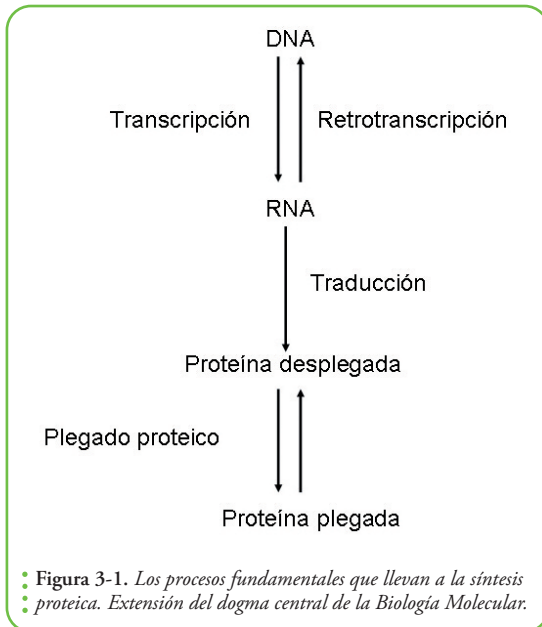


Producción de proteínas en el laboratorio. Tercera Parte

El código genético. Genes. Transcripción y traducción

Los organismos vivos, como por ejemplo, nosotros o las bacterias, y los que están en el límite de lo vivo y lo no vivo, los virus, requieren para poder reproducirse o autoreplicarse una maquinaria biológica, increíblemente, compleja y organizada tanto espacial como temporalmente. La información para llevar adelante estos procesos y las instrucciones para fabricar esta maquinaria debe estar almacenada y debe poder ser transmitida a la descendencia. Para almacenarla y poder recuperarla y utilizarla es necesario que exista un código. Probablemente, uno de los más asombrosos descubrimientos haya sido la capacidad de los ácidos nucleicos, DNA y RNA de portar información. Gran parte de esa información debe ser transcrita en mensajes o mRNAs (la **m** es de mensajero) para que, posteriormente, pueda ser traducida en secuencias de proteínas por los ribosomas, que son una parte importante de la maquinaria de biosíntesis de proteínas (ver Figura 3-1).



Estas proteínas realizarán una gran variedad de funciones, por ejemplo, transportar oxígeno como en el caso de la hemoglobina, otras tendrán la función de sintetizar DNA, RNA, el caso de las DNA y RNA polimerasas y otras, por ejemplo, formarán parte de los ribosomas. Necesitamos proteínas para sintetizar proteínas.

La información genética para sintetizar proteínas está codificada, entonces, en moléculas de DNA doble cadena en la mayor parte de los organismos; aunque ciertos virus usan DNA o RNA de cadena simple o RNA de cadena doble. Las secuencias están codificadas por cuatro tipos de nucleótidos: A (adenina), T (timina), G (guanina), C (citósina) en el DNA. En el RNA la T es reemplazada por U (uracilo). Las regiones del material genético que codifican se llaman genes. Un gen puede codificar para

una o más proteínas. Pero, también, un gen puede codificar para un RNA. Existen RNAs con funciones específicas. Por ejemplo, RNAs ribosomales, RNA de transferencia y las ribozimas. Las regiones a ambos lados de cada gen, a veces incluyendo regiones muy distantes en la secuencia, las distancias, en este caso se miden en número de bases o kilobases, suelen tener funciones regu-

latorias de la expresión de esos genes. La regulación de la expresión genética es un tema apasionante y central en biología molecular.

En este punto es importante recordar que los ácidos nucleicos también son polímeros y que poseen una estructura particular. En el caso del DNA de doble cadena, la estructura más estable, por lo general, es una doble hélice. La doble hélice tiene una característica físico-química notable que soluciona entre otras cuestiones el problema de la autoreplicación de la información porque ambas hebras son complementarias. Con la información molecular contenida en una de ellas podemos reconstruir la otra y viceversa.

Cada hebra tiene un extremo 5' y un extremo 3', cada hebra entonces tiene polaridad.

Los genes se leen en el sentido 5' a 3' y existen secuencias de iniciación y secuencias de finalización que nos permiten definir qué secuencia es un gen y qué secuencia no lo es.

A su vez, una de las hebras se define como la hebra codificante o con sentido, de esta se puede deducir qué secuencia tendrá la proteína que codifica en forma directa. La otra hebra, de polaridad opuesta *posee secuencia complementaria* se la conoce como hebra no codificante, antisentido o hebra molde.

En organismos eucarióticos, entre ellos nosotros, la parte de un gen que codifica para una proteína suele estar dividida en fragmentos llamados exones.

Los exones están separados entre sí por intrones. Los intrones, son regiones no codificantes y pueden ser muy extensos, por ejemplo 10 kilobases (10.000 bases). En cambio, en organismos procarióticos (por ejemplo las bacterias), el DNA que codifica para una cadena polipeptídica lo hace en un segmento continuo sin intrones.

Volviendo un poco para atrás, el primer paso en la biosíntesis de una proteína es la producción de moléculas mensajeras (los mRNAs). La transcripción de la secuencia de nucleótidos del DNA en un mRNA de secuencia complementaria. El descubrimiento de la existencia de los mensajeros también fue uno de esos descubrimientos importantes y maravillosos, porque en células eucariotas alguna molécula debía transportar la información codificada en el DNA, cuya ubicación es nuclear (está en el núcleo) hasta el citoplasma, donde están los ribosomas y se realiza la síntesis de proteínas. La transcripción o síntesis de un mRNA ocurre sólo si una serie de proteínas reguladoras de este proceso están correctamente posicionadas en la región del DNA que regula la transcripción del gen conocida como región promotora. Cuando esto ocurre el gen es transcrito por una

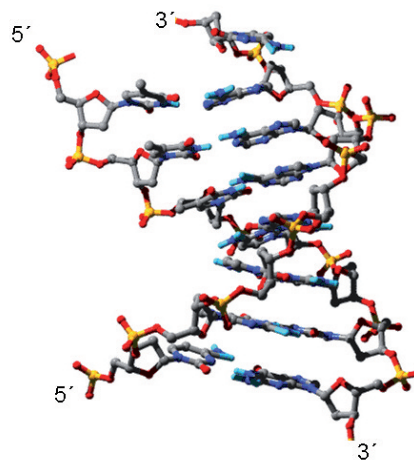


Figura 3-2. Un pequeño fragmento de DNA.

Se muestran los extremos 5' (fosfato) y 3' (-OH). Notar la estructura helicoidal del DNA doble cadena y la polaridad opuesta de ambas hebras, químicamente, complementarias. Los átomos se representan de la siguiente manera: en naranja los fosfatos, en rojo los átomos de oxígeno, en azul los de nitrógeno, en gris los de carbono y en celeste los átomos de hidrógeno, que participan en puentes de hidrógeno. (ver Figura 3-2)

RNA polimerasa. También hay proteínas que pueden inhibir la transcripción, estos reciben el nombre de represores y silenciadores.

La transcripción es iniciada usando a la hebra no codificando como molde (por esto también se la llama hebra molde) es decir se sintetiza en este proceso desde 5' a 3' un polímero de RNA complementario a la hebra molde. Si comparamos este proceso en organismos procarióticos y eucarióticos, podemos ver que existen importantes diferencias. En eucariotas existen dos modificaciones del mRNA conocidas como modificaciones post transcripcionales: 5' cap y poliadenilación del extremo 3', se les agraga una cola de poliA, muchas adeninas. Ambas modificaciones son importantes porque le otorgan al mRNA mayor estabilidad y entonces es más difícilmente degradable. Además,

en el caso de los RNAs eucariotas los intrones deben ser removidos. El proceso de remoción de intrones se llama *splicing* (ver Figura 3-3).

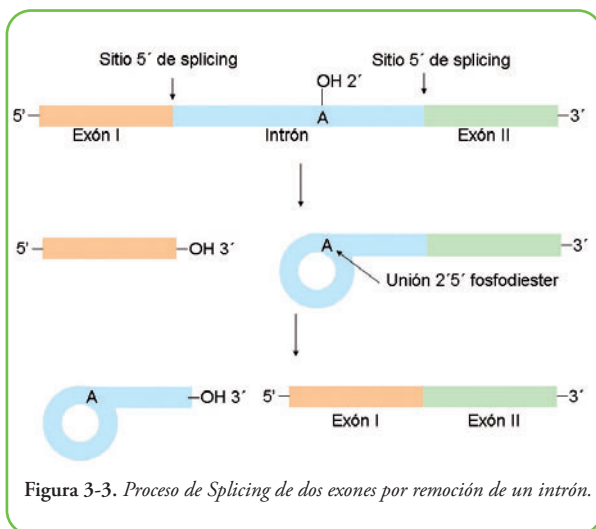


Figura 3-3. Proceso de Splicing de dos exones por remoción de un intrón.

El extremo 5' del intrón reacciona con el oxidrilo 2' de un nucleótido de adenina propio del intrón. El grupo oxidrilo 3' del exon I reaccionará atacando la unión fosfodiéster existente entre el intrón y el exón II para generar un nuevo enlace fosfodiéster con el fosfato 5' del exón II.

de los mRNA eucariotas ya sin intrones, son traducidos en cadenas polipeptídicas en el citoplasma por una maquinaria, espectacularmente, compleja y gigantesca formada por tRNAs (RNA de transferencia) rRNAs (RNA ribosomales) y una gran cantidad de proteínas. La estructura de esta maquinaria, los ribosomas y el mecanismo de biosíntesis se conoce con gran detalle.

Hasta hace relativamente poco tiempo, parecía prácticamente imposible pensar en obtener estructuras con alta resolución de los ribosomas completos.

Hoy en día, es factible, para cualquiera de los lectores de este párrafo, que cuente con una computadora conectada a Internet, obtener modelos experimentales de alta resolución de las estructuras atómicas de ambas subunidades ribosomales.

Inclusive, se pueden obtener de la PDB estructuras en las que se han incorporado al mRNA y tRNA.

Una maquinaria muy compleja llamada spliceosoma es la que se encarga de hacerlo. Algunas variaciones en este proceso conocidas como *splicing* alternativo combinan a los exones de manera alternativa, por ejemplo eliminando uno. Por lo que un único gen puede codificar, en realidad, una gran variedad de proteínas. Se cree que en el ser humano cerca del 50 % de los genes son afectados por algún tipo de *splicing* alternativo.

Este mecanismo contribuye de manera muy importante con la diversidad de proteínas existentes en el interior de una célula. Algunas de ellas pueden tener diferencias funcionales importantes.

Los mRNA, maduros en el caso

La traducción (biosíntesis ribosomal de proteínas) puede reproducirse, experimentalmente, en sistemas simplificados a partir de extractos libres de células obtenidos por la ruptura mecánica de las células y separación y/o purificación de los componentes necesarios que forman parte de la maquinaria biosintética, precursores y la fuente de energía necesaria. La traducción del mRNA en proteínas es realizada mediante *la lectura* (no hay lector sino moléculas que interaccionan entre sí reconociéndose) del código genético: cada triplete o codón de la región codificante de nucleótidos de un mRNA tiene su significado: información para un aminoácido o para terminar la síntesis de la proteína. Cada uno de los cuatro nucleótidos (A, G, U, o C) puede ocupar cada una de las tres posiciones en un codón. Así, hay $4^3=64$ posibles combinaciones de secuencias de tres nucleótidos. Se ve, fácilmente, que hay más combinaciones que los 20 aminoácidos con los que se sintetizan las proteínas. La asignación sistemática de cada codón a un aminoácido o a una señal de terminación (**STOP**) de la traducción fue un paso importantísimo en el campo de la biología molecular y permitió dilucidar por completo el código genético (Figura 3-4).

Con esta asignación y la secuenciación de DNA se hizo posible determinar cuál es la secuencia de aminoácidos que codifica un gen directamente a partir de la secuencia de nucleótidos. Entonces la secuencia de DNA en sentido 5´ hacia 3´ pero leída en tripletes corresponde a la secuencia de aminoácidos de la proteína codificada leída en dirección desde el N-terminal hacia el C-terminal.

Una característica notable del código genético es que se trata de un código degenerado. La mayor parte de los aminoácidos pueden ser codificados por más de un codón. Salvo metionina y triptofano, que son las únicas dos excepciones, codificados únicamente por un triplete, AUG y UGG, respectivamente. Los codones que codifican para el mismo aminoácido se llaman codones sinónimos.

dones sinónimos. Por ejemplo, si empleamos un organismo procariota para producir en forma recombinante una proteína eucariota puede ocurrir que alguno de los codones sea muy poco utilizado y la síntesis de proteínas a partir de mRNAs que contengan este tipo de codones sea ineficiente. Como resultado produciremos poca proteína. Los codones que representan el mismo o un aminoácido con propiedades fisicoquímicas relacionadas, suelen tener secuencias similares. En muchos casos la tercera base del codón suele no ser significativa: los cuatro codones, difiriendo sólo en la tercera base, pueden representar al mismo aminoácido. En otros casos, queda determinada la asignación del aminoácido si

		SEGUNDO NUCLEÓTIDO				
		U	C	A	G	
PRIMER NUCLEÓTIDO	U	Phe	Ser	Tyr	Cys	U
		Phe	Ser	Tyr	Cys	C
		Leu	Ser	STOP	STOP	A
		Leu	Ser	STOP	Trp	G
	C	Leu	Pro	His	Arg	U
		Leu	Pro	His	Arg	C
		Leu	Pro	Gln	Arg	A
		Leu	Pro	Gln	Arg	G
	A	Ile	Thr	Asn	Ser	U
		Ile	Thr	Asn	Ser	C
		Ile	Thr	Lys	Arg	A
		Met	Thr	Lys	Arg	G
	G	Val	Val	Asp	Gly	U
		Val	Val	Asp	Gly	C
		Val	Val	Glu	Gly	A
		Val	Val	Glu	Gly	G

Figura 3-4. El código genético. Cada aminoácido está codificado por uno o más codones (grupos de 3 bases).

Un punto importante, particularmente para los ingenieros y productores de proteínas, es que existen frecuencias de uso de co-

se trata de una purina (A o G) o una pirimidina (U o C) como tercera base del codón, AGU y AGC codifican para el aminoácido serina; AGA y AGG para arginina. La reducción de especificidad en la última posición es conocida como degeneración de la tercera base.

Por otro lado hay que resaltar que 3 de las 64 posibilidades que corresponden a los codones UAA, UAG y UGA no representan aminoácidos sino señales de terminación de la traducción o señales **stop**. Estos codones especifican la finalización de la síntesis de la proteína que está siendo sintetizada.

La síntesis de proteínas se basa en el ensamblado de una línea de producción en la que los ribosomas proceden con un elevado consumo de energía (provista por la hidrólisis de GTP a lo largo del mensajero) trayendo los aminoacil-tRNAs que proveerán los bloques de construcción del producto proteico.

El proceso es complejo y están involucradas muchas reacciones químicas. La iniciación, que es la **primera etapa**, involucra reacciones que preceden la formación del enlace peptídico entre el primer y el segundo aminoácido de la proteína. Esta etapa requiere que el ribosoma se una al mRNA para formar un complejo, conocido como complejo de iniciación que contiene al primer aminoacil-tRNA. Este paso, por lo general, determina la velocidad de la traducción porque es un paso lento. La secuencia en el mRNA en la que ocurre la iniciación en *Escherichia coli* es el sitio de unión del ribosoma (en inglés RBS) o Shine-Dalgarno; es una secuencia corta (5' AGGAGG 3') que precede al AUG en menos de 10 bases (AAACAGGAGGAUUACCCAUG). Esta distancia es muy importante para que la iniciación tenga lugar y **debe** tenerse muy en cuenta al diseñar construcciones genéticas para producción de proteínas recombinantes en bacterias. La secuencia RBS es reconocida por la subunidad menor del ribosoma mediante la interacción con un rRNA (RNA ribosomal 16S) que contiene la secuencia complementaria 3' UCCUCC 5' al mRNA formando el complejo de iniciación, entonces se une la subunidad mayor generando el ribosoma completo.

La **segunda etapa** es la elongación. Esta etapa incluye todas las reacciones desde la síntesis del primer enlace peptídico entre el primer aminoácido que siempre es una metionina y el segundo hasta la adición del último aminoácido.

Los aminoácidos se van agregando de a uno a la cadena polipeptídica que se va sintetizando. La adición de un aminoácido a la cadena es un paso rápido en la síntesis proteica. El aminoacil-tRNA entrante, que aportará el próximo aminoácido, ocupa una región del ribosoma conocida como sitio A. Este sitio expone el codón del mRNA complementario al anticodón del tRNA. Por otro lado en otro sitio, conocido como sitio P, está el peptidil-tRNA (la proteína que se está sintetizando). El enlace peptídico se forma como consecuencia de una reacción en la cual el polipéptido que es llevado como peptidil-tRNA es transferido al aminoacil-tRNA que está posicionado en el sitio A. Esta transferencia es catalizada por componentes de la subunidad ribosomal mayor.

Así, un nuevo peptidil-tRNA fue creado en el sitio A. Este peptidil-tRNA es más largo (un aminoácido más largo) que el peptidil-tRNA existente en el ciclo anterior. El tRNA, desde el que se transfiere el peptidil luego de esta reacción de transferencia queda sin peptidil (deacilado) y ya no es afín por el sitio P pero deja el ribosoma desde un sitio transitorio llamado sitio E, pero mientras este t-RNA deacilado está en el sitio

E la afinidad del sitio A por un nuevo aminoacil-tRNA es muy baja, esto no permite que otro aminoacil-tRNA ingrese hasta que el ribosoma esté listo.

Luego se produce un movimiento molecular por el cual el ribosoma queda posicionado en el siguiente codón. Este proceso es la translocación. Con este movimiento el nuevo peptidil-tRNA queda ubicado en el sitio P y el sitio A expone el siguiente codón vacante y con afinidad incrementada para la entrada del siguiente aminoacil-tRNA. Así se repite el ciclo.

Por último, la **tercera etapa** involucra la liberación del polipéptido que se sintetizó completo y, al mismo tiempo, la disociación del complejo ribosoma-mensajero para dar ribosoma libre más mensajero libre. La biosíntesis ribosomal de proteínas es un proceso relativamente rápido y depende en gran medida de la temperatura. Por ejemplo, a temperaturas bajas (10°C) es muy lento. En bacterias a 37°C se agregan unos 15 aminoácidos por segundo al polipéptido. La dependencia de la velocidad de síntesis con de la temperatura es otra variable que los proteinólogos tienen siempre en cuenta a la hora de producir una proteína de interés. Por ejemplo, a veces, se obtienen buenos rendimientos cuando se disminuye la velocidad de síntesis, cultivando las bacterias a 30°C, en este caso los cultivos deben crecerse por más tiempo.

Producción de proteínas recombinantes en bacterias

Para hacer los estudios estructurales básicos de una proteína de interés son necesarios, aproximadamente, unos 10 ó 20 mg de proteína pura. Esto, que parece un problema sencillo, ha sido en el pasado uno de los problemas centrales en la biología estructural. La mayor parte de las proteínas se produce en muy bajas concentraciones dentro de la célula. Por otro lado dentro de una célula coexisten miles de proteínas diferentes, cerca de unas 10000 proteínas diferentes.

Probablemente, una de las cuestiones más intrigantes es:

¿Cómo podemos hacer para producir una proteína?

La respuesta hoy en día no es tan difícil de responder. Sin embargo, hace unos 60 años podría haberse tratado de un capítulo de ciencia ficción.

La clave está en que podemos usar a otros organismos como fábricas. En este ejemplo vamos a usar una cepa de bacterias de la conocidísima especie *Escherichia coli* y vamos a ver que se puede cultivar a estas bacterias y controlarlas, de tal manera, que produzcan la proteína de interés. Pero para producir proteínas en *E. coli* necesitamos saber algunas cosas.

Tecnología del DNA recombinante.

Como hemos visto, la maquinaria de biosíntesis de proteínas de cualquier organismo funciona reconociendo señales que están codificadas en los ácidos nucleicos en el DNA y en el RNA. Así que *para poder pedirle a estas bacterias algunos favores* tendremos que entender muy bien el lenguaje molecular. Concretamente, el favor será que después de haber incorporado un gen foráneo utilice la información codificada en la secuencia del gen para la síntesis de una proteína de interés. La tecnología para llevar a delante este proceso hasta la biosíntesis de la proteína se llama tecnología del DNA recombinante; combinaremos DNA

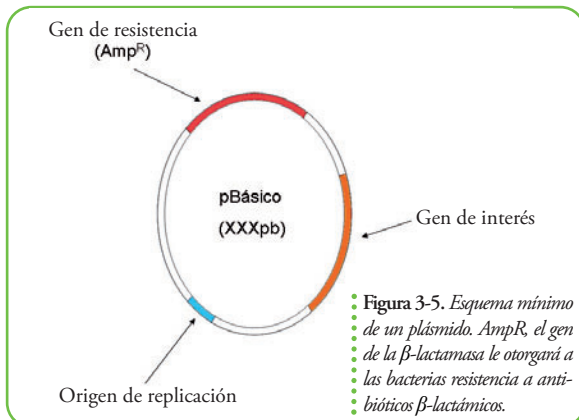
de origen bacteriano con DNA sintetizado en el laboratorio. Cualquier error en nuestras construcciones moleculares o en el dictado de las instrucciones llevaría a que la maquinaria de biosíntesis no lleve adelante el proceso biosintético que nosotros deseamos que realice.

Por esto, como primera medida debemos conocer la estructura mínima que debe tener el mensaje para ser realmente interpretado como mensaje.

El mensaje, como vimos, recibe el nombre de gen. Si bien no todos los genes codifican mensajes para la síntesis de proteínas, aquí nos enfocaremos en entender la estructura de un gen destinado a la síntesis de proteínas en un sistema de producción sencillo y comúnmente usado.

El gen de interés que no es otra cosa más que un fragmento de DNA con características especiales, debe introducirse en un sistema que se replique y que se distribuya cada vez que una bacteria de nuestro cultivo se divida en las bacterias de la siguiente generación. Esto se logra insertando el gen en un sistema que se autorreplique de manera sincronizada con la replicación del cromosoma bacteriano. Los sistemas más comúnmente usados son los plásmidos. Cada plásmido está formado por una molécula de DNA doble cadena circular y consta de un conjunto pequeño de genes propios del plásmido. Para replicarse dentro de la bacteria los plásmidos deben tener un sitio llamado origen de la replicación que debe ser reconocido por la maquinaria de replicación de DNA de la bacteria en la que se replica, en este caso *E. coli*. En la mayor parte de los plásmidos tiene algún gen que le otorga a la bacteria en la que ingresa alguna ventaja con respecto a las bacterias que no lo tienen.

Típicamente, esta ventaja es un gen para la síntesis de alguna proteína que le da resistencia contra un antibiótico presente en el medio en el que se la cultiva, en nuestro ejemplo es la resistencia a la ampicilina, esta resistencia estará dada por la presencia del gen de la β -lactamasa. Esto garantiza que todas las bacterias del cultivo contengan en su interior al plásmido. Las bacterias además de sintetizar la proteína de interés sintetizarán entre otras macromoléculas a la β -lactamasa.



En la **Figura 3-5** se muestra un esquema mínimo de la organización de un plásmido.

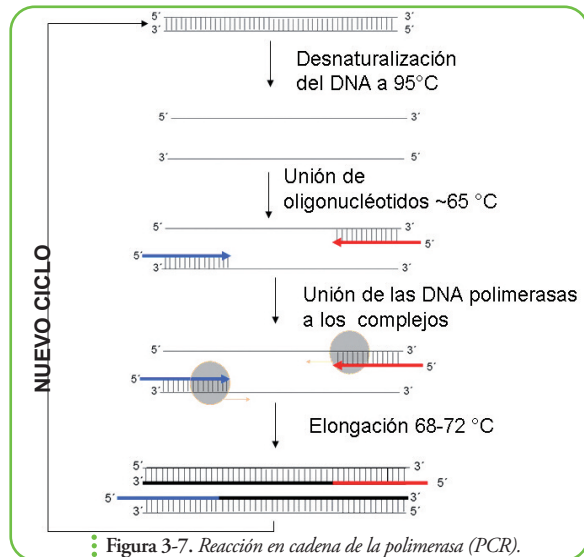
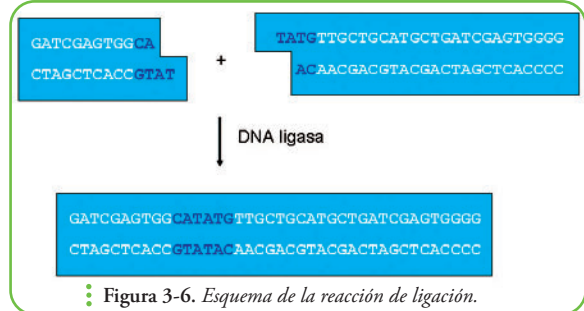
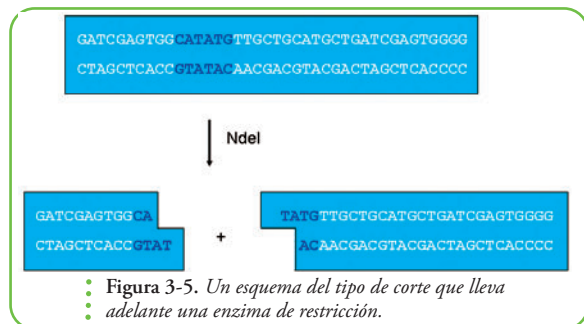
Por suerte contamos con algunas herramientas moleculares que llevaron muchos años de descubrimientos y de desarrollo porque, para poder insertar genes de interés en plásmidos, vamos a necesitar herramientas de diversos tipos:

- **Tijeras moleculares de DNA:** necesitaremos enzimas que catalicen el corte de la doble cadena de DNA en sitios extremadamente específicos. Estas enzimas se conocen con el nombre genérico de **enzimas de restricción** y las más comunes reconocen secuencias palindrómicas en el DNA, por ejemplo Nde I reconoce la secuencia 5' CATATG 3' 3' GTATAC 5'

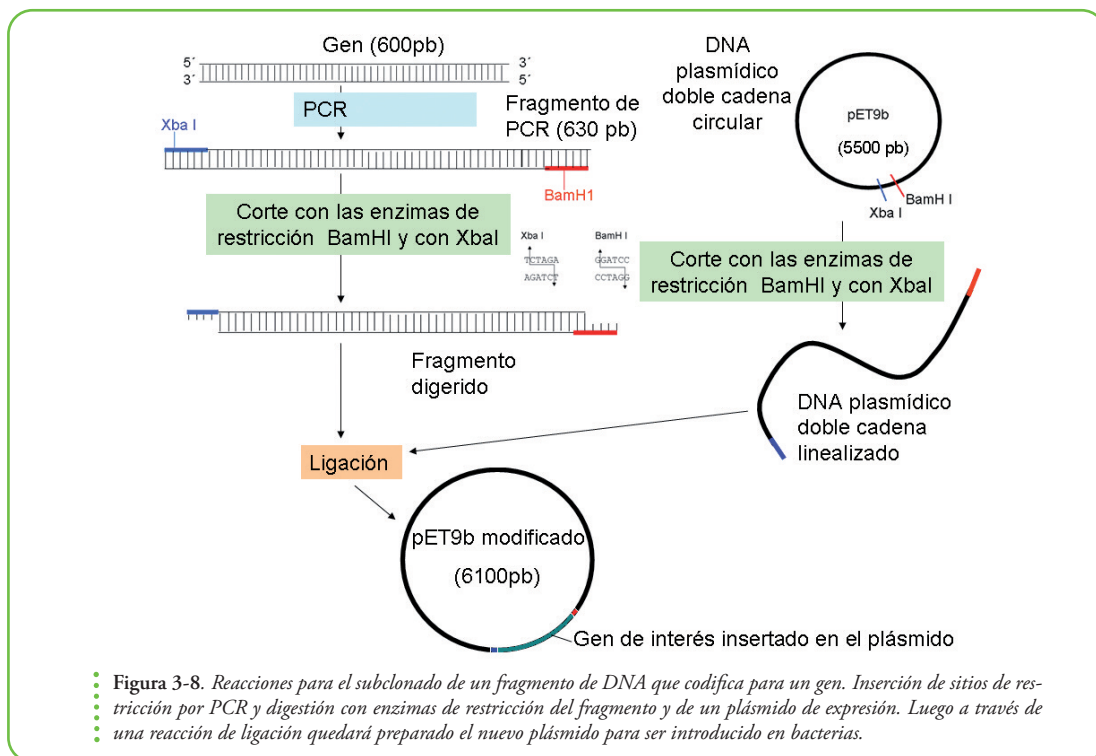
- **Pegamento:** vamos a necesitar enzimas que sean capaces de detectar los extremos de dos fragmentos de DNA y unirlos estas enzimas son las **DNA ligasas**.

Por otro lado, vamos a necesitar producir (amplificar) muchas moléculas idénticas de DNA para poder trabajar. Las bacterias se encargarán, por lo general, de producir plásmido y, así, podremos purificar gran cantidad de moléculas de plásmido con relativa facilidad. Sin embargo, cuando queramos producir en gran cantidad fragmentos pequeños o introducir mutaciones puntuales en el sitio específico en el gen de interés, usaremos una técnica, ahora muy común, conocida con el nombre de PCR, la reacción en cadena de la polimerasa. Vale la pena resaltar que la invención de la técnica de la PCR ha marcado un punto de inflexión en la biología molecular. Se trata de una reacción en la que se utilizan pequeños oligonucleótidos (conocidos con el nombre de *primers*) sobre los que se inicia el proceso de síntesis de DNA (elongación) en el que la DNA polimerasa usa como molde una de las dos hebras de DNA generando una copia fiel de la hebra complementaria. Esta reacción se lleva adelante en varios ciclos (típicamente 30 ciclos) y el número de moléculas de DNA se eleva con cada ciclo aproximadamente en forma exponencial (ver **Figura 3-7**). La clave de este proceso fue el descubrimiento y la aplicación de **DNA polimerasas** termoestables, que resisten al proceso de desnaturalización del DNA y funcionan a temperaturas cercanas a los 70 °C.

Así por PCR, entre otras cosas, podemos introducir sitios de restricción adecuados en el gen de interés, para poder insertarlo en un sitio específico. Usando enzimas de restricción cortaremos el fragmento de PCR para dejar los extremos preparados para su ligación en el plásmido. Para poder ligar el fragmento, el plásmido deberá digerirse con enzimas que generen extremos compatibles, esto se ejemplifica en la **Figura 3-8**.



Un ciclado típico. En la reacción deben estar presentes un conjunto de reactivos; entre ellos los desoxinucleótido trifosfato (dATP, dTTP, dGTP y dCTP), iones magnesio, un buffer para la regulación de un pH óptimo, la polimerasa, los oligonucleótidos iniciadores o *primers*; la complementación de las hebras de DNA se marca con pequeñas rayas negras verticales, los oligonucleótidos sintéticos se señalan con una flecha azul y con una roja y se indica la polaridad de cada hebra de DNA. (ver **Figura 3-7**)



Aquí, tanto el fragmento producto de la amplificación por PCR como el plásmido se cortaron con dos enzimas Xba I y BamH I esto determina que el fragmento únicamente se pueda ligar en el sentido correcto y, así, orientamos correctamente el inserto regenerando por complementación de bases del DNA los sitios de restricción.

Ahora debemos reintroducir la construcción química en bacterias. Sin embargo existen algunos detalles importantes que debemos tener muy en cuenta. De lo contrario no será posible inducir la expresión de proteína de interés (proteína recombinante):

a. Los genes que se expresan en bacterias requieren regiones promotoras de la transcripción del RNA mensajero (mRNA) correspondiente. Sin un promotor adecuado no habrá síntesis de mRNA. Estas regiones promotoras serán reconocidas por las RNA polimerasas que son las enzimas encargadas de la síntesis del mRNA.

b. También debemos tener presente que los genes a su vez pueden tener regiones de silenciamiento o represión de la transcripción. En estas regiones se unen proteínas que reconocen estas secuencias específicas e impiden la transcripción porque bloquean

Esto, muy comúnmente, se hace por electroporación de las bacterias. Se las somete a voltajes extremos (2500 V) durante un muy pequeño instante, esto origina micro poros por los que pueden ingresar las moléculas de DNA plasmídico.

la unión de las RNA polimerasas o impiden su avance. Ambas regiones (promotor más sitio de unión de represores) se conocen con el nombre de región operadora u operador, o región regulatoria de la transcripción.

c. Además, el gen debe contar con una secuencia de terminación de la transcripción, estas son secuencias que causan el desensamblado del complejo mRNA:RNA polimerasa y limitan la longitud del mRNA. Suelen ser regiones ricas en bases GC. Esto determina la presencia de regiones estructuradas en el RNA.

d. Otro punto importante es que, para que se lleve adelante la traducción, debe existir un sitio de iniciación que será reconocido por el ribosoma. Unas 10 pares de bases hacia el extremo 5' del codon ATG (AUG en el mRNA) que codifica para el primer residuo de la cadena polipeptídica (una metionina) debe estar presente la secuencia RBS o sitio de unión del ribosoma, también conocida como secuencia Shine-Dalgarno: AGGAGA. Si esta secuencia no estuviera presente el ribosoma bacteriano no reconocería el sitio de iniciación y tampoco sería posible inducir la expresión de proteína de interés.

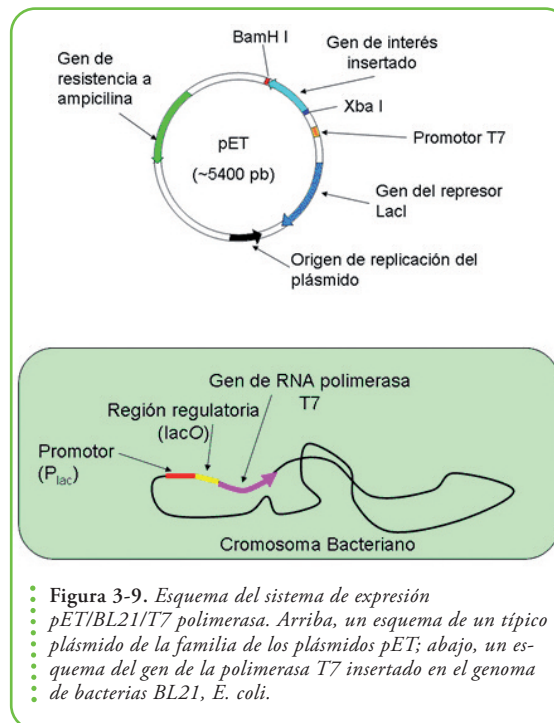
Debemos recordar que los ribosomas, que son la maquinaria que va a estar encargada de sintetizar la proteína utilizarán al RNA mensajero en la síntesis en mRNA estará codificada la información para ser traducida en una secuencia de aminoácidos.

e. Además, debe haber un codón de terminación de la traducción (*STOP*, en inglés) al finalizar la secuencia de codones que codifican para la proteína.

Hace ya unos cuantos años fue desarrollado un sistema que, actualmente, sigue siendo muy comúnmente usado para la producción de proteínas recombinantes en *E. coli*. Es un sistema de complejidad intermedia. Se trata del sistema de expresión basado en los plásmidos pET.

En este caso, se utiliza una cepa de bacterias que tiene algunas modificaciones genéticas. Entre ellas una de las más relevantes es que en el cromosoma bacteriano se introdujo el gen de una polimerasa de origen viral (del virus bacteriofago T7).

La transcripción del gen de la RNA polimerasa viral está en condiciones normales impedida, controlada por un represor; por técnicas de ingeniería genética se ha insertado en la región regulatoria del gen una secuencia que es reconocida por un represor, el represor LacI (lac de lactosa, el gen que codifica para este represor también ha sido insertado en el plásmido, ver Figura 3-9). Sin embargo, en presencia de lac-



tosa, o moléculas miméticas (la más común es el IPTG) el represor deja de unirse al sitio de regulación (operador) en el DNA y se desreprime la transcripción del gen de la RNA polimerasa T7.

Dicho de otra manera, el gen de la RNA polimerasa del bacteriofago se transcribe, fuertemente, en presencia de lactosa o IPTG generando gran cantidad de copias del mRNA. Este mRNA se traducirá mediante la maquinaria de traducción bacteriana, generando una elevada concentración de la proteína RNA polimerasa del fago.

Un esquema de algunos de estos aspectos se muestra en la **Figura 3-10**.

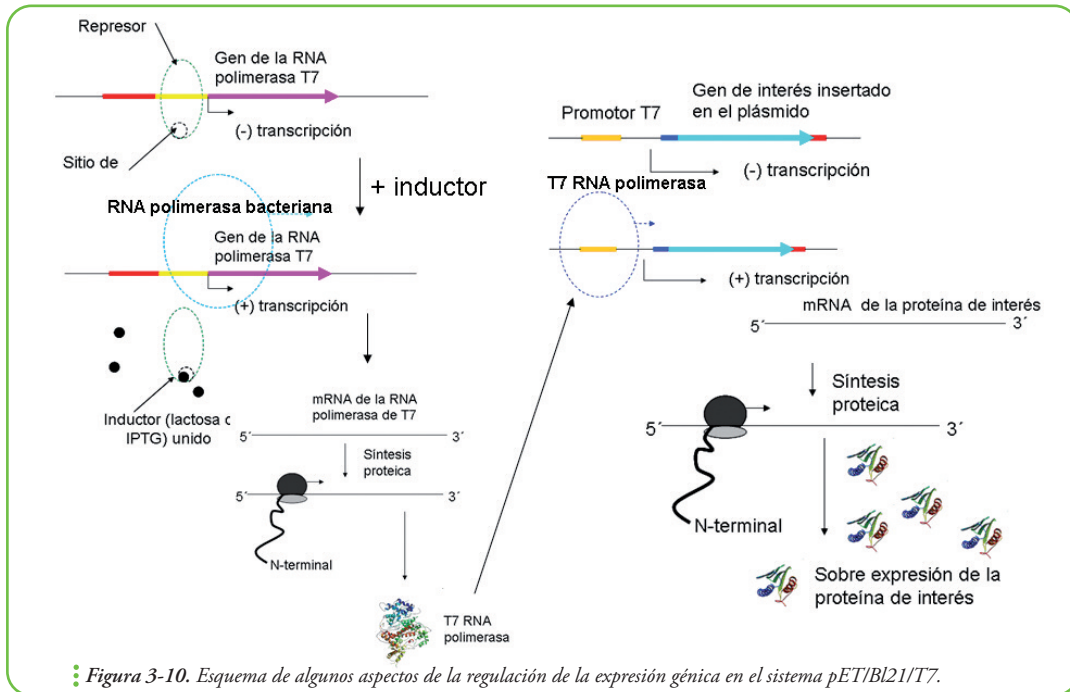


Figura 3-10. Esquema de algunos aspectos de la regulación de la expresión génica en el sistema pET/BL21/T7.

Es decir que el gen únicamente va a ser transcrito por esta RNA polimerasa y no por las RNA polimerasas endógenas de la bacteria, por esto se dice que es *inducible*. En unas pocas horas, después de haber agregado el inductor IPTG, se habrán transcrito suficientes moléculas del mRNA correspondiente a la proteína de interés y, a partir de este, se habrá traducido el mensaje dando como resultado la expresión (producción) de miles de moléculas de proteína.

Para aumentar la eficiencia de la producción se utilizan medios de cultivo ricos en fuentes de carbono y nitrógeno. Además se asegura la presencia de todos los minerales necesarios para el crecimiento adecuado de las bacterias. Los cultivos se realizan en medios líquidos y se garantiza una excelente transferencia de oxígeno requerido para el crecimiento aeróbico con una fuerte agitación de los frascos (típicamente se usan unos

Lo importante es que el gen de interés insertado, adecuadamente, en el plásmido queda bajo el control de una región promotora específica para la unión de la RNA polimerasa del bacteriofago T7 (promotor T7).

frascos especiales conocidos con el nombre de erlenmeyers). Después de unas tres o cuatro horas de cultivo con agitación a 37° C, podremos cosechar las bacterias. La forma más común de hacerlo es la centrifugación.

Para poder separar las bacterias del medio de cultivo usaremos una centrífuga. La centrifugación es un método por el que pueden separarse partículas presentes en líquidos. El aparato que hace este proceso es la centrífuga. La centrífuga aplica a la mezcla un movimiento circular; que provoca la sedimentación del sólido o de las partículas de mayor densidad. Para poder centrifugar las bacterias habrá que transvasar el cultivo a otro tipo de recipientes resistentes a las fuerzas centrífugas. A medida que las partículas que, en nuestro caso, son las bacterias comiencen a percibir la aceleración, producto del movimiento circular de un rotor (unas 5000 revoluciones por minuto (rpm) por 15 minutos serán suficientes), se irán depositando en el fondo de estos recipientes. Finalmente, si eliminamos el medio líquido remanente obtendremos una especie de pasta llamada *pellet*, o material insoluble. El *pellet*, en este caso, está formado efectivamente por bacterias que en su interior contienen la proteína de interés. Por litro de cultivo de bacterias, en condiciones normales, puede obtenerse entre 3 y 5 g de *bacterias inducidas*.

Mediante geles de poliacrilamida en presencia de SDS (ver Capítulo IV) podremos verificar, fácilmente, la expresión (producción) de la proteína.

Para esto tenemos que romper las bacterias y evaluar si la proteína de interés se conserva en forma soluble.

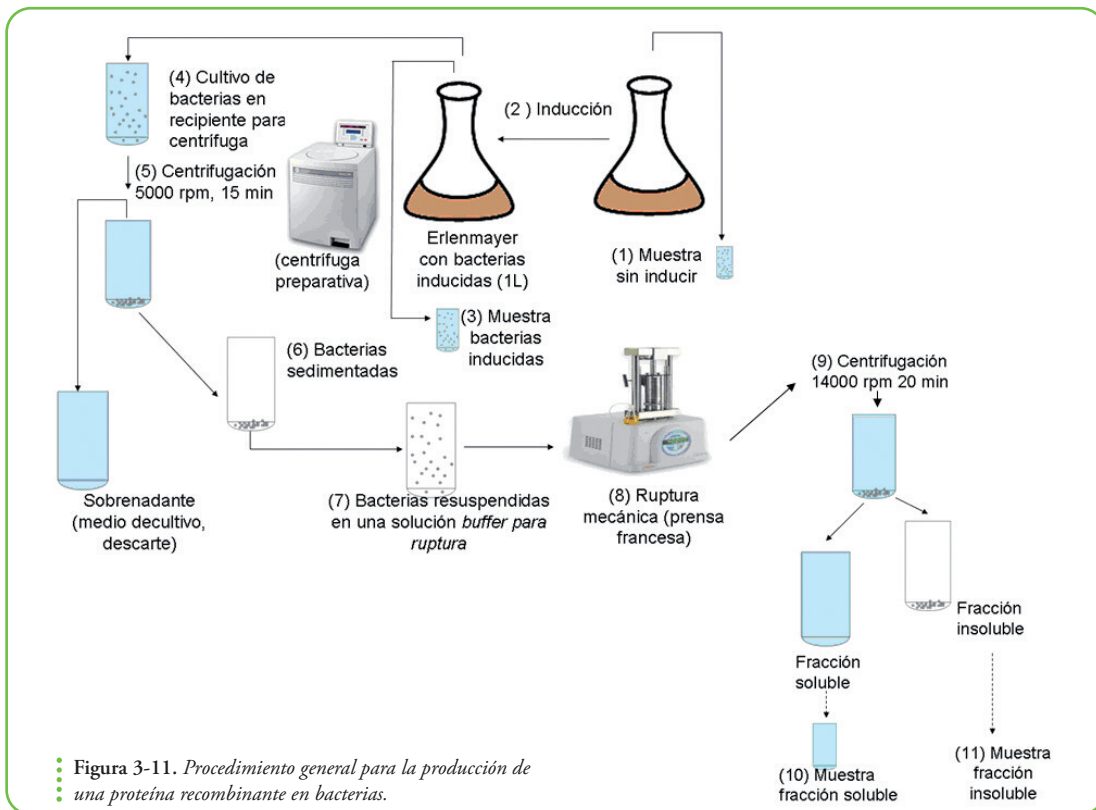
Hay diversas maneras de romper las bacterias:

- a. con cambios bruscos de presión y descompresión (el equipo se conoce como prensa francesa);
- b. con ultrasonido (con un sonicador).

Cuando rompemos las bacterias, nuevamente, tenemos que separar los fragmentos de bacterias y las proteínas insolubles de las proteínas solubles. Para esto, recurrimos a la centrifugación, esta vez deberemos centrifugar a mayor velocidad unas 14000 rpm por 20 minutos). Después de centrifugar podremos separar una fracción soluble, el *sobrenadante* con el cuidado de no resuspender el material insoluble. En un gel de poliacrilamida podremos sembrar muestras de estas fracciones e identificar a la proteína de interés y, finalmente, evaluar si se encontraba o no en la fracción soluble. En la **Figura 3-11** se muestra un diagrama del procedimiento general.

Será muy importante tomar una muestra del cultivo antes de agregar el inductor de la expresión, IPTG 1 mM, concentración final en nuestro caso. Esta muestra de bacterias sin inducir permitirá identificar por comparación con las bacterias en las que se agregó inductor cuál es la proteína de interés y cual es su nivel de expresión.

Será muy importante evaluar si la proteína de interés se encuentra soluble en el interior de las bacterias o no.



Por supuesto aún falta uno de los procesos más importantes y más complicados para poder contar con unos 10 ó 20 miligramos de proteína con que hacer estudios sobre su estructura. Debemos purificarla del resto de las otras proteínas y macromoléculas. Típicamente, después de la ruptura, ambas fracciones soluble e insoluble suelen quedar contaminadas con ácidos nucleicos y la fracción insoluble suele contener lípidos, fosfolípidos y fragmentos de pared bacteriana.

Será muy importante haber definido, correctamente, la ubicación de la proteína de interés porque dependiendo de esto, si se ha mantenido soluble o no, las estrategias de purificación serán completamente distintas e, incluso, si se halla en la fracción insoluble tendremos que intentar replegarla. Como vemos, nuevamente, la biosíntesis de la cadena polipeptídica no garantiza que la conformación de ésta sea la correcta.

¿Por qué la proteína de interés puede hallarse en la fracción insoluble?

Durante la sobre expresión, tanto la velocidad de síntesis (de la traducción) como la cantidad de moléculas producidas, son parámetros que pueden afectar, profundamente, la solubilidad de la proteína de interés. En muchos casos las proteínas no logran plegarse correctamente e interaccionan entre sí (se agregan) formando precipitados intracelulares insolubles que se conocen como cuerpos de inclusión.

En estos casos nuestra tarea será, primero, purificar los cuerpos de inclusión y, luego, di-

solverlos con un agente desnaturalizante (urea, por ejemplo), desarmando los agregados y llevando a las moléculas a la conformación desplegada.

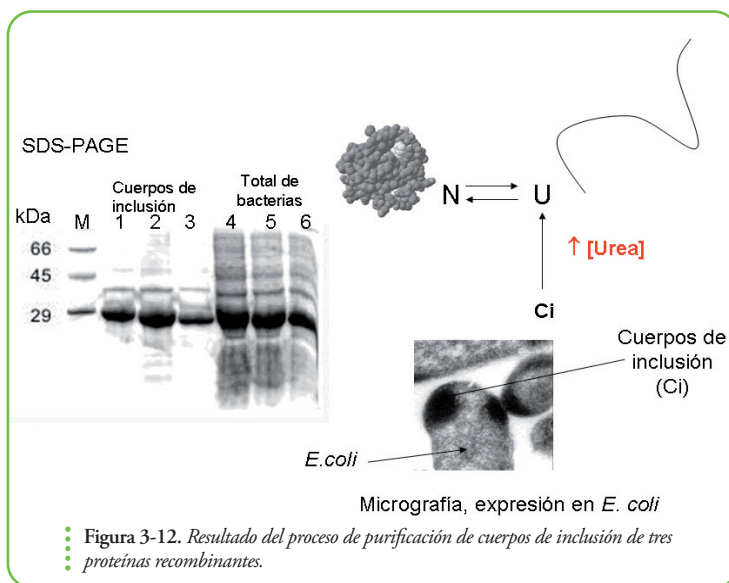
Como los cuerpos de inclusión suelen estar formados, básicamente, por la proteína recombinante más algunos pocos contaminantes, como fragmentos de pared bacteriana (peptidoglicano), lípidos y DNA, después de una serie de tratamientos quedarán relativamente enriquecidos en agregados de la proteína recombinante. Los tratamientos consistirán en lavados con soluciones que contengan enzimas que degradan el DNA (nucleasas), detergentes que permiten solubilizar los lípidos (sin solubilizar la proteína de interés) y lisozima, de clara de huevo que permite degradar los fragmentos de pared bacteriana.

En la **Figura 3-12**, se muestra el resultado de la purificación de cuerpos de inclusión de tres proteínas recombinantes (calles 1, 2 y 3 del gel). También, para poder comparar se sembraron muestras de bacterias después de su ruptura pero sin separar fracciones (calles 4, 5 y 6). Comparando la banda de interés con el resto de las bandas proteicas puede, verse, además, que el nivel de sobreexpresión en este caso fue muy elevado. Además se sembraron marcadores de peso molecular (M) para poder distinguir el tamaño de las proteínas recombinantes.

Como se mencionó anteriormente, en este caso la siguiente etapa es desplegar la proteína presente en los cuerpos de inclusión por la adición de agentes desnaturalizantes y llevarla al estado desplegado (U) para, posteriormente, continuar su purificación y replegado hacia la forma nativa (N).

Una de las estrategias de purificación más común es la cromatografía de intercambio iónico. La cromatografía de intercambio iónico permitirá separar proteínas de acuerdo con su carga. Un cálculo, relativamente, simple nos permitirá determinar la carga de una proteína o un péptido en función del pH (ver pH en propiedades de los aminoácidos). A un pH determinado la proteína no tendrá carga neta, ese es el punto isoelectrico (pI).

Las proteínas aniónicas tienen un pI <7.0 mientras que las catiónicas tendrán un punto isoelectrico >7.0. (**Figura 3-13**). Esto se traduce en que las proteínas aniónicas tendrán carga negativa incluso a pHs menores que pH 7.0 mientras que las catiónicas tendrán cargas positivas incluso a pHs ma-



También se muestra una micrografía de bacterias E. coli que han expresado proteína recombinante y se ha depositado en forma insoluble en cuerpos de inclusión (fotografía, gentileza del Dr. Rubén Iacono, IDEHU, UBA) y un esquema de la reacción de desplegado y solubilización en urea de las proteínas que formaban parte de cuerpos de inclusión. (ver Figura 3-12)

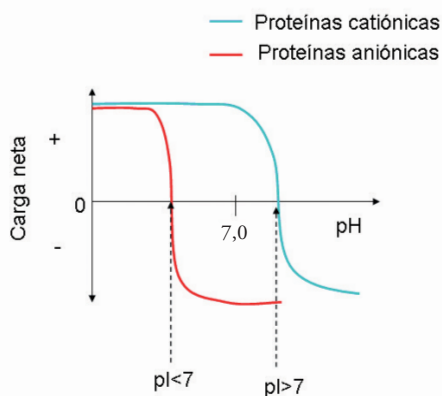


Figura 3-13. La estimación de la carga neta de una proteína en función del pH es importante en el proceso de purificación. Proteínas catiónicas poseen carga positiva a pHs superiores a pH 7,0. Las aniónicas poseen carga negativa a pHs menores a pH 7,0.

yores a pH 7.0. Esta propiedad podrá ser utilizada para separar proteínas en función de su carga: **las proteínas cargadas positivamente se unirán en forma reversible** a grupos cargados negativamente presentes en las matrices cromatográficas. A su vez, aquellas proteínas cargadas negativamente no se unirán y, por lo tanto, no serán retenidas por este tipo de matrices.

La incubación de la matriz con sales, por ejemplo cloruro de sodio, permitirá la competencia entre las proteínas cargada positivamente y los cationes sodio (Na^+) que funcionarán como contra ión al equilibrar las cargas negativas presentes en la matriz. Si elevamos suficientemente la concentración de sales podremos despegar la proteína unida y eluirla. Entonces la habremos separado los péptidos aprovechando la unión del péptido catiónico (Figura 3-14). Posteriormente, podemos lavar la columna cromatográfica y reestablecer las condiciones iniciales.

Se puede ver que, mediante un cambio en el pH, podemos producir un cambio en la carga neta de la proteína; también podríamos

Como ejemplo sencillo podemos ver lo que ocurre en el caso de la separación de dos péptidos pequeños

Péptido 1

LSKGQLKEFLDANLA

Péptido 2

LKKGKLKEFLRANLA

¿Cuál será la carga neta de ambos péptidos a pH 7,0?

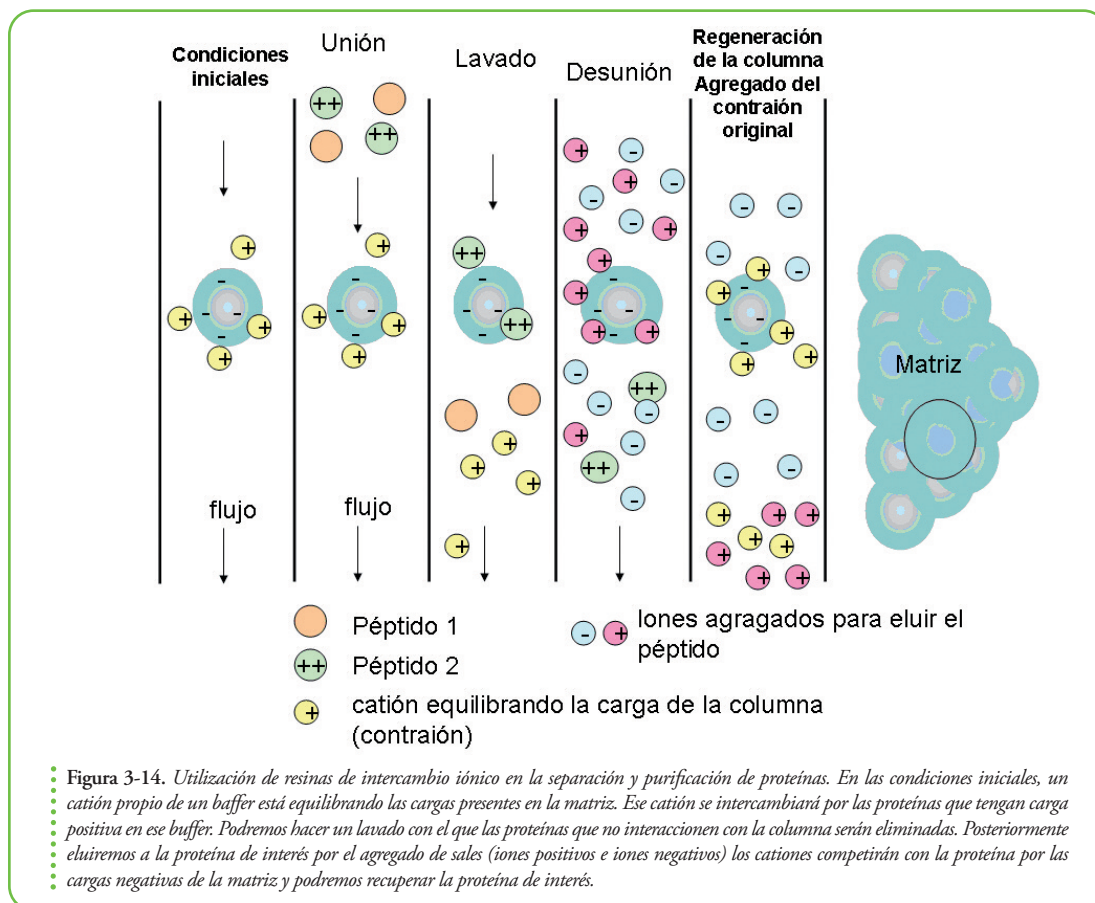
Teniendo en cuenta los valores de pK de cada residuo podemos ver que las lisinas (K) y argininas (R) aportarán una carga neta positiva por residuo mientras que los residuos de ácido aspártico (D) y ácido glutámico (E) aportarán una carga negativa a este pH.

También habrá que tener en cuenta que el grupo amino terminal aporta una carga neta positiva y que el carboxilato terminal aporta una carga neta negativa.

Entonces:

	Péptido 1	Péptido 2
E	1	1
D	1	0
C- -terminal	1	1
Q negativas	3	2
K	2	4
R	0	1
N-terminal	1	1
Q positivas	3	6
Q neta	0	+4

Así la carga neta será la suma de las cargas positivas menos la suma de las cargas negativas. A pH 7,0 el péptido 1 tendrá una carga neta $Q=0$ mientras que el péptido 2 tendrá una carga neta $Q=+4$. El péptido 2 quedaría retenido en una matriz de intercambio catiónico, una matriz con grupos cargados negativamente en su superficie, mientras que el péptido 1 no se unirá a esta matriz en este pH.



despegarla y eluirla. Por esto, debemos tener presente que en esta técnica hay dos factores realmente claves.

Estos factores son:

- la concentración de iones en la solución y,
- el pH de la solución.

Existen, en la actualidad, muchas clases de matrices disponibles. En la **Figura 3-15** se muestran dos ejemplos muy usados en el laboratorio, una matriz de intercambio catiónico (sulfopropil) y una matriz de intercambio aniónico (dientilaminoetil).

Si se recurre al intercambio iónico, la proteína purificada eluye, típicamente, con el agregado de sales típicamente en el rango de 100 a 1000 mM. Estas elevadas concentraciones podrían perjudicar estudios posteriores. Existen varias formas de remover sales. Dos de ellas son:

- mediante una cromatografía de exclusión molecular (ver Capítulo IV);
- la diálisis.

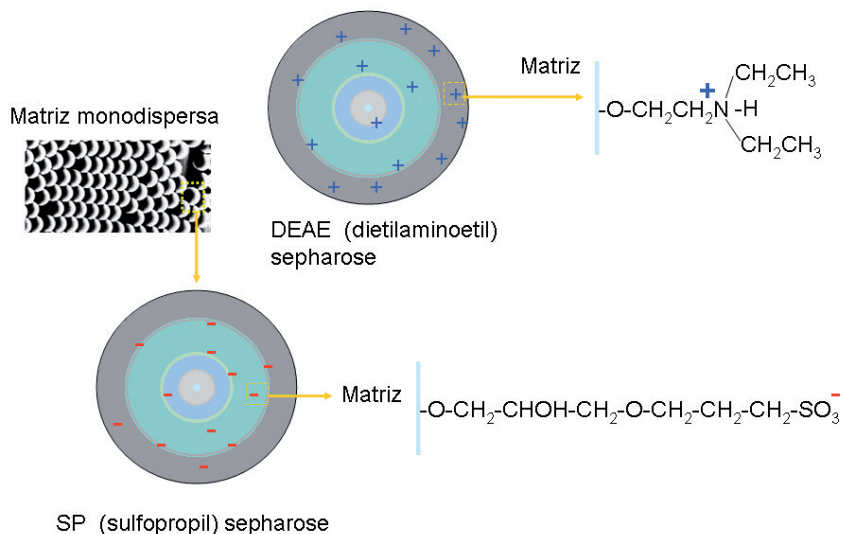
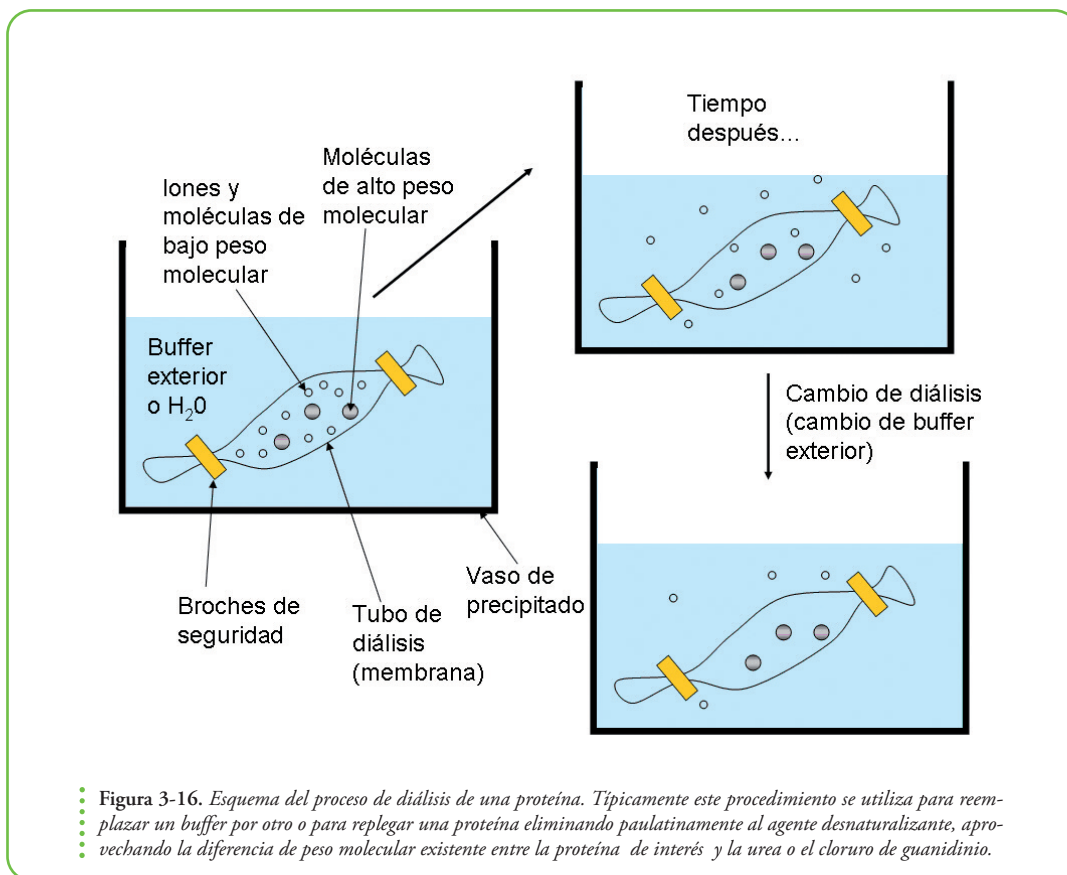


Figura 3-15. Esquema de la estructura química de dos tipos de matrices muy comúnmente usadas para el intercambio de iones. DEAE y SP sepharose. Además, se muestra una micrografía electrónica de una matriz monodispersa típica (extraído de GE Healthcare Life Sciences) que se caracteriza por la homogeneidad en el diámetro de los gránulos de matriz; esto garantiza un muy eficiente empaquetamiento de la matriz y por lo tanto en la elución de las proteínas

En este último caso, la base del recurso es la utilización de una membrana semi-permeable, la membrana de diálisis que permite la difusión de pequeñas moléculas, como por ejemplo iones sodio o cloruro, pero al mismo tiempo, previene la difusión de moléculas de mayor tamaño. Las moléculas pequeñas difundirán atravesando la membrana hasta que las concentraciones de estas moléculas, dentro y fuera (ver Figura 3-16) del tubo de diálisis sean iguales. En esta situación de equilibrio entrarán y saldrán del tubo de diálisis el mismo número de pequeñas moléculas por unidad de tiempo. Pero si hacemos suficientes cambios del *buffer* exterior podemos disminuir, suficientemente, la concentración de sales.

Es importante destacar que, no solamente podemos eliminar de esta forma iones, también otras moléculas pequeñas. Particularmente, en procesos de replegado nos va a interesar eliminar la urea que usamos para desplegar a la proteína.

Por último, vale la pena mencionar que, la etapa de purificación, es una etapa clave en el estudio de la estructura. Debemos conocer y, esto se logra con experiencia en el laboratorio, las particularidades de cada proteína. La estabilidad conformacional de la proteína de interés dictará los próximos pasos a seguir ya que es clave para el éxito subsecuente determinar cómo conservaremos a las moléculas. Como hemos visto en el capítulo II la desnaturalización es un proceso que ocurre a altas y a bajas temperaturas. No siempre será ventajoso congelar la solución que contiene la proteína. En algunos casos preferiremos mantenerla en presencia de agentes estabilizadores como glicerol al 50 %.



En otros casos podremos conservar a la proteína en condiciones desnaturizantes y replegarla posteriormente para su estudio.

Claro que replegar una proteína no siempre es sencillo. Si bien, como hemos visto el proceso de plegado es un proceso espontáneo, en ciertas condiciones experimentales puede ser muy poco eficiente. Además existen otro tipo de procesos paralelos como el de agregación que compiten y reducen aún más los niveles de proteína correctamente plegada. Por todos estos motivos será crucial conocer aspectos termodinámicos de la proteína de interés y tener a nuestra disposición herramientas que permitan evaluar su estado conformacional.