

ORGANIZACIÓN MOLECULAR DEL GENOMA DE LEVADURA

José E. Pérez Ortín

Departamento de Bioquímica y Biología Molecular

Universitat de València e Instituto de Agroquímica y Tecnología de Alimentos,
C.S.I.C.

1. EL PROYECTO DE SECUENCIACIÓN DEL GENOMA DE *Saccharomyces cerevisiae*

En los años 80 el desarrollo de las técnicas de secuenciación de DNA llegó a tal nivel de eficacia, tanto en términos de productividad por investigador como en los de coste por par de bases que, en distintos lugares del mundo se empezó a plantear la posibilidad de secuenciar de forma sistemática genomas completos de organismos. En Estados Unidos se lanzó la idea de la secuenciación del genoma humano (3000 Mb). En Europa las ideas fueron más modestas, pero más realistas también, y se pensó que era preferible la secuenciación de organismos modelo con genomas más pequeños. La ventaja del enfoque europeo residía en que, con una inversión muy inferior, se podía obtener casi tanta información sobre los genes de un organismo eucariota como la que se pensaba obtener con la secuenciación del genoma humano. Por otro lado la secuenciación de genomas pequeños podía servir para poner a punto las técnicas necesarias para afrontar proyectos, como el del genoma humano, de una envergadura miles de veces superior a cualquier otro proyecto de secuenciación realizado hasta esas fechas.

Los investigadores que tienen como organismo de estudio la levadura del pan, *S. cerevisiae* (que también es la que interviene en la producción de la cerveza, el vino y otros procesos biotecnológicos de gran interés), son uno de los grupos científicos mejor organizados tanto a nivel europeo (unos 300 laboratorios) como mundial (unos 4000 investigadores) y, por tanto, estaban en la mejor disposición para responder al reto de secuenciar un genoma completo. Por otro lado esta levadura tiene un número importante de ventajas como organismo modelo sobre otros posibles (Levy, 1994). En primer lugar es el organismo eucariótico mejor conocido con un importante nivel de desarrollo en

las técnicas de Biología Molecular y manipulación genética. En segundo lugar su genoma es pequeño (13 Mb) y compacto, es decir, con muy poco DNA no informativo. En tercer lugar, un número importante de sus genes eran ya conocidos a nivel genético o, incluso, estaban secuenciados, y se disponía de un elevado número de mutantes lo que hacía mucho más sencillo el trabajo de mapeo previo al de secuenciación. Finalmente, a pesar de todas estas ventajas, *S. cerevisiae* es un eucariota similar en casi todas sus características a todos los demás eucariotas y por lo tanto la información que se obtuviera de él se podría aplicar directamente a todos los demás organismos eucariotas (humanos incluidos) con una relación, por tanto, coste/beneficio muy baja.

Dado que la Unión Europea tiene desde hace años un sistema de proyectos de investigación cooperativos entre laboratorios de distintos países se pensó en la posibilidad de organizar un conjunto de grupos de investigación para la secuenciación del genoma de levadura. El proyecto empezó en enero de 1989, liderado por André Goffeau (Universidad de Lovaina la Nueva). Inicialmente se propuso la secuenciación del cromosoma III que es uno de los más pequeños (315 kb) y el mejor conocido en aquella época. Este cromosoma fue secuenciado por un grupo de 35 laboratorios coordinados por Steve Oliver (Universidad de Manchester). La idea era que laboratorios con experiencia previa en levadura y en secuenciación utilizaran parte de su potencial científico en el proyecto en lugar de crear nuevos laboratorios especializados en el tema. En 1992 se publicó la secuencia de este cromosoma (Oliver *et al.*, 1992), que fue la primera secuencia completa de un cromosoma publicada y el fragmento de secuencia contigua de DNA más largo que se había obtenido hasta esa fecha. El éxito de este proyecto piloto llevó a la secuenciación de nuevos cromosomas siguiendo el mismo esquema organizativo. De esta forma aparecieron publicadas las secuencias de los cromosomas XI (Dujon *et al.*, 1994) y II (Feldman *et al.*, 1994). El interés del proyecto se propagó rápidamente entre la comunidad de biólogos moleculares de levadura y pronto dos grupos de Canadá y Japón iniciaron la secuenciación de los cromosomas I y VI, respectivamente. Sin embargo lo que realmente cambió el panorama fue la entrada en este tema de tres laboratorios de nueva construcción especialmente pensados para secuenciación a gran escala: el Sanger Centre de Cambridge, Reino Unido; el laboratorio de Mark Johnston de la Universidad Washington de Saint Louis y el de Ronald Davis de la Universidad de Stanford. Estos tres laboratorios son capaces de secuenciar al ritmo de varias megabases por mes por lo que cromosomas de levadura enteros fueron secuenciados por un (relativamente) pequeño grupo de investigadores y técnicos en pocos meses. La constatación por parte de los científicos del proyecto de la Unión Europea de que la situación de partida había cambiado y que no todo el genoma de levadura estaba a su disposición para ser secuenciado forzó un acuerdo global entre todos los países y grupos. De esta forma todos los cromosomas se repartieron y se secuenciaron mucho antes de lo previsto: el 24 de abril de 1996 se anunció públicamente que el genoma de la levadura se había secuenciado completamente. El tamaño exacto es de 12.147.183 pb, exceptuando las repeticiones del rDNA y otras repeticiones menores. Teniendo en cuenta las repeticiones el tamaño es de unas 13 Mb. El porcentaje de genoma secuenciado por el consorcio europeo ha sido finalmente el 55%.

2. ANÁLISIS DE LOS RESULTADOS

A primera vista lo que más llama la atención del genoma de levadura es la gran densidad de genes que tiene: el 72% del genoma son genes (excluyendo el rDNA), lo cual deja muy poco espacio para DNA no codificante y otros elementos funcionales (Dujon, 1996). La longitud media de una pauta de lectura abierta (ORF) es de 1450 pb (483 codones). La más larga encontrada es una ORF de función desconocida del cromosoma XII (4092 codones). La más pequeña por ahora es *PMP1*, de 40 codones, que codifica para una proteína de membrana. En el caso de las ORFs pequeñas es difícil asegurar cuál es el límite inferior pues existen muchas ORFs cortas de las que no se sabe a ciencia cierta si son genes o no. En general se ha seguido el criterio de tomar como límite inferior de una ORF posible, el de 100 codones, a menos que, como en el caso comentado antes, se sepa con certeza que es un gen auténtico.

Dado el patrón de tamaños de ORFs, la compactación del genoma se debe a la rareza y pequeño tamaño de los intrones (el más largo es de 1 kb, pero habitualmente son mucho más pequeños) y a las cortas distancias intergénicas. En promedio, ORFs divergentes están separadas sólo por 618 nucleótidos, mientras que ORFs convergentes lo están por 316 nucleótidos. En el caso de ORFs orientadas en tandem la distancia de separación es intermedia. Por lo tanto, un gen típico de levadura tiene 1450 pb de pauta de lectura, una región promotora de 309 pb y una región terminadora de 163 pb, lo que hace un total de 1922 pb (Dujon, 1996).

La extrapolación de estos datos a todo el genoma de levadura da un valor de 6200 ORFs predichas. A partir de estudios de uso de codones y otros criterios se ha calculado que el 6-7% de ellas no son genes reales, con lo cual queda un total de 5800 genes codificantes de proteínas. Aparte de éstos existen unos 270 genes de tRNA, muchos de ellos con intrones pequeños (Dujon, 1996) y algunos genes de otros RNAs pequeños.

En cuanto a otros elementos genéticos se han encontrado un cierto número de retrotransposones Ty de varias clases, aunque no son muy comunes. Frecuentemente están integrados en la zona adyacente al promotor de un gen de tRNA, probablemente por su interacción con la maquinaria de la RNA polimerasa III. Muchos de ellos son sólo restos de una antigua integración que tienen las regiones LTR del transposón. Otros elementos comunes son las repeticiones X e Y' de las zonas subtelo méricas. Las últimas pertenecen a la familia de los elementos repetitivos LINE. Hay muy pocos pseudogenes, la mayor parte de ellos en regiones subtelo méricas. Existen también orígenes de replicación distribuidos a lo largo del genoma pero sólo en el caso del cromosoma III se sabe con cierta exactitud cuáles son funcionales y cuáles no. Se han encontrado 13 ARS funcionales en una región central de 210 kb (Newlon *et al.*, 1991), lo cual da una densidad de orígenes de replicación de uno por cada 15 kb aproximadamente.

La densidad de genes no es uniforme a lo largo de los cromosomas (Dujon, 1996). En muchos cromosomas hay regiones de alta densidad de genes (85%) separadas por otras de baja densidad (50%). En muchos cromosomas

ambas regiones se alternan de una manera bastante regular (Figura 1). Las regiones centroméricas y teloméricas siempre son de baja densidad. La orientación de las ORFs es aleatoria salvo en algunas regiones donde algunas agrupaciones en tandem pueden estar relacionadas con relación funcional entre ellas.

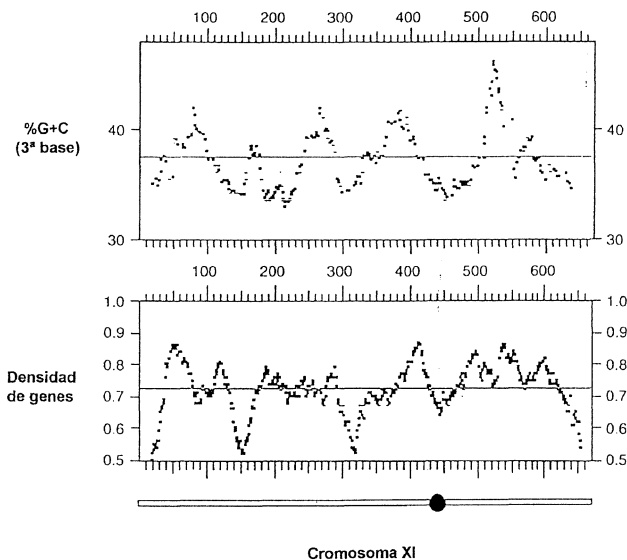


Figura 1.- Perfil del contenido en G+C y de densidad de genes a lo largo del cromosoma XI (tomado de Dujon, 1996).

También se han detectado variaciones de rango largo en la composición de bases (Figura 1). Así se ha visto que a lo largo de un cromosoma existen picos de alto contenido en G+C separados por regiones de mayor contenido en A+T (las pericentroméricas y subtelo méricas son siempre de este tipo). Aunque la correlación de estos picos con los de densidad de genes no es muy buena si que existe una cierta correlación (Dujon, 1996). El genoma de levadura es especialmente pobre en secuencias repetidas (Dujon, 1996). Aparte de las repeticiones de genes sólo se encuentran las repeticiones teloméricas y cierto número de repeticiones de poli (A)/poli (T) o poli(AT) alternante (Valle, 1993).

En resumen, un cromosoma de levadura se podría describir como una pieza de DNA de entre 150 y 1500 Mb cerrado en sus dos extremos por dos estructuras especiales, lo telómeros, constituidas por repeticiones TG₁₋₃ de unas 150-350 pb que forman estructuras tridimensionales especiales (ver más adelante). A estas regiones le siguen las zonas subtelo méricas, de unas 25 kb, pobres en G+C y en genes, que contienen elementos repetitivos X e Y' (Louis, 1995). En una posición más o menos central se encuentra la región centromérica,

que es muy pequeña (125 pb), también con una composición rica en A+T incluyendo algunas secuencias consenso, una estructura cromatínica especial y una vecindad pobre en genes (Clarke, 1990). Entre estos elementos se distribuyen de forma bastante regular, pero con orientación aleatoria, genes de unas 2 kb de longitud promedio ocupando la mayor parte del espacio disponible. De forma alternante con los genes, y con un espaciado bastante irregular, se disponen los orígenes de replicación (ARS) cuya funcionalidad parece no interferirse mutuamente con la de los genes adyacentes salvo en el caso de las repeticiones del rDNA, donde se ha visto que las horquillas de replicación avanzan sólo en la dirección de transcripción y son detenidas antes de que entren en un gen en dirección contraria al avance de la RNA polimerasa I.

3. DUPLICACIONES Y REDUNDANCIAS EN LOS GENES

Una cuestión importante es si el número estimado de genes, ~5800 como he dicho antes, es el necesario para el funcionamiento de la célula. De hecho los estudios de interrupción de genes han demostrado que sólo un pequeño número (15%) de los genes de levadura son esenciales (para el crecimiento en medios ricos). Sin embargo este dato puede ser engañoso porque muchos pueden desempeñar funciones no esenciales en todo momento, otros pueden resultar sólo esenciales en condiciones de competencia natural entre cepas o con otras especies de microorganismos y, finalmente, otros muchos pueden tener copias más o menos homólogas dentro del propio genoma de levadura. Este último dato se ha podido establecer con bastante precisión con el conocimiento de la secuencia completa del genoma.

De hecho, un alto porcentaje del genoma es redundante. Se han encontrado casos de repeticiones de genes casi (o totalmente) idénticos, casos de homólogos parciales con funciones iguales o distintas y finalmente duplicaciones de regiones enteras de cromosomas. La duplicación de buena parte de los genes y, en cierta manera, de su ordenación en los cromosomas ha sugerido que el genoma de *S. cerevisiae* es el resultado de una antigua duplicación de todo el genoma en un tiempo lejano y una posterior evolución con pérdida de parte de los genes duplicados y algunas reorganizaciones cromosomales (Wolfe y Shields, 1996). Un caso que ejemplifica bien todas estas posibilidades es el de la familia de transportadores de hexosas. Se conocen 20 genes con homología que varían entre el 25 y el 100% (Bargues *et al.*, 1996). Algunos de ellos parecen funcionalmente idénticos mientras que en otros casos se ha demostrado que su función es la de transportar diferentes monosacáridos (*GAL2*, transporta galactosa y *HXT1-4* transportan glucosa). Varios de ellos (*HXT3,6,7* y *HXT4,1,5*) están organizados en tandem en regiones de diversos cromosomas lo que sugiere duplicaciones dentro de pequeñas regiones (Figura 2). 9 de los 20 genes están en regiones subteloómicas (a menos de 25 kb desde el telómero) y otros 9 se localizan a unas 250-300 kb del telómero en diferentes cromosomas (ver Fig. 2). Esta curiosa disposición sugiere que la evolución de esta familia de genes implica recombinaciones entre diferentes cromosomas, especialmente en las regiones subteloómicas. Varios de ellos son tan parecidos que es difícil pensar que no sean completamente

redundantes a nivel funcional, sin embargo otros se expresan de formas distintas y producen proteínas con variantes de secuencia lo que podría indicar que desempeñan funciones sólo similares.

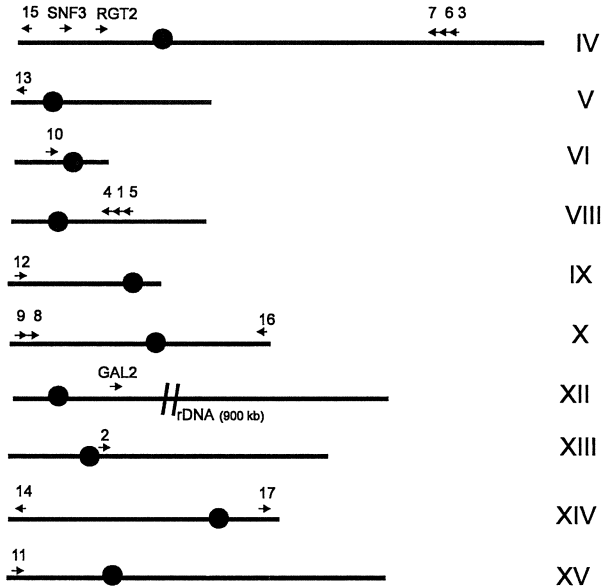


Figura 2.- Localización cromosómica de los genes de los transportadores de hexosas. Los genes llamados *HXT* se representan sólo por su número. *RGT2*, *GAL2* y *SNF3* están indicados por su nombre completo. Los cromosomas están dibujados a escala, identificados por su número (a la derecha) y representados desde el telómero izquierdo hacia el derecho. El centrómero se representa con un círculo negro. El cromosoma XII contiene una zona de 900 kb de repeticiones del rDNA que no se ha dibujado.

Aunque, de acuerdo con lo dicho anteriormente, parece claro que un cierto porcentaje de la redundancia génica es, en realidad, sólo aparente, también podría ser útil para la célula el que ciertos genes estén completamente duplicados a nivel funcional. De esa forma se dispondría de un cierto reservorio de genes repetidos que impedirían que mutaciones puntuales o deleciones producidas por recombinaciones entre cromosomas fueran peligrosas para la célula. Por otro lado la duplicación de genes es un bien conocido método de evolución para la adquisición de nuevas funciones sin perder las antiguas y su existencia en levadura puede ser una prueba de que los eucariotas utilizan profusamente este medio para adaptarse evolutivamente al medio. Finalmente es conocido que algunas duplicaciones, como las de los genes *SUC*, *MAL* o *CUP1*, puede ser un medio de aumentar la expresión de ciertos genes en condiciones ecológicas en

que esto sea beneficioso: crecimiento en sustratos ricos en disacáridos (*SUC*, *MAL*) o en presencia de metales tóxicos (*CUPI*).

En cuanto a la función de los genes sólo un 30% de ellos han sido caracterizados por métodos convencionales, aproximadamente otro 30% tienen homólogos en levadura o en otros organismos con funciones conocidas por lo que se puede suponer que conocemos también su función (aunque se han producido algunas sorpresas en este tema). Queda un 30-35% de genes que, o bien tienen homólogos de función desconocida (“parejas de huérfanos”) o no tienen homología ni función conocida (“huérfanos solitarios”) (Dujon, 1996). En algunos de estos casos los ordenadores pueden dar cierta idea sobre ellos, pues se puede predecir con cierta seguridad la existencia de hélices transmembrana, dedos de zinc, sitios de fosforilación etc... Sin embargo este tipo de predicciones no aportan más que algunas pistas, no siempre en la dirección correcta, al problema de encontrar la función de un gen. Este tipo de planteamiento, “desde el gen a la función”, es el inverso al habitualmente utilizado por la Genética y, por ello, ha recibido el nombre de “Genética inversa”. Con el progreso de los proyectos de secuenciación de genomas y la secuenciación de cDNAs al azar (las “ESTs”) ha pasado a ser mucho más frecuente tener el gen y buscar su función que tener la función y buscar el gen. Parece razonable pensar que debía haber más genes que los conocidos hasta antes de tener toda la secuencia del genoma de la levadura. Sin embargo cabe preguntarse si los genes huérfanos tienen alguna característica que los haya hecho “invisibles” al genético o al biólogo molecular. La respuesta no está clara por el momento pero seguramente la razón de su existencia no es otra que el hecho de que todavía estemos muy lejos de saber como funciona una célula y quedan todavía muchas funciones vitales insospechadas por descubrir.

Por otro lado la publicación reciente de la secuencias genómicas completas de dos bacterias (Fleischmann *et al.*, 1995) y una arqueobacteria (Bult *et al.*, 1996) ha permitido comparar los genomas de los tres grandes grupos de seres vivos. La organización genómica es mucho más parecida entre bacterias y arqueobacterias (un único cromosoma circular, genes en operones etc...) pero en otros muchos aspectos los eucariotas y arqueobacterias son más parecidos. Por ejemplo comparten la existencia de histonas, sistemas de transcripción y replicación similares y otras muchas similitudes que sugieren que están más relacionados entre sí que con las bacterias (Bult *et al.*, 1996). La complejidad de la célula eucariótica hace que, sin embargo, el porcentaje de genes dedicados a funciones informativas, energéticas y de comunicación sea diferente al de procariontes. La levadura dedica mayor porcentaje de su genoma a tareas de comunicación y menos a las de tipo energético que los procariontes (Ouzonis *et al.*, 1996).

Actualmente tanto la Unión Europea (proyecto EUROFAN) como otros laboratorios (como el de Mark Johnston en St. Louis) están llevando a cabo proyectos de deleción sistemática de los genes huérfanos y análisis del fenotipo de los mutantes con los que se pretende averiguar su función o, al menos, disponer de una colección completa de mutantes que pueda servir a la comunidad científica, tanto de levadurólogos como de investigadores en otros

eucariotas, para estudiar en el futuro la función de cualquier gen eucariótico y de las relaciones funcionales entre ellos.

4. ORGANIZACIÓN CROMATÍNICA DE GENES Y OTROS ELEMENTOS FUNCIONALES

Además de la secuencia del DNA el genoma de la levadura tiene otros niveles superiores de organización molecular. Como en todos los organismos eucarióticos el DNA está compactado en una estructura especial llamada cromatina. Aunque el objetivo inicial de este empaquetamiento pueda ser la reducción del tamaño que ocupa una cantidad tan grande de DNA y conseguir que quepa en un núcleo, indudablemente la evolución de los organismos ha ido proporcionando nuevas funciones a esta estructura y se puede decir que no es posible comprender el funcionamiento del material hereditario de un eucariota (replicación, transposición, recombinación, reparación o transcripción) sin tener en cuenta el papel activo que juega la cromatina (Tordera *et al.*, 1993).

En *S. cerevisiae* la cromatina está formada, como en todos los eucariotas, por nucleosomas cuyos componentes, las histonas H2A, H2B, H3 y H4, son muy similares a los del resto de los eucariotas. Los nucleosomas ocupan cualquier secuencia de DNA que no esté bloqueada por su propia estructura anómala (DNA-Z, cruciformes u otras) o por la unión de una proteína específica. Puesto que los nucleosomas empaquetan el DNA de una manera muy efectiva dificultan mucho los procesos que impliquen reconocimiento de secuencias o apertura de la doble hélice. La consecuencia de todo esto es que allá donde se quiera realizar una operación sobre el DNA la célula debe disponer los nucleosomas de tal forma que permitan el acceso de las maquinarias enzimáticas a las regiones sobre las que éstas deben actuar.

Como en muchas otras facetas de la Biología Molecular el estudio de la cromatina de levadura ha sido, y es, un campo de estudio pionero en muchos de los aspectos relacionados con las funciones de esta estructura típicamente eucariótica. La estructura cromatínica de los orígenes de replicación, centrómeros y telómeros es especial y se caracteriza por la existencia de otras proteínas estructurales además de las histonas. En los orígenes de replicación (ARS) hay una región de unos 150 pb sensible a nucleasas que está permanentemente cubierta por un complejo de proteínas llamado ORC ("origin recognition complex") (revisado en Diffley, 1995). En el caso de los centrómeros la zona funcional en cromatina está organizada como una región de 160-220 pb resistente a nucleasas flanqueada por nucleosomas ordenados a ambos lados (Figura 3). Esta organización molecular recibe el nombre de quinetócoro. Se ha encontrado que varias proteínas componentes del quinetócoro tienen homología de secuencia y estructura con las histonas por lo que el quinetócoro podría ser un tipo de nucleosoma modificado (revisado en Basrai y Hieter, 1995). El caso de los telómeros es más complejo. Las secuencias puramente teloméricas (repeticiones TG₁₋₃) tienen una estructura anómala de su DNA con apareamientos de cuatro hebras a través de las G y asociada a proteínas especiales que forman la estructura llamada telosoma que cubre unos

300 pb (Wright *et al.*, 1992). Las secuencias subteloméricas X e Y' están organizadas en nucleosomas normales pero ordenados de forma dependiente de la telosoma (Figura 3). Además la organización cromatínica de estas zonas subteloméricas es especial pues tiene efectos represores sobre la expresión de los genes situados en su vecindad (Louis, 1995).

La estructura cromatínica de los genes de levadura responde a un patrón común (Pérez-Ortín *et al.*, 1989). Un gen típico tiene tres regiones diferenciadas. El promotor contiene nucleosomas distribuidos de forma irregular y dependiente de las secuencias reguladoras. Es muy común que las secuencias UAS (equivalentes a los "enhancers") estén más o menos libres de nucleosomas y sean hipersensibles al ataque de nucleasas. Sin embargo la caja TATA y el inicio de transcripción suelen estar ocupadas por nucleosomas "posicionados", es decir que ocupan posiciones bastante invariables. Esta estructura puede cambiar cuando el gen se activa (en el caso de genes regulados). El cambio más frecuente es el ensanchamiento de la zona de sensibilidad de la UAS con la posible eliminación de los nucleosomas adyacentes que, en algunos casos, dejan libre a la caja TATA (Pérez-Ortín *et al.*, 1993). La segunda zona es la región transcribible que suele estar ocupada por nucleosomas, más o menos posicionados. La tercera zona es el flanco 3' que suele contener también nucleosomas posicionados y, muchas veces, un sitio hipersensible a nucleasas en la región inmediatamente posterior al final de la ORF. Estas dos últimas regiones no sufren cambios apreciables con el cambio la tasa de transcripción del gen (Figura 3).

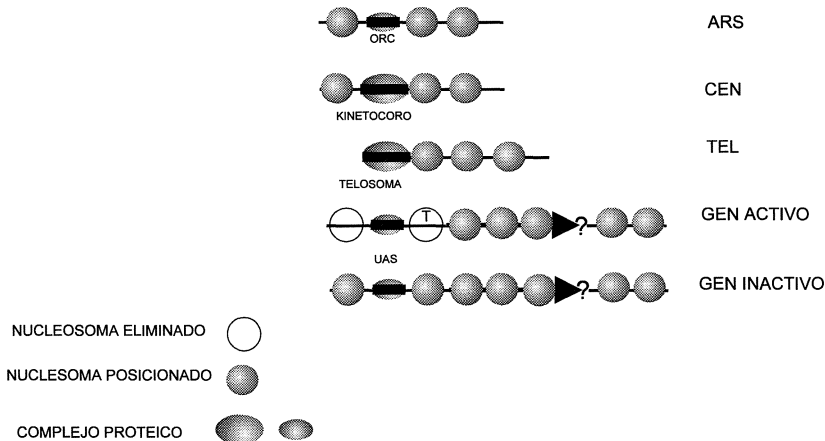


Figura 3.- Organización cromatínica de distintos elementos del genoma de la levadura. Se representa mediante símbolos la disposición característica de los nucleosomas y de otros complejos nucleoproteicos sobre la secuencia de DNA (línea horizontal). Los rectángulos negros representan las secuencias mínimas necesarias para la función de cada elemento. La ORF del gen se representa mediante una flecha gris. La T señala la caja TATA y el signo ? la existencia de una estructura alterada no caracterizada.

Es muy importante tener en cuenta que la organización en nucleosomas influye directamente sobre la transcripción y a su vez está influida por ella. Esta es la razón de la peculiar organización cromatínica de los genes descrita más arriba. La entrada de la RNA polimerasa se ve condicionada por la organización nucleosomal del promotor y, por lo tanto, los cambios de organización cromatínica del promotor responden a las diferentes posibilidades del aparato transcripcional: “represión” (nucleosomas bloqueando las secuencias reguladoras del promotor), “preparado” (huecos libres de nucleosomas para permitir un futuro acceso de la RNA polimerasa o bien la propia RNA polimerasa formando un complejo de pre-iniciación) y “en transcripción” (todos los factores transcripcionales unidos a sus secuencias diana la RNA polimerasa en forma de complejo activo). Este antagonismo entre nucleosomas y transcripción hace que sea necesario que la célula contenga factores proteicos para “remodelar” la cromatina de los genes en función de las necesidades transcripcionales. En levadura se han descrito dos complejos proteicos, llamados complejo SWI-SNF (por los nombres de los genes de las proteínas que lo forman) y RSC (Cairns *et al.* 1996), que tienen como misión desorganizar la estructura nucleosomal, mediante consumo de ATP, de los promotores de los genes en transcripción para permitir la acción de la RNA polimerasa (Peterson y Tamkun, 1995).

El movimiento de la RNA polimerasa, a su vez, produce cambios en la cromatina pues el molde de transcripción debe ser DNA desnudo de nucleosomas. No está claro si la lectura de la RNA polimerasa de un gen produce la eliminación total de los nucleosomas o sólo una desorganización parcial de las histonas que vuelven a reorganizarse en forma de nucleosoma una vez ha pasado el aparato transcripcional. Únicamente hay carencia completa de nucleosomas en los genes activos del rDNA debido a su alta tasa de transcripción, pero éste no es el caso de los genes de proteínas. En cualquier caso, el movimiento de la RNA polimerasa produce cambios en el superenrollamiento del DNA por delante y por detrás del avance de la polimerasa. Por delante de ella se genera superenrollamiento positivo que desestabiliza los nucleosomas y por detrás superenrollamiento negativo que favorece su re-formación. El final de la transcripción implica, en este orden, la detención de la polimerasa (por mecanismos no determinados) el corte y poliadenilación del mRNA y la liberación del complejo transcripcional. Todo esto debe suceder en un espacio muy corto dada la cercanía que existe entre los genes en levadura puesto que la entrada de un complejo transcripcional en el promotor o zona codificante del gen adyacente podría tener consecuencias negativas para éste, dada la capacidad desorganizadora de la cromatina que tiene el avance de la RNA polimerasa. Por otro lado, incluso aunque se disponga de un mecanismo eficaz de detención de la polimerasa no parece sencillo evitar la propagación de ondas de superenrollamiento positivo hacia los genes situados después del gen en transcripción. Estas ondas, si se propagaran dentro de los genes adyacentes, podrían tener efectos sobre su estructura cromatínica y, por tanto, sobre su tasa de transcripción. A pesar de estas consideraciones todavía no se ha encontrado cuál es el mecanismo que “aisla” los genes para evitar sus interferencias transcripcionales. Nosotros pensamos que el estudio de la

estructura cromatínica del flanco 3' de genes de levadura puede aportar las claves sobre este tema.

5. ALGUNOS EJEMPLOS DE ESTUDIO DE LA REGIÓN 3' DE GENES DE LEVADURA

Como he dicho anteriormente en muchos genes de levadura se ha encontrado un sitio hipersensible a nucleasas (SH) en cromatina en la zona proximal del flanco 3'. Estos sitios no sufren ningún cambio en función de la transcripción del gen pero suelen estar situados en la región del gen que se transcribe pero no se traduce (cola o "trailer" del mRNA). En un estudio que llevamos a cabo en nuestro laboratorio sobre la cromatina del gen de la Fructosa 1,6-bisfosfatasa (*FBP1*) encontramos que existía un SH a unos 100 pb del triplete de parada (Del Olmo *et al.*, 1993). El SH estaba flanqueado por nucleosomas posicionados en ambas direcciones. Estudiando la secuencia de DNA de la zona hipersensible encontramos que existía una secuencia (TA)₁₄ alternante casi perfecta capaz de formar un cruciforme por su carácter semipalindrómico (Del Olmo y Pérez-Ortín, 1993). Hemos encontrado que esta secuencia es necesaria, aunque no suficiente, para producir la poliadenilación del mRNA, que sucede unos 50 pb después (Aranda *et al.*, en preparación). ¿Qué relación puede tener un SH con el proceso de terminación/poliadenilación? Si tenemos en cuenta que la maquinaria de corte y poliadenilación de mensajeros actúa a nivel del propio mRNA y no del DNA parece que sólo el proceso de pausa de la RNA polimerasa, que se produce durante la lectura del gen, podría estar afectado por estructuras a nivel de cromatina. Hemos estudiado que estructura tiene este SH *in vivo* y hemos encontrado que no forma un cruciforme pero las hebras están separadas (Aranda *et al.*, 1996). La separación de la hebras (desnaturalización) es dependiente de la energía acumulada en forma de superenrollamiento negativo y, aplicando algoritmos de predicción de secuencias desapareables (Benham, 1993) esta secuencia (TA)₁₄ sería la única desapareada en esta región en las condiciones fisiológicas de superenrollamiento (Aranda *et al.*, 1996). Puesto que el desapareamiento de una secuencia es más sencillo cuanto más superenrollamiento negativo hay (al revés con el positivo) y el hecho de que desaparearse "consume" superenrollamiento se puede postular que la existencia de un sitio desapareable podría servir para contrarrestar el superenrollamiento positivo producido por el avance de la RNA polimerasa que debe ser eliminado para no acabar bloqueando su propio avance. Otra posibilidad sería que este sitio sirva para "absorber" el exceso de superenrollamiento negativo que deja detrás la RNA polimerasa cuando lo sobrepasa. En este último caso la disminución brusca de estrés superhelicoidal produciría una señal para la parada de la RNA polimerasa unos cientos de pb más abajo, es decir, podría ser la razón de la pausa que precede a la poliadenilación/terminación (Figura 4).

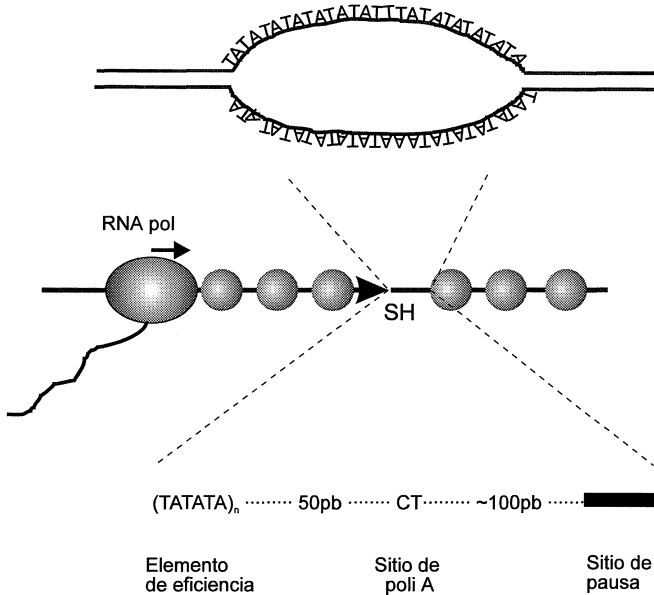


Figura 4.- Esquema de la región 3' del gen *FBPI* indicando sus elementos funcionales. La flecha indica el final de la ORF del gen. El sitio hipersensible a nucleasas se señala con SH. En las parte superior e inferior se representa a mayor escala la región desapareada *in vivo* y las secuencias implicadas en la poliadenilación/terminación, respectivamente.

Si alguna de estas hipótesis fuera cierta deberían existir sitios fácilmente desestabilizables (DS) en las regiones 3' próximas a cada gen. Para comprobar esta hipótesis hemos hecho un análisis de desestabilización del cromosoma III entero (315 kb) usando el algoritmo de C. Benham (Pérez-Ortín *et al.*, en preparación). Este análisis demuestra que en el 82% de los genes de este cromosoma existe un DS en su flanco 3', a menos de 200 pb del triplete de terminación. Es frecuente también (64%) la existencia de este tipo de sitios en la región del promotor (hasta 400 pb desde el ATG) y muy infrecuente (23%) que los DS se encuentren dentro de la zona codificante de la ORF. Por lo tanto se puede concluir que un gen típico de levadura está flanqueado por uno o dos sitios DS y evita tener este tipo de regiones en su región codificante. Un ejemplo de una región del cromosoma se da en la Figura 5. Aunque todavía no podemos dar una explicación definitiva de este resultado la correlación entre DS y flancos de genes en levadura es altamente significativa. Los sitios DS también se correlacionan con los orígenes de replicación: los 6 ARS funcionales identificados en el cromosoma III (Newlon *et al.*, 1991) coinciden con sitios DS lo cual está de acuerdo con el hecho de que una de las propiedades requeridas para un origen de replicación es su capacidad de ser desapareable.

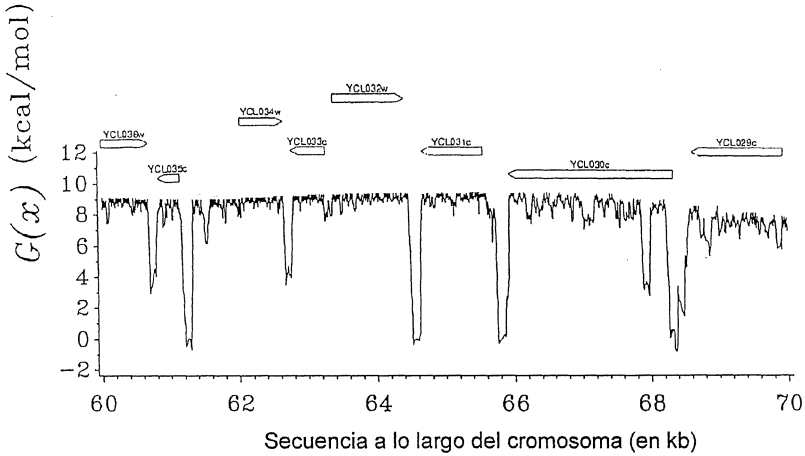


Figura 5.- Predicción de sitios desestabilizados en una región del cromosoma III entre 60 y 70 kilobases desde el telómero izquierdo. Cuanto más baja es la energía de desestabilización [$G(x)$] más fácil es el desapareamiento. Las cajas vacías indican la zona de cada ORF y su dirección de lectura, su situación a diversas alturas es sólo por motivos de claridad y no refleja ninguna característica distintiva. Obsérvese la coincidencia de picos con los extremos 3' de todas las ORFs.

En bastantes casos la proximidad de dos genes es tan grande que parece difícil que no se produzcan interferencias transcripcionales entre ellos. Se han descrito varios casos de genes cuya dirección de transcripción es convergente y que están tan cerca uno del otro que sus mensajeros solapan. Nosotros hemos estudiado el caso de la región del gen *POT1*. En la región 3' de este gen existe otro gen de función desconocida (llamado YIL161w en la nomenclatura sistemática posicional del genoma) cuyo triplete de terminación se encuentra a sólo 56 pb del de *POT1*. El análisis de los sitios de poliadenilación de ambos genes indica que los mRNAs solapan más de 100 nucleótidos, es decir, cuando la RNA polimerasa transcribe uno de ellos entra dentro de la zona de transcripción del otro y viceversa. El análisis de la estructura cromatínica de la región ha revelado que en ella existen nucleosomas posicionados pero, cuando los dos genes se están transcribiendo simultáneamente, el nucleosoma situado en la zona de confluencia de ambos genes pierde su posicionamiento. Nuestra hipótesis (ver Figura 6) es que la acumulación de superenrollamiento positivo causado por la transcripción simultánea y concurrente de ambos genes no puede ser disipado eficazmente por los sistemas habituales de la célula y ocasiona una desestabilización de los nucleosomas de la región. Este ejemplo demostraría que la acumulación de superenrollamiento positivo es un problema real en el caso del genoma de levadura pero que, incluso en genes muy próximos, la célula tiene

algún sistema para evitar las interferencias transcripcionales que ello ocasionaría.

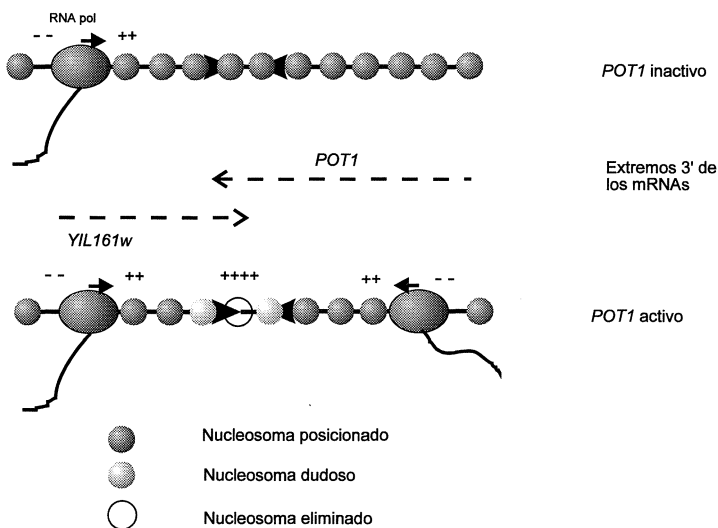


Figura 6.- Esquema de la organización cromatínica de la región de los genes *YIL161w* y *POT1*. Usando una simbología similar a la de las figuras precedentes se representa la estructura cromatínica de esa región en los estados de transcripción y represión de *POT1*. El avance de las moléculas de RNA polimerasa produce superenrollamiento positivo por delante de ellas (++) y negativo por detrás (-). En el caso de que ambos genes se transcriban a la vez el superenrollamiento positivo se acumula en la región terminal de ambos. Los extremos solapantes de los mRNAs se indican en el centro de la figura.

6. PERSPECTIVAS DE FUTURO

El conocimiento de la secuencia de DNA completa de una célula viva es algo nuevo para la Biología. Muchos genes tienen funciones totalmente desconocidas. En muchos casos, seguramente, se descubrirán nuevas funciones para genes que hoy se cree que ya son perfectamente conocidos. Todavía menos se sabe sobre las razones moleculares que gobiernan la organización de los genes dentro del genoma eucariótico y de los mecanismos evolutivos que han construido un genoma. En el caso particular de la levadura la superposición de los nuevos conocimientos con todo el que ya teníamos sobre su Biología Molecular y Fisiología así como la posibilidad de utilizar las herramientas de la Ingeniería Genética abre un horizonte de enormes posibilidades, muchas de ellas insospechadas hoy en día, para los próximos años.

AGRADECIMIENTOS

Quiero agradecer a mis compañeros de laboratorio Emilia Matallana, Agustín Aranda, Marcel.Í del Olmo y Sergio Puig su colaboración en los resultados que aquí se resumen. El trabajo en el laboratorio del autor ha sido posible gracias a las ayudas concedidas por la C.I.C.Y.T.: PB91/0329, BIO94-1271-CE y BIO96-2052-CE; y por la Unión Europea: BIO2-CT94-2071 y BIO4-CT95-0080.

7. REFERENCIAS

- Aranda, A.; Pérez-Ortín, J.E.; Benham, C.J. y del Olmo, M. (1997). Analysis of the *in vivo* structure of a natural alternating d(TA)_n sequence in yeast chromatin. *Yeast* 13 (en prensa).
- Bargues, M.; Salom, D.; Gómez, A.; Paricio, N.; Pérez-Alonso, M. y Pérez-Ortín, J.E. (1996). Sequencing analysis of a 4.1 kb subtelomeric region from yeast chromosome IV identifies *HXT15*, a new member of the hexose transporter family. *Yeast* 12:1005-1011.
- Basrai, M.A. y Hieter, P. (1995). Is there a unique form of chromatin at the *Saccharomyces cerevisiae* centromeres ? *Bioessays* 17: 669-672.
- Benham, C. J. (1993). Sites of predicted stress-induced DNA duplex destabilization occur preferentially at regulatory loci. *Proc. Natl. Acad. Sci. USA* 90: 2999-3003.
- Bult, C.J. *et al.* (1996). Complete genome sequence of the methanogen archeon, *Methanococcus jannaschii*. *Science* 273, 1058-1073.
- Cairns, B. R. *et al.* (1996). RSC, an essential, abundant chromatin-remodeling complex. *Cell* 87:1249-1260.
- Clarke, L. (1990). Centromeres of budding and fission yeast. *Trends Genet.* 6: 150-154.
- Del Olmo M. y Pérez-Ortín, J.E. (1993). A stable A/T-rich cruciform from the 3' end of yeast *FBP1* gene exists in *E. coli* cells. *Plasmid* 29: 222-232.
- Del Olmo, M.; Sogo, J.M.; Franco, L. y Pérez-Ortín, J.E. (1993). Chromatin structure of the yeast *FBP1* gene. *Yeast* 9: 1229-1240.
- Diffley, J.F. (1995). The initiation of DNA replication in the budding yeast cell division cycle. *Yeast* 11, 1651-1670.
- Dujon, B. (1996). The yeast genome project: what did we learn ? *Trends Genet.* 12: 263-270.

- Dujon, B. *et al.* (1994). Complete DNA sequence of yeast chromosome XI. *Nature* 369: 371-378.
- Feldman, H. *et al.* (1994). Complete sequence of chromosome II of *Saccharomyces cerevisiae*. *EMBO J.* 13: 5795-5809.
- Fleischmann, R.D. *et al.* (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269: 496-503.
- Levy, J. (1994). Sequencing the yeast genome: an international achievement. *Yeast* 10: 1689-1706.
- Louis, E.J. (1995). The chromosome ends of *Saccharomyces cerevisiae*. *Yeast* 11: 1553-1573.
- Newlon, C. S. *et al.* (1991). Analysis of a circular derivative of *Saccharomyces cerevisiae* chromosome III: a physical map and identification and location of ARS elements. *Genetics* 129: 343-357.
- Oliver, S. *et al.* (1992). The complete DNA sequence of yeast chromosome III. *Nature* 357: 38-46.
- Ouzonis, C.; Casari, G.; Sander, C.; Tamames, J. y Valencia, A. (1996). Computational comparisons of model genomes. *Trends Biotech.* 14: 280-285
- Pérez-Ortín, J.E., Matallana, E. y Franco, L. (1989). Chromatin structure of yeast genes. *Yeast* 5: 219-238.
- Pérez-Ortín, J.E., Igual, J.C., Del Olmo, M. y Matallana, E. Chromatin structure of glucose-repressible yeast genes. En "*Current Topics in Molecular Genetics*". Council for Scientific Research Integration. 1993. pp. 307-315.
- Peterson, C.L. y Tamkun, J.W. (1995). The SWI-SNF complex: a nucleosome remodeling machine ? *Trends Biochem. Sci.* 20: 143-146.
- Tordera, V., Sendra, R. y Pérez-Ortín, J.E. (1993). The informative content of the non-DNA moiety of chromatin. *Experientia* 49: 780-788.
- Valle, G. (1993). TA-repeat microsatellites are closely associated with ARS consensus sequences in yeast chromosome III. *Yeast* 9: 753-759.
- Wolfe, K. y Shields, D. (1996). Molecular evidence for an ancient duplication of the entire *Saccharomyces cerevisiae* genome. 1996 *Yeast Genetics & Molecular Biology Meeting*. Madison.
- Wright J.H.; Gottschling, D.E. y Zalian, V.A.(1992). *Saccharomyces* telomeres assume a non-nucleosomal chromatin structure. *Genes & Develop.* 6: 197-210.