

## **GUÍA PARA LA SECUENCIACIÓN AUTOMÁTICA DE DNA**

### **CONTENIDOS:**

- 1.- Secuenciación automática de DNA: conceptos básicos
- 2.- Tipos de molde
- 3.- Preparación del molde
- 4.- Diseño de primers para secuenciación
- 5.- Interpretación de resultados. Problemas más frecuentes

## 1.- SECUENCIACIÓN AUTOMÁTICA DE DNA: CONCEPTOS BÁSICOS

La mayoría de los métodos de secuenciación automática utilizados actualmente son modificaciones del método desarrollado por Sanger y colaboradores a finales de los años 70. En este método, el DNA que va ser secuenciado funciona como un molde para la síntesis enzimática de un nuevo DNA que comienza en un sitio definido por la unión de un primer. En la reacción se utiliza una mezcla de desoxinucleótidos y didesoxinucleótidos a unas concentraciones que crean una probabilidad finita de que un didesoxinucleótido se incorpore en lugar del correspondiente desoxinucleótido en cada posición de la cadena que está siendo sintetizada. La incorporación de un didesoxinucleótido bloquea la elongación de la cadena y esto da como resultado una población de fragmentos de DNA truncados de diferente longitud. La identidad del nucleótido que termina la cadena en cada posición se puede determinar bien realizando 4 reacciones de elongación separadas cada una de las cuales con un didesoxinucleótido (ddATP, ddCTP, ddTTP ddGTP) o con una única reacción de elongación combinando los 4 didesoxinucleótidos pero marcando cada uno de estos específicamente. La población de moléculas resultante se separa por tamaños mediante electroforesis en acrilamida y la secuencia se obtiene correlacionando el orden de los fragmentos en la electroforesis con el didesoxinucleótido que termina cada uno de ellos.

En secuenciación automática, la utilización del marcaje fluorescente permite la detección del DNA durante la electroforesis. Si se utilizan 4 fluorocromos diferentes para cada uno de los 4 didesoxinucleótidos, las reacciones se pueden llevar a cabo en un sólo tubo y analizar mediante electroforesis capilar. Además, los datos de la secuencia se recogen durante la electroforesis, contribuyendo a la rapidez del método de secuenciación automática.

Una característica de la secuenciación automática, que la distingue de la secuenciación manual, es la utilización para la extensión de las nuevas cadenas de DNA de una polimerasa termoestable Taq. Es lo que se denomina secuenciación cíclica. En este método, rondas sucesivas de desnaturalización, hibridación y extensión en un termociclador dan como resultado una amplificación lineal de los productos de extensión. Estos productos se inyectan en un capilar para la obtención de su secuencia de bases. Las ventajas de la secuenciación cíclica frente a otros métodos son varias:

- los protocolos son más fáciles de realizar.
- se requiere mucho menos DNA molde que con los métodos de extensión a una única temperatura.
- no se necesita un paso de desnaturalización química del DNA de doble cadena.
- las temperaturas altas reducen la estructura secundaria en el DNA, así como sitios secundarios de hibridación del primer.
- se puede utilizar el mismo protocolo para la secuenciación de DNA de simple cadena, de doble cadena o para la secuenciación directa de productos de PCR.

## 2.- TIPOS DE MOLDE PARA SECUENCIACIÓN AUTOMÁTICA:

Los tipos de DNA que pueden ser usados como moldes para su secuenciación son los siguientes:

### Vectores de clonación:

1.- Vectores de simple cadena: casi todos basados en el bacteriófago M13. Son muy útiles y generan muy buenas secuencias, pero tienen el inconveniente que sólo permiten la secuenciación en una única dirección.

2.- Vectores plasmídicos de doble cadena: son los más utilizados y en los que centraremos esta guía.

3.- Vectores fagémidos: vectores híbridos, plásmidos de doble cadena que contienen un origen de replicación de un bacteriofago de simple cadena. Si es necesario, vectores del tipo pBluescript y pGem 18/19 pueden funcionar como fagémidos. Dependiendo de la forma del fagémido se utilizarán como vectores de doble cadena o de cadena sencilla.

4.- Cósmidos, YACs, PACs y BACs: son vectores capaces de replicar grandes fragmentos de DNA. Son ideales para estrategias de “primer walking”.

### Fragmentos de PCR:

1.- Productos de PCR de cadena sencilla: producidos a partir de reacciones de PCR que contienen un exceso de un primer.

2.- Productos de PCR de doble cadena: obtenidos a partir de reacciones de PCR que contienen cantidades iguales de cada primer.

Se pueden amplificar por PCR fragmentos de DNA a partir de tejidos o amplificar insertos directamente de colonias. Cuando los fragmentos de PCR se clonan pueden producirse problemas de errores en la secuencia, que pueden ser evitados si se realiza la secuenciación directamente del producto de PCR, puesto que la tasa de error para cada nucleótido es muy baja y no es detectable en una población grande del fragmento.

Cuando se secuencian fragmentos de PCR es muy importante que los primers utilizados para amplificar el fragmento sean eliminados eficientemente antes de la secuenciación.

### 3. PREPARACIÓN DEL DNA MOLDE

#### A.- Obtención del DNA molde

La calidad del DNA molde es el factor más importante para una secuenciación exitosa. Cualquier resto de proteínas, RNA, sales, carbohidratos, fenol, cloroformo, etanol, detergentes, PEG, EDTA, etc., va a contribuir a la obtención de una mala secuencia. A continuación, detallamos una serie de recomendaciones para obtener un molde de calidad:

- **DNA plasmídico:**

- Para la purificación del DNA plasmídico pueden utilizarse diferentes protocolos, pero funcionan mejor los que conllevan métodos de purificación con minicolumnas.
- El DNA molde debe tener una ratio entre A260 y A280 nm de 1.8-2.0.
- Idealmente, el DNA molde debe obtenerse resuspendido en agua. Hay que evitar el TE pues el EDTA puede interferir en la reacción de secuenciación.
- Si en la purificación del DNA se ha utilizado algún solvente orgánico como fenol o cloroformo, el DNA debe purificarse al menos dos veces mediante precipitación con etanol, para eliminar cualquier traza del solvente orgánico.
- Si se utiliza etanol o isopropanol para precipitar el DNA, realizar la precipitación utilizando acetato de amonio en lugar de acetato de sodio. Lavar el precipitado después de la centrifugación al menos una vez con etanol al 70%. Secar el precipitado el máximo posible y resuspender en agua. La presencia de trazas de etanol en el molde es una de las principales causas de fallo de la secuenciación.
- Si se utiliza una mini columna para la purificación del DNA, se recomienda centrifugar la columna dos veces después del paso de lavado y a continuación eluir el DNA con agua. Este procedimiento también permite la eliminación del etanol presente en la solución de lavado de la muestra.
- Existen una serie de factores que pueden influir en la calidad del molde
  - Cultivo a partir del que se obtiene el DNA: son dos los elementos de particular importancia i) las condiciones de crecimiento del cultivo y ii) la cepa hospedadora seleccionada. Los cultivos bacterianos para la purificación de plásmidos deben obtenerse de colonias frescas crecidas en medios selectivos o de pequeños precultivos obtenidos de stocks de glicerol o de agar. Respecto a las cepas utilizadas como hospedadoras, se recomienda utilizar las cepas DH1, DH5 $\alpha$ , DH10B, C600 y XL1-Blue para la purificación del plásmido. La cepa HB101 y sus derivados no están recomendadas puesto que contienen una gran cantidad de carbohidratos y poseen el locus endA con lo que producen relativamente grandes cantidades de nucleasa.

- Contaminación por nucleasas: por ejemplo, la DNasa causa un cambio de la forma del plásmido superenrollada a lineal lo que reduce gradualmente la intensidad de la señal y la longitud de la secuencia obtenida. Por lo tanto, hay que evitar las cepas hospedadoras que produzcan nucleasas.
- Contaminación por RNA: esto ocurre más frecuentemente cuando los cultivos han crecido demasiado tiempo y el tampón de lisis es insuficiente para lisar todas las células. Las muestras contaminadas con RNA muestran un ruido de fondo alto. El tratamiento del DNA molde con RNasa (libre de DNasa) degradará el RNA con lo que se pueden mejorar los resultados de la secuenciación.
- Contaminación por sales: la presencia de sales inhibe la reacción de secuenciación, puesto que disminuye la procesividad de la polimerasa Taq. Sin embargo, el tipo de sal influye en la severidad del efecto: por ejemplo, el cloruro de sodio o de potasio tiene un efecto mayor que la misma concentración de acetato de sodio. Una concentración de acetato de sodio mayor de 20mM inhibe severamente la reacción de secuenciación. Las causas más comunes de contaminación por sales son i) la coprecipitación de éstas con el etanol en la precipitación del DNA, sobre todo si esta se realiza a baja temperatura, ii) una eliminación insuficiente del sobrenadante o iii) un lavado con etanol 70% insuficiente. Por lo tanto, la precipitación del DNA molde y los lavados deben realizarse muy cuidadosamente y a temperatura ambiente.
- Contaminación por etanol: se produce por la resuspensión del DNA molde tras la precipitación sin que se haya secado suficientemente. La presencia de pequeñas cantidades de etanol (5%) puede causar la terminación prematura de las cadenas y por lo tanto la obtención de secuencias cortas. Una contaminación con etanol superior al 10% tiene como consecuencia normalmente el fallo total de la reacción de secuenciación.
- Contaminación con fenol: algunos métodos de preparación de DNA plasmídico utilizan un paso de extracción con fenol. La contaminación con este reactivo tiene como resultado la degradación severa de la secuencia resultante, especialmente en lecturas largas. Más de un 1,4% de fenol en el DNA molde produce secuencias totalmente inutilizables.

- **Productos de PCR:**

- Los productos de PCR deben verse como una única banda en un gel de agarosa.
- La secuenciación de fragmentos de pequeño tamaño es algo problemática por lo que se recomienda que los productos sean al menos de 150 pb. Los fragmentos más pequeños se pueden comportar como primers y son difíciles de purificar con buen rendimiento.
- Los productos de PCR contienen restos de nucleótidos, primers, dNTPs y enzimas procedentes de la amplificación que pueden interferir en la reacción de secuenciación. Por este motivo es necesario incluir un paso purificación del producto previo a la secuenciación. Existen varias opciones y aquí sólo vamos a detallar algunas:
  - Utilización de kits de purificación directa del producto. Suelen ser kits de filtración o mini columnas. Se pueden utilizar sólo cuando el resultado de la amplificación es una única banda.
  - Purificación del producto de PCR de un gel de agarosa. Se trata de correr el producto de PCR en un gel de agarosa de bajo punto de fusión, cortar la banda correspondiente y extraer el DNA con un kit comercial.
  - Tratamiento del producto de PCR con Exonucleasa I y SAP (Shrimp Alkaline Phosphatase). El tratamiento con Exonucleasa I degrada los primers de PCR residuales mientras que la SAP defosforila los dNTPs restantes. Tras la inactivación por calor de ambas enzimas puede utilizarse el producto de PCR para secuenciación automática. Evidentemente, este método no puede aplicarse en el caso de productos de PCR contaminados con subproductos secundarios.
  - Dilución directa del producto de PCR. Este método implica la dilución del producto de PCR entre 5 y 10 veces en agua y utilizarlo directamente en la reacción de secuenciación. Para que este método sea efectivo, es necesario emplear bajas concentraciones de dNTPs y primers en la reacción de PCR (concentración final de los primers 0,2  $\mu$ M y de los dNTPs 100  $\mu$ M o menos). La principal desventaja de este método es que requiere una optimización individual de las condiciones para cada producto de PCR.
  - Precipitación selectiva del producto de PCR
- Al final de la purificación los productos de PCR deben ser resuspendidos en agua. Por el mismo motivo que con el DNA plasmídico debe evitarse el uso de TE.

## B.- Cuantificación del DNA molde

La obtención de unos buenos resultados de secuenciación está condicionada a una cuantificación precisa del DNA molde. Poca cantidad de DNA en la reacción de secuenciación produce datos de secuencias con baja señal, ruido de fondo, errores y ambigüedades. Demasiado DNA molde también puede producir efectos similares.

Por lo tanto, la cuantificación de la muestra es un paso previo muy importante para la secuenciación. Puede realizarse mediante la medida de la absorbancia a 260 nm en un espectrofotómetro. Puesto que la contaminación del DNA molde con RNA o proteínas pueden afectar negativamente a los resultados de la secuenciación, hay que medir también la absorbancia de la muestra a 280 y 230 nm y comprobar las ratios A260/A280 y A260/A230.

Es aconsejable además confirmar la cuantificación mediante electroforesis en gel de agarosa y comparando con un marcador de concentración conocida. Este gel nos permitirá además detectar otros posibles problemas presentes en nuestras muestras, como la presencia de más de un amplificado en el caso de fragmentos de PCR o la detección de dímeros de los primers utilizados en la amplificación.

En la siguiente tabla se indican las cantidades de DNA necesarias para cada tipo de molde

**Tabla I. Cantidad de molde utilizado en las reacciones de secuenciación.**

<b>Molde</b>	<b>Cantidad</b>
Producto de PCR:	
100-200 pb	10 ng
200-500 pb	20 ng
500-1000 pb	30 ng
>1000 pb	100 ng
AND cadena sencilla	100 ng
ADN cadena doble	250 ng
Cósmido, BAC	1 µg

#### 4.- DISEÑO DE PRIMERS PARA SECUENCIACIÓN

La secuencia de los primers utilizados en la secuenciación tiene una influencia importante en la especificidad y la sensibilidad de la reacción de secuenciación. Por ello es importante seguir una serie de recomendaciones a la hora de diseñar los primers.

1. Asegurarse de que la secuencia utilizada para el diseño de los primers es correcta.
2. La longitud del primer debe ser al menos de 18 bases y la longitud óptima entre 20-25 bases.
3. El contenido en G-C del primer debe ser entre 40-60% siendo óptimo el 50%. Los primers con un menor contenido en GC deben diseñarse algo más largos para conseguir una Tm por encima de la recomendada.
4. La Tm del primer debe estar entre 55 y 70 °C. Se puede utilizar la siguiente formula para calcular la Tm.  
$$Tm=69.3+0.41x(\%GC)-650/\text{longitud del primer}$$
5. La presencia de una G o C en el extremo 3' contribuye a la estabilidad del final del primer.
6. Evitar repeticiones de la misma base (3 o más seguidas). Su presencia puede producir hibridaciones secundarias sobre otras secuencias que puedan contener un motivo complementario.
7. Evitar los primers que pueden formar dímeros o estructuras secundarias.
8. Diseñar el primer al menos 50 bases "upstream" de la secuencia de interés. La secuencia inmediatamente posterior al primer o no se obtiene o puede ser bastante imprecisa.

Los primers deben resuspenderse en agua estéril y no en TE. La disolución original de los primers se puede almacenar a largo plazo a -70 °C pero se debe guardar una disolución de trabajo a -20°C a una concentración 5 µM que puede ser utilizada durante 1 o 2 meses como máximo.

## 5.- INTERPRETACIÓN DE RESULTADOS: PROBLEMAS MÁS FRECUENTES Y POSIBLES SOLUCIONES.

La calidad de los datos obtenidos en la secuenciación de un DNA puede verse afectada negativamente por distintos factores tanto en la reacción de secuenciación, así como en cualquiera de las fases posteriores a esta en el proceso de secuenciación. A continuación, vamos a describir los problemas más frecuentes que se detectan, así como sus posibles causas y soluciones.

1. No se obtiene secuencia o la señal decae poco después de comenzar. Esto puede deberse a alguna de las siguientes causas:
  - La cuantificación del DNA molde no es correcta, no hay DNA o mucho menos del necesario.
  - En la reacción de secuenciación no hay primer o mucho menos del indicado. El primer puede estar degradado.
  - El primer no se une al molde. Hay que asegurarse de que el diseño del primer es correcto.
  - El sitio de unión del primer en el vector de clonación se ha perdido o alterado. Durante la clonación pueden crearse artefactos con una delección cerca del sitio de inserción del DNA a clonar o delecciones que eliminan la secuencia del primer del vector.
  - Si se trata de productos de PCR, puede que los amplificados no sean el producto de la amplificación con los dos primers. Si el amplificado se genera a partir de uno de los dos primers sólo se obtendrá secuencia con éste. Este efecto se puede producir cuando se utilizan para secuenciar los mismos primer que se han utilizado en la amplificación del DNA molde y alguno de ellos no funciona correctamente. Es mejor diseñar primers distintos para amplificar y secuenciar.
2. La señal es muy baja debido a una pobre unión del primer. Esto puede deberse a alguna de las siguientes causas:
  - La  $T_m$  del primer es demasiado baja porque es demasiado corto o tiene un bajo contenido en GC.
  - El primer tiene estructura secundaria, particularmente en el extremo 3'.
  - El DNA molde tiene estructura secundaria en la zona de hibridación con el primer. En este caso como en los anteriores es conveniente diseñar un nuevo primer.
  - La calidad o la cantidad del DNA molde no es adecuada.

3. La secuencia tiene mucho ruido de fondo. Esto puede deberse a alguna de las siguientes causas:
  - El DNA molde tiene un sitio secundario de unión al primer.
  - Si el primer está mal sintetizado y existen fragmentos de menor longitud (N-1) se producen secuencias superpuestas e ilegibles. Lo mismo ocurre si el primer no está bien purificado.
  - La muestra contiene más de un amplificado.
  - El fragmento de PCR utilizado como molde no ha sido purificado y los primers de amplificación que quedan en el medio interfieren generando productos de extensión a partir de más de un primer.
4. Se obtienen dos secuencias superpuestas:
  - El primer tiene varias dianas dentro del DNA molde.
  - Hay más de un DNA molde distinto en la muestra. Si se trata de un producto de PCR habrá que verificar que no existen amplificaciones secundarias inespecíficas. Si el molde está clonado la purificación debe realizarse a partir de una única colonia.
  - Si la muestra es un DNA que está subclonado, durante la subclonación se ha arrastrado un fragmento del polilinker y como resultado la muestra tiene dos dianas para el primer universal que hemos utilizado.
  - El producto de PCR no está purificado o la purificación no ha eliminado totalmente los restos de primer que no se utilizaron en la amplificación.
  - El producto de PCR está generado por un único primer (diana del primer presente en ambos extremos).
5. La señal al principio es intensa, pero decae rápidamente:
  - Esto puede deberse a la presencia de algún contaminante en la muestra como EDTA, sales, etanol, fenol, etc.
  - No están equilibradas las cantidades de DNA molde y primer.
6. La secuencia obtenida es normal hasta un punto en el que súbitamente cae la señal:
  - El DNA está cortado o digerido y la reacción finaliza.
  - La propia secuencia del DNA produce determinados motivos que provocan ese efecto en la reacción. La causa más frecuente es la existencia de una estructura secundaria en el molde debida por ejemplo a la presencia de repeticiones invertidas o palindromes, que no puede ser superada por la polimerasa. Existen varias posibles soluciones a este problema como son: secuenciar la cadena complementaria, digerir el molde, subclonarlo y secuenciar los subclones o utilizar un kit alternativo para las reacciones de secuenciación como el kit dGTP Big Dye de ABI. También se puede

añadir a la reacción de secuenciación el reactivo "Sequencer Enhancer Solution A" de GIBCO BRL.

7. La señal en la secuencia disminuye antes de lo esperado:
  - La muestra puede contener demasiadas sales u otros contaminantes arrastrados durante los procesos de extracción, purificación, amplificación, etc.
8. La parte inicial de la secuencia presenta grandes picos:
  - Esto puede ocurrir cuando los primers empleados en la amplificación forman dímeros. Estos se comportan como un molde sobre el que se producirá la reacción de secuenciación. Cuando termina la secuencia de los dímeros la lectura es correcta.
9. La secuencia comienza a leerse correctamente, pero a partir de un punto aparece una mezcla de picos o un gran "background":
  - Esto ocurre cuando el molde es una mezcla de más de un clon. La secuencia del vector se lee perfectamente hasta el sitio de clonaje, pero luego se observa una mezcla de picos. Es muy importante por lo tanto que en la obtención de DNA molde nos aseguremos de que se parte de una única colonia.
10. La secuencia se lee correctamente hasta llegar a una zona de poliA o poliT:
  - Este efecto se debe a que la polimerasa se desliza sobre el DNA molde. Este problema puede producirse durante la reacción de secuenciación, o en el caso de un producto de PCR también durante el proceso de amplificación. También se ha descrito este fenómeno durante la amplificación de muestras que presentan repeticiones de dinucleótidos. Aunque los productos de deslizamiento no son aparentes cuando el fragmento se visualiza en un gel de agarosa, sí lo son en los datos de secuenciación. Tras la región homopolimérica la secuencia se convierte en ilegible.
  - Para solucionar este problema se pueden utilizar dos estrategias: secuenciar la cadena complementaria o utilizar los denominados primers anclados. Son primers que se unen al homopolímero, tipo T25-(N) y que terminan en una base degenerada para obligar al oligo a pegarse en la posición 3'.
11. La secuencia empeora después de una zona rica en GT, GA u homopolímeros de G:
  - Se trata de zonas difíciles de secuenciar debido a su composición nucleotídica. Lo mejor es secuenciar la cadena complementaria o repetir la secuenciación con un kit alternativo (dGTP Big Dyes).
12. Existen una serie de tipos de DNA molde que son difíciles de secuenciar que son los siguientes: DNA ricos en GC, secuencias ricas en AT, DNA con estructuras secundarias, repeticiones, palindromes, regiones de homopolímeros. Para cada uno de estos tipos de secuencias existen una serie de alternativas que describimos a continuación:
  - Secuenciación de zonas con alto contenido en GC (>60%):
    - a. Añadir 5% DMSO a la reacción de secuenciación.

- b. Secuenciar la cadena complementaria puede ayudar a resolver ambigüedades.
- c. Linearizar el DNA con un enzima de restricción.
- d. Probar una combinación de las siguientes modificaciones:
  - i. Aumentar la temperatura de desnaturalización a 98 °C.
  - ii. Añadir más TAQ
  - iii. Incluir un paso de pre-desnaturalización de 96°C de 4-5 min.
- Secuencias ricas en AT: generalmente no son tan problemáticas como las secuencias ricas en GC.
  - a. Si el primer es rico en AT, lo mejor es aumentar su longitud hasta 24-26 bases para conseguir una Tm cercana a los 55 °C.
  - b. A veces funciona mejor secuenciar con primers marcados que con terminadores marcados.
- Zonas de estructura secundaria, repeticiones invertidas, palíndromes: una pérdida brusca de señal normalmente significa un problema de estructura secundaria en el DNA, debido a la incapacidad del enzima a continuar sobre la región del molde en cuestión. Posibles soluciones son:
  - a. Añadir 5% DMSO a la reacción de secuenciación.
  - b. Alguna de las sugerencias para zonas ricas en GC.
- Zonas con repeticiones:
  - a. Las zonas de repeticiones de di, tri o tetranucleótidos no son problemáticas en material clonado. Sin embargo, en el caso de productos de PCR se puede producir el deslizamiento de la polimerasa y no se puede obtener secuencia de esa región.
  - b. Las repeticiones largas como VTRs de 30 o más bases también suelen ser difíciles de secuenciar y lo mejor es obtener delecciones seriadas de esa zona y secuenciarlas.
  - c. Las repeticiones de AG también pueden ser problemáticas pues la Taq produce una señal débil de G después de A cuando se utilizan terminadores marcados. En este caso conviene secuenciar la cadena complementaria.
- Zonas de homopolímeros: cuando una base se repite más de 10 veces seguidas en una zona de DNA pueden aparecer problemas en la secuenciación de esta región. Cuando la base repetida es G o C incluso se toleran menos repeticiones. Para zonas de poli-A o poli-T se puede utilizar la estrategia de primers anclados que hemos descrito anteriormente. Para los homopolímeros de poli-G y poli-C puede funcionar el utilizar un primer que hibride 20-30 bases antes del homopolímero para forzar a la polimerasa a pasar sobre él. También se pueden realizar delecciones seriadas sobre esa región y secuenciarlas.