRNA association rate constant with the ribosomal 30S subunit because these types of structures represent a thermodynamic barrier to form a stable mRNA-30S complex [de Smit and van Duin, 1990a, b; Studer and Joseph, 2006]. Early evidence of the effect of RNA structures on translation initiation has been reported for the lysozyme gene of bacteriophage T4 [Knight et al., 1987], *E. coli galE* [Merrill et al., 1981] and *lamB* genes [Hall et al., 1982], among others, and systematic and qualitative analyses of the influence of the stability and position of RNA secondary structures near the SD sequence on translation initiation have been performed by de Smit and van Duin [1990a, b], Studer and Joseph [2006], and Osterman et al. [2013]. Last year, Mahalik et al. [2014] revised a list of variables required to improve gene expression in *E. coli* and demonstrated that translation initiation in the presence of stable mRNA secondary structures could be a bottleneck for protein expression.

The influence of the secondary structures near the SD and start codon on translation initiation efficiency is not restricted to *E. coli* and seems to be a universal trait in all organisms. In a bioinformatics screening of 340 organisms including bacteria, archaea, fungi, plants, insects, fish, birds and mammals, Gu et al. [2010] identified a statistically significant tendency for reduced mRNA stability near the translation initiation site. These authors also proposed that this universal tendency of genes to present low stable local mRNA secondary structures near their translation initiation region is accomplished by a selective pressure on synonymous codons of the coding regions near the amino terminus of the genes.

In relation to previously reported studies regarding the effect of RNA secondary structures near the translation initiation region of the GFP gene, the results obtained by Kudla et al. [2009], who analyzed the effect of synonymous mutations on GFP gene expression, are worth mentioning. The authors synthesized a library of GFP genes that varied randomly in their codon usage but that encoded the same amino acid sequence. These authors found a 250-fold difference in the fluorescence intensity across the library, which mainly correlated with the free energy of potential RNA secondary structures that could be formed in the region -4 to +79 with respect to the GFP coding region, demonstrating that small fluctuations in the nucleotide sequence of this gene might influence its rate of translation initiation and the observed cellular fluorescence, even in gene variants coding for identical amino acid sequences of GFP [Kudla et al., 2009].

Here, we presented data regarding the effect of codon deletions near the amino terminus of GFP genes that are in good agreement with the aforementioned studies. We demonstrated that the improvement in cell fluorescence by deletion of the codon for Gly4 observed by Arpino et al. [2014] could be observed not only in cells expressing the construct T7-EGFPΔG4, but also in cells expressing other GFP constructs such as Trc-rsEGFPΔG4 and Trc-rsGFPΔG4 (fig. 1, 3). This enhanced effect on cell fluorescence was originally explained by these authors to be the result of a better protein folding. However, in our study we presented data supporting the idea that the effect of ΔGly4 on cells expressing GFP depends on the RNA context where the SD is located (fig. 1, 5). In addition, Western blot analysis indicated that the fluorescent phenotype of the cells harboring the eight different constructs correlated with the total amount of GFP produced (fig. 2).

These results indicate that efficient expression of any recombinant gene must consider not only the sequence of its beginning coding region, but also the sequence of its untranslated leader region associated to different promoters (fig. 4). For instance, in our study the mutant gene EGFPΔG4 under the control of the T7 promoter produced more protein than the parental gene EGFP. A similar result was observed with the genes rsEGFPΔG4 and rsGFPΔG4 under the control of the Trc promoter. However, EGFPΔG4 under the control of the Trc promoter yielded inverted results.

In recent years, a number of studies aiming to optimize the synthesis of recombinant proteins have indicated that one strategy to increase the translation initiation rate is to make substitutions in the region surrounding the initiation codon that generates unstructured RNA regions

[Care et al., 2008; Komarova et al., 2005; Seo et al., 2013]. As an alternative to codon substitutions and based on the results shown in our research, we propose that the expression of any gene could be notably improved by performing a fine scanning of codon deletions for the first 15 codons of the gene, using 15 site-specific oligonucleotides synthesized independently and combined before the cloning process, or using a library of combined deletions such as that created by the COBARDE mutagenesis approach, which generates all possible combinations of single to multiple amino acid deletions on a focused region of the gene.

Experimental Procedures

All reagents and enzymes used for PCR amplification and cloning were purchased from New England BioLabs (NEB), including the enzyme Vent DNA polymerase. The kits for PCR product purification, agarose gel extraction and plasmids were purchased from Bio Basic Inc.

Genes expressed under the control of the T7 promoter were cloned as *NcoI/XhoI* inserts in the plasmid pET-28a (Novagen), whereas genes expressed under the control of the Trc promoter were cloned as *NdeI/XhoI* inserts in the noncommercial plasmid termed pJOQ. This 2,001-bp multicopy plasmid contains the kanamycin-resistance gene, the pBR322 origin of replication, the Trc promoter and the cloning sites *NdeI* and *XhoI* [Torres Tejerizo et al., 2015]. Genes cloned in pJOQ are expressed constitutively.

Construction of T7-EGFP

The EGFP gene was PCR amplified under standard conditions using the plasmid pEGFP-N3 (Clontech) as the template and the oligonucleotides Pr1_Fw and Pr2_Rv as primers. These oligonucleotides, which are shown in online supplementary table S1 (see www.karger.com/doi/10.1159/000448786 for all online suppl. material), amplify the complete gene and introduce the *NcoI* and *XhoI* restriction sites at the 5' and 3' ends, respectively, for cloning purposes. The PCR product was purified using an EZ-10 Spin Column Purification Kit to remove any remaining salts, dNTPs, primers and enzymes from the PCR reaction. The DNA was recovered with 60 µl of water and subjected to overnight digestion at 37°C with the combined restriction enzymes *NcoI/XhoI* using buffer 2 from NEB and BSA. The next day, the digested product was purified by agarose gel using an EZ-10 Spin Column DNA Gel Extraction Kit, and the pure product was immediately ligated into the plasmid pET-28a, which had been previously digested and purified by agarose gel. For the ligation reaction, we mixed 200 ng of digested plasmid, 200 ng of digested PCR product, 1.5 units of T4 DNA ligase (Thermo Fisher Scientific), 2 µl of ligase buffer and water to a total volume of 20 µl. The reaction was incubated for 2 h at 22°C, and then 2 µl of the reaction was directly employed to transform 100 µl of electrocompetent *E. coli* BL21-Gold (DE3) cells by electroporation. An aliquot of the transformants was plated on LB plates supplemented with 25 µg/ml of kanamycin and 0.1 mM of IPTG, and the plates were incubated for 18 h at 37°C. Plasmids were isolated and sequenced from three different fluorescent colo-

nies to confirm the correct amplification and cloning of the gene. This construct was termed T7-EGFP because the recombinant EGFP gene is expressed under the control of the T7 promoter.

Construction of T7-EGFPΔG4

This mutant was constructed as T7-EGFP, replacing the primer Pr1_Fw with the primer Pr3_Fw (online suppl. table S1), which lacks the codon ggc that encodes glycine 4 (G4). The numbering was according to the original avGFP.

Construction of Trc-EGFP, Trc-EGFPΔG4, Trc-rsGFP, Trc-rsGFPΔG4 and Trc-rsEGFPΔG4

These mutants were constructed in a manner similar to T7-EGFP using the pairs of primers and templates shown in online supplementary table S1. However, the different genes were cloned as *NdeI/XhoI* inserts in the plasmid pJOQ, and *E. coli* strain MC1061 cells were transformed with these constructs.

Construction of Trc-rsEGFP

The rsEGFP gene was constructed to encode EGFP from the rsGFP gene. Compared to the original avGFP, rsGFP contains the mutations F64L, S65C, I167T, L231H and K238N, as well as an additional amino acid numbered Ala-1b, whereas EGFP carries the mutations F64L and S65T, and an additional residue numbered Val-1b. Therefore, the conversion of rsGFP to rsEGFP was achieved by site-directed mutagenesis using the overlap extension PCR approach. In the first step, rsGFP was amplified in three segments using the primer pairs Pr13_Fw/Pr14_Rv, Pr15_Fw/Pr16_Rv and Pr17_Fw/Pr18_Rv (shown in online suppl. table S1), which contained the target mutations, and then the expected products (228, 339 and 249 bp, respectively) were independently purified by agarose gel to remove the original template rsGFP. In the fourth PCR reaction, 20 ng of these fragments was mixed and amplified with the outermost primers Pr13_Fw and Pr18_Rv, yielding a unique 750-bp product. This amplicon was cloned and expressed under the control of the Trc promoter in a manner similar to the previously described constructs.

Protein Analysis

E. coli MC1061 cells that were transformed with the Trc-based constructs were grown at 37°C for 18 h in liquid LB supplemented with kanamycin, whereas *E. coli* BL21-Gold (DE3) cells that were transformed with T7-based constructs were grown in a similar manner with IPTG addition when the cultures reached a cellular density of 0.6 OD. After the cultures were incubated, 200 μl of culture was centrifuged at 13,000 rpm for 5 min and the pellet was

resuspended and boiled in 20 μl of denaturing loading buffer for 5 min. Five-microliter aliquots of total extract were analyzed in duplicate gels by SDS-PAGE under denaturing conditions. One gel was analyzed by Coomassie blue staining, whereas the other gel was analyzed by Western blot using alkaline phosphatase-conjugated anti-GFP for specific detection of the GFP bands, which were revealed by the addition of the blue liquid substrate system BCIP®/NBT [Flores-Ramírez et al., 2007].

Kinetic Fluorescence Analysis

E. coli strains transfected with the indicated constructs were grown overnight in liquid kanamycin-supplemented LB broth at 37°C and then diluted 1:100 in fresh LB broth media. These cultures were grown to 0.5 OD and then immediately chilled on ice. Next, 200 μl of culture was added to each well of a black 96-well cell culture plate with a flat clear bottom and a lid. Excitation and emission spectra were measured for 24 h with reads taken every 25 min using a Synergy 2 Multi-Mode Reader BioTek instrument. The fluorescence levels of the cultures were monitored by fixing the excitation at 474 nm and recording the emission at 508 nm. As an indirect estimation of cellular growth, cultures were monitored by their absorbance at 600 nm. The data were collected with Gen5™ Reader Control and Data Analysis Software. Each of our constructs was grown in 12 replicates in a 200-μl total volume, starting with 0.05 OD per well.

Computer Analysis of mRNA Secondary Structure

Secondary structure prediction and analysis were performed with the ViennaRNA Package 2.0 [Lorenz et al., 2011] and the ViennaRNA Web Services [Andronescu et al., 2010; Gruber et al., 2008] using their default value parameters. Our secondary structure analysis considered the leader sequences of the T7 or Trc promoters and 60 bp within the ORF of our different constructs.

Acknowledgments

The authors acknowledge Eugenio López, Jorge Yáñez, Santiago Becerra and Ana Yanci Alarcón for helpful technical assistance with oligonucleotide synthesis and DNA sequencing, as well as Leopoldo Güerica for helpful advice regarding protein analysis by electrophoresis. José Luis Rodríguez-Mejía is a doctoral student from Programa de Doctorado en Ciencias Biomédicas (PDCB) at the Universidad Nacional Autónoma de México (UNAM) and received fellowship 176337 from CONACyT.

References

- Amann E, Ochs B, Abel KJ: Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. *Gene* 1988;69:301–315.
- Andronescu M, Condon A, Hoos HH, Mathews DH, Murphy KP: Computational approaches for RNA energy parameter estimation. *RNA* 2010;16:2304–2318.
- Arpino JA, Reddington SC, Halliwell LM, Rizkallah PJ, Jones DD: Random single amino acid deletion sampling unveils structural tolerance and the benefits of helical registry shift on GFP folding and structure. *Structure* 2014; 22:889–898.
- Care S, Bignon C, Pelissier MC, Blanc E, Canard B, Coutard B: The translation of recombinant proteins in *E. coli* can be improved by in silico generating and screening random libraries of a -70/+96 mRNA region with respect to the translation initiation codon. *Nucleic Acids Res* 2008;36:e6.
- Cormack BP, Valdivia RH, Falkow S: FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* 1996;173:33–38.

- Cramer A, Whitehorn EA, Tate E, Stemmer WP: Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat Biotechnol* 1996;14:315–319.
- de Smit MH, van Duin J: Control of prokaryotic translational initiation by mRNA secondary structure. *Prog Nucleic Acid Res Mol Biol* 1990a;38:1–35.
- de Smit MH, van Duin J: Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. *Proc Natl Acad Sci USA* 1990b;87:7668–7672.
- Delagrave S, Hawtin RE, Silva CM, Yang MM, Youvan DC: Red-shifted excitation mutants of the green fluorescent protein. *Biotechnology (NY)* 1995;13:151–154.
- Dubendorff JW, Studier FW: Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *J Mol Biol* 1991;219:45–59.
- Flores-Ramírez G, Rivera M, Morales-Pablos A, Osuna J, Soberón X, Gaytán P: The effect of amino acid deletions and substitutions in the longest loop of GFP. *BMC Chem Biol* 2007;7:1.
- Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL: The Vienna RNA websuite. *Nucleic Acids Res* 2008;36:W70–W74.
- Gu W, Zhou T, Wilke CO: A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. *PLoS Comput Biol* 2010;6:e1000664.
- Hall MN, Gabay J, Débarbouillé M, Schwartz M: A role for mRNA secondary structure in the control of translation initiation. *Nature* 1982;295:616–618.
- Heim R, Cubitt B, Tsien RY: Improved green fluorescence. *Nature* 1995;373:663–664.
- Heim R, Prasher DC, Tsien RY: Wavelength mutations and posttranslational autooxidation of green fluorescent protein. *Proc Natl Acad Sci USA* 1994;91:12501–12504.
- Jones DD: Triplet nucleotide removal at random positions in a target gene: the tolerance of TEM-1 β -lactamase to an amino acid deletion. *Nucleic Acids Res* 2005;33:1–8.
- Knight JA, Hardy LW, Rennell D, Herrick D, Poeteete AR: Mutations in an upstream regulatory sequence that increase expression of the bacteriophage T4 lysozyme gene. *J Bacteriol* 1987;169:4630–4636.
- Komarova AV, Tchufistova LS, Dreyfus M, Boni IV: AU-rich sequences within 5' untranslated leaders enhance translation and stabilize mRNA in *Escherichia coli*. *J Bacteriol* 2005;187:1344–1349.
- Kudla G, Murray AW, Tollervey D, Plotkin JB: Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* 2009;324:255–258.
- Li X, Zhang G, Ngo N, Zhao X, Kain SR, Huang CC: Deletions of the *Aequorea victoria* green fluorescent protein define the minimal domain required for fluorescence. *J Biol Chem* 1997;272:28545–28549.
- Liu S, Wei X, Dong X, Xu L, Liu J, Jiang B: Structural plasticity of green fluorescent protein to amino acid deletions and fluorescence rescue by folding-enhancing mutations. *BMC Biochem* 2015;16:17.
- Lorenz R, Bernhart SH, Höner zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL: ViennaRNA Package 2.0. *Algorithms Mol Biol* 2011;6:26.
- Mahalik S, Sharma AK, Mukherjee KJ: Genome engineering for improved recombinant protein expression in *Escherichia coli*. *Microb Cell Fact* 2014;13:177.
- Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH: Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci USA* 2004;101:7287–7292.
- Merril CR, Gottesman ME, Adhya SL: *Escherichia coli* gal operon proteins made after prophage lambda induction. *J Bacteriol* 1981;147:875–887.
- Murakami H, Hohsaka T, Sisido M: Random insertion and deletion of arbitrary number of bases for codon-based random mutation of DNAs. *Nat Biotechnol* 2002;20:76–81.
- Osterman IA, Evfratov SA, Sergiev PV, Dontsova OA: Comparison of mRNA features affecting translation initiation and reinitiation. *Nucleic Acids Res* 2013;41:474–486.
- Osuna J, Yáñez J, Soberón X, Gaytán P: Protein evolution by codon-based random deletions. *Nucleic Acids Res* 2004;32:e136.
- Prasher DC, Eckenrode VK, Ward WW, Prendergast FG, Cormier MJ: Primary structure of the *Aequorea victoria* green-fluorescent protein. *Gene* 1992;111:229–233.
- Raghunathan G, Soundarajan N, Sokalingam S, Yun H, Lee SG: Deletional protein engineering based on stable fold. *PLoS One* 2012;7:e51510.
- Seo SW, Yang JS, Kim I, Yang J, Min BE, Kim S, Jung GY: Predictive design of mRNA translation initiation region to control prokaryotic translation efficiency. *Metab Eng* 2013;15:67–74.
- Shaner NC, Patterson GH, Davidson MW: Advances in fluorescent protein technology. *J Cell Sci* 2007;120:4247–4260.
- Studer SM, Joseph S: Unfolding of mRNA secondary structure by the bacterial translation initiation complex. *Mol Cell* 2006;22:105–115.
- Torres Tejerizo G, Banuelos LA, Cervantes L, Gaytán P, Pistorio M, Romero D, Brom S: Development of molecular tools to monitor conjugative transfer in rhizobia. *J Microbiol Methods* 2015;117:155–163.
- Zimmer M: Green fluorescent protein (GFP): applications, structure, and related photophysical behavior. *Chem Rev* 2002;102:759–781.

13. Tabla suplementaria del manuscrito

Promoter-Construction	Template	Primers (Forward and Reverse)	PCR fragment size	Cloning vector (Restriction Sites)
E7-EGFP	pEGFP-N3	Pr1_Fw: aacagcccatggtagcaaggcggaggagctg Pr2_Rv: ctctctcaggaactgtacagctcgtccatgccag	740 bp	pET28a (NcoI/XhoI)
E7-EGFPΔG4	pEGFP-N3	Pr3_Fw: aacagcccatggtagcaaggcggaggagctgtcaecggg Pr4_Rv: ctctctcaggaactgtacagctcgtccatgccag	737 bp	pET28a (NcoI/XhoI)
Tre-EGFP	pEGFP-N3	Pr5_Fw: cacaggaacacagcatatgtagcaaggcggaggagctg Pr6_Rv: ctctctcaggaactgtacagctcgtccatgccag	747 bp	pJOQ (NdeI/XhoI)
Tre-EGFPΔG4	pEGFP-N3	Pr7_Fw: cacaggaacacagcatatgtagcaaggcggaggagctgtcaecggg Pr8_Rv: ctctctcaggaactgtacagctcgtccatgccag	744 bp	pJOQ (NdeI/XhoI)
Tre-rsGFP	pQB125-f	Pr9_Fw: cacaggaacacagcatatgtagcaaggcggaggagctg Pr10_Rv: cgggctctcaggtatcaggttacaggtcaccatgccatg	750 bp	pJOQ (NdeI/XhoI)
Tre-rsGFPΔG4	pQB125-f	Pr11_Fw: cacaggaacacagcatatgtagcaaggcggaggagctg Pr12_Rv: cgggctctcaggtatcaggttacaggtcaccatgccatg	747 bp	pJOQ (NdeI/XhoI)
Tre-EGFPecol	pQB125-f	Pr13_Fw: cctttccgcttcataatgtagcaaggcggaggagctg Pr14_Rv: gcatggaacacataggtcagagtagtgaclag Pr15_Fw: ctatgcaactctgacctatgggttcaatgc Pr16_Rv: ttcatgtgtggcgaactcaggttcaactt Pr17_Fw: aaagtgaactcaagatccacaaacatgaa Pr18_Rv: cgggctctcaggtatcaggttacaggtcaccatgccatg	228 bp 339 bp 249 bp (750 bp)	pJOQ (NdeI/XhoI)
Tre-EGFPecolΔC4	Tre-EGFPecol	Pr19_Fw: ctccaggaacacagcatatgtagcaaggcggaggagctg Pr20_Rv: cgggctctcaggtatcaggttacaggtcaccatgccatg	747 bp	pJOQ (NdeI/XhoI)

Pr2_Rv = Pr4_Rv = Pr6_Rv = Pr8_Rv

Pr10_Rv = Pr12_Rv

Pr18_Rv = Pr20_Rv

14. Bibliografía

- Amann, E., Ochs, B. & Abel, K.J., 1988. Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. *Gene*, 69, pp.301–315.
- Arpino, J. a J. et al., 2014. Random single amino acid deletion sampling unveils structural tolerance and the benefits of helical registry shift on GFP folding and structure. *Structure*, 22(6), pp.889–898.
- Berg, L. et al., 2012. Exploring the 5'-UTR DNA region as a target for optimizing recombinant gene expression from the strong and inducible P_m promoter in *Escherichia coli*. *Journal of Biotechnology*, 158(4), pp.224–230.
- Campbell, R.E. et al., 2002. A monomeric red fluorescent protein. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12), pp.7877–82.
- Cormack, B.P., Valdivia, R.H. & Falkow, S., 1996. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene*, 173, pp.33–38.
- Flores-Ramírez, G. et al., 2007. The effect of amino acid deletions and substitutions in the longest loop of GFP. *BMC chemical biology*, 7, p.1.
- Gruber, A.R. et al., 2008. The Vienna RNA websuite. *Nucleic acids research*, 36(April), pp.70–74.
- Gu, W., Zhou, T. & Wilke, C.O., 2010. A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. *PLoS Computational Biology*, 6(2), pp.1–8.
- Ho, S.N. et al., 1989. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*, 77(1), pp.51–59.
- Hofacker, I.L., 2014. Chapter 4: Energy-Directed RNA structure Prediction. In *RNA Sequence, Structure, and Function: Computational and Bioinformatic Methods, Methods in Molecular Biology*. pp. 71–84.
- Hofacker, I.L. & Lorenz, R., 2014. Chapter 1 Predicting RNA Structure : Advances and Limitations. In *RNA Folding: Methods and Protocols, Methods in Molecular Biology*. pp. 1–19.
- Jones, D.D., 2005. Triplet nucleotide removal at random positions in a target gene: The tolerance of TEM-1 β -lactamase to an amino acid deletion. *Nucleic Acids Research*, 33(9), pp.1–8.
- Katz, L. & Burge, C.B., 2003. Widespread Selection for Local RNA Secondary Structure in Coding Regions of Bacterial Genes. *Genome Research*, (13), pp.2042–2051.
- Kozak, M., 2005. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene*, 361, pp.13–37.
- Kudla, G. et al., 2009. Coding-Sequence Determinants of Gene Expression in *Escherichia coli*. *Science*, 324(April), pp.255–258.
- Levin-Karp, A. et al., 2013. Quantifying translational coupling in *E. coli* synthetic operons using RBS modulation and fluorescent reporters. *ACS Synthetic Biology*, 2(6), pp.327–336.
- Li, G.W., 2015. How do bacteria tune translation efficiency? *Current Opinion in Microbiology*, 24, pp.66–71.

- Li, X. et al., 1997. Deletions of the *Aequorea victoria* green fluorescent protein define the minimal domain required for fluorescence. *Journal of Biological Chemistry*, 272(November), pp.28545–28549.
- Lorenz, R. et al., 2011. ViennaRNA Package 2.0. *Algorithms for Molecular Biology*, 6, p.26.
- Mahalik, S., Sharma, A.K. & Mukherjee, K.J., 2014. Genome engineering for improved recombinant protein expression in *Escherichia coli*. *Microbial Cell Factories*, 13(1), p.177.
- Marzi, S. et al., 2007. Structured mRNAs Regulate Translation Initiation by Binding to the Platform of the Ribosome. *Cell*, 130, pp.1019–1031.
- Murakami, H., Hoshika, T. & Sisido, M., 2002. Random insertion and deletion of arbitrary number of bases for codon-based random mutation of DNAs. *Nature biotechnology*, 20(January), pp.76–81.
- Osuna, J. et al., 2004. Protein evolution by codon-based random deletions. *Nucleic acids research*, 32(17), p.e136.
- Pédelacq, J.-D. et al., 2006. Engineering and characterization of a superfolder green fluorescent protein. *Nature biotechnology*, 24(1), pp.79–88.
- Rasband, W.S., 2016. ImageJ.
- Rekas, A. et al., 2002. Crystal structure of venus, a yellow fluorescent protein with improved maturation and reduced environmental sensitivity. *Journal of Biological Chemistry*, 277, pp.50573–50578.
- Rose, A.S. et al., 2016. Web-based molecular graphics for large complexes. In *Proceedings of the 21st International Conference on Web3D Technology - Web3D '16*. New York, New York, USA: ACM Press, pp. 185–186.
- Schulz, V.P. & Reznikoff, W.S., 1990. In vitro secondary structure analysis of mRNA from lacZ translation initiation mutants. *Journal of molecular biology*, 211, pp.427–445.
- Seo, S.W. et al., 2014. Predictive combinatorial design of mRNA translation initiation regions for systematic optimization of gene expression levels. *Scientific reports*, 4, p.4515.
- Shaner, N.C. et al., 2004. Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nature biotechnology*, 22(12), pp.1567–1572.
- Shaner, N.C., Patterson, G.H. & Davidson, M.W., 2007. Advances in fluorescent protein technology. *Journal of cell science*, 120, pp.4247–4260.
- Sharma, R.C. & Schimke, R.T., 1996. Preparation of electro-competent *E. coli* using salt-free growth medium. *BioTechniques*, 20(1), pp.42–44.
- Shine, J. & Dalgarno, L., 1975. Terminal-Sequence Analysis of Bacterial Ribosomal RNA: Correlation between the 3' Terminal-Polypyrimidine Sequence of 16-S RNA and Translational Specificity of the Ribosome. *European Journal of Biochemistry*, 57(1), pp.221–230.
- Siemering, K.R. et al., 1996. Mutations that suppress the thermosensitivity of green fluorescent protein. *Current biology : CB*, 6, pp.1653–1663.
- de Smit, M.H. & van Duin, J., 2003. Translational standby sites: How ribosomes may deal with the rapid folding kinetics of mRNA. *Journal of Molecular Biology*, 331(3), pp.737–743.
- Steitz, J. & Jakes, K., 1975. How ribosomes select initiator regions in mRNA: base

- pair formation between the 3' terminus of 16S rRNA and the mRNA during initiation of protein synthesis in *Escherichia coli*. *Pnas*, 72(12), pp.4734–4738.
- Studer, S.M. & Joseph, S., 2006. Unfolding of mRNA Secondary Structure by the Bacterial Translation Initiation Complex. *Molecular Cell*, 22, pp.105–115.
- Tsien, R.Y., 1998. The green fluorescent protein. *Annual review of biochemistry*, 67, pp.509–544.
- Yang, F. et al., 1996. The molecular structure of green fluorescent protein. *Nat. Biotechnol.*, 14(10), pp.1246–1251.
- Zimmer, M., 2002. Green fluorescent protein (GFP): Applications, structure, and related photophysical behavior. *Chemical Reviews*, 102, pp.759–781.