



Memoria del XLVIII Taller de Actualización Bioquímica, Facultad de Medicina; UNAM

Expresión de proteínas recombinantes en un sistema heterólogo.

Expression of recombinant proteins in a heterologous system.

Hernández-Alcántara, Gloria^{1*}; García-Torres, Itzhel²; Alba-Martínez, Zoe¹ y Ramírez-Silva, Leticia¹.

1. Laboratorio de péptidos y proteínas, Departamento de Bioquímica, Facultad de Medicina, UNAM.
2. Grupo de estudio en Biomoléculas y Salud Infantil, Laboratorio de Errores Innatos del Metabolismo y Tamiz. Instituto Nacional de Pediatría. Secretaría de Salud.

*Correspondencia. Torre de Investigación, Facultad de Medicina, Circuito Escolar, Ciudad Universitaria, Coyoacan, CDMX, México. CP. 04510 Tel. +52 (55) 56 23 21 33, ghernandez@bq.unam.mx

Resumen

Los sistemas de expresión en organismos procariontes y eucariontes se utilizan comúnmente para la sobreexpresión de proteínas recombinantes solubles. La elección de un sistema de producción adecuado es la base para lograr un buen rendimiento en la purificación de la proteína recombinante de interés. Aspectos como: la optimización del uso de codones, la elección de la cepa hospedera, el vector de expresión, la etiqueta o proteína de fusión a utilizar para la purificación, la composición del medio de cultivo, las condiciones de inducción, entre otros factores, son la clave para el resultado final. El nivel de expresión del gen blanco, así como los procesamientos post-traduccionales de la proteína recombinante sintetizada son algunos de los principales retos a los que se enfrentan la mayoría de los estudios realizados. Uno de los sistemas de sobreexpresión más utilizados sin duda es el procarionte, en los que se incluyen principalmente a *Escherichia coli*. Esto debido a la facilidad de su manipulación genética, crecimiento celular, gran cantidad de biomasa, bajo costo, y la gran cantidad de cepas hospederas disponibles. De tal forma, que *E. coli* ha superado la barrera de realizar algunas modificaciones post-traduccionales que requieren algunas proteínas de origen eucariota para su

Abstract

The expression systems in prokaryotic and eukaryotic organisms are commonly used for the overexpression of soluble recombinant proteins. The choice of suitable production system is the basis for achieving good purification yields of the recombinant protein of interest. Aspects such as: optimization of the target gene, the choice of host strain, the expression vector, the tag or fusion proteins required for purification, the composition of the cellular growth media, the induction conditions, and other factors are the key to the final result. The expression level of the target gene, as well as the post-translational processing of the recombinant protein are some of the main challenges faced by most of the studies carried out. One of the most widely used overexpression system is the prokaryote, which mainly includes *Escherichia coli*. The reasons attributed to the popularity of *E. coli* is due to the ease genetic manipulation, cell growth, large amount of biomass, low cost, and the large number of hosts strains available. Thus, *E. coli* has overcome the barrier of carrying out some post-translational modifications that some proteins of eukaryotic origin require for their production in this system. This work, discusses some strategies that have helped to obtain active and functional recombinant proteins in

producción en este sistema. En este trabajo se discuten algunas estrategias que han ayudado a la obtención de proteínas recombinantes activas y funcionales en un sistema comercial de sobreexpresión como el BL21 y así mismo, se sugieren algunos protocolos para determinar las mejores condiciones de sobreexpresión de las proteínas recombinantes de interés.

Palabras claves: Plásmidos de sobreexpresión, sistema de expresión BL21, expresión recombinante, proteína recombinante soluble, cuerpos de inclusión..

a commercial overexpression system such as BL21 and also suggests some protocols to determine the best condition for overexpression of the recombinant proteins of interest.

Keywords: overexpression plasmids, BL21 expression system, recombinant expression, soluble recombinant protein, inclusion bodies.

Introducción

Las proteínas recombinantes (PR), son biomoléculas que generalmente se expresan en microorganismos fácilmente manipulables diferentes al de su origen (heterólogos), con el fin de producirlas en grandes cantidades. Existen numerosos sistemas heterólogos para la expresión de PR tanto de importancia científica, biológica y/o farmacéutica, la bacteria *E. coli* es uno de los sistemas más populares que se ha utilizado durante décadas y al que constantemente se le siguen realizando modificaciones genéticas para su mayor uso [1-5]. Éste y otros sistemas de expresión, tanto en células procariotas como eucariotas, tienen la función de maximizar la expresión de la proteína de interés a partir de un gen clonado en un plásmido o vector de expresión [6].

Uno de los objetivos principales de cualquier experimento de sobreexpresión de una PR, es combinar: *a)* una producción máxima de proteína soluble por célula, *b)* con una mayor densidad celular para obtener de la manera más eficiente altos rendimientos de la PR de interés. De tal forma que, casi siempre se deben optimizar una gran cantidad de parámetros para lograr el objetivo [7]. Algunos de los parámetros que se han descrito, que tienen un fuerte efecto tanto en la productividad por célula como en la densidad celular son los medios de cultivos (nutrientes), la oxigenación, las condiciones de inducción, el tipo de promotor del plásmido de clonación, la optimización del uso de codones, la elección del hospedero, entre otros. Sin embargo, se ha visto que para cada proteína (aún en genes homólogos), es necesario optimizar cada uno de estos parámetros. Por ejemplo [8-11]:

a) En algunos casos la alta densidad celular lleva a la obtención de un mayor rendimiento de proteína, pero se puede producir de forma insoluble, como

producto de una rápida producción y a una gran cantidad de proteína producida. Lo que promueve que la PR no se pliegue correctamente (formando cuerpos de inclusión).

- b)* En algunos casos la proteína se pliega correctamente pero el rendimiento es muy bajo.
- c)* La densidad celular es adecuada pero no se expresa la PR.
- d)* En el menor de los casos la proteína puede ser tóxica por lo cual el crecimiento celular se ve disminuido.

La mayoría de las veces para lograr la expresión óptima de una PR, se usa un proceso de prueba y error, y muy pocas veces se utilizan enfoques estadísticos para evaluar las variables que tienen la mayor influencia en la producción de una PR. En cuanto al proceso de prueba y error (experimental), es una selección basada en el conocimiento, que utiliza lo publicado en la literatura o en investigaciones propias de cada tema de investigación. Este proceso, implica la optimización de una variable a la vez; esto es variar un factor en cada experimento manteniendo el resto constante. Sin embargo, a pesar de que se pueden requerir muchos experimentos y mucho tiempo, es la estrategia mayormente utilizada en los laboratorios de investigación.

Por otra parte, los enfoques estadísticos que incluyen métodos computacionales requieren de una optimización *in silico*, en donde se optimizan: factores basados en la secuencia de aminoácidos de la proteína y en la secuencia del gen, estabilidad del RNAm, así como la selección del vector y la cepa hospedera. A pesar de que los enfoques estadísticos han demostrado mejorar los niveles de expresión en las PRs no son muy utilizados, quizá debido a la poca disponibilidad

de los programas computacionales que se utilizan [12-14].

A continuación, se describen algunos de los componentes más importantes que se requieren para la obtención de una PR.

Obtención de una proteína recombinante

Obtención de un constructo en un plásmido recombinante

La construcción del plásmido o vector recombinante que contenga el gen de interés (constructo) para el proceso de traducción de una PR es el primer paso a seguir. El constructo se puede obtener de dos maneras: a) mediante amplificación y clonación del gen de interés y b) diseño y síntesis del gen de interés en una compañía comercial (Fig. 1). Ambos caminos son ampliamente utilizados, aunque en algunos casos la elección dependerá de la disponibilidad del material genético que servirá como molde; de lo contrario la síntesis del gen es la mejor elección.

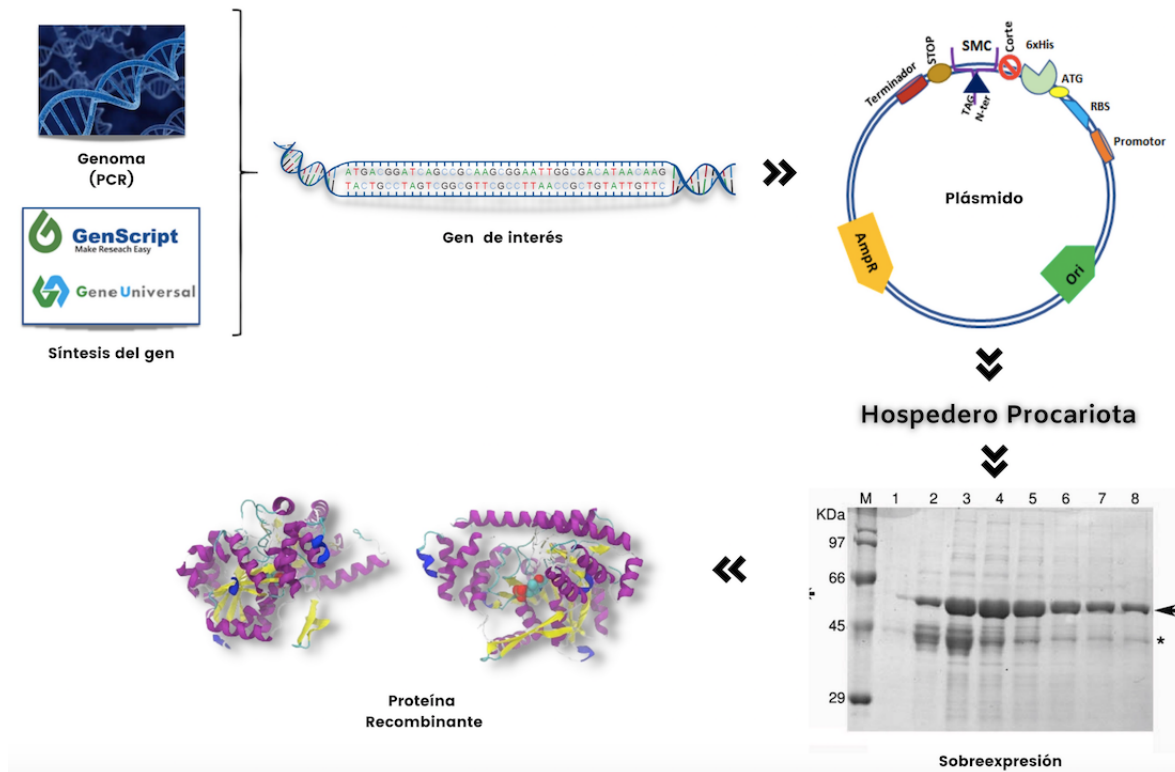


Figura 1. Esquema general para la obtención de un DNA recombinante para la sobreexpresión de una proteína.

a) Amplificación y clonación

La amplificación del gen de interés generalmente se realiza por la técnica de PCR o RT-PCR (que se detallan en esta misma revista en el tema de Ingeniería de proteínas y PCR en Tiempo Real) y su posterior clonación en el plásmido de elección [15]. La amplificación del gen incluye una búsqueda de la secuencia del gen de interés en alguna base de datos como el NCBI (<https://www.ncbi.nlm.nih.gov>) o KEEG (<https://www.genome.jp/kegg/>), y el análisis del marco de lectura abierto. Con base en la secuencia del gen reportada, se diseñan y sintetizan los oligonucleótidos que flanquean los extremos 5' y 3' del gen (esta técnica se detalla en esta misma revista en síntesis química de oligonucleótidos: método de

fosforamiditas). Cabe mencionar que la secuencia de los oligonucleótidos pueden incluir los sitios de corte para las enzimas de restricción que flanquean los extremos 5' y 3'. Una vez que se tiene, el material genético purificado y los oligonucleótidos sintetizados se realiza la amplificación del gen. Posteriormente se realiza la clonación, está puede realizarse utilizando los sitios para enzimas de restricción que se encuentran en el Sitio Múltiple de Clonación (SMC) del plásmido y su posterior ligación con una enzima T4 DNA ligasa o bien utilizando la técnica de Ligación Independiente de Clonación (LIG) [16]. Esto dependerá del plásmido de elección. Finalmente, al obtener un constructo éste se debe secuenciar para corroborar la secuencia de nucleótidos (esta técnica se detalla en esta misma revista en Secuenciación de

ADN por el método de terminación de la cadena de Sanger).

b) Diseño y síntesis del gen

[<https://www.genscript.com>,
<https://www.geneuniversal.com>].

Si se opta por la síntesis comercial del gen, primeramente se debe hacer una búsqueda de la secuencia del gen en alguna base de datos como NCBI (<https://www.ncbi.nlm.nih.gov>) o KEEG (<https://www.genome.jp/kegg/>). Se analiza el marco de lectura abierto y se diseñan e incluyen sitios de restricción en los extremos 5' y 3' del gen para su clonación.

En este punto se puede hacer una optimización del uso de codones para el sistema de expresión que se utilizará (<https://www.idtdna.com/pages/tools/codon-optimization-tool>), o si se prefiere las compañías encargadas de la síntesis pueden hacerla. También existe la opción de conservar la secuencia de nucleótidos tal y como está reportada del genoma. Además, se le pueden hacer mutaciones silenciosas para eliminar algún sitio de corte para alguna enzima de restricción que interfiera con las del SMC. Generalmente, las compañías de síntesis clonan el gen en un plásmido de alta copia como el pUC.

Actualmente estas compañías tienen la opción de que sea clonado en alguno de los plásmidos de expresión, tanto para sistemas procariotas como eucariotas. Adicionalmente, se puede solicitar que el constructo sea transformado en la cepa de expresión en donde se desea expresar la PR.

Elección del sistema de expresión

La elección del sistema de expresión está conformada principalmente por el organismo hospedero (cepa huésped) y el vector de expresión. De tal forma que ambos componentes deben ser compatibles con la maquinaria celular requerida para llevar a cabo dicho proceso. Esta maquinaria consiste básicamente en que el plásmido de expresión contenga los elementos génicos necesarios para llevar a cabo los procesos de replicación, transcripción y traducción del gen de interés en la cepa hospedera [17]. La elección del sistema de expresión definirá los requerimientos en cuanto a infraestructura, disponibilidad de reactivos y materiales que se requerirán en el proyecto de investigación. Existen sistemas de expresión con hospederos procariotas o eucariotas, entre los cuales se encuentran bacterias, levaduras, hongos filamentosos, algas unicelulares, plantas, insectos, así

como organismos transgénicos. La selección del hospedero dependerá del origen de la proteína a expresar, ya que algunas proteínas requieren de modificaciones post-traduccionales o bien requieren ser dirigidas a diferentes compartimentos celulares en donde llevan a cabo su función (Fig. 2). Dentro del sistema procariota, *E. coli* constituye sin duda uno de los hospederos que más reportan su uso, tanto en la investigación científica como a nivel industrial [18]. Incluso algunas cepas de *E. coli* se han modificado genéticamente para poder llevar a cabo algunas modificaciones post-traduccionales. Dentro del sistema eucariota entre los modelos más utilizados se encuentran *Pichia pastoris* y *Saccharomyces cerevisiae*, en los cuales también se han expresado cientos de proteínas (Fig. 2). Estos sistemas tienen la ventaja de realizar modificaciones post-traduccionales, necesarios para la mayoría de proteínas eucariotas [19-21].

Como se mencionó, en *E. coli* existen algunas cepas que en combinación con ciertos plásmidos pueden realizar algunas modificaciones post-traduccionales, lo que ha permitido expresar proteínas de organismo eucariotas con éxito [20]. Uno de los sistemas procariotas mayormente utilizados es el binomio pET- BL21 [18]. El sistema BL21(DE3) es un sistema que contiene el gen que codifica para la RNA polimerasa del bacteriófago T7, bajo el control de un promotor inducible por IPTG. La T7 RNA polimerasa es una enzima muy activa, que sintetiza RNA a una velocidad varias veces superior a la de la RNA polimerasa de *E. coli* [22]. Aunado a la cepa hospedera se encuentra el promotor uno de los componentes más importantes de los plásmidos. Éste juega un papel muy importante en el inicio de la transcripción de los genes clonados [22, 23], por lo que un promotor debe ser lo suficientemente fuerte para permitir la acumulación de hasta el 50% de las proteínas celulares totales, y debe estar regulado para prevenir la toxicidad del producto. Los promotores mayormente utilizados provienen principalmente de bacterias: Lac, tac, trp, araBAD o de bacteriófagos: T7, T5, SP6. Uno de los promotores más utilizados en los sistemas de expresión en *E. coli* es el T7/lac, presente en una gran cantidad de plásmidos, entre los más utilizados se encuentran los de la serie pET (Fig. 3, Tabla I) [24]. De ahí que el binomio pET-BL21 sea uno de los sistemas más populares a nivel de investigación.

Es importante mencionar que la selección del plásmido de expresión permite incluir etiquetas y/o proteínas de fusión que favorecen la sobreexpresión de la proteína de forma soluble y que además permiten la purificación de la proteína de interés mediante

cromatografía de afinidad. Así mismo, los plásmidos contienen secuencias de reconocimiento para diferentes proteasas que permiten remover las

etiquetas o proteínas de fusión de la proteína de interés posterior a su purificación (Tabla I).

Tipo de Proteína Recombinante	Sistema de Sobreexpresión
Proteínas sin N-glicosilaciones de organismos procariotas, termófilos y eucariotas	<i>E. coli</i> y levadura (<i>Saccharomyces cerevisiae</i>)
Proteínas con N-glicosilaciones	Células de mamífero (CHO), <i>Pichia pastoris</i> , <i>Aspergillus niger</i>
Proteínas de secreción, expresión homóloga, expresión de proteasas	<i>Bacillus megaterium</i> , <i>B. subtilis</i> , <i>B. licheniformis</i> , <i>B. brevis</i>
Proteínas glucosiladas y con enlaces disulfuro	<i>Pichia pastoris</i> (levadura metilotrófica)
Proteínas de mamífero con diversas modificaciones post-traduccionales como fosforilaciones, N y O-glicosilaciones, acilación, palmitoilación, carboximetilación, prenilación, formación de puentes disulfuro	Células de insecto (<i>Spodoptera frugiperda</i>), utilizando como vector baculovirus

Figura 2. Sistemas de sobreexpresión de proteínas recobinantes. Se presentan algunos de los diferentes organismos que se utilizan para sobreexpresar una PR de acuerdo con el tipo de requerimiento de la proteína de estudio. CHO (células de ovario de hámster Chino)

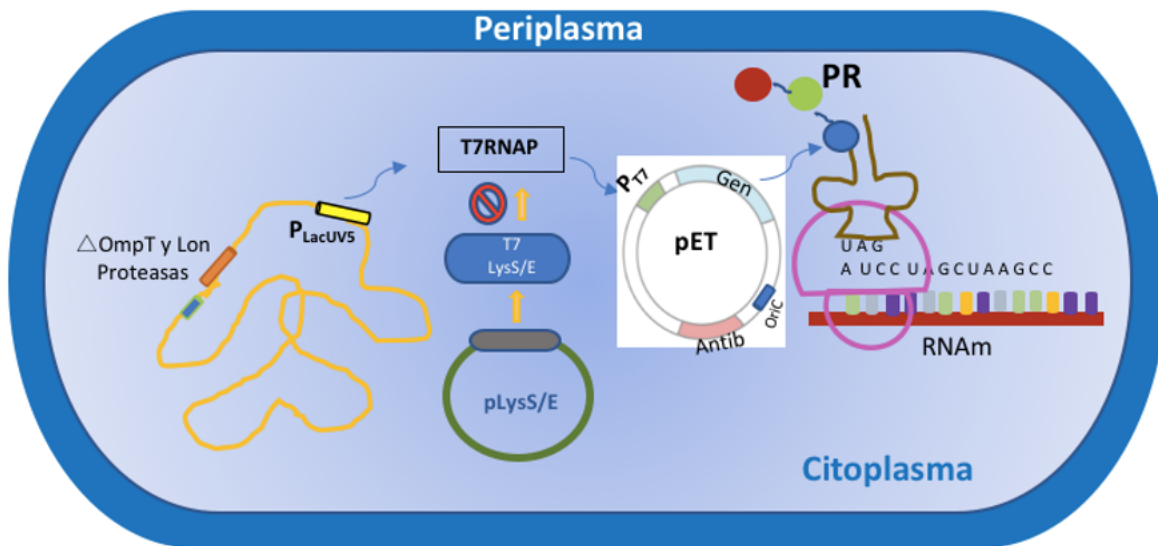


Figura 3. Esquema general del proceso de expresión, representado en la cepa BL21DE3pLysS o pLysE. Por lo general las cepas BL21(DE3) son deficientes de proteasas OmpT y Lon, para disminuir la degradación proteolítica de la PR. (DE3) indica que la cepa hospedera contiene el lisógeno λDE3 y por lo tanto porta una copia cromosómica del gen de la T7 RNA polimerasa, bajo el control del promotor lacUV5. El gen de interés se encuentra bajo el control de un promotor, éste se transcribe a RNAm y se traduce la PR.

Tabla I. Ejemplo de algunos plásmidos de la serie pET compatibles con las cepas BL21

Plásmido o vector	Algunas características del plásmido y de la PR obtenida
pET-31b(+)	Producción con alto rendimiento para péptidos y proteínas pequeñas fusionadas a la proteína cetosteroido isomerasa (KSI). Un sitio único de clonación para AlwN I, que permite la inserción unidireccional de regiones codificantes de péptidos repetidos en tándem.
pET-32a(+), pET-32b(+), pET-32c(+)	Producción de proteínas solubles. Contiene 6xHis, TrxA, S-Tag, sitio para enterocinasa y trombina.
pET-33b(+)	Producción de proteínas apropiadas para el marcaje con P ³² sitio específico. Proteína cinasa A (PKA), sitio para trombina, 6xHis.
pET-39b(+), pET-40b(+)	Contiene una etiqueta Dsb para exportación y plegamiento periplásmico de la proteína de interés, 8xHis y 6xHis. Sitio para enterocinasa y trombina.
pET-41a(+), pET-41b(+), pET-41c(+), pET-42a(+), pET-42b(+), pET-42c(+)	Contiene una etiqueta de fusión GST para mejorar la producción y solubilidad. 6xHis, S-Tag, sitio para enterocinasa y trombina.
pET-43.1a(+), pET-43.1b(+), pET-43.1c(+)	Diseñado para la clonación y alta expresión de proteínas fusionadas con la proteína NusA de 495 aa (Nus-Tag) que estabiliza la PR durante la traducción, 6xHis, sitio para enterocinasa y trombina.
pET-44a(+), pET-44b(+), pET-44c(+)	Además de secuencia Nus-Tag o HSV-Tag (virus herpes simple), 6xHis A y C terminal, S-Tag, sitio para enterocinas y trombina.
pET-45b(+)	Contiene secuencias 6xHis, S-Tag y un SCM con otras enzimas de restricción diferentes a los otros pET. Sitio para enterocinasa.
pET-46 EK/LIC	Contiene un sitio de clonación independiente de ligación (LIC) y 6xHis.
pET-47b(+), pET-48b(+), pET-49b(+), pET-50b(+)	Contienen un sitio de corte para la proteasa HRV 3C (proteasa 3C del rinovirus humano), para la eliminación de la etiqueta de fusión, 6xHis, sitio para trombina.

Etiquetas de fusión utilizadas para la expresión y purificación de proteínas solubles

Proteínas de fusión Tag	Potencian el plegamiento, promoviendo la adquisición de la estructura correcta: -MBP: Proteína de unión a maltosa. -GST: Glutation-S-transferasa. -TRXA: Tiorredoxina A. -NuSa: Terminación de la transcripción/Proteína de antiterminación.	
Poli Tag para la purificación	6x His-tag 6xArg-Tag	Para purificación en un sólo paso, por columna de afinidad a níquel e intercambio catiónico SP-sephadex
Eliminación de la Tag	Proteasas -Enterocinasa -Trombina -Factor Xa -TEV (proteasa del virus del tabaco) -HRV 3C (rinovirus humano)	Eliminan la etiqueta después de la purificación. La proteasa dependerá del plásmido que lleve este sitio de corte.

Ejemplos de algunas cepas del sistema BL21 utilizadas en la expresión de la PR

Las cepas de *E. coli* BL21(DE3) y sus derivados se han modificado genéticamente con el paso de las décadas, para producir grandes cantidades de proteínas solubles y funcionales [22, 23, 25-28]. En su mayoría las diferentes cepas de BL21(DE3) son deficientes en las proteasas Lon (citoplásmicas) y OmpT (exterior de la membrana), tienen baja producción de acetato en altas concentraciones de glucosa y tienen mejorada la permeabilidad. Algunas de estas cepas se ilustran en la Tabla II.

Estandarización de las condiciones óptimas de sobreexpresión

Lo descrito anteriormente permite la obtención de un constructo que lleve a la obtención de una PR. Con la finalidad de obtener la proteína de interés soluble, prevenir la acumulación de la proteína en cuerpos de inclusión y tener buenos rendimientos de purificación, se recomienda estandarizar las condiciones de sobreexpresión de la PR. Por lo general, la expresión de la PR induce una carga metabólica en la cepa hospedera por lo que puede haber acumulación de proteínas en agregados insolubles o hasta efectos de

toxicidad. En la Tabla III, se resumen algunos de los factores que deben optimizarse.

Tabla II. Características de algunas cepas de sobreexpresión del sistema BL21

Cepa	Característica genómica	Usos
BL21StartDE3	Mutación en RNasaE	Reduce la degradación del RNAm, no recomendada para genes tóxicos
BL21trxBDE3	Mutación en la tiorredoxina reductasa. Resistencia a kanamicina	Promueve la formación de enlaces disulfuro, mejora el plegamiento
RossetaDE3	tRNAs para AGG (R), AGA (G), AUA (I), CUA (L), CCC (P), GGA (G). En plásmidos resistentes a cloranfenicol	Expresión de proteínas eucariotas que contienen codones raros
Lemo21DE3	lysY, inhibidor de la T7RNA polimerasa	Permite la expresión variando el nivel de lysY, se modula añadiendo L-ramnosa. Produce proteínas solubles y plegadas
BL21AI BL21SI	La T7 RNA polimerasa se encuentra bajo el control del promotor: AI = araBAD (Inductor arabinosa). SI = proU (Inductor sales -NaCl-)	Regulación estricta de la expresión, ideal para obtener proteínas tóxicas
BLRDE3	Deficiente de RecA	Ampliamente recomendada para plásmidos con secuencias repetitivas. Expresión de proteínas inestables, tóxicas o insolubles.
C41DE3 C43DE3	Poseen una mutación (C41) o dos mutaciones (C43), que confiere tolerancia a proteínas tóxicas.	Proteínas de membrana y/o tóxicas, tanto procariontas como eucariotas (insectos, plantas, levaduras, mamíferos). Evitan la muerte celular por proteínas tóxicas
TKB1	Contiene un plásmido con un gen para tirosina cinasa (<i>elk</i>)	Genera proteínas fosforiladas
T7 express	No restringe el DNA metilado	Expresión de proteínas de mamífero tóxicas en <i>E. coli</i>

Problemas comunes y soluciones al seleccionar el sistema de sobreexpresión

Problema	Posible explicación	Soluciones
Proteína no se sobreexpresa o tiene baja sobreexpresión	Tóxica antes de la inducción	Control basal: usar cepas pLysS/pLysE, promotores con regulación estricta, medios definidos con glucosa.
	Tóxica después de la inducción	Controlar el nivel de inducción: Plásmido con bajo número de copias, utilizar cepas que controlen los niveles de inducción, dirigir la proteína al periplasma.
	Optimizar el uso de codones	Uso preferencial de codones para la cepa hospedera. Utilizar cepas con tRNA extras para codones raros.
Formación de cuerpos de inclusión	Mal plegamiento o formación incorrecta de enlaces disulfuro	Dirigir la proteína al periplasma, utilizar cepas con citoplasma oxidativo, co-expresar con chaperonas. Cambiar el medio al inducir, bajar la temperatura
Proteína inactiva	Mut. en transcripción Baja solubilidad Mal plegamiento Modificaciones post-traduccionales	Relacionada con el uso apropiado de codones Utilizar potenciadores Cambiar parametros de inducción (temperatura) Utilizar organismos hospederos adecuados

Tabla III. Factores que contribuyen a la obtención de una proteína recombinante soluble y activa

Parámetro	Variante	Efecto
Concentración del Inductor Relacionada con promotores fuertes.	Concentración baja	Resulta en una inducción ineficiente.
	Concentración en exceso	Puede producir efectos tóxicos: -crecimiento celular reducido. -Una concentración de PR menor.
	Se recomienda evaluar concentraciones de IPTG entre 0.4 – 1.0 mM	
Temperatura de Inducción	Temperaturas altas 25°C – 37°C	Favorecen la agregación debido a la dependencia de las interacciones hidrofóbicas por la temperatura.
	Temperaturas bajas 4°C – 20°C	Limitan la agregación.
Densidad óptica (D.O)	Las D.O recomendadas son: -D.O _{600nm} = 0.4 a 1.0. A mayores densidades la actividad metabólica de la célula disminuye debido a la disminución de nutrientes, acumulación de acetato, menor disponibilidad de oxígeno y altos niveles de dióxido de carbono.	
Medios de Cultivo	Medios ricos Terrific Broth	Favorecen el crecimiento rápido, pueden favorecer agregación.
	Medio LB	Favorecen la producción de proteínas solubles.
	Medio Mínimo	Favorecen el crecimiento lento de la célula. Proteínas solubles
Aditivos en el medio	Glucosa	Del 0.5 -1 % de glucosa limita la expresión basal.
	Cofactores o grupos prostéticos	Favorecen el plegamiento, estabilidad y rendimiento de la proteína.
	Sorbitol o sacarosa	Para proteínas que son secretadas en el periplasma.
	Etanol	Induce la expresión de proteínas de heat-shock.
	NaCl	Mantiene la fuerza iónica del medio.
	PEG	Potencia la estabilidad de la proteína.
Chaperonas	GroE1/GroES DNAK/DNAI/GrpE	Co-expresión de chaperonas facilitan el plegamiento. Impiden la degradación.

Estrategias metodológicas para la producción eficiente de proteínas heterólogas en *E. coli*

El constructo debe ser transformado tanto en una cepa de mantenimiento (ejemplos: DH5 α , Top-10) como en la de expresión (sistema de expresión establecido), para su almacenamiento a -70 °C (en glicerol al 9% o al 50%) y su posterior extracción de DNA cuando se requiera, o para utilizar como inóculo para expresar la PR.

NOTA: En algunos laboratorios se opta por almacenar el constructo solamente en la cepa de mantenimiento, y a partir de ese glicerol extraer DNA y transformar en células competentes cada vez que se vaya a expresar la PR. En otros casos el constructo puede ser inestable en la cepa de expresión y de igual

forma se opta por almacenarlo en la cepa de mantenimiento.

Protocolo inicial de sobreexpresión en cultivos pequeños

Se debe transformar el DNA del constructo en alguna célula competente que sea de sobreexpresión. Por lo general, se elige la cepa de sobreexpresión considerando lo publicado en la literatura, o la experiencia propia de cada laboratorio, o bien si la proteína a sobreexpresar presenta alguna característica en particular. Sin embargo, si se cuenta con un banco de cepas de sobreexpresión, lo recomendado es transformar el DNA del constructo en varias cepas y probar un protocolo estándar, como el que se menciona a continuación:

1. En condiciones de esterilidad se extrae una asada del glicerol almacenado a $-70\text{ }^{\circ}\text{C}$, se siembra en medio sólido (caja) con antibiótico a $37\text{ }^{\circ}\text{C}$. De una colonia aislada se inocula un medio de cultivo líquido en presencia del antibiótico, se crece durante toda noche (precultivo) a $37\text{ }^{\circ}\text{C}$ en agitación constante entre 150 a 250 rpm.
 2. A partir del precultivo se inoculan unos 10 o 20 mL de medio de cultivo hasta llegar a una $D.O_{600\text{nm}}=0.1$; se incuba a $37\text{ }^{\circ}\text{C}$ en agitación entre 150 a 250 rpm hasta llegar a una $D.O_{600\text{nm}}=0.6$ (fase logarítmica). En este momento se agrega una concentración fija del inductor y se establece una temperatura fija y un tiempo definido para expresar la PR (en un principio se siguen las recomendaciones establecidas por el proveedor de la cepa hospedera y/o del plásmido o por experiencia propia de cada investigación).
 3. Después de la inducción se toma una alícuota (25 μl) del cultivo para cargarlos en un gel de poliacrilamida, para observar una banda del peso esperado que en teoría debe tener una mayor intensidad con respecto al resto de las demás proteínas propias de cada cepa hospedera. También, es recomendable tomar una alícuota del cultivo antes de la inducción y así poder comparar el nivel de sobreexpresión de la proteína de interés. Si se tiene éxito, alguna o algunas de las cepas monitoreadas mostraran un mayor grado de sobreexpresión.
 4. De las cepas que muestren mayor grado de sobreexpresión, el resto del cultivo se puede procesar para verificar si la proteína se encuentra en la fracción soluble. Entonces, el cultivo se centrifuga para obtener el botón celular, se resuspende en algún amortiguador de lisis seleccionado y se lisa por sonicación. El cultivo sonificado se centrifuga a 15 000 rpm para separar la fracción soluble (o sobrenadante) y el precipitado (que contiene restos celulares y proteínas agregadas). Se corren en un gel de poliacrilamida tanto el sobrenadante como el precipitado; con el fin de hacer un ensayo cuantitativo, previamente se puede hacer una cuantificación de proteína para cargar la misma concentración de proteína (que se detallan en esta misma revista en el tema de purificación de proteínas). El resultado esperado será encontrar la mayor parte de la PR en la fracción soluble de lo contrario se dice que la PR se encuentra en forma de cuerpos de inclusión.
 5. Si la proteína se encuentra sobreexpresada y soluble en alguna o algunas de las cepas monitoreadas, entonces se procederá a escalar el cultivo como se describe a continuación: a partir del precultivo se inocularán 1 L o más de medio de cultivo, hasta llegar a una $D.O_{600\text{nm}}=0.1$. De igual forma se incuba a $37\text{ }^{\circ}\text{C}$ en agitación constante entre 150 a 250 rpm hasta llegar a la D.O ya establecida a la cual se va a agregar el inductor y se dejará el tiempo establecido para su expresión.
- Si como resultado de este primer ensayo resultará que la proteína no se sobreexpresa, se sobreexpresa muy poco, o se forman cuerpos de inclusión, entonces se deben buscar las condiciones para mejorar el proceso (Tabla IV). Para ello a continuación se describen algunos protocolos que están basados en la optimización de un factor a la vez (Tabla IV).

Conclusiones

El primer paso para el éxito en la producción de la PR es la selección del sistema y las condiciones de expresión apropiadas, no obstante, algunas proteínas se expresan a niveles muy bajos o no se expresan en absoluto. Por tal razón a lo largo de varias décadas se han desarrollado varias estrategias para expresar estas proteínas, bajo un estricto control de los niveles de transcripción, traducción y de los componentes necesarios para llevar a cabo el proceso. De esta forma, el desarrollo de cada experimento requiere de un análisis exhaustivo de las características propias de cada proteína, de la cepa hospedera y del plásmido de sobreexpresión. En este trabajo, se hace un análisis de los problemas más comunes a los que se enfrenta un sistema de sobreexpresión procariota y se recomiendan algunas alternativas para una obtención eficiente de una PR.

Agradecimientos

Investigación realizada gracias al Programa PAPIIT-UNAM IN205821 (GH-A) y IN201021 (LR-S).

Tabla IV. Protocolos sugeridos para una mejor expresión de la PR

A partir de un precultivo se inoculan 50 o 100 mL de medio de cultivo que contenga el antibiótico, hasta llegar a una D.O _{600nm} =0.1. Incubar a 37 °C en agitación constante entre 150 a 250 rpm a 37 °C hasta llegar a la D.O deseada.	PROTOCOLO 1	
	Determinación de la D.O_{600nm} a la que se agregará el inductor	
	<i>Modificar (un factor)</i>	<i>Constante (otros factores)</i>
	<i>Dividir los 50 o 100 mL del cultivo en fracciones de 10 o 20 ml incubar hasta llegar a la D.O que se requiera. Densidades ópticas establecidas: 0.4, 0.6, 0.8 y 1.0</i>	Concentración fija de inductor Temperatura de inducción fija Tiempo de inducción fijo Agitación constante fija
	PROTOCOLO 2	
	Determinación de la concentración óptima de inductor	
	<i>Modificar</i>	<i>Constante</i>
	<i>Al llegar a la D.O ya establecida, dividir los 50 o 100 mL del cultivo en 10 o 20 ml. A cada tubo o matraz agregar una concentración de inductor, por ejemplo para IPTG: 0.4, 0.6, 0.8 y 1.0 mM</i>	Densidad óptica fija Temperatura de inducción fija Tiempo de inducción fijo Agitación constante fija
	PROTOCOLO 3	
	Determinación de la temperatura óptima de inducción	
	<i>Modificar</i>	<i>Constante</i>
	<i>Al llegar a la D.O ya establecida, dividir los 50 o 100 mL del cultivo en 10 o 20 ml. Cada tubo o matraz se va a incubar a distintas temperaturas, por ejemplo: 15, 20, 25, 30 o 37 °C</i>	Densidad óptica fija Concentración de inductor fijo Tiempo de inducción fijo Agitación constante fija
	PROTOCOLO 4	
	Determinación del tiempo óptimo de inducción	
	<i>Modificar</i>	<i>Constante</i>
<i>Al llegar a la D.O ya establecida, dividir los 50 o 100 mL del cultivo en 10 o 20 ml. Cada tubo o matraz se va a incubar a distintos tiempos, por ejemplo: 1, 3, 8, 15 hrs</i>	Densidad óptica fija Concentración de inductor fijo Temperatura de inducción fija Agitación constante fija	
PROTOCOLO 5		
Elección del medio de cultivo		
Si alguno de los protocolos mencionados anteriormente da un resultado favorable, pero aún se requiere mejorar la producción, otra alternativa es crecer en otros medios de cultivos. Por lo general, LB es el medio de elección para el crecimiento de <i>E. coli</i> , pero se puede usar un medio más rico como el Terrific Broth y/o uno más limitado de nutrientes como Medio Mínimo (u otros). Así también se pueden agregar aditivos al medio (Tabla III).		
En todos los protocolos para monitorear los resultados, los cultivos se procesan como en los pasos 3 y 4 mencionados en el protocolo inicial de sobreexpresión en cultivos pequeños.		

Referencias

- Shatzman, A.R. (1990) Gene expression using gram-negative bacteria. *Curr Opin Biotechnol* 1, 5-11. doi: 10.1016/s0958-1669(98)80027-x.
- Dyson, M.R., Shadbolt, S.P., Vincent, K.J et al. (2004) Production of soluble mammalian proteins in *Escherichia coli*: identification of protein features that correlate with successful expression. *BMC Biotechnol* 4: 32. doi: 10.1186/1472-6750-4-32.
- Welch, M., Govindarajan, S., Ness, J.E., Villalobos, A. Minshull, J., Gustafsson, C. (2009) Design parameters to control synthetic gene expression in *Escherichia coli*. *PLoS One* 4:e7002. doi.org/10.1371/journal.pone.0007002.
- Huang, C.J., Lin, H., Yang, X. (2012) Industrial production of recombinant therapeutics in *Escherichia coli* and its recent advancements. *J Ind Microbiol Biotechnol* 39:383–399. doi: 10.1007/s10295-011-1082-9.
- Jia, B., Jeon, C.O. (2016) High-throughput recombinant protein expression in *Escherichia coli*: current status and future perspectives. *Open Biol.* 6, 160196. doi.org/10.1098/rsob.160196.
- Jonasson, P., Liljeqvist, S., Nygren, P-AA., Stahl, S., Stahl, S. (2002) Genetic design for facilitated production and recovery of recombinant proteins in *Escherichia coli*. *Biotechnol Appl Biochem* 35:91-105. doi: 10.1042/BA20010099.
- Rosano, G. L., Morales, E.S., Ceccarelli, E.A. (2019) New tools for recombinant protein production in *Escherichia coli*: a 5-year update. *Protein Sci.* 28, 1412-1422. doi: 10.1002/pro.3668.
- Chen, H.C., Hwang, C.F., Mou, D.G. (1992) High-density *Escherichia coli* cultivation process for hyperexpression of recombinant porcine growth hormone. *Enzyme microb Technol.* 14, 321-326. doi: 10.1016/0141-0229(92)90159-1.
- Studier, F.W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* 41:207–234. doi.org/10.1016/j.pep.2005.01.016.
- Vera, A., Gonzalez-Montalban, N., Aris, A., Villaverde, A. (2007) The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. *Biotechnol Bioeng* 96:1101–1106. doi:10.1002/bit.21218.

11. Liu M, Feng X, Ding Y, Zhao G, Liu H, Xian M (2015) Metabolic engineering of *Escherichia coli* to improve recombinant protein production. *Appl Microbiol Biotechnol* 99:10367–10377. doi: 10.1007/s00253-015-6955-9.
 12. Uhoraningoga, A., Kinsella G. K., Henehan, G.T., Ryan, B.J. (2018) The goldilocks approach: a review of employing design of experiments in prokaryotic recombinant protein production. *Bioengineering* 5,89. doi:10.3390/bioengineering5040089.
 13. Smialowski, P., Doose, G., Torkler, P., Kaufmann, S., Frishman, D. (2018) PROSO II- a new method protein solubility prediction. *FEBS Journal* 279, 2192–2200. doi:10.1111/j.1742-4658.2012.08603.x.
 14. Packiam, K.A.R., Ramanan, R.N., Ooi, C. W., Krishnaswamy, L. L., Tey, B. Y. (2020) Stepwise optimization of recombinant protein production in *Escherichia coli* utilizing computational and experimental approaches. *Appl. Microbiol. Biotechnol.* 104, 3253-3266. doi: 10.1007/s00253-020-10454-w.
 15. Finney, M., Nisson, P.E., Rashtchian, A. (2001) Molecular cloning of PCR products. *Curr Protoc Mol Biol.* Chapter 15: Unit 15.4. doi: 10.1002/0471142727.mb1504s56.
 16. Aslanidis, C., de Jong, P.J. (1990) Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res* 18,6069–6074. doi: 10.1093/nar/18.20.6069.
 17. Tabor, S., Richardson, C.C. (1985) A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proc. Natl Acad Sci USA* 82, 1079-1080. doi: 10.1073/pnas.82.4.1074.
 18. Terpe, K. (2006) Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol* 72,211-222. doi:10.1007/s00253-006-0465-8.
 19. Ahmad, M., Hirz, M., Pichler, H., Schwab. (2014) Protein expression in *Pichia pastoris*: recent achievements and perspectives for heterologous protein production. *Appl Microbiol Biotechnol* 98(12), 5301-5317. doi: 10.1007/s00253-014-5732-5.
 20. Karste, K., Bleckmann, M., van den Heuvel, J. (2017) Not limited to *E. coli*: versatile expression vectors for mammalian protein expression. In Burgess-Brown N. (eds) *Heterologous gene expression in E. coli. Methods in Molecular Biology*, vol 1586. Humana Press, New York, NY. doi.org/10.1007/978-1-4939-6887-9_20.
 21. Baghban, R., Farajnia, S., Rajabibaz, M., Ghasemi, Y., Mafi, A., Hoseinpoor, R., Rahbarnia, L., Aria, M. (2019) Yeast Expression systems: overview and recent advances. *Mol Biotechnol.*61, 365-384. doi: 2443/10.1007/s12033-019-00164-8.
 22. Studier, F.W., Moffatt, B.A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* 189:113–130. doi: 10.1016/0022-2836(86)90385-2.
 23. Studier, F.W. (2018) T7 expression system for inducible production of proteins from cloned genes in *E. coli*. *Current Protocols in Molecular Biology.* 124, e63. doi:10.1002/cpmb.63.
 24. Guzman, L.M., Belin, D., Carson, M.J., Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose pBAD promoter. *J. Bacteriol.* 177 (14), 4121–4130. doi: 10.1128/jb.177.14.4121-4130.1995.
 25. Chao, Y.P., Chiang, C.J, Hung, W.B (2002) Stringent regulation and high-level expression of heterologous genes in *Escherichia coli* using T7 system controllable by the araBAD promoter. *Biotechnol Prog* 18:394–400. doi:10.1021/bp0101785.
 26. Choi, J.H., Lee, S.Y. (2004) Secretory and extracellular production of recombinant proteins using *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 64 (5), 625–635. doi: 10.1007/s00253-004-1559-9.
 27. Giacalone, M.J., Gentile, A.M., Lovitt, B.T., Berkley, N.L., Gunderson, C.W., Surber, M.W. (2006) Toxic protein expression in *Escherichia coli* using a rhamnose-based tightly regulated and tunable promoter system. *Biotechniques* 40 (5), 355-364. doi.org/10.2144/000112112.
 28. Daegelen, P., Studier, F.W., Lenski, R.E., Cure, S., Kim, J.F. (2009) Tracing ancestors and relatives of *Escherichia coli* B, and the derivation of B strains REL6006 and BL21(DE3). *J Mol Biol* 394:634-643. doi.org/10.1016/j.jmb.2009.09.022.
-



**DRA. GLORIA HERNÁNDEZ
ALCÁNTARA**

ID ORCID: 0000-0002-7029-6033

Realizó sus estudios de Biología en la Facultad de Biología de la Universidad Veracruzana. Posteriormente, realizó sus estudios de Maestría y Doctorado en el Posgrado en Ciencias Bioquímicas de la Facultad de Química, UNAM; mientras que la

fase experimental la desarrolló en el Instituto de Fisiología Celular, UNAM.

Es profesora de la asignatura de Bioquímica y Biología Molecular de la carrera de Médico Cirujano, en la Facultad de Medicina. Adicionalmente, ha impartido cursos del Taller de la Facultad de Ciencias en la carrera de Biología, así como diversos cursos en los posgrados en Ciencias Biológicas y en Ciencias Bioquímicas, respectivamente. Ha formado alumnos a nivel de licenciatura.

Actualmente es Profesor Titular en el Departamento de Bioquímica de la Facultad de Medicina, UNAM. Su producción científica consta de 18 artículos científicos en revistas de prestigio internacional, 2 capítulos de libro internacionales y 5 artículos de divulgación Nacional. Ha participado en diversos Congresos Nacionales e Internacionales.

Su trabajo de investigación ha sido sobre la estructura y función de algunas enzimas de la vía glucolítica de algunos organismos de importancia médica como *Trypanosoma brucei*, *T. cruzi*, *Giardia lamblia* y *Vibrio cholerae*.