

MÉTODOS EXPERIMENTALES PARA EL ESTUDIO DE LOS GENOMAS

María Isabel González Siso y Ana María Rodríguez Torres
Departamento de Biología Celular y Molecular
Universidad de A Coruña

1. INTRODUCCIÓN

Durante la última década hemos asistido no sólo al desarrollo de una gran variedad de nuevas técnicas aplicables al estudio de los genomas, sino también a su automatización, lo que ha posibilitado que la información disponible acerca de la secuencia y organización molecular de los genomas eucariotas haya aumentado notablemente en un corto período de tiempo.

En este sentido, la efectividad alcanzada en las técnicas relacionadas con la elaboración de mapas físicos de los genomas y con la secuenciación del DNA ha permitido que en el presente año se descifrara por primera vez la secuencia completa del genoma de un organismo eucariota: la levadura *Saccharomyces cerevisiae*. Este proyecto, liderado por investigadores de la Unión Europea, comenzó en 1989 con la secuenciación del cromosoma III (Oliver et al., 1992) uno de los más pequeños de los 16 de que consta el genoma de la levadura, y su ejecución se prolongó 7 años. Nuestro laboratorio se incorporó a dicho proyecto en el año 1994 y ha participado secuenciando un total de 17kb del cromosoma VII (Rodríguez-Belmonte et al., 1996; Tizón et al., 1996).

Utilizaremos la secuenciación del genoma de la levadura como modelo para describir como se puede abordar de un modo global y sistemático el estudio del genoma de un organismo eucariota sencillo. Es importante tener en cuenta que para este proyecto se coordinaron más de 100 laboratorios que ya estaban previamente estudiando la Biología Molecular de *S. cerevisiae* y que se disponía, por tanto, de numerosos genes clonados, secuenciados y cartografiados, líneas mutantes, marcadores, etc. además de experiencia, lo que facilitó notablemente la tarea (Dujon, 1996; Johnston, 1996; Williams, 1996). De hecho, cada grupo participante utilizó su propia metodología, lo que además de estimular la evolución de las técnicas ha posibilitado establecer comparaciones entre las diferentes estrategias utilizadas y valorar su efectividad.

En la actualidad están en marcha otros proyectos de secuenciación de genomas eucariotas como el de *Arabidopsis*, *Drosophila* o el Humano.

2. PERFIL CROMOSÓMICO DEL GENOMA. ELECTROFORESIS DE CAMPO PULSANTE

La investigación con genomas eucariotas requiere manipular fragmentos y moléculas de DNA de tamaño muy superior al que puede separarse mediante las técnicas de electroforesis convencionales, en gel de agarosa o poliacrilamida y bajo un campo eléctrico estático, ya que en estas condiciones, es imposible separar moléculas de más de 20kb de tamaño.

Los cromosomas de levaduras, cuyo tamaño oscila entre 200 y 10.000kb, se separan mediante la técnica denominada electroforesis de campo pulsante o PFG ("pulsed field gel electrophoresis"), descrita originalmente en 1982 (Smith et al., 1990).

La PFG se basa en forzar a las grandes moléculas de DNA a cambiar la dirección en la que se van moviendo en el gel exponiéndolas a campos eléctricos alternantes con distinta orientación (Figura 1). La duración de cada campo eléctrico es un pulso. La velocidad a la cual una molécula de DNA cambia de dirección en estas condiciones depende de su peso molecular. Así pues, el rango de tamaños que se separan en una PFG es directamente proporcional a la duración de cada pulso. La resolución y movilidad de los fragmentos en una PFG depende, además, de otros factores como son: ángulo de reorientación y homogeneidad del campo, gradiente de voltaje, temperatura, tipo de agarosa, concentración del tampón; así como concentración de DNA, sales y proteínas contaminantes en la muestra.

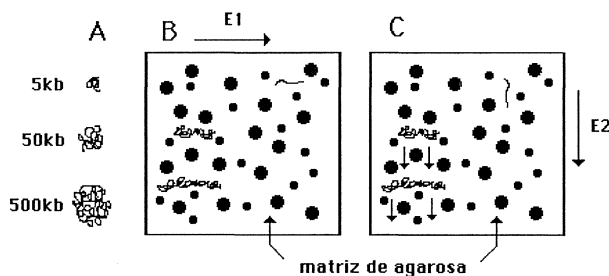


Figura 1.- Esquema del proceso de separación de moléculas de DNA por electroforesis convencional y de campo pulsante. (A) moléculas de DNA de diferentes tamaños en ausencia de un campo eléctrico externo. (B) En presencia de un campo eléctrico estático E1, todas las moléculas se alinean con el campo pero las de mayor tamaño presentan la misma movilidad. (C) En un gel de campo pulsante, cuando el primer campo eléctrico E1 se para y se conecta un segundo campo eléctrico E2 en otra dirección, las moléculas más pequeñas se alinean rápidamente en la nueva dirección pero las grandes se reorientan más lentamente cuanto mayor es su peso molecular. Las flechas indican los posibles caminos que las moléculas grandes de DNA pueden escoger para reorientarse en el nuevo campo eléctrico. (Modificado de Birren y Lai, 1993)

Se han desarrollado muchas variantes de la PFG referidas a la geometría original de los electrodos, homogeneidad y método de reorientación de los campos eléctricos, algunas de las cuales se recojen en la tabla I.

Tabla I.- Tipos de electroforesis de campo pulsante (Birren y Lai, 1993)

Acrónimo	Sistema de electroforesis
PFGGE	Pulsed field gradient gel electrophoresis
OFAGE	Orthogonal field alternation gel electrophoresis
TAFE	Transverse alternating field electrophoresis
FIGE	Field inversion gel electrophoresis
CHEF	Contour clamped homogeneous electric field
RGE	Rotating gel electrophoresis
	Crossed-field gel electrophoresis
Rotaphor	Rotating electrodes gel electrophoresis
Waltzer	Crossed field gel electrophoresis
PACE	Programmable autonomously controlled electrodes
ZIFE	Zero integrated field electrophoresis
ST/RIDE	Simultaneous tangential/rectangular inversion decussate electrophoresis

En la figura 2 se observa el resultado de la separación mediante electroforesis de campo pulsante de los 16 cromosomas en que se organiza el genoma de *Saccharomyces cerevisiae*.

Dado el gran tamaño de las moléculas a separar, es imposible preparar el DNA para la PFG por los métodos convencionales puesto que se fragmentaría mecánicamente durante el pipeteo o la extracción con fenol, por ejemplo. Para solventar este problema, se han desarrollado técnicas sencillas para extraer el DNA de las células incluídas en agarosa. Las muestras de DNA se cargan en los geles incluídas en agarosa.

Mediante técnicas de "Southern blot" y posterior hibridación con una sonda marcada conteniendo un determinado gen, es posible localizar en que cromosoma se situa dicho gen (Figura 3). El "Southern blot" de los fragmentos separados por PFG presenta particularidades especiales ya que los fragmentos de DNA mayores de 20kb no se transfieren eficientemente del gel a la membrana, siendo necesario escindirlos previamente, lo que generalmente se hace por irradiación UV o por despurinización con ácidos.

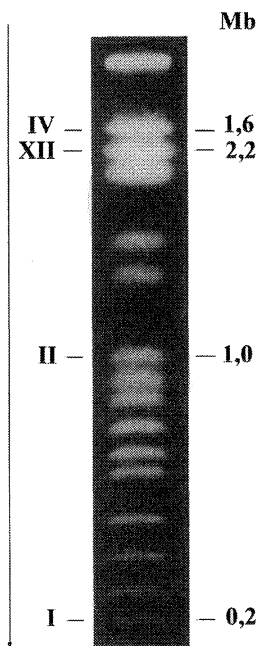


Figura 2.- Separación de cromosomas de levadura (*S. cerevisiae*) por electroforesis de campo pulsante (Modificado de Stryer, 1995).

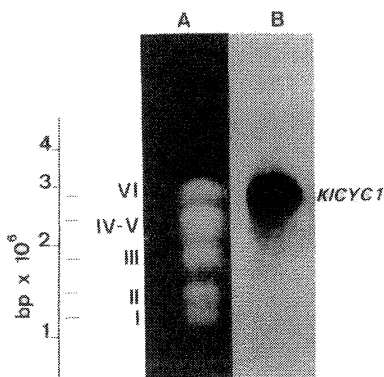


Figura 3.- Localización cromosómica del gen *CYC1* de *Kluyveromyces lactis*. (A) gel teñido con bromuro de etidio mostrando los cromosomas separados por electroforesis CHEF. (B) Autoradiografía después de hibridar el blot cromosómico con una sonda marcada conteniendo el *KICYC1* (González Siso et al., 1994)

Existen técnicas que permiten rescatar el DNA del gel una vez separado por PFG, las más utilizadas son la digestión del gel con agarosa y la electroelución que es efectiva para moléculas menores de 500kb, no se pueden emplear otros métodos tradicionales como el "kit GeneClean" pues se fragmentan las moléculas de DNA. Los geles preparativos se hacen con agarosa de bajo punto de fusión.

La PFG también puede realizarse en modo bidimensional. En este caso las moléculas de DNA como los cromosomas de levadura se separan en un primer gel. A continuación se digieren con enzimas de restricción (el DNA puede digerirse incluido en agarosa) y los fragmentos se separan por electroforesis en una segunda dirección en diferentes condiciones. De este modo se pueden identificar los fragmentos de restricción de cada cromosoma. Estos métodos son muy útiles para elaborar de un modo bastante rápido mapas de restricción de genomas pequeños sin necesidad de disponer del genoma clonado.

El análisis del perfil cromosómico mediante PFG puede utilizarse también para diferenciar especies de levaduras. Así, tras la introducción de las técnicas de biología molecular en la sistemática microbiana, se han reclasificado especies que en función de caracteres morfológicos y fisiológicos se habían incluido en un Género y dado su perfil cromosómico o el patrón de restricción del DNA se ha visto que pertenecen a otro (Guillamón et al., 1996).

3. CONSTRUCCIÓN DE BIBLIOTECAS GENÓMICAS. EL CASO DE LOS TELÓMEROS

Si bien genomas pequeños como los bacterianos pueden secuenciarse directamente mediante la estrategia denominada "shot gun" que se explica en otro apartado, la secuenciación de genomas de mayor tamaño como los de células eucariotas requiere un paso intermedio de clonación del DNA en fragmentos. Por lo tanto, la naturaleza y calidad de las genotecas (todo o una parte determinada del genoma de un organismo fragmentado y clonado), así como la estrategia utilizada para elaborar los mapas físicos (situación de cada clon en el genoma del organismo) son de la máxima importancia.

En el caso del proyecto de secuenciación del genoma de la levadura, la mayor parte de las genotecas se realizaron en cósmidos y en vectores derivados del fago-lambda, sólo una pequeña parte en plásmidos. La ventaja de los cósmidos sobre los otros tipos de vectores es que admiten insertos de DNA de mayor tamaño (hasta 50kb), lo que simplifica la elaboración del mapa físico. Los insertos en vectores lambda son más pequeños pero los clones son más estables y la representación del genoma en las bibliotecas suele ser superior. Las genotecas para la secuenciación de genomas de mayor tamaño como el humano se construyen fundamentalmente en YACs que admiten insertos todavía mayores (hasta 1000kb) si bien su manipulación es más dificultosa.

Por ejemplo, para el cromosoma III se construyeron genotecas rescatando el DNA cromosómico por electroelución a partir de geles de electroforesis en campo pulsante (FIGE), digiriéndolo parcialmente con *Sau3AI*, y ligando los

fragmentos del tamaño adecuado en el sitio *Bam*HI del vector lambda-EMBL4 (Yosikawa e Isono, 1990). Para la secuenciación completa de este cromosoma se utilizaron tres genotecas de diferente origen (Oliver et al, 1992). Para otros cromosomas como los X, XI, XV y VII se utilizaron mayoritariamente bibliotecas del genoma total construídas en el sitio *Bam*HI de los cósmidos pEW15 y pOU61cos (Figura 4) (Thierry et al., 1995). Los regiones de DNA no representadas se secuenciaron recurriendo a otras genotecas o mediante la estrategia del "walking" o paseo, que se describe en otro apartado de este capítulo. Los coordinadores de cada cromosoma distribuyeron el DNA a secuenciar a cada uno de los laboratorios participantes en cósmidos o en vectores lambda. Cada laboratorio subclonó su inserto en fragmentos menores, en el tipo de vector adecuado según su estrategia de secuenciación.

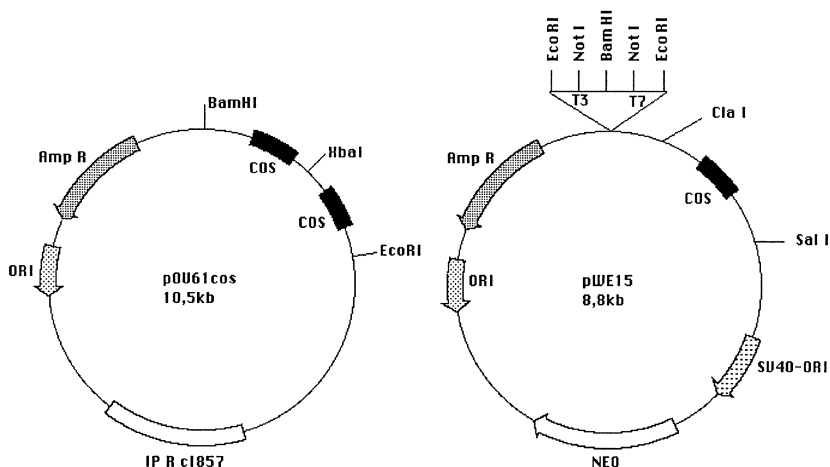


Figura 4.- Mapa de los cósmidos usados para construir algunas bibliotecas genómicas de *S. cerevisiae* (Thierry et al, 1995).

Los procedimientos para seleccionar los clones que componen una subgenoteca de un cromosoma o de un fragmento cromosómico, de entre todos los clones que componen una genoteca de mayor cobertura, se basan en técnicas de hibridación "in situ", como la que aplicaron Scholler et al. (1995) en el caso del brazo izquierdo del cromosoma XII, por ejemplo. En primer lugar, para construir la genoteca, dicho fragmento cromosómico se aisló a partir de un gel de electroforesis de campo pulsante pero junto con el DNA correspondiente al cromosoma IX ya que es del mismo tamaño. A continuación, se realizó una digestión parcial con *Mbo*I, se integraron los fragmentos generados en el cósmido Lawrist7, y se transformó *Escherichia coli*. La eficiencia obtenida fue de 20.000 transformantes/microgramo de DNA. De los clones obtenidos, solamente 768 se guardaron a -70°C individualizados en dos placas de microtitulación de 384 pocillos. Se estimó que este número de clones era

suficiente para obtener todo el DNA correspondiente al cromosoma IX y al brazo izquierdo del cromosoma XII con una redundancia de 30 veces, teniendo en cuenta que el tamaño de los insertos generados por la digestión parcial es de aproximadamente 37kb. Mediante un replicador de 384 puntas, se transfirió por duplicado el material bacteriano a membranas. El DNA se hibridó "in situ" con dos sondas correspondientes a los cromosomas XII y IX, respectivamente. Se seleccionaron los clones que sólo daban hibridación positiva con la sonda del cromosoma XII y de esta manera se elaboró la subgenoteca correspondiente. A continuación, ordenaron los clones según se describe en el apartado siguiente.

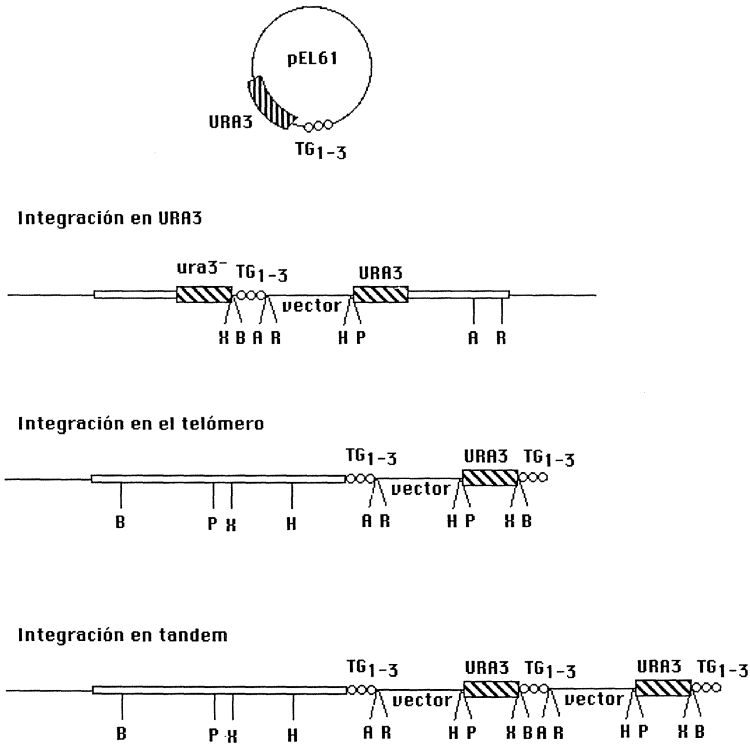


Figura 5.-Vector pEL61 y su integración en los telómeros (Louis y Borts, 1995).

Los telómeros presentan una problemática especial. Por los métodos convencionales de clonación, no se obtienen los fragmentos teloméricos los cuales poseen un extremo rico en G (regiones TG₁₋₃) que no se puede ligar sin una modificación enzimática previa. Por tanto, los telómeros de levaduras no suelen estar representados en las genotecas estándar que se usan en los proyectos

de secuenciación. Además, las regiones teloméricas contienen repeticiones que son comunes a varios cromosomas con lo cual la asignación de un clon a un determinado telómero de un cromosoma concreto resulta complicada.

Para solventar este problema, durante el proyecto de secuenciación de la levadura se ha desarrollado un método (Louis, 1994; Louis y Borts, 1995) basado en el diseño de un vector marcado que se integra en los telómeros y que permitió la clonación de cada uno de ellos específicamente por el procedimiento de "plasmid rescue" (rescate del plásmido). Dicho vector es un plásmido integrativo denominado pEL61 que contiene el marcador *URA3* y aproximadamente 200pb de una región TG₁₋₃. Al transformar la levadura con este vector, la integración puede tener lugar en el *URA3* o en la región TG₁₋₃, pudiendo en este caso ser única o en tandem (Figura 5). La eficacia de la transformación se incrementó creando mellas al azar en el vector con DNAsa I.

En primer lugar, para averiguar en que cromosoma de un determinado transformante se había producido la integración, se aisló DNA genómico total de cada uno de los transformantes (un total de 327), se separaron los cromosomas por electroforesis de campo pulsante (CHEF) y se transfirieron a membranas. Como la mayor parte del vector integrado corresponde a secuencias del plásmido bacteriano pBR322, se utilizó este plásmido como sonda. En 45 transformantes, se obtuvieron asignaciones a los 16 cromosomas. En el caso de los transformantes asignados a los cromosomas V y VIII, donde residen las copias cromosómicas del *URA3*, era necesario discriminar si la integración se había producido a nivel del telómero o del propio gen *URA3*, lo cual se consiguió mediante análisis de restricción. En segundo lugar, la localización en el telómero del brazo izquierdo o derecho de la integración se determinó mediante digestiones del DNA de cada cromosoma con *NotI* o *SfiI*, electroforesis CHEF, transferencia a membrana e hibridación con pBR322 marcado, comparando el patrón de bandas obtenido con el mapa físico del cromosoma correspondiente. Por último, para la fase de clonación, se seleccionaron enzimas que den fragmentos de tamaños clonables como plásmidos en *E.coli* (<20kb), se cortaron con ellos los cromosomas con la integración, se ligaron a baja concentración de DNA y se seleccionaron los clones por transformación de *E.coli*.

De este modo se clonaron 20 de los 32 telómeros de la levadura. La secuencia del telómero derecho del cromosoma II, por ejemplo, se describe en Feldman et al., 1994. Otros 3 telómeros se secuenciaron usando clones existentes previamente. Los 9 restantes, los cuales contienen elementos Y' y pueden ser desventajosos para *E.coli* o delecionan espontáneamente como los del cromosoma VIII, resultaron imposibles de clonar. Estas 9 regiones teloméricas se secuenciaron a partir de productos de amplificación por PCR directamente del genoma, usando una mezcla de Taq y Pwo que es correctora de errores y genera fragmentos largos. También se usó la PCR para secuenciar algunos huecos que quedaron entre los telómeros y las regiones adyacentes (Louis y Underwood, 1995).

4. ELABORACIÓN DE MAPAS FÍSICOS DEL GENOMA

Gran parte de los clones que componen una genoteca contienen insertos solapantes. Por tanto, una manera de optimizar el trabajo de secuenciación, reduciendo costes y tiempo, consiste en elaborar un buen mapa físico y elegir el número mínimo de clones que deben ser secuenciados.

La elaboración de un mapa genómico implica la generación de una genoteca ordenada de clones que sea representativa de un genoma en su totalidad (o de una parte determinada del genoma), junto con los suficientes marcadores genéticos que permitan el alineamiento preciso de los mapas físico y genético (Coulson y Sulston, 1990).

Cuando se comenzó el proyecto de secuenciación del genoma de la levadura ya se disponía de un mapa genético detallado de *S.cerevisiae*, que fue elaborado recopilando los datos aportados por numerosos investigadores individuales que los habían obtenido por métodos clásicos como los de hibridación de un gen clonado a un blot cromosómico, análisis de tétradas, etc. (Mortimer et al., 1992). Para el proyecto de secuenciación, se hizo además necesario elaborar mapas físicos detallados de cada cromosoma lo que se abordó mediante técnicas clásicas de "restriction fingerprinting" (huellas de restricción) o de hibridación, y también mediante métodos nuevos que se desarrollaron durante el proyecto de secuenciación.

La técnica de "restriction fingerprinting" consiste en obtener, a partir de cada clon, un número adecuado de fragmentos marcados y separables por electroforesis en gel de poliacrilamida de alta resolución de modo que, los cósmidos que se solapen entre 1/2 y 1/3 de su longitud, se reconozcan sin ambigüedades. En esquema, consta de las siguientes etapas: corte con una enzima de restricción, marcaje de los extremos, corte con segunda enzima de restricción y separación de fragmentos por electroforesis. El análisis de la autorradiografía revelará los solapamientos entre los clones en función de los patrones de restricción. Para la localización de los solapamientos y edición de la contigüidad de los clones se requiere un densitómetro de barrido acoplado a un sistema de análisis de imagen y "software" adecuado. Mediante esta técnica se cartografiaron genotecas en cósmidos de *Arabidopsis* y de *Caenorhabditis*. En ambos casos se utilizó como primera enzima *Hind*III y como segunda enzima *Sau*3AI (Coulson y Sulston, 1990). En el caso de la genoteca en lambda-EMBL4 del cromosoma III de la levadura, se utilizó una variante de esta metodología, en primer lugar se realizó digestión parcial de cada clon con ocho enzimas de restricción, seguida de separación de las bandas por electroforesis en agarosa y transferencia a membrana de nitrocelulosa, finalmente hibridación con un fragmento marcado del brazo derecho del vector y posterior análisis informático de las bandas obtenidas en la autorradiografía (Yoshikawa e Isono, 1990).

El desarrollo del proyecto de secuenciación de la levadura dió lugar a la aparición de nuevas metodologías para la elaboración de mapas físicos, como por ejemplo la ordenación de cósmidos llevada a cabo por el equipo del Dr. J.D. Hoheisel en el brazo izquierdo del cromosoma XII, basada en técnicas de hibridación (Scholler et al., 1995). Los clones previamente seleccionados como

pertenecientes a la subgenoteca de este fragmento según se explica en el apartado anterior, se transfirieron a membranas desde una placa de microtitulación mediante un replicador, a continuación se desnaturalizó y fijó el DNA y se hibridó "in situ" con sondas procedentes de clones elegidos al azar entre la genoteca. Cada hibridación proporciona información acerca de todos los clones que son solapantes con la sonda. A fin de conseguir el mapa con el menor número posible de hibridaciones, las sondas siguientes se eligen entre las que todavía no han dado positivos en las hibridaciones anteriores. Con esta estrategia y el "software" desarrollado por Mott et al. (1993) que ordena las sondas en vez de los clones, sólo fueron necesarias 19 hibridaciones para obtener el "contig" completo.

Otra técnica surgida durante el desarrollo del proyecto de secuenciación del genoma de la levadura, para ordenar los clones en las genotecas es la denominada "nested chromosomal fragmentation" (fragmentación cromosómica anidada) que posee la ventaja de que utiliza el propio DNA de la levadura y no los clones para construir el mapa, con lo que se eliminan muchos artefactos. Así se cartografió, entre otras, la genoteca de cósmidos sobre la que se secuenció el cromosoma XI de 620kb (Thierry y Dujon, 1992).

En esencia, el método se basa en la fragmentación del cromosoma con la meganucleasa *I-SceI*, la primera disponible de una clase nueva de endonucleasas con secuencias de reconocimiento muy largas, en este caso de 18pb. La probabilidad de encontrar un sitio de corte *I-SceI* en un genoma es de, aproximadamente, una vez cada 20 veces el tamaño del genoma humano. En el genoma de la levadura, de hecho, está ausente. Gracias a *I-SceI* se puede cortar un cromosoma en un sitio único, introducido artificialmente por recombinación homóloga usando "cassettes" específicos con marcadores, o al azar, usando transposones.

La estrategia tal como se aplicó al cromosoma XI de la levadura es la siguiente:

- Integración de sitios *I-SceI* y selección de líneas de levaduras transgénicas: la integración se realizó por recombinación homóloga usando un "cassette" *URA3-I-SceI*, en siete puntos diferentes del cromosoma. Dos de los sitios se insertaron en genes cartografiados genéticamente y los otros en posiciones desconocidas a partir de cinco cósmidos no solapantes de la genoteca tomados al azar.
- Digestión de los cromosomas de cada una de las siete levaduras transgénicas con *I-SceI* y separación por electroforesis de campo pulsante. Así, se dedujo la posición de cada sitio *I-SceI* en cada línea transgénica en función del tamaño de los fragmentos, sin tener en cuenta la orientación derecha o izquierda.
- Selección de una sonda que hibride con el fragmento más corto del conjunto para orientar todos los fragmentos y deducir el mapa primario *I-SceI*. Se seleccionó la línea transgénica con el sitio *I-SceI* que generaba el fragmento más pequeño y que estaba, por tanto, más próximo a un extremo del cromosoma. El cósmido que se había usado para hacer esa integración, se utilizó como sonda frente a todos los

fragmentos resultantes de las digestiones I-SceI de los cromosomas XI de las siete líneas transgénicas. Hibridó solamente con uno de los dos fragmentos de cada cromosoma transgénico, permitiendo así establecer la orientación izquierda o derecha de los sitios I-SceI con relación a la sonda y elaborar un mapa I-SceI con una resolución de 80kb.

- Purificación de cada uno de los fragmentos cromosómicos a partir del gel de electroforesis en campo pulsante para usarlos como sondas para ordenar la genoteca del cromosoma XI. Los clones de la genoteca se hibridaron secuencialmente con los fragmentos cromosómicos seleccionados en orden creciente de tamaños desde el extremo izquierdo del cromosoma. Así se posicionó cada clon en el mapa I-SceI. A continuación, los clones en cósmidos ya ordenados, se digirieron con *EcoRI* y se hibridaron con las mismas sondas. Para cada sonda se distinguen dos casos, clones en los que hibridan todos los fragmentos y clones en los que sólo hibridan algunos fragmentos. En el primer caso, el inserto está contenido enteramente en un fragmento cromosómico, en el segundo caso está solapando un sitio I-SceI. De este modo se pudo ordenar con mayor precisión los clones en el mapa I-SceI y además colocar algunos sitios *EcoRI*. A continuación se compararon los perfiles de restricción *EcoRI* de los cósmidos más próximos entre sí y se aumentó la resolución del mapa físico del cromosoma hasta 3kb.
- Mediante hibridación con sondas conteniendo genes previamente cartografiados genéticamente se alinearon los mapas físico y genético, observando alguna inversión en un extremo, lo que recalcó la necesidad de un mapa físico elaborado independientemente del genético.

Una ventaja de este método frente a los de "fingerprinting" o hibridación es que permite medir la distancia física de un sitio dado a un extremo del cromosoma, independientemente de que haya un "contig" completo de clones que cubran ese segmento.

Más recientemente se ha publicado un método para el cartografiado de bibliotecas genómicas usando colecciones de oligonucleótidos (Sapolsky y Lipshutz, 1996). Implica una serie de pasos enzimáticos para capturar un grupo de marcadores cortos dispersos en un DNA de alto peso molecular. Estos marcadores, una vez amplificados y marcados por PCR, pueden ser hibridados y detectados en una colección de alta densidad de oligonucleótidos portando sondas complementarias a todos los posibles marcadores. Para cada clon en una biblioteca genómica, se puede determinar un grupo característico de marcadores. En base a la similitud entre los grupos de marcadores de cada par de clones, se puede medir su solapamiento relativo y así ordenar la genoteca. Esta nueva metodología no requiere métodos basados en geles o información previa de la secuencia, además, las manipulaciones que implica permiten su automatización.

5. TÉCNICAS DE SECUENCIACIÓN

La meta de muchos proyectos de secuenciación (Proyecto Genoma Humano y de otros organismos como *Saccharomyces cerevisiae*) para descifrar

la secuencia completa, requiere el desarrollo rápido y eficiente de métodos para la obtención de secuencias nucleotídicas. En la actualidad existen 2 métodos para la obtención de secuencias nucleotídicas, el método enzimático desarrollado por Sanger et al. (1977) el cual usa DNA polimerasas y terminación de cadenas de didesoxinucleótidos; y el método químico desarrollado por Maxam y Gilbert (1977) el cual usa reacciones químicas base-específicas. Aunque en principio son muy diferentes, estos dos métodos generan poblaciones separadas de oligonucleótidos marcados que comienzan en un punto fijo y terminan al azar en otro residuo fijo o combinación de residuos. Luego estos oligonucleótidos se separan por electroforesis bajo condiciones tales que pueden discriminar entre fragmentos de DNA individuales que difieren tan sólo en un nucleótido de longitud.

De estos 2 métodos, el enzimático ha resultado el más adecuado para la automatización, y es por ello que ha recibido más atención en el estudio del uso de diferentes polimerasas y reactivos para las mismas. La T7 DNA polimerasa descrita por Tabor y Richardson (1987) y la polimerasa del organismo *Thermus aquaticus* (Taq) (Saiki et al., 1988) han reemplazado la DNA polimerasa I (Fragmento Klenow) de *Escherichia coli* y la transcriptasa reversa, como las enzimas de elección para la secuenciación enzimática.

A) Método de Sanger

La técnica se basa en la síntesis enzimática a partir de un cebador ó "primer", de secuencia complementaria a una molécula de DNA molde, por acción de una DNA polimerasa que va incorporando, al extremo 3' de la cadena en crecimiento, dNTPs (2'-deseoxinucleótidos, de los cuales uno debe estar marcado con isótopo radioactivo o marcadores fluorescentes) y ddNTPs (2',3'-deseoxinucleótidos) que difieren de los anteriores en que les falta el grupo hidroxilo del carbono 3' del azúcar, de manera que evita la formación del enlace fosfodiéster entre la cadena en crecimiento y el siguiente nuevo nucleótido.

Las moléculas de DNA molde ó "templates" pueden ser de cadena sencilla, aquellas provenientes del bacteriofago M13, o de cadena doble en un vector plasmídico, cósmido o lambda DNA, y desnaturalizadas por acción del calor o álcalis. La calidad del DNA debe ser obtenida tras una gran purificación generalmente por centrifugación a través de gradientes de cloruro de cesio. Es esencial que la cantidad de moléculas de RNA se reduzca al máximo debido a que sus grupos 5'-OH, así como su pequeño tamaño, son los sustratos preferidos al competir efectivamente con las moléculas de DNA durante la reacción de marcaje terminal. La eliminación de estas moléculas de RNA es bastante difícil, y el uso de RNasa (Ribonucleasa A) sólo origina un aumento de su concentración debido a que después del tratamiento se obtienen pequeños ribonucleótidos.

Los cebadores ó "primer" son oligonucleótidos de entre 15 a 30 residuos y que deben ser de secuencia complementaria a la molécula de DNA molde, y con el extremo 5' bloqueado para que la síntesis se dirija en la dirección 3'. En el caso de moléculas de DNA de secuencia desconocida, éstas se introducen en

vectores de clonación que llevan a ambos lados del MCS (Sitio de Clonaje Múltiple) unas secuencias conocidas sobre las que hibridan unos cebadores comerciales: Universal y Reverse, que permiten la secuenciación en ambas direcciones y de las dos hebras de DNA.

Los dNTPs radioactivos más utilizados son [α -³²P]dNTPs y [³⁵S]dATP (Biggin et al., 1983). Los derivados de ³²P fueron los primeros en usarse, pero las partículas β que emiten tienen el inconveniente de que las bandas que se observan son muy grandes y difusas, además de la degradación del DNA ya que se produce radiolisis con el tiempo.

La simetría "Dyad" (debido a un alto contenido en G+C) puede formar estructuras secundarias internas que no son totalmente desnaturalizadas durante la electroforesis. Esto no se arregla sólo cambiando la polimerasa sino también con el uso de nucleótidos análogos convencionales de dNTPs como: dITP (2'-desoxiinosina-5'-trifosfato) y 7-deaza-dGTP (7-deaza-2'-desoxiguanosina-5'-trifosfato), que son buenos sustratos para las Sequenasas.

Para conseguir lecturas de secuencias más largas, del rango de las 700-800 bases, se utilizan soluciones de terminación que lleven una relación de ddNTPs:dNTPs del orden de 1:30 frente a la relación 1:10 en una secuenciación de lectura de 400-500 bases.

Las reacciones se cargan en un gel de poliacrilamida al 6% aproximadamente, que lleva urea como agente desnaturalizante, al que se le aplica un voltaje entre 1500-2000 voltios durante un par de horas. La concentración de acrilamida utilizada para preparar el gel dependerá de la resolución que se quiera obtener. Para conseguir una mejor resolución en los primeros 50 nucleótidos a partir del cebador la concentración de acrilamida puede oscilar entre 12-20%. Entre 25-400 nucleótidos a partir del extremo 5' del cebador se puede utilizar acrilamida al 6%. Si la concentración de acrilamida disminuye se puede obtener información de secuencias distantes, pero puede dar problemas a la hora de manipular el gel por su fragilidad. Los componentes de los geles son: acrilamida, N,N-metilen-bisacrilamida, solución amortiguadora Tris-Borato-EDTA (TBE), urea, persulfato amónico y N,N,N,N-tetrametil-etil-diamina (TEMED). Estos geles necesitan ser secados al vacío, como paso previo a la exposición de la película autorradiográfica. Después de la separación electroforética y posterior revelado por autorradiografía, la longitud del fragmento observado es igual a la distancia entre el punto de corte y el punto de marcaje. La lectura de los geles se realiza del extremo 5' al 3', es decir, de abajo hacia arriba de la autorradiografía.

En la lucha por obtener lecturas de hasta 1.000 bases se han conseguido aparatos de electroforesis termostatzados, de mayores dimensiones que los convencionales y de menor espesor (Garoff y Ansorge, 1981; Ansorge y Barker, 1984).

B) Método de Maxam y Gilbert

La secuenciación química fué desarrollada por A. Maxam y W. Gilbert. Se basa en la capacidad de determinadas sustancias de modificar específicamente bases nitrogenadas de la molécula de DNA. El éxito del método depende de la especificidad de estas reacciones de rotura, las cuales se llevan a cabo en 2 etapas:

- 1.- Modificación química de bases específicas (o tipos de bases).
- 2.- La base modificada es separada del azúcar con rotura de los enlaces fosfodiéster 5' y 3'. Los mecanismos químicos de las primeras reacciones son los siguientes:
 - G: DMS (Dimetil sulfato) metila el N7 de la guanina.
 - G+A: El ácido fórmico debilita las uniones entre A y G, protonizando los nitrógenos del anillo de las purinas.
 - C: En presencia de Cloruro sódico, la hidrazina sólo reacciona con citosina.
 - C+T: La hidrazina divide el anillo de T y C.

En una segunda reacción la Piperidina elimina las bases modificadas o fragmentos de bases. Una vez que la cadena de DNA se ha cortado, un solo fragmento es radioactivo ya que únicamente se ha marcado un extremo de la misma. Este tipo de secuenciación requiere un conocimiento del mapa de restricción del fragmento a secuenciar. Esto permite la generación de subfragmentos que pueden ser marcados en cualquiera de sus extremos por la enzima polinucleótido quinasa o por el fragmento Klenow de la DNA Polimerasa I de *E. coli*. Este método es óptimo para secuencias de DNA que tienen menos de 250 nucleótidos a partir del extremo marcado.

Este método ha sido desplazado por el método enzimático tras el descubrimiento de las nuevas polimerasas: T7 y Taq; y también debido a que las reacciones conllevan problemas: los reactivos a usar deben llevar un control de calidad y se necesitan DNAs de gran pureza. Pero se utiliza para la obtención de información de secuencias de regiones internas de grandes insertos de DNA, como la obtención de IPs, para el estudio de conformaciones de DNA e interacciones DNA-Proteína.

C) Secuenciación Automática.

La automatización de diferentes aspectos de las reacciones de secuenciación nucleotídica (purificación del DNA, reacciones enzimáticas, y transferencia de la información) ha sido objeto de numerosos trabajos, y estos pasos están siendo mejorados continuamente. Algunos de estos laboriosos aspectos de la secuenciación del DNA han sido automatizados parcialmente, tal como el uso de las estaciones de trabajo robóticas que realizan reacciones enzimáticas, el uso de lectores de films, y transferencia de información usando marcadores fluorescentes.

En la secuenciación automática el método utilizado es el enzimático de Sanger, lo que se automatiza es la separación electroforética y la detección de los productos. De esta manera obtenemos los electroforetogramas que presentan las bandas (fragmentos de DNA marcado) como picos en el eje de las "Y", y el tiempo de la electroforesis en el eje de las "X". Cada pico es identificado como: A, T, G, ó C dependiendo del sistema de detección, que en estos aparatos es de DNA marcado con fluorescencia. Las compañías que ofrecen secuenciadores automáticos, se basan en la detección de fragmentos de DNA marcados con fluorescencia y usan laser para leer el gel y detectar los fragmentos de DNA, pero se diferencian en la química de los marcadores fluorescentes utilizados. Unos utilizan un grupo de 4 marcadores diferentes en los terminadores ddNTPs (Prober et al., 1987), y otros emplean un grupo de 4 cebadores marcados diferentes, un cebador para cada una de las reacciones de secuenciación, A, T, G, y C (Smith et al., 1986). El grado de fluorescencia se detecta por 2 tubos fotomultiplicadores (PMT1 y PMT2) que miden a distintas longitudes de onda y el software integra ambos valores.

Esta técnica basada en la secuenciación enzimática mediante un secuenciador automático utiliza un cebador que se divide en cuatro fracciones y en cada una de éstas el marcaje en su extremo 5' es con un fluorocromo de color diferente (Fluoresceína, NBD, Rojo de Texas y Tetrametilrodamina). Posteriormente, a cada una de estas fracciones se le añade un único ddNTP y se realiza una incubación. Todos aquellos fragmentos que hayan incorporado el mismo ddNTP estarán marcados por el mismo color. Posteriormente las 4 reacciones son mezcladas y colocadas en un gel de secuenciación. Cada banda observada corresponde a un conjunto de fragmentos del mismo tamaño y estará marcada con el fluorocromo correspondiente al cebador que se ha utilizado en la síntesis de los fragmentos.

La electroforesis se realiza de forma continua en un solo carril y el color de la banda se determina cuando el fluorocromo es excitado por un rayo láser situado en la base del gel, produciendo así una señal que es recibida por un fotomultiplicador y transmitida a un ordenador (Figura 6). El análisis de las señales recibidas en el ordenador permite establecer con gran precisión la secuencia del fragmento en estudio, incluso cuando hay cierta ambigüedad debida a la presencia de bandas muy cercanas.

Una alternativa a este proceso consiste en marcar todos los cebadores con el mismo fluorocromo y realizar la reacción de secuenciación en cuatro incubaciones independientes de igual forma que la secuenciación clásica manual. Posteriormente las bandas obtenidas en los cuatro carriles del gel son analizadas por unas baterías de sistemas excitadores de los fluorocromos por rayos láser y fotodiodos.

Esta variante tiene dos ventajas respecto a la anterior. En primer lugar supone un ahorro económico, ya que es más caro marcar los cebadores con los diferentes fluorocromos que con el mismo fluorocromo. En segundo lugar, el fijar los fluorocromos a los cebadores produce interferencias en la migración, si el marcaje es siempre con el mismo fluorocromo, la alteración de la migración

siempre será la misma a diferencia de si se usan varios fluorocromos diferentes, en los cuales la alteración en la migración será distinta.

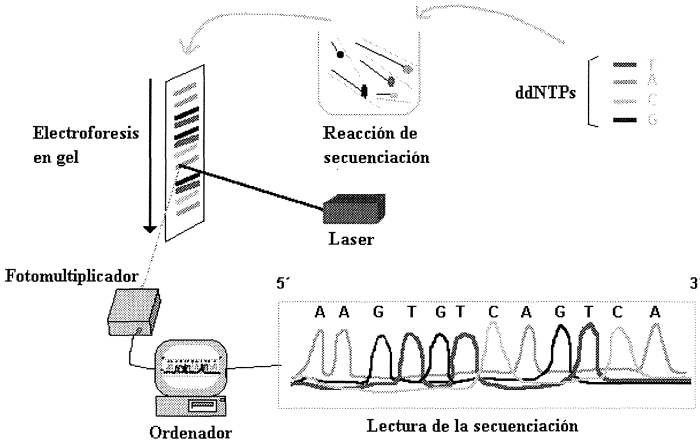


Figura 6.-Sistema de secuenciación automática.

Otra alternativa consiste en marcar con los cuatro fluorocromos los ddNTPs en lugar de los cebadores. Esta alternativa supone un mayor ahorro económico y el color solamente depende del didesoxinucleótido incorporado y el desarrollo electroforético se realiza en un solo carril como en el primer caso.

Una de las utilidades presentes en el sistema de control del ordenador de un secuenciador automático es que permite directamente comparar en la pantalla diferentes grupos de datos (Amorese y Hochberg, 1989). Así, en la búsqueda de secuencias polimórficas, los datos actuales así también como las secuencias derivadas de varios geles pueden ser cargados en ventanas de la pantalla. Posibles polimorfismos pueden ser visualizados y confirmados por este método de comparación. Otra de las utilidades es la lectura correcta de zonas de compresión que son difíciles de visualizar en una secuenciación clásica manual.

D) Estrategias de Secuenciación.

Antes de comenzar a secuenciar, es importante desarrollar una estrategia general que tenga en cuenta: el tamaño de la región a ser secuenciada, la exactitud de la secuencia requerida, y las facilidades de las que se disponga.

Solamente una pequeña proporción de proyectos tratan sobre la secuenciación de fragmentos nuevos de DNA. Su uso más frecuente es para el mapeo e identificación de mutaciones (puntuales ó deleciones) o para verificar la orientación y estructura de las construcciones de DNA recombinante.

Para la confirmación de la secuencia de DNA en mutantes se usan oligonucleótidos próximos a la mutación. Para plantearnos la secuenciación de un fragmento de DNA nuevo se pueden seguir varias estrategias de las cuales destacan:

- 1- Secuenciación Directa de deleciones seriadas ó "*Nested deletions*": que comienzan en un punto común, generalmente en un extremo de la sonda de DNA, manteniéndose el otro extremo constante, y se interna progresivamente dentro de la región a secuenciar, mediante el uso de distintas nucleasas como: Bal31, exonucleasa III y DNasaI entre otras. En esta estrategia se utiliza el mismo "primer" para todas las reacciones, y que suele ser uno de los cebadores universales que hibridan en el MCS del vector.
- 2- Secuenciación "Al Azar" ó "*Shot-gun*": donde se procede a secuenciar distintos fragmentos al azar y luego se introducen los datos en el ordenador para su solapamiento. Esto se realiza con enzimas de restricción de corte múltiple, sonicación o con DNasa pancreática, siendo los 2 últimos los de mejor resultado.
- 3- Secuenciación por Paseo ó "*Primer Walking*": donde el cebador se desplaza a lo largo del fragmento de DNA a secuenciar progresivamente, es decir, se van construyendo nuevos cebadores a medida que avanza la secuenciación en el fragmento de DNA a estudio. Esta última estrategia requiere la posibilidad de disponer de un sintetizador de oligonucleótidos.
- 4- Secuenciación por Hibridación ó "*Sequencing by hybridization*" (SBH): que conlleva el uso de un grupo de combinaciones de los 4 nucleótidos (256 tetranucleótidos) que se fijan a una placa, se marcan con fluorescencia, y a través de los cuales se hace pasar el DNA sonda a secuenciar.
- 5- Secuenciación desde varios "IPs" ó Puntos de Iniciación: En los primeros métodos se consume demasiado tiempo, este problema puede subsanarse obteniendo mas puntos de secuenciación (internos dentro del inserto) del cual iniciar la secuenciación de doble hebra; estos puntos se refieren como Puntos de Iniciación (IPs). Este método puede ser usado para cualquier inserto de DNA clonado. Existen diversos métodos para la obtención de IPs:
 - a) El uso del procedimiento químico de secuenciación (Maxam y Gilbert, 1977).
 - b) El uso de "Splinkers" (covalently linked sequencing primer linkers) para añadir una secuencia conocida de cebador a un sitio de enzima de restricción (Kalisch et al., 1986).
 - c) El más recientemente desarrollo del uso de la Exonucleasa III (Sorge y Blinderman, 1989).
 - d) El uso de oligonucleótidos cebadores para hibridación al azar a lo largo de la sonda de DNA (Studier, 1989; Siemieniak y Slightom, 1990).

Cualquiera de estos métodos es apropiado para la obtención de IPs adicionales. La segunda meta a conseguir es obtener lecturas de gran extensión que nos permita reducir: el número de oligo-cebadores necesarios, el número de reacciones enzimáticas, el número de sondas de DNA, la cantidad de secuencias solapadas, y lo más importante el número de geles de secuenciación necesarios.

Cada IP se extiende en ambas direcciones 5' y 3', dando lugar a regiones de secuencia de DNA de crecimiento continuo que se refiere como "de secuencia contigua" (*sequencing contig*). Una vez obtenidos los IPs de una manera al azar, su localización relativa se puede llevar a cabo mediante amplificación por PCR. Los cebadores 5' y 3' de cada IP se usan en series de amplificaciones diseñadas, incluyendo los oligos de las regiones extremas del MCS.

E) Ensamblaje de Secuencias de DNA.

Durante la secuenciación de genes se obtienen lecturas o secuencias contiguas que pueden abarcar una extensión que normalmente oscila de 200 a 800 nucleótidos según el método empleado. Si la estrategia de secuenciación que se ha seguido corresponde a delecciones seriadas, puede ser factible realizar una asignación de las distintas lecturas solapantes mediante una mera inspección visual, aunque el trabajo se simplifica mucho si disponemos de un programa de ensamblaje automático. En algunas estrategias de secuenciación como el *shot-gun* en el que se obtienen lecturas de fragmentos al azar del gen, disponer de un programa de este tipo resulta imprescindible.

El conjunto de programas que a continuación se describen como necesarios para llevar a cabo un ensamblaje forman parte del paquete informático GCG (Genetics Computer Group Inc., University Research Park, 575 Science Drive, Suite B, Madison, WI, 53711, U.S.A.). Sin embargo, son numerosos los programas que pueden utilizarse tanto con esta finalidad como con la de análisis posterior de la secuencia una vez ensamblada.

Los pasos fundamentales para poder realizar la operación de ensamblaje desde GCG se resumen a continuación:

- 1.- Preparación de la base de datos que contiene las secuencias a ensamblar.

Se utilizan los programas *GelStart* ó *NewgelStart* para la preparación de la base de datos. *GelEnter* se utiliza para añadir nuevos fragmentos, bien a partir del teclado, de un lector óptico o de un fichero.

- 2.- Eliminación sistemática de secuencias erróneas.

Conviene que dispongamos de la secuencia del vector en un fichero para así poder editar estas secuencias, limpiándolas de aportaciones exógenas, antes de proceder al ensamblaje. En el paquete GCG existe también un modo automatizado para poder visualizar y por tanto eliminar estas secuencias: se define desde el programa de *GelEnter*; también permite visualizar a través de los programas *MapSort* ó *FindPatterns* sitios reconocidos por determinadas enzimas de

restricción u otros patrones de secuencias que nos puedan servir de puntos de control para verificar un ensamblaje correcto.

3.- Reconocimiento de solapamientos.

El programa *GelOverlap* realiza el reconocimiento de solapamientos, compara cada una de las secuencias frente a todas las demás de la base de datos generada en el paso 1. El resultado es un fichero de salida que puede ser utilizado por *GelAssemble* para cargar secuencias en el editor en la posición probable de ensamblaje. *GelMerge* es una alternativa a *GelOverlap* (modificación introducida a partir de la versión 7.2) que lo supera en muchos aspectos ya que proporciona directamente los ensamblajes realizados a los que se denomina "contigs" (abreviatura de secuencia contigua). Estos ensamblajes pueden visualizarse y editarse desde *GelAssemble*.

4.-Confirmación de solapamientos y ensamblaje definitivo.

Los falsos solapamientos pueden crear tal magnitud de conflictos en el ensamblaje final que es necesario contar con un programa editor que nos permita aceptar los solapamientos sólo si son claramente correctos. *GelAssemble* es el programa editor que utiliza el resultado de los programas de búsqueda de overlaps para inspeccionar el alineamiento realizado y así tener la oportunidad de aceptarlo o rechazarlo. Los conflictos de disparidad de nucleótidos entre las distintas secuencias aparecen remarcados en la pantalla. Si el alineamiento realizado se considera correcto, el comando *meld* permite salvarlo en la base de datos. Para escribir la secuencia consenso se puede utilizar el comando *seqout*.

En el Proyecto Europeo de secuenciación del genoma de la levadura *Saccharomyces cerevisiae*, durante el curso de la secuenciación completa de un cromosoma, las secuencias son sucesivamente clasificadas como Fase 1, 2 y 3 después de la aplicación de criterios apropiados. Las secuencias FASE 1 son aquellas determinadas por los laboratorios de secuenciación. Las secuencias deben ser determinadas en ambas hebras con suficientes solapamientos y resultados de todas las posibles incertidumbres. Las estrategias de secuenciación detalladas son examinadas por el coordinador. Estas secuencias son así examinadas al comprobar los sitios de restricción predichos con el mapa físico, detección del orden de los fragmentos, y solapamientos con los fragmentos vecinos o con material publicado previamente (si existe). La secuencia del cromosoma entero debe ser determinada por los *contractors* de la red europea, exista o no material previamente publicado. Las secuencias que completan el criterio superior son clasificadas como FASE 2. Problemas menores se avisan al autor para su verificación y secuencias que muestran demasiados problemas se devuelven al autor después de una verificación general. Las secuencias de FASE 2 son comparadas con secuencias de verificación enviadas independientemente. Si existen discrepancias, se devuelven a ambos autores (el de la secuencia original y el de la secuencia de verificación) para que las revisen y resuelvan los problemas experimentalmente. Luego se hacen las correcciones apropiadas y así las secuencias se clasifican de la FASE 3, que ya están listas para el ensamblaje del cromosoma entero.

6. ENSAYOS DE “FOOTPRINTING”

Se define “*footprinting*” ó “huella” a los perfiles de fragmentos de DNA obtenidos despues de diversos tratamientos como: digestiones con enzimas de restricci3n, nucleasas, productos qu3micos, etc., en presencia y ausencia de prote3nas de uni3n a DNA. En este apartado nos referiremos s3lo a la determinaci3n de los sitios de interacci3n de prote3nas de uni3n a secuencias espec3ficas de DNA como factores de transcripci3n, y tambi3n para conocer la estructura cromat3nica, en cuanto a la disposici3n de las prote3nas histonas de los nucleosomas.

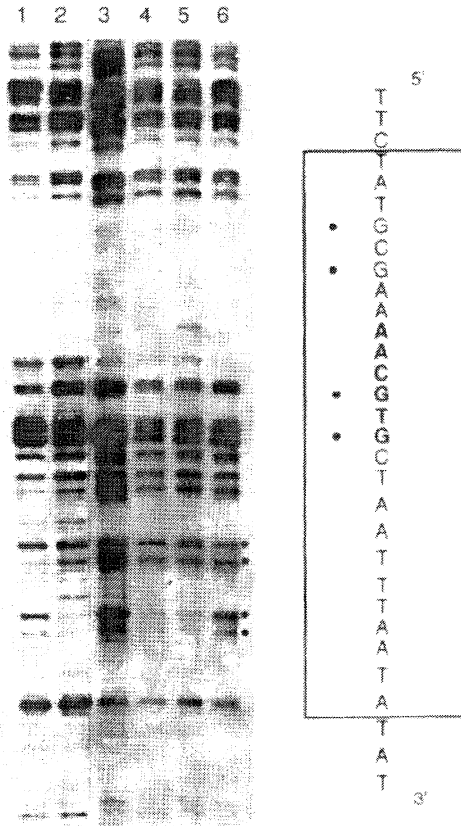


Figura 7.-Footprinting análisis con Dimetil sulfato (DMS) *in vitro* (1,2) e *in vivo* (3-6) de un fragmento de DNA “desnudo” (1,3) y en presencia de una prote3na espec3fica de uni3n a ese DNA. La secuencia del sitio de uni3n de la prote3na aparece en el recuadro donde se se3alan con puntos las G sobre las que actúa el DMS (modificado de Venter et al., 1994).

En los últimos años las técnicas de "*footprinting*" han ido incrementando su importancia en el estudio de las interacciones proteína-ácido nucleico. Hay 2 grandes grupos de sondas utilizadas en estas técnicas: las sondas químicas y las enzimáticas. Estas últimas, como la DNasaI, o la MNasa I tienen la ventaja de actuar específicamente sobre el DNA, tanto en condiciones "*in vitro*" como "*in vivo*". Las sondas químicas son menos específicas y algunas de ellas reaccionan también con la proteína, modificando la correcta interacción de la proteína y el DNA. La sonda química más utilizada, y que produce menos interacciones es el DMS (Dimetil Sulfato) (Figura 7). Para estudios de complejos DNA-proteína se prefiere el uso combinado de estas técnicas.

En estos métodos la visualización de los fragmentos obtenidos después del tratamiento correspondiente se puede llevar a cabo de distintas maneras:

- 1.- Marcaje directo de la sonda de DNA utilizada en la reacción.
- 2.- Amplificación de los fragmentos obtenidos mediante técnicas de PCR utilizando "primers" marcados.
- 3.- Técnicas de hibridación de los productos de la reacción con sondas de DNA marcadas complementarias a la región a estudio.

ENSAYOS "*in vitro*"

En esta técnica un fragmento de DNA de secuencia conocida interacciona con una proteína de unión específica a ese DNA, y posteriormente es parcialmente digerido con DNasaI. La proteína unida protege la región del DNA al que se encuentra unido del ataque de la DNasa. El análisis posterior del DNA degradado por electroforesis y posterior autorradiografía identifica la región protegida como un hueco "*gap*" entre los productos de digestión del DNA.

La DNasa I es una proteína de aproximadamente 40 Amgstroms de diámetro. Se une al surco menor del DNA y corta los enlaces fosfodiéster de ambas hebras de DNA independientemente. Su gran tamaño ayuda a prevenir la digestión del DNA por debajo y alrededor de una proteína unida. Sin embargo, la proteína generalmente tiene otros efectos sobre el ataque normal por la DNasaI, resultando que algunos sitios del DNA se vuelven hipersensibles a la enzima, así como observar un cambio en el patrón de la diana de DNasa sobre el DNA sin ningún tipo de protección, y todo ello debido a la modificación de la estructura terciaria del DNA al unirse la proteína.

Para localizar la posición de las huellas, se suelen cargar en los geles de electroforesis reacciones de secuenciación química G+A y/o C+T del mismo fragmento de DNA marcado. Se suele realizar el experimento sobre las 2 hebras de DNA para conocer el comportamiento de la proteína unida.

Entre las sondas químicas más utilizadas está el DMS, reactivo químico que produce metilación de las bases Guaninas. El uso posterior de Piperidina, rompe los enlaces de estas bases metiladas de su unión a la cadena de DNA originando unos patrones de secuencia específicos. Después de la adicción de las proteínas de unión al DNA, la protección de regiones específicas de la sonda

evita la acción del DMS, dando lugar a diferencias en los geles de electroforesis donde se separan dichos fragmentos.

ENSAYOS "*in vivo*"

Este tipo de técnicas requiere la permeabilización previa de la membrana celular del organismo a estudio o el aislamiento de núcleos, para permitir la entrada al mismo de los distintos agentes enzimáticos ó químicos. Se suele emplear para el estudio de la distribución de los nucleosomas en la cromatina de un determinado fragmento de DNA en diferentes estados de activación ó represión, para observar el efecto de la maquinaria enzimática de transcripción, así como de factores transcripcionales, que al unirse al DNA provocan el desplazamiento o la distorsión de los nucleosomas presentes en esa zona.

Además del DMS y de la acción de la DNasa I, en los ensayos "*in vivo*" se utiliza otra enzima la MNasa I (Micrococcal nucleasa) que es específica de regiones de DNA espaciador (entre los nucleosomas) donde no existe protección por las proteínas histonas u otros factores transcripcionales. Esta enzima es una endonucleasa tanto de DNA como de RNA.

En todos estos ensayos es necesaria la inclusión de controles de DNA "desnudo" (en ausencia de proteínas) digerido de igual manera, con el fin de detectar puntos de corte debidos a la especificidad de la secuencia que pueden presentar estos enzimas y no originados por la estructura del DNA o de la cromatina. Otra característica a tener en cuenta es la aparición de las llamadas regiones "hipersensibles", que surgen como consecuencia de la presencia de las proteínas que al unirse al DNA, provocan que zonas anteriormente protegidas sean más vulnerables a la acción de las enzimas o agentes químicos.

7. ENSAYOS DE POLIMORFISMO GENÉTICO

Con el desarrollo de la Biología Molecular han surgido numerosas técnicas para la identificación y clasificación de microorganismos, tanto en la discriminación entre especies como dentro de la misma especie. Las secuencias nucleotídicas de los genes de rRNA y regiones espaciadoras se han usado para analizar relaciones filogenéticas entre procariotas y eucariotas. Debido a que la secuenciación nucleotídica es bastante laboriosa y consume mucho tiempo, es mucho mas beneficioso el uso de enzimas de restricción que generan patrones de restricción de las regiones amplificadas de rDNA.

Se conocen numerosas técnicas que producen estos fragmentos de restricción, que no se usan aisladas sino que se intenta comparar resultados de varias de ellas, unas de mejor aplicación que otras segun el organismo a estudio. Entre ellas se mencionan las siguientes: ensayos RAPD (Random Amplification Polymorphic DNA: amplificación al azar de DNA polimórfico); ensayos de PCR-Fingerprinting; análisis por enzimas de restricción de regiones amplificadas ITS (Espaciador Transcrito Interno) y NTS (Espaciador No Transcrito), que son regiones espaciadoras entre el ss rDNA (subunidad pequeña de DNA que

codifica para el rRNA) y el ls rDNA (subunidad grande que codifica para el rRNA). Estas regiones espaciadoras son altamente variables entre especies y dentro de una misma especie, por lo que se utilizan en estos estudios de variabilidad genética (Baleiras-Couto et al., 1995).

Ensayos RAPD

Consisten en una reacción de PCR que utiliza un único primer, de secuencia nucleotídica arbitraria (de unos 10mer), de manera que tiene numerosos sitios de unión por complementariedad a lo largo del DNA genómico utilizado como sonda (Grando et al., 1994). Si algunos de estos sitios de unión están orientados de manera opuesta o palindrómica a lo largo del DNA sonda, y están lo suficientemente cercanos para permitir la amplificación, se generará un producto en la reacción de PCR. En una única reacción de RAPD se generan numerosos productos de amplificación de longitud nucleotídica variada. Estos ensayos RAPD se basan en el empleo de cientos de estos primers de 10 nucleótidos de secuencia arbitraria, donde se eligen aquellas reacciones de PCR cuyos fragmentos polimórficos son de gran intensidad y producen variabilidad genética entre las muestras a estudio. Se emplean geles de electroforesis de agarosa y poliacrilamida para fraccionar los productos, dando lugar a un perfil análogo al DNA fingerprinting.

Ensayos de PCR-Fingerprinting

Se basa en el mismo fundamento que la técnica anterior pero en este caso el único oligonucleótido utilizado puede ser: uno de los micro-satelites (GTG)₅, (GACA)₄, ó una secuencia central del bacteriofago M13. Se generan de esta manera fragmentos de distintos tamaños que se separan en un gel de electroforesis y dan lugar a patrones de DNA que permite discriminar especies entre sí y variedades dentro de la misma especie.

Análisis de fragmentos de restricción de regiones amplificadas

Una vez obtenidas las regiones amplificadas ITS y/o NTS mediante técnicas de PCR y con el uso de “primers” de secuencia conocida de las regiones ss rDNA y ls rDNA, que flanquean dichas regiones espaciadoras, los productos obtenidos se someten a digestiones con enzimas de restricción generando fragmentos que se separan en geles de electroforesis. Con esta técnica también se obtienen patrones característicos que se utilizan en la discriminación de especies.

La Biología Molecular en plena expansión, nos permite utilizar numerosas variaciones de estas técnicas, donde se combinan reacciones de PCR con primers de secuencia diversa y enzimas de restricción que dan lugar a patrones característicos capaces de diferenciar tanto especies diferentes como variabilidad genética dentro de una misma especie. Es necesario hacer estudios previos en cada caso para optimizar las condiciones de ensayo.

8. REFERENCIAS

- Alvarado-Urbina, G., Sathe, G.M., Liu, W.-C., Gillan, M.F., Duck, P.D., Bender, R. and Ogilvie, K.K. (1981). Automated synthesis of gene fragments. *Science* 214: 270-274.
- Amorese, D.A. and Hochberg, A.M. (1989). A fluorescence-labeled DNA analysis system for sequencing and mapping. *Am. Biotech. Lab.* 7: 38-46.
- Ansorge, W. and Barker, R. (1984). System for DNA sequencing with resolution of up to 600 base pairs. *J. Biochem. Biophys. Meth.* 9: 33-47.
- Ausubel, F. and cols. (1989). *Current Protocols in Molecular Biology*. John Wiley and Sons, Inc.
- Baiget, M., Gallano, P. y Tizzano, E. (1995). *Técnicas de Biología Molecular*. Publicación de la Sociedad Española de Bioquímica Clínica y Patología Molecular. Gráficos Vigor, S.A.
- Baleiras Couto, M.M., Eijmsa, B., Hofstra, H., Huis in't Veld, J.H.J. and Van der Vossen, J.M.B.M. (1996). Evaluation of molecular typing techniques to assign genetic diversity among *Saccharomyces cerevisiae* strains. *App. Env. Microbiology* 62 (2): 41-46.
- Benoît, L. and Moss, T.(1994). DNase I footprinting. *Methods in Molecular Biology* 30: 1-9.
- Berger, S.L. and Kimmel, A.R. (1987). *Guide to Molecular Cloning Techniques*. *Methods in Enzymology*, Academic Press.
- Biggin, M.D., Gibson, T.J. and Hong, G.F. (1983). Buffer gradient gels and ³⁵S label as an aid to rapid DNA sequence determination. *Proc. Natl. Acad. Sci.* 80: 3963.
- Birren, B. and Lain, E. (1993). *Pulsed field gel electrophoresis. A practical guide*. Academic Press, Inc. California.
- Cormack, B.P. and Struhl, K. (1993). Regional codon randomization: defining a TATA-binding protein surface required for RNA polymerase III transcription. *Science* 262: 244-248.
- Coulson, A. and Sulston, J. (1990). *Genome mapping by restriction fingerprinting*. En *Genome Analysis. A practical approach*. Ed. K.E. Davies. IRL Press. Oxford
- Davis, L.G., Dibner, M.D. and Battey, J.F. (1986). *Basic Methods in Molecular Biology*. Elsevier, N.Y.
- Dujon, B. (1993). Mapping and sequencing the nuclear genome of the yeast *Saccharomyces cerevisiae*: strategies and results of the European Enterprise.

- Symposia on Quantitative Biology Vol. LVIII, 357-366. DNA and Chromosome, Cold Spring Harbor Laboratory Press.
- Dujon, B. (1996). The yeast genome project: what did we learn?. *Trends in Genetics* 12: 263-270.
- Emery, A.E. and Rimoin, D.L. (1990). Principles and Practice of Medical Genetics. Vol. I. Churchill Livingstone, Medical Division of Longman Group, UK.
- Feldman, H. et al. (1994). Complete DNA sequence of yeast chromosome II. *EMBO J.* 13: 5795-5809.
- Ferl, R.J., Nairn, C.J., She, J.-X., Wakeland, E. and Almira, E. (1991). The application of automated DNA sequence analysis to phylogenetic studies, pag. 45-58. *Phylogenetic Analysis of DNA Sequences*. Oxford University Press.
- Gait, M.J. (1984). Oligonucleotide synthesis: A practical approach. IRL Press, Oxford.
- Garoff, H. and Ansorge, W. (1981). Improvements of DNA sequencing gels. *Anal. Biochem.* 115: 450-457.
- González Siso, M.I., Freire Picos, M.A., Cerdán, M.E., Wésolowski-Louvel, M. and Fukuhara, H. (1994). Chromosomal mapping of the *KICY1* gene from *Kluyveromyces lactis*. *Genome* 37: 515-517.
- Grando, M.S., Ubeda, J. and Briones, A.I. (1994). RAPD analysis of wine *Saccharomyces cerevisiae* strains differentiated by pulsed field gel electrophoresis. *Biotech. Tech.* 8 (8): 557-560.
- Guillamón, J.M., Barro, E. and Querol, A. (1996). Characterization of wine yeast strains of the *Saccharomyces* genus on the basis of molecular markers: relationships between genetic distance and geographic or ecological origin. *System. Appl. Microbiol.* 19: 122-132.
- Johnston, M. (1996). Genome sequencing: The complete code for a eukaryotic cell. *Current Biology* 6: 500-503
- Kalisch, B.W., Krawetz, S.A., Schenwaelder, K.-H. and Van de Sande, J.H. (1986). Covalently linked sequencing primers linkers (spinklers) for sequence analysis of restriction fragments. *Gene* 44: 263-270.
- Kirby, L.T. (1990). *DNA fingerprinting*. Stockton Press.
- Louis, E.J. (1994). Correct sequence for the right telomere of *Saccharomyces cerevisiae* chromosome III. *Yeast* 10: 271-274.

- Louis, E.J. and Borts, R.H. (1995). A complete set of marked telomeres in *Saccharomyces cerevisiae* for physical mapping and cloning. *Genetics* 139: 125-136.
- Louis, E.J. and Underwood, A.P. (1995). The cloning and sequencing of telomeres. EU Biotech Meeting of the Yeast Genome Sequencing Network, Lisboa, Book of Abstracts, 183.
- Maxam, A.M. and Gilbert, W. (1977). A new methods for sequencing DNA. *Proc. Natl. Acad. Sci.* 74: 560.
- Maxam, A.M. and Gilbert, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* 65: 499-559.
- Metzger, W. and Heumann, H. (1994). Footprinting with Exonuclease III. *Methods in Molecular Biology* 30: 11-20.
- Mierendorf, R.C. and Pfeffer, D. (1987). Direct sequencing of denatured plasmid DNA. *Methods Enzymol.* 152: 556.
- Mortimer, R.K., Contopoulou, C.R. and King, J.S. (1992). Genetic and physical maps of *Saccharomyces cerevisiae*, Edition 11. *Yeast* 8: 817-902.
- Mott, R., Grigoriev, A., Maier, E., Hoheisel, J.D. and Lehrach, H. (1993). Algorithms and software tools from ordering clone libraries; application to the mapping of the genome of *Schizosaccharomyces pombe*. *Nucl. Acids Res.* 21: 1965-1974.
- Narang, S.A. (1983). DNA synthesis. *Tetrahedron* 39: 3-22.
- Oliver, S.G. et al. (1992) The complete DNA sequence of yeast chromosome III. *Nature* 357: 38-46.
- Prober, J.M., Trainor, G.L., Dam, R.J., Hobbs, F.W., Robertson, C.W., Zagursky, R.J., Cocuzza, A.J., Jensen, M.A. and Baumeister, K. (1987). A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science* 238: 336-341.
- Rodríguez Belomonte, E., Rodríguez Torres, A., Tizón, B., Cadahía, J.L., González Siso, M.I., Ramil, E., Becerra, M., González Domínguez, M. and Cerdán, E. (1996). Sequence analysis of a 10Kb DNA fragment from yeast chromosome VII reveals a novel member of the DnaJ family. *Yeast* 12:145-148.
- Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239: 487-491.

- Sanger, F. and Coulson, A.R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* 94: 441.
- Sanger, F., Nicklen, S. and Coulson, A.R. (1977). DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci.* 74: 5463.
- Sapolsky, R.J., Lipshutz, R.J. (1996). Mapping genomic library clones using oligonucleotide arrays. *Genomics* 33: 445-456.
- Schmitz, A. and Galas, D.J. (1978). DNase I footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Res.* 5: 3157-3170.
- Scholler, P., Schwarz, S. and Hoheisel, J.D. (1995). High-resolution cosmid mapping of the left arm of *Saccharomyces cerevisiae* chromosome XII; a first step towards an ordered sequencing approach. *Yeast* 11: 659-666.
- Siemieniak, D.R. and Slightom, J.L. (1990). A library of 3342 useful nonamer primers for genome sequencing. *Gene* 96: 121-124.
- Slightom, J.L., Siemieniak, D.R. and Sieu, L.C. (1991). DNA sequencing: strategy and methods to directly sequence large DNA molecules, pag.18-44. *Phylogenetic Analysis of DNA Sequences*. Oxford University Press.
- Smith, C.L., Klco, S.R. and Cantor, C. R. (1990). Pulsed-field gel electrophoresis and the technology of large DNA molecules. En *Genome Analysis. A practical approach*. Ed. K.E. Davies. IRL Press. Oxford
- Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R., Heiner, C., Kent, S.B.H. and Hood, L. (1986). Fluorescence detection in automated DNA sequence analysis. *Nature* 321: 674-679.
- Sorge, J.A. and Blinderman, L. (1989). ExoMeth sequencing of DNA: Eliminating the need for subcloning and oligonucleotide primers. *Proc. Natl. Acad. Sci. USA.* 86: 9208-9212.
- Stothard, J.R. (1995). Computer programs for modeling the RAPD assay on DNA sequences. *J. Hered.* 86: 408-409.
- Stryer, L (1995). *Bioquímica*. Ed. Reverté. Barcelona.
- Studier, F.W. (1989). A strategy for high-volume sequencing of cosmid DNAs: Random and directed priming with a library of oligonucleotides. *Proc. Natl. Acad. Sci.* 86: 6917-6921.
- Tabor, S. and Richardson, C.C. (1987). DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci.* 84: 4767-4771.

- Thierry, A. and Dujon, B. (1992). Nested chromosomal fragmentation in yeast using the meganuclease I-Sce I: a new method for physical mapping of eukariotic genomes. *Nucleic Acid Research* 20: 5625-5631.
- Thierry, A., Gaillon, L., Galibert, F. and Dujon, B. (1995). Construction of a complete genomic library of *Saccharomyces cerevisiae* and physical mapping of chromosome XI at 3.7kb resolution. *Yeast* 11:121-135.
- Tizón, B., Rodríguez Torres, A., Rodríguez Belmonte, E., Cadahía, J.L. and Cerdán, E. (1996). Identification of a putative methylenetetrahydrofolate reductase by sequence analysis of a 6.8kb DNA fragment of yeast chromosome VII. *Yeast* 12: 1047-1051.
- Venter, U., Svaren, J., Schmitz, J., Schmid, A. and Hörz, W. (1994). A nucleosome precludes binding of the transcription factor Pho4 *in vivo* to a critical target site in the PHO5 promoter. *EMBO J.* 13 (20): 4848-4855.
- Williams, N. (1996). Yeast Genome Sequence Ferments New Research. *Science* 272: 481.
- Yosikawa, A. and Isono, K. (1990). Chromosome III of *Saccharomyces cerevisiae*: An ordered clone bank, a detailed restriction map and analysis of transcripts suggest the presence of 160 genes. *Yeast* 6: 383-401.