

ASPECTOS METODOLÓGICOS DE LA EXPRESIÓN DE PROTEÍNAS RECOMBINANTES EN *Escherichia coli**

Andrés González^{1,2,3**} y María F. Fillat^{2,3}

¹Instituto de Investigación Sanitaria Aragón (IIS Aragón). San Juan Bosco 13, 50009 Zaragoza, España.

²Departamento de Bioquímica y Biología Molecular y Celular, Universidad de Zaragoza. Pedro Cerbuna 12, 50009 Zaragoza, España.

³Instituto Universitario de Biocomputación y Física de Sistemas Complejos. Mariano Esquillor, Edificio I + D. Campus Río Ebro, Universidad de Zaragoza. 50018 Zaragoza, España.

**Autor de correspondencia correo E: andresglezrod@gmail.com

RESUMEN

Cantidades significativas de proteínas puras, solubles y biológicamente activas son requeridas de modo creciente tanto en proyectos de investigación como para su empleo con fines terapéuticos, diagnósticos o industriales. La purificación de proteínas a partir de sus fuentes naturales no logra suplir en la mayoría de los casos los requerimientos de cantidad, calidad, facilidad de aislamiento o factibilidad económica del proceso. El desarrollo de la tecnología del ADN recombinante desde mediados de la década de los 70 del pasado siglo marcó el comienzo de la era moderna de la biotecnología. Con los conocimientos actuales en genómica, proteómica y bioinformática, el número de proteínas producidas por vía recombinante se ha incrementado exponencialmente. Dado su rápido crecimiento y altos rendimientos de biomasa en medios relativamente económicos, el amplio conocimiento sobre su fisiología y genética, así como su elevada capacidad de producción de proteínas heterólogas, la enterobacteria *Escherichia coli* continúa siendo el sistema de elección para la expresión recombinante, tanto a escala de laboratorio como en la industria. En el presente trabajo revisamos los aspectos metodológicos más importantes a tener en cuenta para garantizar adecuados niveles de expresión de proteínas recombinantes biológicamente activas, empleando a *E. coli* como organismo hospedero.

ABSTRACT

Large amounts of pure, soluble and biologically active proteins are increasingly required in research projects, diagnosis, therapies and industry. Purification of proteins from natural sources usually fails in guarantee the appropriate amount and quality of these biomolecules, and sometimes diminishes the economic feasibility of the process. The advent of recombinant DNA technology in the mid-1970s marked the beginning of modern biotechnology. Current knowledges in genomic, proteomic and bioinformatics have led to an exponentially increase in recombinant protein production. Due to its fast growth and high cell yields in inexpensive media, its well-characterized genetics and physiology, as well as its high capability for production of heterologous proteins, the enterobacterium *Escherichia coli* remains as the system of choice for the production of recombinant proteins in both on a lab scale and in industry. Here, we review the most relevant methodological aspects to be considered in order to ensure successful expression of biologically active recombinant proteins using *E. coli* as host.

PALABRAS CLAVE:

Escherichia coli, expresión recombinante, optimización, biotecnología, proteínas, bacteria.

KEY WORDS:

Escherichia coli, recombinant expression, optimization, biotechnology, proteins, bacteria.

1. *Escherichia coli*: una biofábrica unicelular de producción de proteínas

El primer aspecto a tener en cuenta para la obtención de una proteína por vía recombinante es la correcta selección del sistema de expresión. Un sistema de expresión lo conforma un organismo hospedero y un vector de expresión que contiene los elementos génicos necesarios para realizar los procesos de transcripción y traducción del gen de interés en dicho organismo hospedero. La elección del sistema óptimo de expresión dependerá de diversos factores incluyendo el origen biológico y las propiedades de la proteína de interés, las características de crecimiento del hospedero, los niveles de expresión y la localización celular de la proteína, la existencia de modificaciones postraduccionales y su implicación en la actividad biológica de la proteína, los aspectos regulatorios derivados de su posterior aplicación, así como el costo económico de todo el proceso.

En la actualidad existe una amplia variedad de sistemas de expresión disponibles para la producción a gran escala de proteínas recombinantes (1-12), los cuales se dividen en dos categorías generales: sistemas procarióticos y sistemas eucarióticos. Dentro de los sistemas procarióticos, *Escherichia coli* constituye sin lugar a dudas el hospedero más ampliamente utilizado a nivel de laboratorio y uno de los más importantes a escala industrial, dado el vasto conocimiento que se tiene de su fisiología y genética, lo cual facilita enormemente los trabajos de clonaje y cultivo (1, 13-15). Su alta velocidad de crecimiento y rendimiento celular en medios de cultivo poco costosos, unido a la capacidad de expresar elevados niveles de proteínas heterólogas, en ocasiones hasta un 30% del contenido proteico total, hacen de esta enterobacteria una auténtica biofábrica. Las cepas de *E. coli* utilizadas para la producción recombinante han sido genéticamente manipuladas delecionando los genes implicados en los mecanismos de patogenicidad, por lo cual son consideradas hospederos seguros para la fermentación a gran escala (1). Además de *E. coli*, han sido empleadas con éxito otros hospederos bacterianos para la producción de proteínas recombinantes, incluidos *Bacillus megaterium*, *Bacillus subtilis*, *Caulobacter crescentus*, *Pseudomonas fluorescens*, *Staphylococcus carnosus*, *Streptomyces sp*, entre otros (4, 16, 17).

Los vectores de expresión en los sistemas procarióticos comúnmente son vectores plasmídicos obtenidos por manipulación genética y configurados cuidadosamente a fin de asegurar los niveles óptimos de síntesis de la proteína de interés. Todos los vectores de expresión tendrán una arquitectura

esencial con elementos necesarios para su replicación en el hospedero y para la transcripción y traducción del gen clonado. Entre estos elementos se encuentran el origen de replicación, el promotor, el sitio de unión al ribosoma, la secuencia codificadora, terminadores de la transcripción, así como marcadores de resistencia a antibióticos que facilitan la selección de los recombinantes y garantizan la estabilidad del plásmido en el hospedero mediante el cultivo en un medio selectivo (Fig. 1). Adicionalmente, como parte de la estructura del vector podrá encontrarse un gen regulador que actúe directa (unión a operador) o indirectamente sobre el promotor, modulando su actividad y permitiendo condicionar los niveles de expresión según los intereses de síntesis de la proteína en cuestión (4, 15). De la misma forma, el vector de expresión podrá incorporar secuencias específicas de codones en el mismo marco de lectura del gen a expresar, que codifican pequeños péptidos o proteínas de fusión, que como veremos en detalle más adelante podrán cumplir diferentes propósitos.

Los sistemas procarióticos, si resultan satisfactorios, por lo general son los de elección para la mayoría de los propósitos dada su fácil manipulación y relativo bajo costo, aunque en ocasiones presentan serias limitaciones para la producción de proteínas eucarióticas (1, 16, 18-20). Muchas proteínas eucarióticas sufren una gran variedad de eventos postraduccionales tales como formación de puentes disulfuro, glicosilaciones, fosforilaciones, entre otras, las cuales son necesarias para su correcto plegamiento, estabilidad y actividad biológica. En hospederos procariotas no se expresan muchas de estas modificaciones postraduccionales; los puentes disulfuro no se forman debido a que el ambiente intracelular es muy reductor, durante la síntesis no enlazan oligosacáridos a las proteínas ni modifican las tirosinas mediante sulfatación (21, 22).

Por otro lado, en bacterias resulta difícil la secreción al medio extracelular de grandes cantidades de la proteína expresada por vía recombinante, por tal motivo en presencia de altos niveles de expresión resulta frecuente la formación de agregados insolubles en el citoplasma denominados cuerpos de inclusión (21, 23). Para la purificación a partir de estos cuerpos de inclusión es necesario el uso de agente caotrópicos que tienden a desnaturalizar la proteína, siendo necesaria la posterior implementación de técnicas de renaturalización, todo lo cual muchas veces resulta en una considerable pérdida de los rendimientos y un incremento sustancial de los costos de producción (23-25). Aunque muchas proteínas eucarióticas biológicamente activas han sido producidas con éxito en sistemas de expresión procarióticos (18, 20, 26, 27), en ocasiones serán

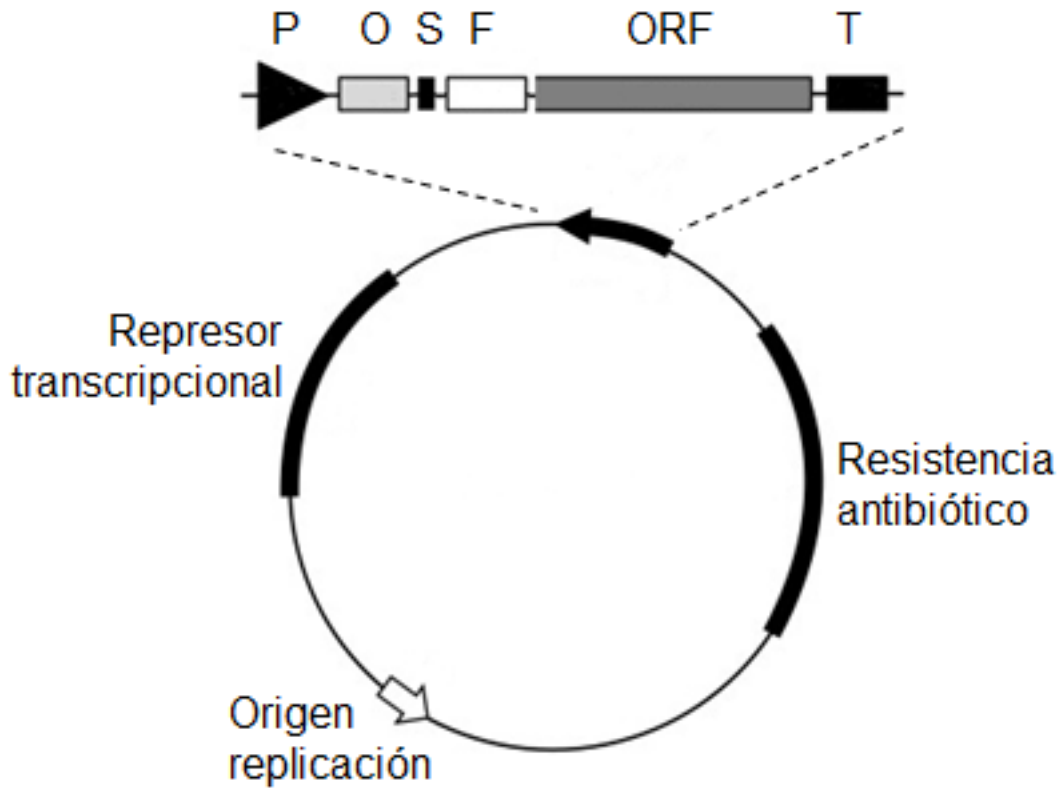


Figura 1. Elementos básicos en la arquitectura de un vector plasmídico empleado para la expresión recombinante en *E. coli*. P: promotor, O: operador, S: sitio de unión al ribosoma, F: secuencia de fusión, ORF: secuencia codificadora del gen de interés, T: terminador.

necesarios sistemas de expresión eucarióticos con vistas a obtener elevados rendimientos de este tipo de proteínas, en especial aquellas que requieren modificaciones postraduccionales para su actividad biológica. Entre los sistemas de expresión eucarióticos más frecuentemente usados en la industria destacan las levaduras (*Saccharomyces cerevisiae*, *Pichia pastoris*, *Hansenula polymorpha* y *Kluyveromyces lactis* (3, 5, 9, 28), pero también son empleados baculovirus/células de insecto (8, 29, 30), células de mamíferos (3, 10, 29), protozoos (31, 32), hongos filamentosos (12) y células de plantas (6, 7, 33). Para la producción recombinante de proteínas integrales de membrana, cuya expresión usualmente resulta altamente tóxica para los hospederos, se han desarrollado sistemas de expresión libres de células (*cell-free expression*), los cuales emplean la maquinaria metabólica de organismos procariotas o eucariotas (2, 11, 34).

2. Factores de interés en el diseño o selección de un sistema de expresión

Como ya hemos comentado, un sistema de expresión lo conforma un organismo hospedero y un

vector de expresión. La elección de un sistema de expresión recombinante determinado en función de la naturaleza de la proteína de interés no asegurará necesariamente la obtención de proteínas heterólogas funcionales, correctamente plegadas y en cantidades significativas. Muchas veces el hospedero es correcto, pero otros muchos factores atentan contra los niveles de expresión, el plegamiento y la estabilidad de la proteína recombinante.

A pesar del gran número de hospederos y sistemas de expresión desarrollados y utilizados con éxito en los últimos años, *E. coli* continúa siendo la principal herramienta en los laboratorios de genómica y biología molecular, tanto a nivel educacional en nuestras universidades como a nivel investigativo en universidades e instituciones científicas, constituyendo el sistema de expresión más utilizado para la producción de proteínas heterólogas (1, 3, 13). Casi el 80% de las proteínas de estructura tridimensional resuelta, remitidas al PDB (*Protein Data Bank*), han sido obtenidas por vía recombinante utilizando a *E. coli* como hospedero (1, 14). Por tal motivo, dada su gran versatilidad, uso generalizado e importancia como herramienta de trabajo en los estudios de genómica y proteómica a nivel de laboratorio,

centraremos esta revisión en discutir algunos de los aspectos metodológicos más relevantes a tener en cuenta durante la expresión recombinante de proteínas empleando este sistema de expresión.

2.1 Selección de la cepa hospedera

Una correcta selección de la cepa a utilizar para la expresión recombinante repercutirá marcadamente en los rendimientos de proteína obtenidos. Las cepas de expresión deberán carecer de las principales proteasas naturales, mantener la estabilidad del vector de expresión y conferir los elementos genéticos necesarios para una expresión eficiente y regulable. Algunas cepas de *E. coli* empleadas como hospederos para la expresión de proteínas recombinantes pueden contener mutaciones en los genes de determinadas enzimas con actividad disulfuro reductasa para favorecer la formación de puentes disulfuro necesarios para la conformación nativa de la proteína de interés. En otros casos, las cepas hospederas pueden co-expresar chaperonas para favorecer la expresión de proteínas recombinantes solubles y disminuir la formación de cuerpos de inclusión. En la actualidad se dispone de una gran variedad de cepas de *E. coli* comerciales que han sido manipuladas genéticamente para aplicaciones individuales (Tabla 1) (1, 13, 15, 24, 35).

2.2 Diseño óptimo de los vectores de expresión

Los elementos que forman parte del vector de expresión incidirán directamente en la eficiencia de la expresión del gen clonado. Aunque existe una gran variedad de vectores de expresión comerciales (Tabla 2), en muchas ocasiones resulta de interés la creación de un nuevo vector o la modificación de vectores disponibles en el laboratorio, a fin de alcanzar un determinado resultado; en tal caso, un óptimo diseño del vector resultará de vital importancia y se deberá tener en cuenta la no omisión de elementos indispensables para su funcionamiento.

2.2.1 Promotores

La expresión génica está fuertemente regulada a nivel transcripcional por lo que la síntesis del ARN mensajero (ARNm) pudiera considerarse el determinante primario del rendimiento final de la proteína de interés (36). Los niveles de ARNm sintetizado estarán primariamente determinados por las propiedades del promotor utilizado en el vector de expresión. Usualmente se requerirá el uso de promotores fuertes, capaces de conseguir niveles de síntesis proteica entre un 10 y un 30% del contenido proteico total de la célula (1, 37).

El promotor deberá producir un nivel mínimo de transcripción basal y su actividad estará regulada según los intereses de síntesis proteica. La expresión a gran escala de una proteína preferiblemente se buscará con una alta densidad de crecimiento celular y mínima actividad del promotor, seguido por la inducción o desrepresión del promotor. Una fina regulación del promotor resultará esencial para la síntesis de proteínas tóxicas para el hospedero (35). Una inadecuada regulación del promotor podrá causar además inestabilidad del plásmido, disminución de la velocidad de crecimiento del hospedero, formación de cuerpos de inclusión y pérdidas en los rendimientos de la proteína recombinante. Algunos promotores muy fuertes como aquellos de los ARN de transferencia de *E. coli* no son extensivamente utilizados dada su difícil regulación (38).

La inducción del promotor deberá producirse por un mecanismo simple y económicamente práctico. Los promotores más ampliamente utilizados para la expresión recombinante a gran escala en *E. coli* emplean inducción térmica o activadores químicos (4, 14, 37). Los promotores inducibles por isopropil- β -D-tio-galactósido (IPTG) son fuertes, eficientes y ampliamente utilizados a escala de laboratorio, pero su empleo a escala productiva en la síntesis de proteínas terapéuticas para uso humano resulta inadecuado dada su toxicidad (3, 38).

Aunque los promotores procarióticos presentan una región esencial para su funcionamiento, conformada por las secuencias hexaméricas -10 y -35 separadas por 15-19 pb, ha sido demostrado que otros elementos ubicados *upstream* a la secuencia -35 actúan como activadores transcripcionales e influyen decisivamente en la actividad del promotor (39). La exclusión de estos elementos durante el diseño y construcción de un vector de expresión pudiera repercutir desfavorablemente en la transcripción y por ende en los niveles de ARNm y proteína recombinante.

2.2.2 Terminadores de la transcripción

Aunque a menudo son pasados por alto durante el diseño y construcción de un vector de expresión, los terminadores de la transcripción son elementos indispensables dado que cumplen importantes funciones. La transcripción sobre la secuencia de un promotor puede inhibir su función, en un fenómeno conocido como oclusión del promotor (38). Esta interferencia puede ser prevenida con la apropiada ubicación de un terminador de la transcripción al finalizar la secuencia codificadora. De forma similar, la ubicación de un terminador de la transcripción en posición *upstream* al promotor será de utilidad para minimizar transcripciones no deseadas. Las

Tabla 1. Algunas cepas de *E. coli* empleadas comúnmente para la expresión de proteínas recombinantes.

Cepa	Características	Uso recomendado
BL21	<ul style="list-style-type: none"> Deficiente en las proteasas Ion y OmpT. Carece de T7 RNA polimerasa. 	Expresión recombinante de genes bajo control de promotores reconocidos por la RNA polimerasa de <i>E. coli</i> : <i>lac</i> , <i>tac</i> , <i>trc</i> , <i>T5</i> .
BL21(DE3)	Derivada de BL21. Contiene el profago λ DE3 que transporta el gen de la T7 RNA polimerasa bajo control del promotor <i>lacUV5</i> , inducible por IPTG.	Expresión recombinante de genes bajo control de promotores T7, T7- <i>lac</i> y promotores reconocidos por la RNA polimerasa de <i>E. coli</i> .
BL21(DE3)pLysS	<ul style="list-style-type: none"> Derivada de BL21. Contiene el profago λDE3 que transporta el gen de la T7 RNA polimerasa bajo control del promotor <i>lacUV5</i>. Contiene el plásmido pLysS que transporta el gen de la T7 lisozima (LysS), un inhibidor de la T7 RNA polimerasa que previene expresión basal en cultivos no inducidos con IPTG. 	Expresión recombinante con vectores basados en promotores T7, donde se requiera un mayor control de la expresión basal.
BL21(DE3)pLysE	<ul style="list-style-type: none"> Derivada de BL21. Contiene el profago λDE3 que transporta el gen de la T7 RNA polimerasa bajo control del promotor <i>lacUV5</i>. Contiene el plásmido pLysE, que expresa mayores niveles de T7 lisozima que pLysS. 	Expresión recombinante con vectores basados en promotores T7, donde se requiera un elevado control de la expresión basal. Especialmente recomendada para la expresión de proteínas tóxicas para <i>E. coli</i> .
BL21-SI y BL21-AI	<ul style="list-style-type: none"> BL21-SI deriva de la cepa "salt-inducible" GJ1158. Contiene el gen de la T7 RNA polimerasa bajo control del promotor proU, inducible por la concentración salina del medio. BL21-AI contiene el gen de la T7 RNA polimerasa bajo control del promotor araBAD, inducible por arabinosa. 	Ambas cepas logran un control estricto de la expresión basal de las proteínas recombinantes. Especialmente recomendadas para la expresión de proteínas tóxicas para <i>E. coli</i> .
CD41(DE3) y CD43(DE3)	<ul style="list-style-type: none"> Derivadas de la cepa BL21(DE3). Contiene el profago λDE3 que transporta el gen de la T7 RNA polimerasa bajo control del promotor <i>lacUV5</i>. Contienen mutaciones no caracterizadas que previenen la muerte celular asociada a la expresión de muchas proteínas recombinantes tóxicas. 	Expresión de proteínas tóxicas y proteínas de membrana de diversos organismos, tanto procariontes como eucariotes, incluyendo insectos, plantas y mamíferos
Rosetta (DE3) y Rosetta 2 (DE3)	<ul style="list-style-type: none"> Derivadas de BL21. Contienen el profago λDE3 que transporta el gen de la T7 RNA polimerasa bajo control del promotor <i>lacUV5</i>. Contienen el plásmido pRARE o pRARE2, que transportan los genes para los tRNA de los codones AGG, AGA, AUA, CUA, CCC, GGA (pRARE) y AGA, AGG, AUA, CUA, GGA, CCC, CGG (pRARE2). 	Expresión recombinante de genes eucarióticos que contienen codones raros para <i>E. coli</i> . Se emplean además para la expresión recombinante de genes de especies procariontes alejadas filogenéticamente de <i>E. coli</i> , que también presenten codones raros.
BL21 CodonPlus y BL21 CodonPlus (DE3)	<ul style="list-style-type: none"> Derivadas de BL21-Gold Contienen los genes para los tRNA de los codones AGG, AGA, AUA, CUA, CCC. 	Expresión recombinante de genes que contienen codones raros para <i>E. coli</i> , tanto procariontes como eucariotes
Origami (DE3)	<ul style="list-style-type: none"> Derivada de K12. Contiene el profago λDE3 que transporta el gen de la T7 RNA polimerasa bajo control del promotor <i>lacUV5</i>. Contiene mutaciones en los genes <i>trxB</i> (tioredoxina reductasa) y <i>gor</i> (glutación reductasa), incrementando la formación de puentes disulfuro en el citoplasma. 	Expresión recombinante de proteínas que requieran la formación de puentes disulfuro para alcanzar su conformación nativa.
ArcticExpress (DE3)	<ul style="list-style-type: none"> Derivada de BL21-Gold. Contiene el profago λDE3 que transporta el gen de la T7 RNA polimerasa bajo control del promotor <i>lacUV5</i>. Contiene los genes de las chaperoninas Cpn10 y Cpn60 de <i>Oleispira antarctica</i>, las cuales comparten gran homología con GroEL y GroES de <i>E. coli</i>, pero son altamente funcionales a bajas temperaturas de cultivo. 	Expresión recombinante de proteínas que tiendan a formar cuerpos de inclusión. La co-expresión de las chaperoninas Cpn10 y Cpn60 unido al cultivo a bajas temperaturas (15-25°C) pueden incrementar significativamente los rendimientos de proteína soluble y biológicamente activa.

Tabla 2. Ejemplos de vectores plasmídicos empleados para la expresión recombinante en *E. coli*.

Vector	Promotor/ repressor	Resistencia	Fusión	Cepa hospedera	Aplicación
pET-28a(+)	T7/LacI	Kanamicina	N-His C-His	BL21(DE3)	Sobreexpresión en citoplasma inducible por IPTG, purificación por IMAC.
pET-20b(+)	T7	Ampicilina	C-His Péptido señal	BL21(DE3)	Sobreexpresión en periplasma inducible por IPTG, purificación por IMAC.
pET-32a(+)	T7/LacI	Ampicilina	N-Trx N-His C-His	Origami(DE3)	Sobreexpresión de proteínas con puentes disulfuro, purificación por IMAC.
pGEX-4T-3	tac	Ampicilina	N-GST	BL21	Purificación por afinidad a glutatión.
pETDuet-1	2×(T7/LacI)	Ampicilina	N-His C-S	BL21(DE3)	Co-expresión de dos proteínas, purificación independiente por IMAC y afinidad a proteína S.

secuencias terminadoras de la transcripción formarán además una estructura secundaria por complementariedad de bases en el ARNm que lo protegerá de la acción de exonucleasas, extendiendo su vida media e incrementando sustancialmente los niveles de proteína producida (14, 38).

2.2.3 Secuencias reguladoras de la traducción

Ubicada en posición *downstream* al promotor se ubica una región denominada sitio de unión al ribosoma (RBS), dentro de la cual se haya una secuencia altamente conservada denominada Shine-Dalgarno (SD) que interactúa por complementariedad con el extremo 3' del ARN ribosomal 16S durante el inicio de la traducción. La distancia entre la secuencia SD y el codón de inicio de la traducción varía entre 5 y 13 nucleótidos e influye en la eficiencia del inicio de la traducción (38). Varios estudios se han llevado a cabo en *E. coli* para determinar la secuencia óptima de nucleótidos de la región SD, así como la distancia más efectiva entre esta región y el condón de inicio. Los resultados demuestran que una región SD con secuencia UAAGGAGG produce de 3-6 veces más proteína que otras secuencias típicas como AAGGA. Por otro lado, cada secuencia SD tendrá un rango de espaciamiento óptimo en relación al codón de inicio, siendo de 4-8 nucleótidos para UAAGGAGG y 5-7 nucleótidos para AAGGA (13). La estructura secundaria del RBS juega un papel crucial en la eficiencia del inicio de la traducción, pudiendo dificultar el acceso del ARNm a la subunidad ribosomal 30S y llegar a inhibir la traducción. El enriquecimiento de esta región con residuos de A/T ha demostrado

incrementar considerablemente la expresión de ciertos genes (38).

El diseño de un vector de expresión frecuentemente incluye la inserción consecutiva de los tres codones de parada (UAA, UGA, UAG) al final de la región de clonaje, a fin de asegurar la terminación de la traducción en cualquiera de los marcos de lectura posibles y prevenir posibles "saltos" del ribosoma. La eficiencia en la terminación varía significativamente con la identidad del nucleótido que precede inmediatamente al codón de parada. La secuencia UAAU es la señal de terminación traduccional más eficiente en *E. coli* (14, 38).

2.2.4 Uso de codones

Aunque el código genético es universal, tanto hospederos procarióticos como eucarióticos harán un uso preferencial de determinados codones sinónimos, lo cual está determinado por la abundancia relativa de los ARN de transferencia (ARNt) complementarios a cada codón. Los codones raros para un determinado hospedero usualmente se encuentran en genes homólogos con un bajo nivel de expresión o genes heterólogos introducidos durante el clonaje. Para *E. coli* pueden constituir fuentes de codones raros algunos genes provenientes de eucariotas, arqueas o incluso eubacterias alejadas filogenéticamente. La expresión de genes con codones raros puede conllevar a errores traduccionales por sustituciones de aminoácidos, disminución de la velocidad de traducción, cambio del marco de lectura e interrupción de la traducción, lo cual resulta frecuente en sistemas de expresión recombinante con genes heterólogos

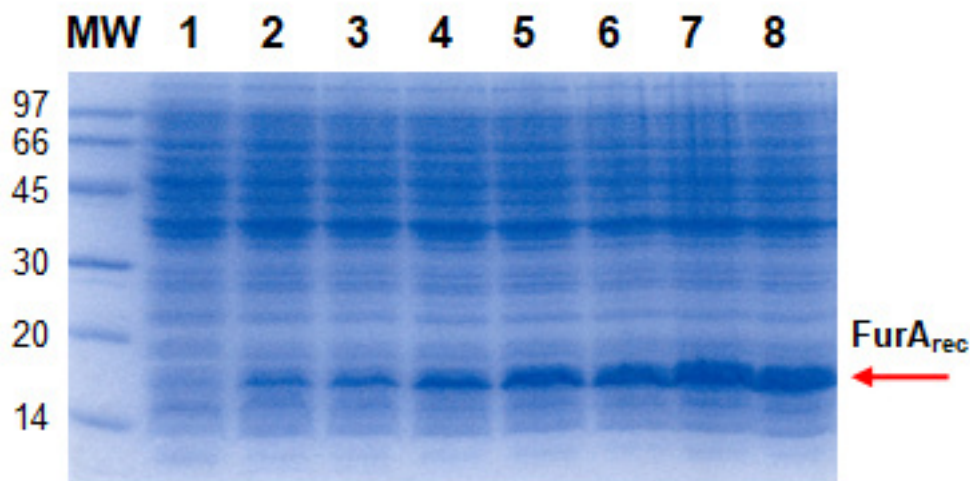


Figura 2. Cambio en el nivel de expresión de la proteína *FurA* de *Anabaena* sp. PCC 7120 tras 30 min (3), 1 hora (4), 2 horas (5), 4 horas (6), 6 horas (7) y toda la noche (8) de inducción con 1 mM de IPTG. La proteína *FurA* fue expresada por vía recombinante en *E. coli* BL21(DE3), empleando el vector de expresión pET-28a(+). Los resultados de expresión fueron evidenciados mediante SDS-PAGE (15% acrilamida) y tinción con azul de Coomassie de un extracto soluble de células completas. Como controles se incorporaron en la electroforesis la cepa BL21(DE3) transformada con el vector pET-28a(+) vacío (1) y la cepa BL21(DE3) con el vector pET-28a-*furA* sin inducir con IPTG (2). Los pesos moleculares en kDa de un patrón de peso comercial (MW) se indican a la izquierda del gel. La migración electroforética de la proteína *FurA* recombinante (~17 kDa) se indica a la derecha.

presentando dobletes o tripletes de codones raros (14, 20, 38, 40).

Existen dos estrategias alternativas para remediar la presencia de codones raros en los genes de interés. La primera consiste en mutagénesis dirigida de la secuencia con cambio hacia codones homólogos preferenciales para el hospedero de expresión. La segunda estrategia consiste en la co-transformación del hospedero con un plásmido que porte los genes para los ARNt deficitarios. En la actualidad existen cepas de expresión comerciales (Ej: cepas *Rosetta* de Novagen o *CodonPlus* de Stratagene) que llevan incorporados estos plásmidos con un origen de replicación distinto a los usados en la mayoría de los vectores de expresión, a fin de garantizar la compatibilidad y estabilidad de ambos vectores (Tabla 1).

3. Estrategias de optimización de la expresión recombinante

Una vez elegido el sistema de expresión adecuado resultará de vital importancia la optimización de las condiciones de expresión. Para definir las condiciones óptimas de expresión se realizarán ensayos o "screenings" de expresión, evaluando el grado de expresión de la proteína recombinante, su solubilidad y estabilidad intracelular mediante ensayos rápidos y sencillos, como la electroforesis desna-

turalizante en gel de poliacrilamida (SDS-PAGE) (41). Previo a la producción de cultivos de la cepa recombinante a gran escala resultará necesario evaluar diferentes condiciones de expresión empleando varias réplicas de cultivos a pequeños volúmenes, tomando muestras del microorganismo crecido bajo estas condiciones y definiendo el valor óptimo de expresión de la proteína de interés para cada condición en estudio mediante SDS-PAGE (Fig. 2). La optimización de la expresión recombinante incluirá la definición de una estrategia de expresión que garantice el máximo rendimiento y estabilidad de la proteína y facilite en lo posible el posterior proceso de purificación. Algunas de las estrategias a tener en cuenta para optimizar la expresión de proteínas recombinantes en *E. coli* se detallan a continuación.

3.1 Variación de la condiciones de cultivo

El medio de cultivo, las condiciones de crecimiento (temperatura, pH, aireación) o el sistema de fermentación influirán sensiblemente en la expresión recombinante. Los cultivos discontinuos o *batch* constituyen la forma más simple y generalizada de producción de proteínas recombinantes a escala de laboratorio, en ellos los nutrientes son suministrados solo al inicio del cultivo y existe un limitado control de las condiciones de crecimiento durante el proceso. A medida que se incrementa la densidad

celular los nutrientes disminuyen, disminuye la concentración de oxígeno disuelto, se incrementa la concentración de residuos metabólicos secretados al medio y se pueden producir variaciones apreciables de pH. Las densidades celulares y los niveles de producción recombinante bajo estas condiciones de cultivo son moderadas, aunque suficientes para la mayoría de los trabajos de investigación (24).

Bajo estas condiciones de fermentación se deberá prestar especial atención a la composición del medio de cultivo, la temperatura o el pH, todo lo cual influirá apreciablemente no solo en los niveles de biomasa sino también en los rendimientos de proteína recombinante (38). Aunque el medio Luria-Bertani (LB) pudiera considerarse el más utilizado para el cultivo de *E. coli* a escala de laboratorio y con él se logran obtener altos niveles de expresión, en ocasiones el empleo de un medio más rico en nutrientes como el *Terrific broth* permite obtener los mismos niveles de proteína recombinante con menor volumen de medio (21). La temperatura de cultivo influirá en la actividad de toda la maquinaria metabólica de la célula, incluyendo los mecanismos de transcripción, traducción, proteólisis, expresión y actividad de chaperonas, etc. (24, 38). En muchas ocasiones será recomendable cultivar hasta fase logarítmica a 37°C y luego realizar la inducción de la expresión a 25-30°C, con lo cual se disminuirá la actividad de proteasas, disminuirá la velocidad de síntesis proteica y aumentará la expresión de chaperonas, todo lo cual favorecerá el correcto plegamiento y estabilidad de la proteína recombinante, evitando en muchos casos la formación de cuerpos de inclusión (1).

En la actualidad están siendo introducidas novedosas tecnologías para aumentar la eficiencia de los cultivos e incrementar los rendimientos celulares y los niveles de expresión recombinante sin excesivo costo económico (15).

3.2 Definición de las condiciones óptimas de inducción

Con el fin de garantizar un mayor control de la expresión recombinante, numerosos vectores de expresión incluyen promotores regulables que dirigen la expresión del gen de interés, los cuales contienen secuencias operadoras reconocidas por reguladores transcripcionales muchas veces codificado en el mismo vector de expresión. De esta forma, los reguladores transcripcionales, mayormente represores, bloquean la transcripción del gen clonado en ausencia de una señal inductora. La señal de inducción puede ser un cambio de temperatura, pH o fuerza iónica en el medio de cultivo, aunque en la mayoría de los casos el inductor es un metabolito que se agrega al medio cuando queremos inducir

la expresión recombinante (42). Por ejemplo, en vectores de expresión con promotores reprimidos por el represor LacI del operón lactosa, la adición de un análogo no metabolizable de la lactosa, el IPTG, libera al represor del sitio operador y promueve la transcripción del gen clonado.

En muchos sistemas de expresión basados en el promotor T7, tanto la expresión de la T7 RNA polimerasa como la expresión del gen clonado serán inducidas tras la adición del inductor IPTG (13). En tal sentido, tanto la concentración del inductor como el tiempo de inducción influirán sensiblemente en los niveles de expresión recombinante. Con el uso de promotores fuertes (aquellos con una alta afinidad y estabilidad de unión de la RNA polimerasa, siendo los más empleados para alcanzar altos niveles de expresión recombinante), la adición del inductor supone un incremento sustancial del gasto metabólico de las células hacia la producción de la proteína heteróloga, lo cual ralentizará apreciablemente el crecimiento del microorganismo hospedero. Por tal motivo, la adición del inductor se realizará preferiblemente una vez las células hayan alcanzado la fase exponencial media-tardía del crecimiento (43), momento en el cual los cultivos aún son jóvenes y presentan una alta densidad celular. La adición del inductor en un estadio muy temprano o demasiado tarde en la fase de crecimiento del cultivo repercutirá negativamente en los rendimientos de proteína recombinante.

3.3 Prevención de cuerpos de inclusión

El principal objetivo de la expresión recombinante es usualmente la obtención de grandes cantidades de la proteína de interés en estado nativo y biológicamente activa. Esta estrategia no siempre es aceptada por la maquinaria metabólica de la célula hospedera y puede desencadenar una respuesta de estrés celular caracterizada por la formación de agregados insolubles en el citosol conocidos como cuerpos de inclusión, en los cuales la proteína sintetizada se encuentra solo parcialmente plegada y por tanto biológicamente inactiva (14, 24). Bajo condiciones de biosíntesis normales las proteínas citoplasmáticas pueden adquirir espontáneamente su estructura tridimensional; en otros casos las cadenas polipeptídicas nacientes interactúan reversiblemente con proteínas chaperonas celulares que previenen la agregación y favorecen un correcto plegamiento. La formación de cuerpos de inclusión durante la sobre-expresión de las proteínas recombinantes pudiera ser resultado de la acumulación de altas concentraciones de intermediarios de plegamiento o de un ineficiente procesamiento por las chaperonas celulares (24).

Maximizar la obtención de la proteína recombinante en forma soluble resulta en la mayoría de los casos la solución más atractiva y recomendable. Diversas estrategias han sido desarrolladas para este fin y en la mayoría de los casos buscan evitar, minimizar o resolver la situación de estrés metabólico que ocurre en la célula. El cultivo o inducción a bajas temperaturas, el empleo de cepas mejoradas (Ej: ArcticExpress (DE3), Tabla 1), la modificación de los medios y/o condiciones de cultivo, la co-expresión de chaperonas, el uso de proteínas de fusión (Ej: proteína de unión a maltosa (MBP), *N-utilizing substance A* (NusA) o la disminución del tamaño e hidrofobicidad del fragmento a expresar son algunas de las estrategias más utilizadas para incrementar la solubilidad en la expresión recombinante (24, 43, 44).

Sin embargo, dada la posibilidad de purificación y renaturalización exitosa de algunas proteínas a partir de cuerpos de inclusión, los mismos pueden ofrecer ciertas ventajas en cuanto a facilidad de aislamiento de la proteína, su alta concentración y el alto grado de pureza (1). En ocasiones la formación de cuerpos de inclusión resulta de elección cuando la proteína recombinante activa en el citosol es letal para la célula huésped (38). A pesar de obvias desventajas como la disminución de los rendimientos, la necesidad de optimización de las condiciones de replegamiento, el riesgo de afectación irreversible de la conformación nativa o el encarecimiento del proceso de purificación, varios métodos de aislamiento y renaturalización de proteínas recombinantes a partir de cuerpos de inclusión han sido desarrollados con éxito (23, 25). Adicionalmente, algunos estudios han demostrado actividad catalítica significativa en cuerpos de inclusión de enzimas recombinantes como la β -galactosidasa de *E. coli* y la dihidrofolato reductasa humana (45). En tal sentido, la purificación de cuerpos de inclusión como nanopartículas naturales de enzimas puras biológicamente activas, con una arquitectura porosa y altamente hidratada que permiten una eficiente difusión del sustrato a través de su estructura, tendría una gran aplicabilidad en la industria alimentaria y farmacéutica (45-47).

3.4 Definición de la localización celular

Las proteínas expresadas por vía recombinante en principio pueden ser dirigidas a tres diferentes localizaciones celulares: citoplasma, periplasma y medio extracelular (medio de cultivo). Cada una de estas localizaciones tendrá ventajas y desventajas (4, 14). La expresión y mantenimiento en el citosol es normalmente preferible dados los mayores rendimientos de proteína usualmente obtenidos,

la mayor simplicidad de los vectores de expresión o en algunos casos la ventaja que ofrecen los cuerpos de inclusión formados; sin embargo, en ocasiones resulta de mejor elección la exportación de la proteína hacia el espacio periplasmático o el medio extracelular.

Algunas proteínas no logran recuperar su plegamiento nativo y actividad biológica tras un procedimiento de purificación desnaturizante a partir de cuerpos de inclusión. En otros casos el ambiente fuertemente reductor del citoplasma bacteriano impide el correcto plegamiento de proteínas heterólogas cuya estructura terciaria depende apreciablemente de la formación de puentes disulfuro. Por otro lado, la obligada incorporación del codón de inicio de la traducción bacteriana (AUG) y con ello una metionina en el extremo N-terminal de proteínas recombinantes eucarióticas, muchas veces afecta de manera apreciable su conformación, actividad biológica y antigenicidad, obstaculizando en muchos casos la aprobación del uso de estas proteínas no nativas en humanos.

El espacio periplasmático ofrece un ambiente oxidante que favorece la formación de puentes disulfuro, disminuye la posibilidad de degradación proteolítica dado el menor número de proteasas celulares presentes y simplifica la purificación al concentrar el producto deseado en un compartimiento mucho menos rico en proteínas celulares que el citosol. El clivaje *in vivo* del péptido señal durante la translocación al periplasma hace más probable la eliminación de la metionina añadida en proteínas eucarióticas rindiendo un extremo N-terminal nativo (38). La exportación de una proteína heteróloga al periplasma requiere la incorporación de una secuencia señal en el extremo N-terminal del gen clonado. Una gran variedad de péptidos señales procariotas han sido utilizados con éxito (4, 38), sin embargo el proceso de translocación es complejo y la presencia del péptido señal no siempre asegura un transporte eficiente al periplasma, lo cual puede repercutir apreciablemente en los rendimientos obtenidos. Varias estrategias para mejorar la translocación al periplasma han sido descritas (14, 38, 48).

La secreción de la proteína de interés al medio extracelular comparte muchas de las ventajas logradas con la translocación al periplasma y favorece sustancialmente el proceso de purificación para algunas proteínas (49). Por lo general la secreción de la proteína recombinante se logra empleando los sistemas de secreción extracelular existentes en el hospedero para otras proteínas celulares (38, 50, 51). *E. coli* normalmente secreta muy pocas proteínas y sus sistemas de secreción pueden ser ineficientes, los rendimientos de proteína recombi-

nante obtenidos usualmente son modestos, aún con la construcción de proteínas híbridas secretables (13). Otros sistemas de expresión bacterianos como *S. carnosus* o *B. subtilis* son capaces de secretar grandes cantidades de proteína al medio de cultivo favoreciendo su purificación en estado soluble y nativo (4, 51). Sistemas de expresión eucarióticos como levaduras o cultivos de células de mamíferos rinden niveles marcadamente superiores de proteínas recombinantes secretadas al medio, en comparación con los hospederos bacterianos (3).

3.5 Generación de proteínas de fusión

Una proteína de fusión o proteína quimérica consiste en la unión de una corta secuencia de aminoácidos, polipéptido o proteína a la proteína que se quiere expresar por vía recombinante a través de un sitio de reconocimiento a proteasas específicas que posteriormente facilitan la separación de nuestra proteína. Usualmente esta fusión se construye a nivel de ADN y la quimera se transcribe y expresa como una sola entidad dentro de la célula. Esta estrategia se utiliza con frecuencia para facilitar el proceso de purificación, favorecer el plegamiento, incrementar la solubilidad y prevenir cuerpos de inclusión, disminuir la degradación proteolítica, dirigir la translocación a periplasma y en algunos casos ha logrado incrementar los rendimientos de proteína recombinante desde niveles no detectables hasta casi el 20% del contenido proteico celular (14, 19, 48, 52-54).

Han sido desarrollados una amplia variedad de patrones de proteínas de fusión. La fusión de una cola de poli-histidinas o de la glutatión S-transferasa permite la purificación por afinidad de la proteína recombinante mediante protocolos rápidos y altamente eficientes, para lo cual existen vectores de clonaje específicos y resinas cromatográficas comerciales (55). La fusión con las proteínas MBP o NusA de *E. coli* incrementan considerablemente la solubilidad y son especialmente recomendadas para evitar o disminuir la formación de cuerpos de inclusión (14). La fusión a SUMO (*small ubiquitin-related modifier*), una pequeña proteína de unos 100 residuos altamente conservada en eucariotas, ha logrado incrementar considerablemente los niveles de expresión y la solubilidad en hospederos procariotas (54). La incorporación del péptido señal de numerosas proteínas procarióticas permite dirigir la translocación hacia periplasma (19, 49, 51). Varios estudios han demostrado que los altos niveles de expresión de una proteína homóloga dada pueden ser transferidos a una proteína heteróloga

pobremente expresada con la fusión del extremo N-terminal de aquella, quizás como resultado de una estabilización del ARNm (14).

3.6 Prevención de la proteólisis

La proteólisis es un proceso selectivo y altamente regulado que juega un importante papel en la fisiología celular. *E. coli* contiene un gran número de proteasas localizadas en el citoplasma, periplasma y membranas interna y externa, que participan en actividades metabólicas y en la selectiva eliminación de proteínas anormales. La maquinaria proteolítica actúa como consecuencia de una gran variedad de condiciones tales como síntesis incompleta, mutaciones por sustitución de aminoácidos, síntesis excesiva de subunidades de complejos multiméricos, daño postraducciona por oxidación o ataque de radicales libres, etc. (38). En principio, una proteína heteróloga podría fácilmente ser reconocida como anormal y ser degradada por la célula, de hecho en muchos casos la degradación proteolítica disminuye apreciablemente los rendimientos de proteína recombinante.

Varias estrategias han sido aplicadas a fin de minimizar la degradación proteolítica durante la expresión recombinante tales como el empleo de cepas deficientes de proteasas (*E. coli* BL21 es deficiente de las proteasas Ion y OmpT) (Tabla 1), construcción de proteínas de fusión, translocación fuera del citosol, crecimiento o inducción a bajas temperaturas (la mayoría de las proteasas tienen su actividad óptima a 37°C), co-expresión con chaperonas, sustitución de aminoácidos específicos en el sitio de clivaje de la proteasa, optimización de las condiciones de fermentación, entre otras (14, 38, 56).

3.7 Co-expresión de proteínas

Ha sido comentado con anterioridad las ventajas de la co-expresión de chaperonas para incrementar la estabilidad conformacional y solubilidad de la proteína de interés o prevenir su degradación proteolítica. En otros casos resulta igualmente conveniente la expresión simultánea de dos o más proteínas clonadas en el mismo vector de expresión o en vectores compatibles diferentes. Tal es el caso de la síntesis recombinante de complejos proteicos con varios componentes o subunidades. La co-expresión de varios componentes a menudo resulta en un incremento de la estabilidad conformacional de cada componente, protegiéndose en ocasiones mutuamente de la degradación proteolítica del hospedero (21, 57, 58). El empleo de vectores policistronicos ha permitido expresar con

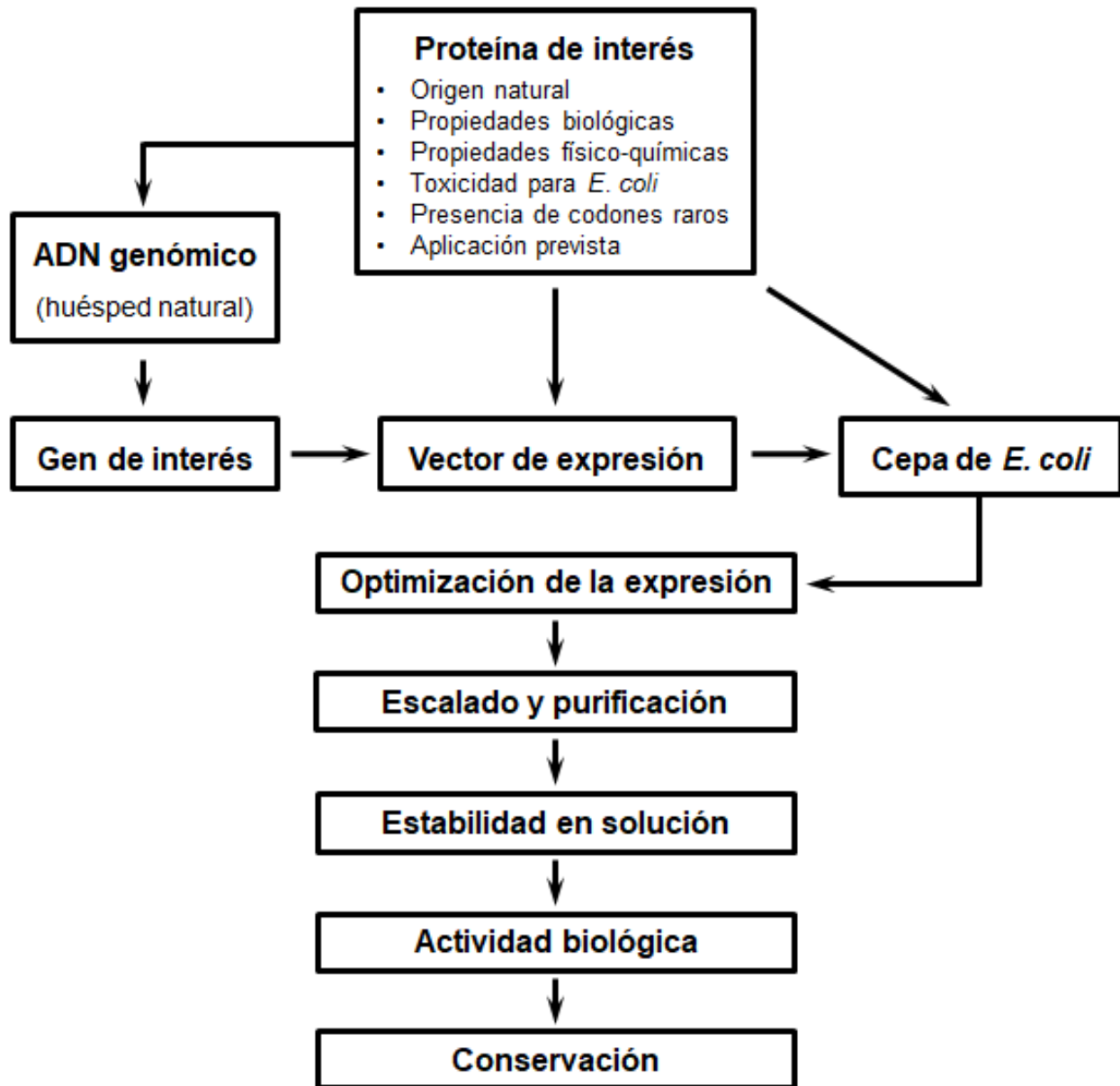


Figura 3. Flujograma de los aspectos metodológicos a tener en cuenta para la expresión de proteínas recombinantes en *E. coli*.

éxito más de dos genes clonados en un mismo vector y complejos binarios y ternarios. El empleo de vectores compatibles diferentes requerirá no solo orígenes de replicación diferentes sino también marcadores de selección diferentes a fin de garantizar la adecuada estabilidad de cada vector (14, 59). Sistemas comerciales de co-expresión consistentes en 4 vectores bicistrónicos diferentes, todos con orígenes de replicación y marcadores de selección diferentes, permiten la expresión simultánea en un mismo hospedero de hasta 8 proteínas diferentes (58).

3.8 Tolerancia o disminución de la toxicidad al hospedero


La expresión recombinante de algunos genes heterólogos puede resultar extremadamente tóxica para el hospedero, interfiriendo con su fisiología y originando una afectación significativa del crecimiento e incluso la muerte del cultivo. La toxicidad de estos genes puede variar desde moderada hasta muy elevada, dependiendo del nivel de expresión y por consiguiente, la cantidad de proteína recombinante que es capaz de interferir con la fisiología

del hospedero y causar afectación del crecimiento o la muerte. Varias estrategias son empleadas para lograr expresar y producir estas proteínas tóxicas, entre ellas el empleo de promotores altamente regulados, la selección de cepas hospederas más resistentes al producto potencialmente tóxico, el empleo de promotores antisentido, la generación de proteínas de fusión que neutralicen la toxicidad, utilización de vectores de bajo número de copias, empleo de sistemas de expresión libres de células, entre otras (2, 35, 60).

4. Conclusiones

La producción de proteínas por vía recombinante ofrece una vía simple, rápida, económicamente rentable y altamente eficiente para suplir los actuales requerimientos de proteínas puras y biológicamente activas, tanto para fines de investigación como para usos terapéuticos o industriales. La selección del sistema de expresión apropiado dependerá del origen y propiedades biológicas de la proteína, las características del hospedero, la futura aplicación de la proteína, entre otros factores (Fig. 3). En los últimos años ha sido desarrollada una gran variedad de sistemas de expresión recombinante basados en hospederos procarióticos, levaduras, hongos filamentosos, protozoos, baculovirus/células de insectos, células de plantas o células de mamíferos.

Sin embargo, *E. coli* continúa siendo el hospedero de elección para la producción de proteínas recombinantes dadas sus reconocidas ventajas. Esta bacteria es empleada con éxito en la actualidad para la producción de proteínas humanas con fines terapéuticos (61).

A pesar del desarrollo de diversos sistemas de expresión y del considerable avance en las tecnologías de expresión y purificación, en muchas ocasiones resulta imprescindible la aplicación de estrategias de optimización de la expresión recombinante, a fin de garantizar un adecuado rendimiento y estabilidad de las propiedades biológicas de la proteína. El empleo de las cepas hospederas adecuadas, el diseño óptimo de los vectores de expresión, la variación de las condiciones de crecimiento o la generación de proteínas de fusión son algunas de las estrategias más utilizadas para lograr alcanzar los estándares de rendimiento y calidad esperados. Una vez optimizada la expresión recombinante empleando el sistema de expresión apropiado, se procederá al escalado del cultivo a grandes volúmenes y la purificación de la proteína de interés. Estudios adicionales podrán ser necesarios con vistas a optimizar las condiciones *in vitro* (pH, fuerza iónica, composición del solvente, crioprotectores) que garanticen la estabilidad físico-química y actividad biológica de la proteína una vez purificada. 

REFERENCIAS

1. Jia B, Jeon CO (2016) High-throughput recombinant protein expression in *Escherichia coli*: current status and future perspectives. *Open Biol* 6: 160196.
2. Klammt C, Schwarz D, Lohr F, Schneider B, Dotsch V, Bernhard F (2006) Cell-free expression as an emerging technique for the large scale production of integral membrane protein. *FEBS J* 273: 4141-4153.
3. Schmidt FR (2004) Recombinant expression systems in the pharmaceutical industry. *Appl Microbiol Biotechnol* 65: 363-372.
4. 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.
5. Martinez JL, Liu L, Petranovic D, Nielsen J (2012) Pharmaceutical protein production by yeast: towards production of human blood proteins by microbial fermentation. *Curr Opin Biotechnol* 23: 965-971.
6. Streatfield SJ (2007) Approaches to achieve high-level heterologous protein production in plants. *Plant Biotechnol J* 5: 2-15.
7. Tiwari S, Verma PC, Singh PK, Tuli R (2009) Plants as bioreactors for the production of vaccine antigens. *Biotechnol Adv* 27: 449-467.
8. Jarvis DL (2009) Baculovirus-insect cell expression systems. *Methods Enzymol* 463: 191-222.
9. Cregg JM, Tolstorukov I, Kusari A, Sunga J, Madden K, Chappell T (2009) Expression in the yeast *Pichia pastoris*. *Methods Enzymol* 463: 169-189.
10. Lalonde ME, Durocher Y (2017) Therapeutic glycoprotein production in mammalian cells. *J Biotechnol* 251: 128-140.
11. Zemella A, Thoring L, Hoffmeister C, Kubick S (2015) Cell-free protein synthesis: pros and

- cons of prokaryotic and eukaryotic systems. *Chembiochem* 16: 2420-2431.
12. Gomez S, Lopez-Esteba M, Fernandez FJ, Suarez T, Vega MC (2016) Alternative eukaryotic expression systems for the production of proteins and protein complexes. *Adv Exp Med Biol* 896: 167-184.
 13. Rosano GL, Ceccarelli EA (2014) Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front Microbiol* 5: 172.
 14. Sorensen HP, Mortensen KK (2005) Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *J Biotechnol* 115: 113-128.
 15. Peti W, Page R (2007) Strategies to maximize heterologous protein expression in *Escherichia coli* with minimal cost. *Protein Expr Purif* 51: 1-10.
 16. Zerbs S, Frank AM, Collart FR (2009) Bacterial systems for production of heterologous proteins. *Methods Enzymol* 463: 149-168.
 17. Chen R (2012) Bacterial expression systems for recombinant protein production: *E. coli* and beyond. *Biotechnol Adv* 30: 1102-1107.
 18. Fedulova N, Hanrieder J, Bergquist J, Emren LO (2007) Expression and purification of catalytically active human PHD3 in *Escherichia coli*. *Protein Expr Purif* 54: 1-10.
 19. Jeiranikhameneh M, Moshiri F, Keyhan Falasafi S, Zomorodipour A (2017) Designing signal peptides for efficient periplasmic expression of human growth hormone in *Escherichia coli*. *J Microbiol Biotechnol* 27: 1999-2009.
 20. Sahdev S, Khattar SK, Saini KS (2008) Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Mol Cell Biochem* 307: 249-264.
 21. Hunt I (2005) From gene to protein: a review of new and enabling technologies for multi-parallel protein expression. *Protein Expr Purif* 40: 1-22.
 22. Demain AL, Vaishnav P (2009) Production of recombinant proteins by microbes and higher organisms. *Biotechnol Adv* 27: 297-306.
 23. Singh SM, Panda AK (2005) Solubilization and refolding of bacterial inclusion body proteins. *J Biosci Bioeng* 99: 303-310.
 24. Sorensen HP, Mortensen KK (2005) Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microb Cell Fact* 4: 1.
 25. Burgess RR (2009) Refolding solubilized inclusion body proteins. *Methods Enzymol* 463: 259-282.
 26. Krishnan B, Hedstrom L, Hebert DN, Gierasch LM, Gershenson A (2017) Expression and purification of active recombinant human alpha-1 antitrypsin (AAT) from *Escherichia coli*. *Methods Mol Biol* 1639: 195-209.
 27. Ma Y, Yu J, Lin J, Wu S, Li S, Wang J (2016) High efficient expression, purification, and functional characterization of native human epidermal growth factor in *Escherichia coli*. *Biomed Res Int* 2016: 3758941.
 28. Wang M, Jiang S, Wang Y (2016) Recent advances in the production of recombinant subunit vaccines in *Pichia pastoris*. *Bioengineered* 7: 155-165.
 29. Kost TA, Condreay JP, Jarvis DL (2005) Baculovirus as versatile vectors for protein expression in insect and mammalian cells. *Nat Biotechnol* 23: 567-575.
 30. Kost TA, Kemp CW (2016) Fundamentals of baculovirus expression and applications. *Adv Exp Med Biol* 896: 187-197.
 31. Niimi T (2012) Recombinant protein production in the eukaryotic protozoan parasite *Leishmania tarentolae*: a review. *Methods Mol Biol* 824: 307-315.
 32. Niimi T (2016) *Leishmania tarentolae* for the production of multi-subunit complexes. *Adv Exp Med Biol* 896: 155-165.
 33. Ahmad A, Pereira EO, Conley AJ, Richman AS, Menassa R (2010) Green biofactories: recombinant protein production in plants. *Recent Pat Biotechnol* 4: 242-259.
 34. Thoring L, Dondapati SK, Stech M, Wustenhagen DA, Kubick S (2017) High-yield production of "difficult-to-express" proteins in a continuous exchange cell-free system based on CHO cell lysates. *Sci Rep* 7: 11710.
 35. Saida F, Uzan M, Odaert B, Bontems F (2006) Expression of highly toxic genes in *E. coli*: special strategies and genetic tools. *Curr Protein Pept Sci* 7: 47-56.
 36. Browning DF, Busby SJ (2016) Local and global regulation of transcription initiation in bacteria. *Nat Rev Microbiol* 14: 638-650.
 37. Tegel H, Ottosson J, Hober S (2011) Enhancing the protein production levels in *Escherichia coli* with a strong promoter. *FEBS J* 278: 729-739.
 38. Makrides SC (1996) Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol Rev* 60: 512-538.
 39. Estrem ST, Ross W, Gaal T, Chen ZW, Niu W, Ebright RH, Gourse RL (1999) Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase alpha subunit. *Genes Dev* 13: 2134-2147.

40. Kane JF (1995) Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Curr Opin Biotechnol* 6: 494-500.
41. Brunelle JL, Green R (2014) One-dimensional SDS-polyacrylamide gel electrophoresis (1D SDS-PAGE). *Methods Enzymol* 541: 151-159.
42. Balbas P (2001) Understanding the art of producing protein and nonprotein molecules in *Escherichia coli*. *Mol Biotechnol* 19: 251-267.
43. Galloway CA, Sowden MP, Smith HC (2003) Increasing the yield of soluble recombinant protein expressed in *E. coli* by induction during late log phase. *Biotechniques* 34: 524-526, 528, 530.
44. Costa S, Almeida A, Castro A, Domingues L (2014) Fusion tags for protein solubility, purification and immunogenicity in *Escherichia coli*: the novel Fh8 system. *Front Microbiol* 5: 63.
45. Garcia-Fruitos E, Gonzalez-Montalban N, Morell M, Vera A, Ferraz RM, Aris A, Ventura S, Villaverde A (2005) Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. *Microb Cell Fact* 4: 27.
46. Nahalka J (2008) Physiological aggregation of maltodextrin phosphorylase from *Pyrococcus furiosus* and its application in a process of batch starch degradation to alpha-D-glucose-1-phosphate. *J Ind Microbiol Biotechnol* 35: 219-223.
47. Villaverde A, Corchero JL, Seras-Franzoso J, Garcia-Fruitos E (2015) Functional protein aggregates: just the tip of the iceberg. *Nanomedicine (Lond)* 10: 2881-2891.
48. Malik A (2016) Protein fusion tags for efficient expression and purification of recombinant proteins in the periplasmic space of *E. coli*. *3 Biotech* 6: 44.
49. Choi JH, Lee SY (2004) Secretory and extracellular production of recombinant proteins using *Escherichia coli*. *Appl Microbiol Biotechnol* 64: 625-635.
50. Zhou Y, Lu Z, Wang X, Selvaraj JN, Zhang G (2018) Genetic engineering modification and fermentation optimization for extracellular production of recombinant proteins using *Escherichia coli*. *Appl Microbiol Biotechnol* 102: 1545-1556.
51. Albinia AM, Matos CF, Branston SD, Freedman RB, Keshavarz-Moore E, Robinson C (2013) High-level secretion of a recombinant protein to the culture medium with a *Bacillus subtilis* twin-arginine translocation system in *Escherichia coli*. *FEBS J* 280: 3810-3821.
52. Ryan DP, Tremethick DJ (2016) A dual affinity-tag strategy for the expression and purification of human linker histone H1.4 in *Escherichia coli*. *Protein Expr Purif* 120: 160-168.
53. Malhotra A (2009) Tagging for protein expression. *Methods Enzymol* 463: 239-258.
54. Butt TR, Edavettal SC, Hall JP, Mattern MR (2005) SUMO fusion technology for difficult-to-express proteins. *Protein Expr Purif* 43: 1-9.
55. Zhao X, Li G, Liang S (2013) Several affinity tags commonly used in chromatographic purification. *J Anal Methods Chem* 2013: 581093.
56. Rozkov A, Enfors SO (2004) Analysis and control of proteolysis of recombinant proteins in *Escherichia coli*. *Adv Biochem Eng Biotechnol* 89: 163-195.
57. Busso D, Peleg Y, Heidebrecht T, Romier C, Jacobovitch Y, Dantes A, Salim L, Troesch E, Schuetz A, Heinemann U, Folkers GE, Geerlof A, Wilmanns M, Polewacz A, Quedenau C, Bussow K, Adamson R, Blagova E, Walton J, Cartwright JL, Bird LE, Owens RJ, Berrow NS, Wilson KS, Sussman JL, Perrakis A, Celie PH (2011) Expression of protein complexes using multiple *Escherichia coli* protein co-expression systems: a benchmarking study. *J Struct Biol* 175: 159-170.
58. Diebold ML, Fribourg S, Koch M, Metzger T, Romier C (2011) Deciphering correct strategies for multiprotein complex assembly by co-expression: application to complexes as large as the histone octamer. *J Struct Biol* 175: 178-188.
59. Haffke M, Marek M, Pelosse M, Diebold ML, Schlattner U, Berger I, Romier C (2015) Characterization and production of protein complexes by co-expression in *Escherichia coli*. *Methods Mol Biol* 1261: 63-89.
60. Schwarz D, Dotsch V, Bernhard F (2008) Production of membrane proteins using cell-free expression systems. *Proteomics* 8: 3933-3946.
61. Kamionka M (2011) Engineering of therapeutic proteins production in *Escherichia coli*. *Curr Pharm Biotechnol* 12: 268-274.