



**Nuevas aportaciones al estudio del minisatélite  
humano MsH42:  
Filogenia, Digestión y Paradojas**

*Miguel González Blanco*

*Departamento de Bioquímica y Biología Molecular*

*Facultad de Biología*

*Universidad de Santiago de Compostela*

TESIS DOCTORAL

Mayo de 2005



**Jaime Gómez Márquez**, Catedrático de Bioquímica y Biología Molecular de la Universidad de Santiago de Compostela, y **Francisco Boán Fernández**, Doctor en Biología e Investigador del Departamento de Bioquímica y Biología Molecular de la Universidad de Santiago de Compostela, certifican:

Que la presente Tesis Doctoral, con el título: "*Nuevas aportaciones al estudio del minisatélite humano MsH42: Filogenia, Digestión y Paradojas*", presentada por el Licenciado en Biología **D. Miguel González Blanco**, ha sido realizada bajo nuestra dirección en el Departamento de Bioquímica y Biología Molecular de la Facultad de Biología de la Universidad de Santiago de Compostela y que, hallándose concluida, autorizamos su presentación para que pueda ser juzgada por el tribunal correspondiente.

Y para que así conste a los efectos oportunos, firmamos el presente certificado en Santiago de Compostela, a 20 de Mayo de 2005.

Dr. Jaime Gómez Márquez

Dr. Francisco Boán Fernández



*A mis padres*



## *Agradecimientos*

Resulta curioso que la escritura de este epígrafe, tanto por ubicación como por importancia en una Tesis Doctoral, se reserve casi por convención justo para el final. En estas líneas me gustaría mostrar mi reconocimiento a aquellas personas que de una u otra forma han contribuido al desarrollo de esta Tesis y a aquellas que de una u otra forma la han sufrido.

Como no, el primero de estos agradecimientos recae en mis Directores, el Dr. Jaime Gómez Márquez y el Dr. Francisco Boán Fernández, quienes después de aceptarme en su grupo de investigación me han inculcado tanto la pasión como los conocimientos necesarios y capacidad de sacrificio para poder llegar hasta este punto, prestándome siempre su consejo, apoyo y, sobretodo, amistad.

También quiero mencionar a todas aquellas personas del Departamento que han puesto su granito de arena tanto para la consecución de este trabajo como para hacer el día a día más llevadero en nuestro microcosmos. Paula, Jorge, Fran, Juan, Ana, Amparo, Macamen, Luís, Mónica, Susana, Mercedes, Manolo, Alicia, Begoña... y perdón a los que se me quedan en el tintero. Los aludidos saben que cuentan con unas cañas y lo que haga falta... También quisiera reseñar el gran saber y mejor trato que me brindó el Dr. José González Castaño y su equipo durante mi estancia en su laboratorio en el Instituto de Investigaciones Biomédicas "Alberto Sols" de la UAM-CSIC.

Debo además agradecer la gentileza de todos aquellos que han aportado diferentes materiales para la realización de esta tesis: al Dr. J. L. B. Caeiro por sus muestras de DNA genómico de diversas poblaciones humanas africanas; a los Drs. J. Bertranpetit, S. Pääbo, A. Blancher, J. Bullerdiel, P. Rogalla, W. Schempp y R. Wimmer por cedernos muestras de DNA genómico de diversos primates; a la Dra. M. Delclaux y a sus colaboradores del Zoo de Madrid por suministrarnos muestras de sangre de diversos primates; al Dr. F. Domínguez por la protimosina  $\alpha$  recombinante y al Dr. Gregor Meiss por la endonucleasa G recombinante.

Por último y mención aparte, gracias a todos los que afortunadamente viven ajenos a este mundillo de la Biología Molecular, mi familia y amigos, por ser estímulo, aliciente y válvula de escape tanto a las duras como a las maduras. Carmen, Álex, David, Anxo, Lemon, Toño, Rubio, Montes, Martín, Moncho, Borja, Aníbal, Manuela, Tino, Ana y el GC en general, gracias por aguantarme cuando estoy y por saber disculparme cuando no puedo estar. A todos un abrazo.

Durante el desarrollo de esta tesis he disfrutado de una beca para realización de estudios de tercer ciclo de la Xunta de Galicia (Octubre 2001-Diciembre 2002) y de una beca del Programa de Formación de Profesorado Universitario del Ministerio de Educación, Cultura y Deporte (Enero 2003- presente). Esta tesis ha sido financiada por el Ministerio de Educación, Cultura y Deporte (BMC2001-3242).

## *Resumen*

Msh42 es una secuencia de tipo minisatélite del genoma humano localizada en la banda cromosómica 15q25.1. Este minisatélite presenta un bajo polimorfismo y ha sido previamente relacionado con procesos recombinatorios *in vitro*. En esta Tesis Doctoral nos hemos propuesto conocer el origen e historia evolutiva de este locus, así como la caracterización de posibles puntos de corte de doble cadena (DSB) que pudiesen estar promoviendo los procesos recombinatorios descritos. Hemos descubierto que Msh42 es un minisatélite específico de primates que se originó durante las primeras etapas de la evolución de este grupo, antes de la divergencia entre monos del Nuevo y del Viejo Mundo, hace aproximadamente 40 millones de años. El polimorfismo existente en las poblaciones humanas actuales podría haber surgido a partir de una forma alélica ancestral común a gorilas, chimpancés y seres humanos. Por otro lado, hemos mostrado que la región Msh42 (minisatélite y secuencias flanqueantes) es degradada por endonucleasas específicas en ensayos de corte (ensayos DSB). En estos experimentos hemos detectado una actividad dependiente de  $Mg^{2+}$  que produce DSBs en tramos de guaninas, tratándose muy probablemente de la endonucleasa G. También hemos observado actividades independientes de  $Mg^{2+}$  que cortan la región Msh42, aunque no podemos afirmar de qué enzima(s) se trata. Durante estos ensayos DSB, hemos puesto de manifiesto que la relación DNA/proteína es la principal responsable de las diferencias entre los ensayos de unión de extremos de DNA (DNA *end-joining*) y los ensayos DSB, es decir, del predominio de los eventos de ligación sobre los de degradación.

## *Summary*

Msh42 is a minisatellite sequence of the human genome located in the chromosomal band 15q25.1. This minisatellite displays a low polymorphism and it has been previously shown its ability to promote recombination events *in vitro*. The aim of this Doctoral Thesis was to study the birth and the evolutionary history of this locus as well as to carry out the characterization of double strand break (DSB) sites that could be involved in the described recombinational processes. We have discovered that Msh42 is a minisatellite originated during the first stages of the evolution of primates, before the divergence between New and Old World monkeys, about 40 millions of years ago. The existing polymorphism in the present human populations could have arisen from an ancestral allelic form common to gorillas, chimpanzees and humans. On the other hand, we have shown that the Msh42 region (minisatellite plus flanking sequences) is degraded by specific endonucleases in DNA cleavage assays (DSB assays). In these experiments, we have detected a  $Mg^{2+}$ -dependent activity that generates DSBs in guanine stretches, which is very likely to be the endonuclease G. We have also observed  $Mg^{2+}$ -independent activities that are also able to cleave the Msh42 region, although we do not know which enzyme(s) is involved in this degradative pathway. Furthermore, during the DSB assays we have shown that the DNA/protein ratio is the main responsible for the differences between the DNA *end-joining* assays and the DSB assays, that is the prevalence of ligation over degradation events.

## *Abreviaturas*

|                                |  |
|--------------------------------|--|
| <b>CEPH</b>                    | Centro para el Estudio del Polimorfismo Humano                 |
| <b>DTT</b>                     | Ditiotreitol   |
| <b>DSB</b>                     | Corte de doble cadena  |
| <b>H1</b>                      | Histona H1   |
| <b>H3</b>                      | Histona H3   |
| <b>HR</b>                      | Recombinación homóloga   |
| <b>IPTG</b>                    | Isopropil- $\beta$ -D-tiogalactósido                           |
| <b>Msh42</b>                   | Minisatélite Msh42   |
| <b>NHEJ</b>                    | Unión de extremos de DNA no homólogos                          |
| <b>PMSF</b>                    | Fluoruro de fenilmetil sulfonilo                               |
| <b>ProT<math>\alpha</math></b> | Protimosina $\alpha$   |
| <b>Región Msh42</b>            | Minisatélite Msh42 con sus secuencias flanqueantes<br>(863 pb) |
| <b>SSA</b>                     | Hibridación de cadena sencilla                                 |
| <b>X-Gal</b>                   | 5-bromo-4-cloro-3-indolil- $\beta$ -D-galactopiranosido        |



# *Índice*

|  |                  |
|--|------------------|
| <b><i>Introducción</i></b>                               | <b><i>15</i></b> |
| El DNA repetitivo  | 17               |
| El DNA minisatélite                                      | 20               |
| El minisatélite Msh42                                    | 25               |
| <b><i>Objetivos</i></b>                                  | <b><i>37</i></b> |
| <b><i>Procedimientos Experimentales</i></b>              | <b><i>41</i></b> |
| Metodología general de clonación                         | 43               |
| Obtención del DNA genómico                               | 43               |
| Preparación de los sustratos de DNA para los ensayos DSB | 43               |
| Preparación de los extractos proteicos                   | 45               |
| Ensayos de corte de doble cadena (DSB) del DNA           | 49               |
| Ensayos DSB/end-joining                                  | 50               |
| Clonaje de productos del ensayo DSB                      | 51               |
| Amplificaciones por PCR                                  | 51               |
| Secuenciación de DNA                                     | 53               |
| Ensayos de DNaseI footprinting                           | 53               |
| Análisis filogenéticos                                   | 55               |

|   |            |
|---|------------|
| <b><i>Resultados y Discusión</i></b>  | <b>57</b>  |
| Localización del minisatélite MsH42 en el genoma humano                             | 59         |
| Nacimiento e historia evolutiva de MsH42  | 63         |
| <i>Origen evolutivo de MsH42</i>  | 64         |
| <i>Estructura de MsH42 en primates</i>  | 67         |
| <i>Interacción de las HCS con proteínas</i>   | 70         |
| <i>Análisis filogenético de los datos de secuencia</i>                              | 72         |
| <i>Historia evolutiva del minisatélite MsH42</i>                                    | 74         |
| Generación y localización de cortes de doble cadena en la región MsH42              | 77         |
| <i>El ensayo DSB in vitro</i>   | 79         |
| <i>Influencia del Mg<sup>2+</sup> en la actividad endonucleásica del ensayo DSB</i> | 82         |
| <i>Caracterización de las actividades endonucleásicas en el ensayo DSB</i>          | 84         |
| <i>Análisis de los productos DSB</i>  | 88         |
| Actividad Mg <sup>2+</sup> -dependiente   | 88         |
| Actividad Mg <sup>2+</sup> -independiente   | 94         |
| <i>Análisis de los productos recombinantes generados en los ensayos DSB</i>         | 97         |
| Paradojas en los ensayos DSB  | 101        |
| <i>Generación de DSBs vs. Ligación de extremos de DNA (end-joining)</i>             | 101        |
| <i>Efecto de la adición de histona H1 al ensayo DSB</i>                             | 104        |
| <i>Diferencias en la relación DNA/proteína en los ensayos DSB</i>                   | 109        |
| <i>Otros modos de favorecer la ligación frente a la degradación</i>                 | 111        |
| <b><i>Conclusiones</i></b>  | <b>115</b> |
| <b><i>Bibliografía</i></b>  | <b>119</b> |
| <b><i>Apéndice</i></b>  | <b>133</b> |

# *Introducción*



Aproximadamente tres mil cien millones es el número de pares de bases que constituyen el manual de instrucciones a partir del cual se desarrolla un ser humano. Durante años se pensó que esta abrumadora cifra de nucleótidos escondía un número de genes codificadores de proteínas relativamente grande, en torno a 100000. Sin embargo, ya con las estimas derivadas de los primeros borradores del genoma humano (International Human Genome Sequencing Consortium, 2001; Venter et al., 2001) esta cifra disminuyó a unos 30000 genes, cantidad que actualmente se ha visto reducida a tan sólo 20000-25000 tras la publicación del último borrador del genoma humano, que cubre más del 99% de nuestra secuencia eucromatínica (International Human Genome Sequencing Consortium, 2004). Hay que indicar que en esta cifra no se incluyen numerosos genes conocidos que no codifican proteínas, como los de los RNA transferentes, RNA ribosómicos, RNA pequeños nucleolares, microRNAs, etc. (Storz, 2002; Bartel, 2004). Sin embargo, un dato que apenas se ha modificado desde la primera versión de la secuencia del genoma humano es que más de un 50% de nuestro DNA es de carácter repetitivo. ¿Qué tipo de secuencias constituyen este DNA repetitivo? ¿Por qué existe esta enorme cantidad de material genético aparentemente inerte? En esta Tesis Doctoral abordaremos el análisis de algunas facetas de una pequeña parte del genoma humano que contiene secuencias repetitivas y que denominamos MSH42.

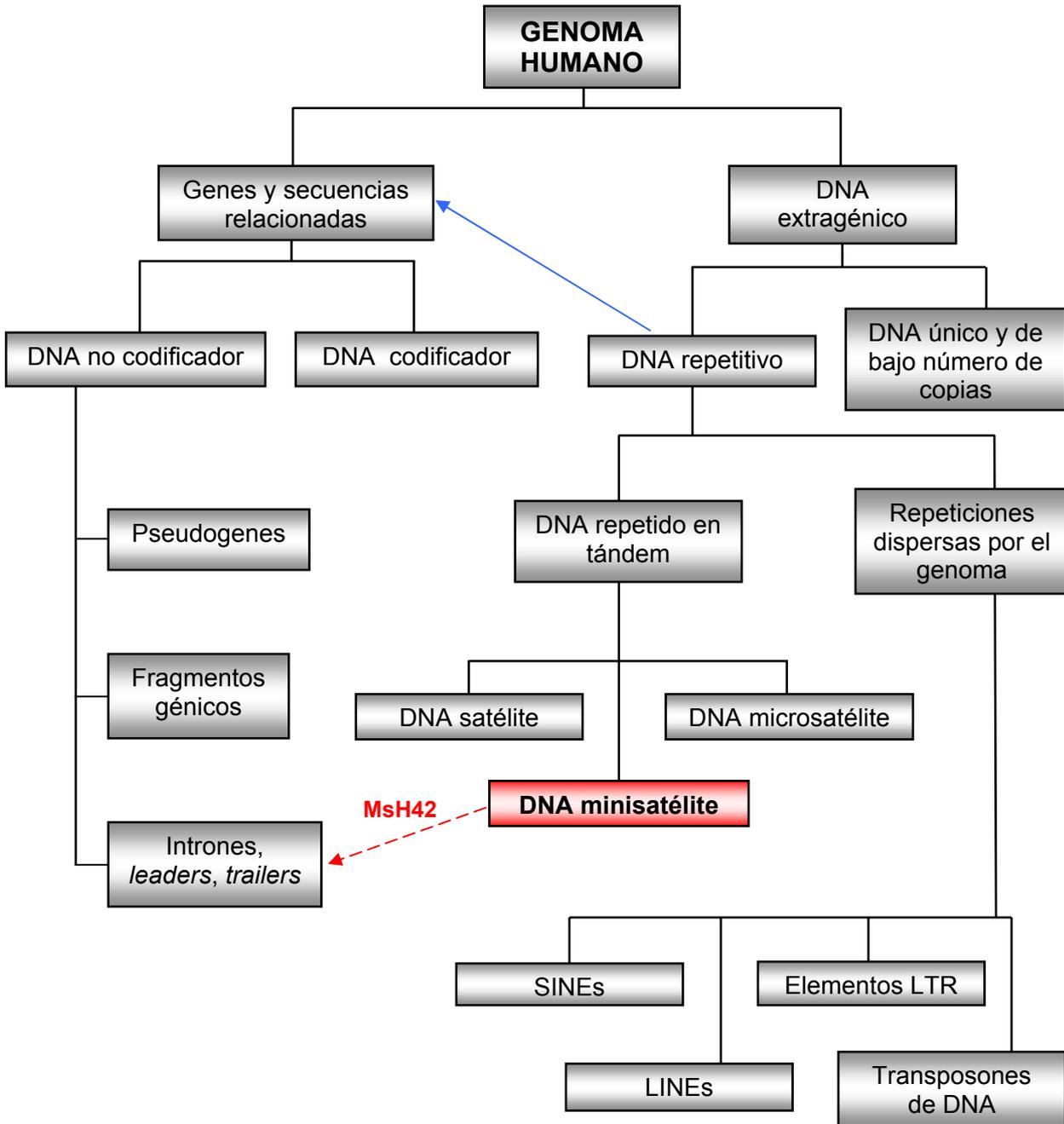
### ***El DNA repetitivo***

La existencia de secuencias repetitivas en el genoma fue inicialmente propuesta a partir de los experimentos de cinética de reasociación de cadenas de DNA (Britten & Kohne, 1968). En los

siguientes años se profundizó en el conocimiento de la estructura de este DNA repetitivo y su localización concreta, manteniéndose la controversia de su posible función e influencia en la dinámica de los genomas. El DNA repetitivo fue considerado por muchos investigadores como una especie de DNA basura, un lastre genético o DNA egoísta que permanecía como remanente de genomas virales u otro tipo de secuencias que se comportaban de forma parasita, sin ello suponer una desventaja fenotípica severa (Doolittle & Sapienza, 1980; Orgel & Crick, 1980). Sin embargo, investigaciones con diversas clases de DNA repetitivo realizadas posteriormente, han permitido llegar a la conclusión de que muchas de estas secuencias, lejos de ser una carga en el genoma, están implicadas en importantes procesos biológicos como la transcripción, la estabilidad de los cromosomas o la impronta genética (Vergnaud & Denoeud, 2000).

Desde un punto de vista funcional, las secuencias presentes en el genoma humano se pueden subdividir en dos grandes bloques: las secuencias génicas y las extragénicas. Dentro de este último grupo, se incluye la mayor parte del DNA repetitivo, aunque hay que señalar que esta clase de DNA también puede encontrarse en las regiones génicas. En la Figura 1 se muestra un esquema general de los distintos tipos de secuencias que componen el genoma (Brown, 1999).

Este tipo de DNA puede clasificarse en dos categorías según la disposición de las unidades repetitivas: repeticiones en tándem o dispersas a lo largo del genoma. Dentro del DNA repetitivo disperso podemos realizar una segunda subdivisión de las secuencias, basada principalmente en su mecanismo de transposición:



**Figura 1.** Clasificación de los distintos tipos de secuencias que componen el genoma humano (Adaptado de Brown, 1999). Se resalta la posición de los minisatélites en esta clasificación. La flecha azul indica la presencia de DNA repetitivo en algunas regiones génicas. De hecho, el minisatélite que nos ocupará en este trabajo, MsH42, se encuentra situado en el intrón 5 del gen Q9ULM1 (flecha roja).

**1) Retroelementos o transposones de RNA:** caracterizados por la existencia de un RNA intermediario que se retrotranscribe previamente a la reintegración en el genoma.

**1.1) Elementos LTR (long terminal repeats):** se caracterizan por la presencia de largas repeticiones terminales en sus extremos. Aquí se incluyen:

**1.1.1) *Retrovirus*.** Virus de RNA en fase lisogénica con plena capacidad de generar nuevas partículas infectivas.

**1.1.2) *Retrovirus endógenos (ERVs) y elementos tipo retrovirus (RTVLs)*.** Similares en estructura al grupo anterior, se diferencian de ellos en que la mayoría han perdido su capacidad para formar nuevos virus.

**1.1.3) *Retrotransposones*.** De estructura similar a los ERVs, son especialmente abundantes en invertebrados y plantas. En determinados vegetales incluso llegan a formar agrupaciones en tándem y pueden representar hasta el 50% del genoma.

**1.2) *Retroposones*:** se diferencian básicamente del grupo anterior en la ausencia de repeticiones terminales largas en sus extremos. Hay dos grandes clases:

**1.2.1) *LINES* (long interspersed nuclear elements).** Los elementos nucleares dispersos largos poseen un gen tipo transcriptasa inversa probablemente asociado con su transposición.

**1.2.2) *SINES* (short interspersed nuclear elements).** Los elementos nucleares dispersos cortos carecen de una transcriptasa inversa propia y utilizan aquella que es producida por otros retroelementos para su retrotranscripción. El ejemplo más destacable es la familia *Alu*, derivada del RNA 7SL y con un número de copias por genoma haploide humano próximo al millón.

**2) *Transposones de DNA*:** Estos elementos, mucho menos abundantes que los transposones de RNA en los eucariotas, utilizan una transposasa para realizar una interacción directa DNA-DNA entre el transposón donante y el

lugar aceptor, bien de forma conservativa o replicativa.

Por otro lado tenemos las secuencias formadas por unidades repetitivas dispuestas en tándem, poco abundantes en procariotas, pero muy comunes en eucariotas. Atendiendo al tamaño de la unidad que se repite, pueden distinguirse tres tipos principales entre estas secuencias:

**a) *DNA satélite*.** Esta clase de DNA se caracteriza por la repetición de bloques de más de 100 pb, llegando a comprender regiones de hasta 100 Mb. Estas secuencias reciben su nombre de los primeros experimentos donde el DNA genómico fragmentado se sometía a centrifugación en gradientes de densidad. El DNA satélite, dada su peculiar composición en G+C comparado con el resto del genoma y, por tanto, su diferente densidad de flotación, aparecía separado de la banda principal formando un grupo de bandas "satélite". Uno de los casos más estudiados de este tipo de DNA, el DNA satélite alfoide, forma parte de los centrómeros de los cromosomas humanos.

**b) *DNA minisatélite*.** Por analogía con el DNA satélite, se denominan minisatélites a las secuencias de DNA constituidas por unidades de repetición entre 6 y 100 pb, aunque estos márgenes son relativamente arbitrarios y pueden variar en función de cada autor. Los minisatélites habitualmente comprenden desde ~ 0.5 hasta varios miles de kb.

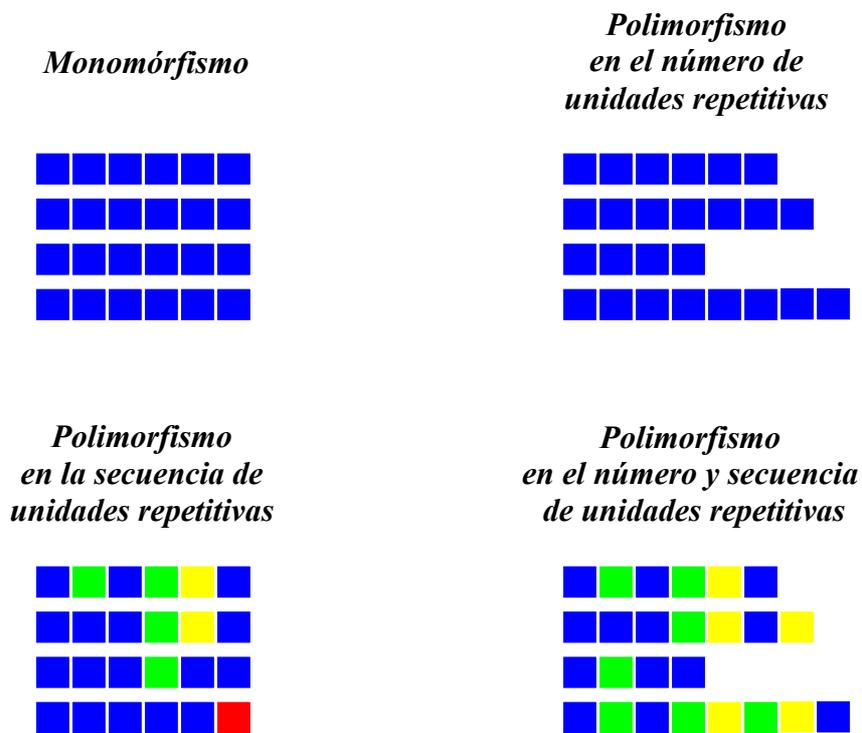
**c) *DNA microsatélite*.** De forma análoga al DNA minisatélite, los microsatélites se definen como las secuencias DNA compuestas por la repetición en tándem de bloques de 1-5 pb, abarcando normalmente entre 100-200 pb.

### El DNA minisatélite

Una vez presentados los distintos tipos de DNA repetitivo, profundizaremos algo más en las características del DNA minisatélite, en el que se incluye el locus MsH42, hilo conductor de este trabajo.

El término minisatélite fue empleado por primera vez en 1985 para referirse a secuencias formadas por la repetición de bloques cortos de DNA, ricos en G+C, muy polimórficos y con una secuencia consenso similar al motivo recombinogénico *Chi* de procariontes (Jeffreys et al., 1985a). Posteriormente, este término se universalizó abarcando un conjunto mucho más amplio de secuencias de DNA repetidas en tándem que pueden ser tanto muy polimórficas (revisado en Vogt, 1990; Bois & Jeffreys, 1999; Vergnaud & Denoeud,

2000), como poco polimórficas (Boán et al., 1997) o incluso monomórficas (Ali & Wallace, 1988), además de minisatélites ricos en A+T (Berg & Olaisen, 1993; Ellsworth et al., 1995; Yu et al., 1997; Jobling et al., 1998). Desde su descubrimiento (Wyman & White, 1980), el alto grado de polimorfismo de muchos minisatélites, también llamados VNTRs -*Variable Number Tandem Repeats*- (Nakamura et al., 1987) debido a las diferencias en el número de unidades repetitivas que los componen (Bell et al., 1982), los convirtió en valiosos marcadores moleculares. En este sentido, las secuencias minisatélite han sido empleadas en el desarrollo de las técnicas de DNA *fingerprinting* o huella genética para estudios de identificación personal (Jeffreys et al., 1985b), así como también en la construcción de mapas genéticos de ligamiento (O'Connell et al., 1987). En



**Figura 2.** Diferentes tipos de polimorfismo presentes en las secuencias minisatélite. Cada bloque representa una unidad repetitiva. Los colores diferentes indican variantes de la unidad repetitiva. Según esta clasificación, MsH42 pertenecería al grupo de minisatélites con polimorfismo en el número y secuencia de sus unidades repetitivas.

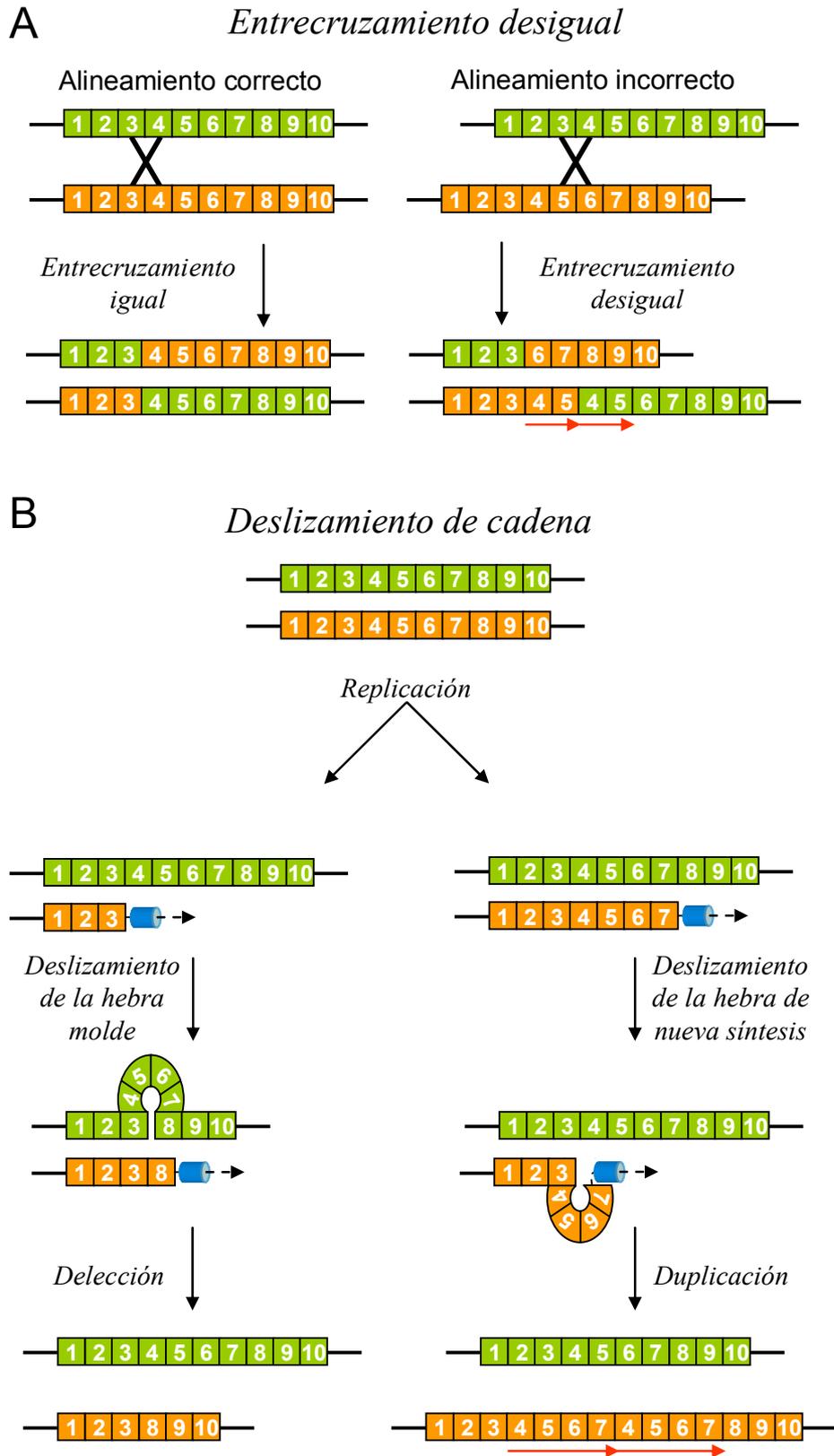
algunos casos, el alto nivel de polimorfismo que presentan estas secuencias se encuentra asociado a una elevada tasa de mutación, que puede llegar hasta un 20% por locus y gameto, constituyendo así unas de las regiones más hipervariables del genoma (Jeffreys et al., 1988; revisado en Bois, 2003). La variabilidad de los minisatélites puede venir dada no sólo por cambios en el número de sus unidades repetitivas, sino también por variaciones en la secuencia de las mismas y en el orden en que se disponen (Jeffreys et al., 1990). A estas variantes se las denominó MVRs (*Minisatellite Variant Repeats*) y para su discriminación se han desarrollado técnicas como la MVR-PCR (Jeffreys et al., 1991). En la Figura 2 se muestran de forma esquemática los diferentes tipos de polimorfismo que se pueden encontrar entre las secuencias tipo minisatélite.

Se ha postulado que los principales procesos que llevan a la generación del polimorfismo que presentan muchos minisatélites son el entrecruzamiento desigual y el deslizamiento de cadena durante la replicación o *slippage*. El entrecruzamiento desigual es un mecanismo propuesto, en principio, para explicar la evolución de cualquier tipo de secuencia de DNA repetida en tándem (Smith, 1976). Consiste en un alineamiento imperfecto de la región repetitiva entre las cromátidas hermanas en mitosis o entre cromosomas homólogos en meiosis. Este tipo de alineamiento incorrecto puede provocar un entrecruzamiento entre repeticiones no equivalentes, lo que concluye con una duplicación y el consiguiente alargamiento del minisatélite o con una delección y, por tanto, con el acortamiento del mismo (Figura 3A). Por otro lado, el *slippage* implica la aparición de emparejamientos imperfectos entre las dos hebras del

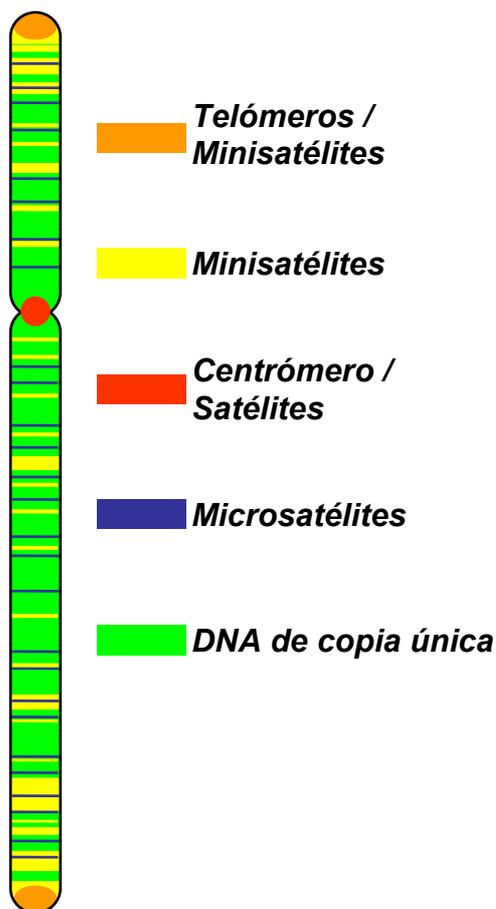
dúplex de DNA que se está replicando y, de este modo, la generación de bucles de cadena sencilla. Estos podrían dar lugar a delecciones o duplicaciones de la secuencia dependiendo de si el deslizamiento se ha producido en la hebra que sirve como molde o en la de nueva síntesis, respectivamente (Figura 3B). Este es posiblemente uno de los procesos más importantes que determinan la generación y evolución del DNA repetitivo en tándem (revisado en Levinson & Gutman, 1987). Otros mecanismos como la conversión génica también pueden contribuir a la generación de polimorfismo en los minisatélites (Jeffreys et al., 1994).

Existen secuencias tipo minisatélite en una gran variedad de organismos, incluyendo levaduras (Andersen & Nillson-Tillgren, 1997), hongos (Giraud et al., 1998), plantas (Dallas, 1988) y animales (Jeffreys et al., 1985, 1987; Burke & Bruford, 1987; Jeffreys & Morton, 1987; Buitkamp et al., 1991; Canestro et al., 2002). Su ubicación cromosómica es dispersa, hallándose distribuidos en toda su longitud (Figura 4), aunque en el genoma humano se ha constatado una concentración de aproximadamente el 90% en las regiones subteloméricas, valor que desciende al 66% en el cerdo, 30% en la rata y 15% en el ratón (Royle et al., 1988; Amarger et al., 1998; Bois et al., 1998).

¿Cuáles son las funciones que pueden desempeñar en las células las secuencias minisatélite? Se ha demostrado que algunos minisatélites son reconocidos específicamente por proteínas celulares (Collick & Jeffreys, 1990; Wahls et al., 1991; Wahls & Smith, 1994; Fukuda et al., 2001; Enokizono et al., 2003), apoyando la implicación de este tipo de secuencias en determinadas funciones celulares. Veamos algunos ejemplos que ilustran la implicación del DNA



**Figura 3.** Mecanismos generadores de polimorfismo en secuencias repetidas en tándem. **A.** Modelo de entrecruzamiento desigual. Las X señalan los lugares de entrecruzamiento. **B.** Modelo de deslizamiento de cadena durante la replicación. Los cilindros y las flechas discontinuas representan la polimerasa y el sentido de la síntesis de DNA. Los bloques que representan las unidades repetitivas se hallan numerados para facilitar el seguimiento de los procesos. Las flechas rojas indican grupos de repeticiones duplicadas.



**Figura 4.** Esquema de la distribución de los distintos tipos de DNA repetitivo en tándem a lo largo de un cromosoma humano. La cantidad relativa de cada color no refleja la abundancia real de cada tipo de secuencia.

minisatélite en algunos procesos biológicos. Se sabe que en la región 5' del gen de la insulina existe un minisatélite (Bell et al., 1982) cuyo polimorfismo parece estar relacionado con la diabetes mellitus insulino-dependiente, puesto que algunas variantes alélicas de este minisatélite influyen en el descenso de la transcripción de este gen en el timo (Bennet et al., 1995; Kennedy et al., 1995). Este descenso del nivel de producción de insulina en el timo se traduciría en una presentación antigénica insuficiente de esta proteína por parte de

las células dendríticas, con la consiguiente supervivencia de los linfocitos T autoinmunes capaces de reconocer a la insulina al no llevarse a cabo el proceso de selección negativa de los mismos. Tales linfocitos T autoinmunes son los causantes del desarrollo de la patología al mediar en la destrucción de las células  $\beta$  del páncreas (Pugliese et al., 1997; Vafiadis et al., 1997; Anjos & Polychronakos, 2004). Otros genes cuya expresión puede estar influida por la presencia de minisatélites en su entorno son el proto-oncogen HRAS1 (Phelan et al., 1996), los genes de las cadenas pesadas de las inmunoglobulinas (Trepicchio & Krontiris, 1993) o el gen de la cistatina B, que se considera involucrado en la aparición de epilepsia mioclónica progresiva tipo 1 (Laloti et al., 1997). Algunos minisatélites se encuentran en regiones codificantes y su polimorfismo es capaz de generar variantes alélicas de una proteína. En este sentido, expansiones o contracciones en el número de las unidades repetitivas provocan la aparición o desaparición de un determinado número de aminoácidos en la cadena polipeptídica y, por tanto, la aparición de distintas alozimas con propiedades bioquímicas diferentes. Ejemplos de este tipo son los genes de las apolipoproteínas (Boguski et al., 1986) o el receptor D4DR de la dopamina (Van Tol et al., 1992; Lichter et al., 1993), donde el polimorfismo de los minisatélites asociados a estos genes se ha relacionado con la generación de diferentes variantes alélicas que muestran cambios en su afinidad por la unión a lípidos y a la dopamina, respectivamente. En este último caso, se ha establecido una correlación entre dichas variaciones en la interacción receptor-ligando y determinados trastornos cognitivos y emocionales (Ebstein et al., 1996; Benjamin et al., 1996). Otros minisatélites han sido

localizados en intrones, estando algunos de ellos relacionados con procesos que pueden afectar al *splicing* (Turri et al., 1995). También se ha descrito que algunos minisatélites pueden estar implicados en el control de la impronta genética de determinados loci (Chaillet et al., 1995; Neumann et al., 1995). Asimismo, la expansión de determinados minisatélites, especialmente los ricos en A+T, ha sido asociada con la aparición de puntos frágiles infrecuentes en los cromosomas (Handt et al., 2000) y, de hecho, se conoce la existencia de minisatélites cerca de lugares donde se dan rupturas que generan translocaciones cromosómicas de forma recurrente, así como en los lugares donde ocurre la recombinación de cambio de clase isotípica en los genes de la cadena pesada de las inmunoglobulinas (Brusco et al., 1999). Finalmente, se ha demostrado que determinados minisatélites tanto hipervariables como

no hipervariables pueden potenciar la recombinación homóloga *in vitro* (Wahls et al., 1990; Boán et al., 1998; Wahls & Moore, 1998), lo que sugiere que estas secuencias puedan estar funcionando como *hotspots* de recombinación *in vivo*, es decir, lugares donde la tasa de recombinación es mayor que la media del resto del genoma (Wahls, 1998). La Figura 5 sintetiza este conjunto de funciones propuestas para los minisatélites.

La asociación entre determinadas secuencias minisatélite y la recombinación es muy estrecha, existiendo trabajos que, en este sentido, relacionan la generación del polimorfismo de minisatélites y su inestabilidad genética con procesos recombinatorios. Algunas de estas investigaciones se han basado en el análisis de nuevas variantes alélicas generadas en gametos individuales, lo

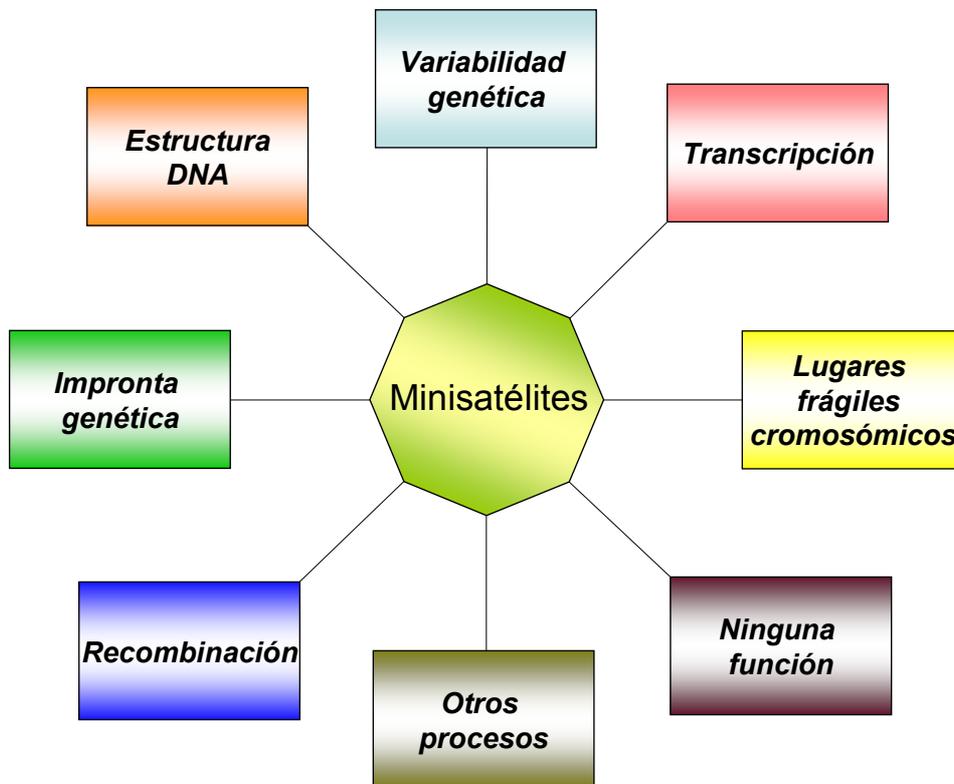


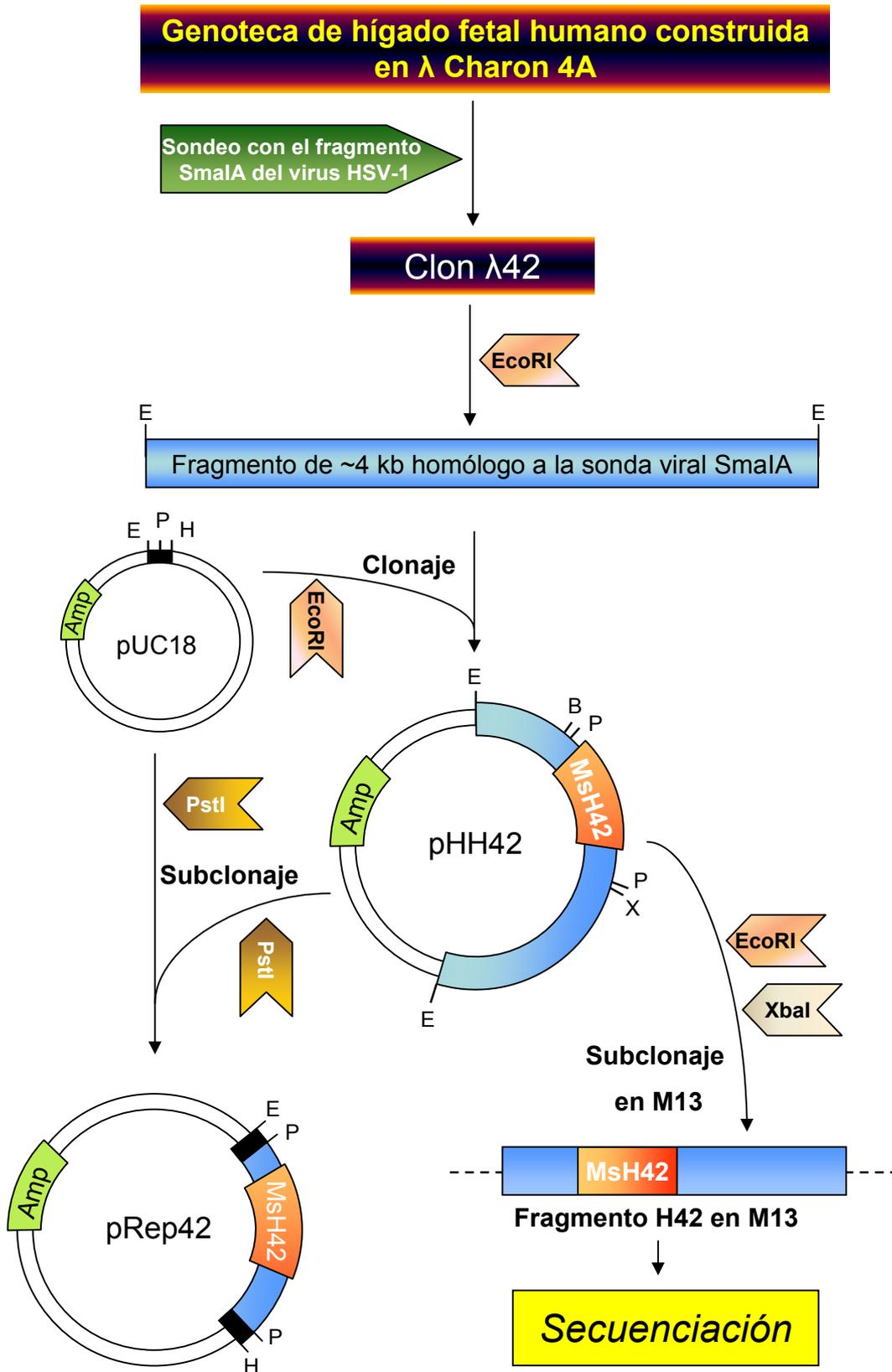
Figura 5. Representación de los principales procesos biológicos en los que pueden estar implicados los minisatélites.

cual ha sido posible gracias al desarrollo de técnica de la SP-PCR (*small-pool PCR*) (Jeffreys et al., 1994). La combinación de MVR-PCR y SP-PCR, así como la introducción de minisatélites humanos en los genomas del ratón y de la levadura, ha permitido la investigación pormenorizada de los procesos de inestabilidad tanto somática como germinal en algunos minisatélites hipervariables (revisado en Jeffreys et al., 1999). Algunas de las conclusiones más relevantes de estos estudios son: i) en determinados loci minisatélite, la tasa de mutación en línea germinal entre los distintos sexos es diferente, siendo generalmente más elevada la masculina (Buard & Vergnaud, 1994; May et al., 1994; Jeffreys et al., 1998); ii) la existencia de una polaridad dentro de algunos minisatélites para la aparición de nuevas mutaciones, tanto en su zona 5' (Jeffreys et al., 1994, 1998) como 3' (Armour et al., 1993); iii) el tamaño y el grado de homogeneidad de los alelos de un minisatélite es un factor determinante de su grado de inestabilidad, principalmente de aquella debida a reorganizaciones intralélicas (Buard et al., 1998); iv) la tasa de generación de mutantes de tamaño aumenta con el grado de divergencia entre los alelos parentales (He et al., 2002; Jauert & Kirkpatrick, 2005); v) existen elementos cercanos a los minisatélites que pueden regular en *cis* su nivel de inestabilidad (Monckton et al., 1994; Andreassen et al., 1996; He et al., 1999; Debrauwère et al., 1999). Además, en algunos de estos estudios también se han descrito alelos mutantes que parecen originarse mediante eventos recombinatorios más complejos que los representados en la Figura 3, incluyendo conversión génica con o sin duplicaciones adyacentes asociadas, conversión génica seguida de reorganizaciones internas o duplicaciones intralélicas múltiples. Para explicar estas reorganizaciones

complejas, se han propuesto diversos modelos que combinan deslizamiento de cadena en la replicación y recombinación, iniciados por cortes del DNA dentro del minisatélite (Bois & Jeffreys, 1999; Buard et al., 2000; Boán et al., 2002) o en su entorno (Debrauwère et al., 1999; He et al., 2002).

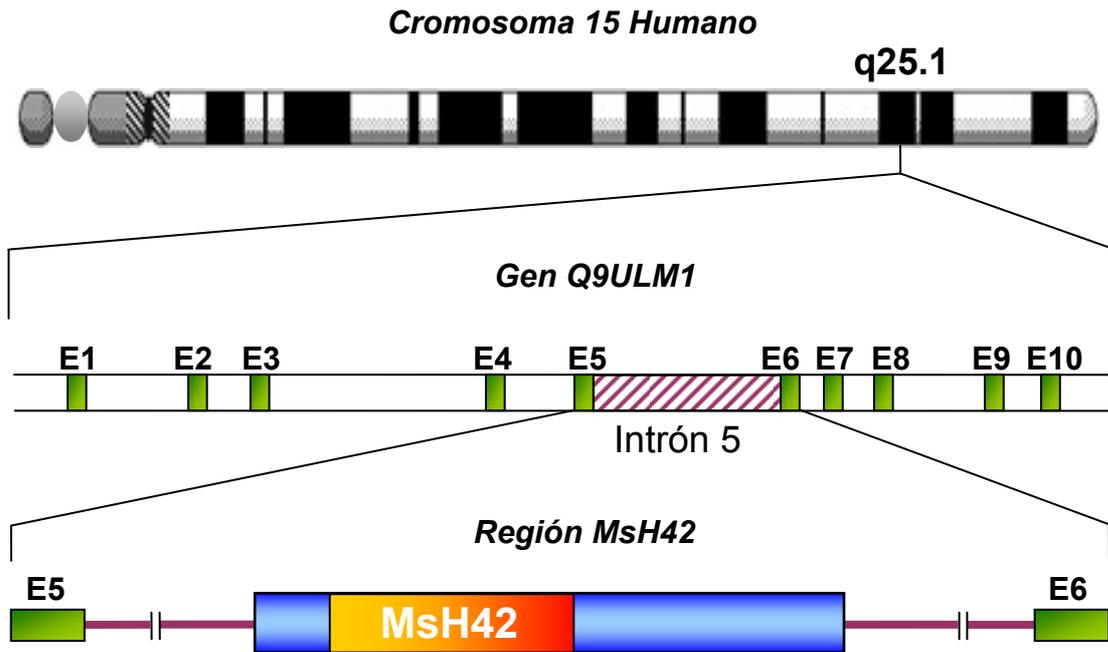
### ***El minisatélite humano MsH42***

El aislamiento y caracterización estructural del minisatélite MsH42 tuvo lugar en nuestro laboratorio a finales de los años 80. Los antecedentes fueron los experimentos realizados a principios de esa década donde se demostraba la existencia de secuencias homólogas entre determinadas regiones del genoma de mamíferos y de algunos miembros de la familia de los Herpesvirus (Puga et al., 1982; Peden et al., 1982; Heller et al., 1982). La hibridación de una genoteca de ratón con la sonda SmaIA obtenida del virus *Herpes simplex* tipo 1 (HSV-1), muy rica en G+C (75%) y que contiene parte de una de las repeticiones terminales del genoma viral, produjo una serie de clones positivos con secuencias que presentaban un alto contenido en G+C y estructura repetitiva (Gómez-Márquez et al., 1985). Posteriormente, ya en nuestro laboratorio, se empleó este mismo fragmento viral para sondear una genoteca de hígado fetal humano. El análisis de los insertos de tres de los fagos recombinantes positivos mostró que la homología con la sonda viral se debía a repeticiones en tándem ricas en G+C (González, 1993), al igual que ocurría con los clones murinos, mencionados anteriormente. De entre estos clones positivos, uno de ellos denominado  $\lambda$ 42, englobaba un fragmento *EcoRI* de  $\sim 4$  kb que estaba implicado en la hibridación detectada con la sonda SmaIA viral. Este fragmento fue subclonado en pUC18,



**Figura 6.** Esquema del aislamiento y caracterización inicial del minisatélite MsH42 a partir de una genoteca de hígado fetal humano (González, 1993). B, *Bgl*II; E, *Eco*RI; H, *Hind*III; P, *Pst*I; X, *Xba*I; Amp, gen de la  $\beta$ -lactamasa (resistencia a ampicilina).





**Figura 8.** Esquema general de la ubicación del minisatélite Msh42 en el intrón 5 del gen Q9ULM1 en el cromosoma 15 humano. El recuadro azul que incluye al minisatélite Msh42 (naranja) representa la denominada “región Msh42”. Los exones se representan como cajas verdes. Esta es la localización publicada en Boán et al. (2004), artículo derivado de esta tesis. Entre la publicación del artículo y la escritura de esta tesis, se han producido modificaciones de esta región en las bases de datos del genoma humano, como indicaremos en *Resultados*.

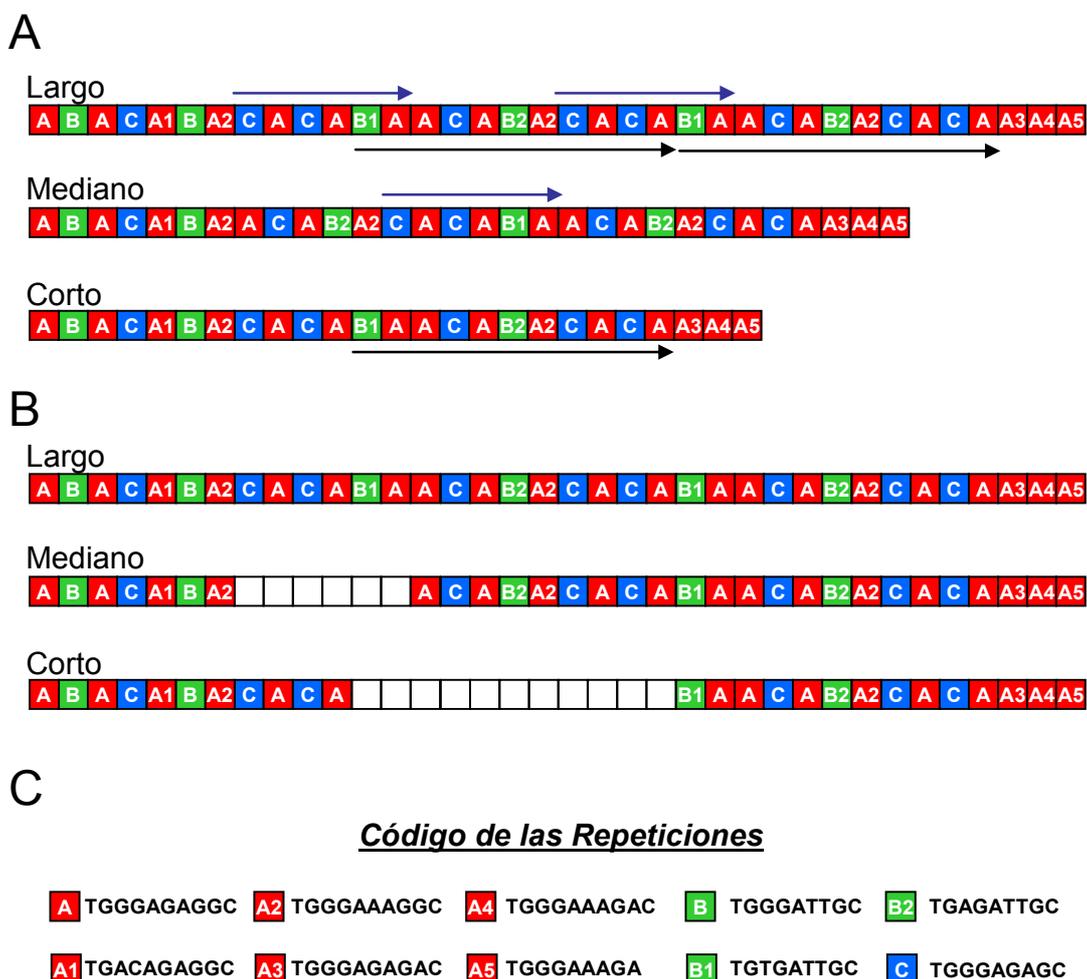
El contenido en G+C de Msh42 es del 63,2% y la asimetría entre purinas y pirimidinas en cada hebra es del 75%. En el entorno de este minisatélite existen distintos motivos relacionados con recombinación, como dos secuencias *Chi* de recombinación de procariontas (GCTGGTGG) (Smith et al., 1981), una secuencia similar a *Chi* localizada a 5' de Msh42 (GCTGGAGG) y dos oligómeros (TGTTTTTGG y CACAGTG) implicados en la recombinación V(D)J en los genes de las inmunoglobulinas (Sakano et al., 1981). La secuencia de Msh42 presenta una importante similitud (~ 65%) con distintas regiones involucradas en la recombinación de cambio de clase en las inmunoglobulinas, debido a la abundancia en Msh42 de los motivos GAGCT y TGGG, que son muy similares a los oligómeros (G)AGCT(G) y TGGG(G) repetidos en tándem en dichas regiones (revisado en Dunnick et

al., 1993). Cabe señalar que en las 1208 pb del fragmento *BglII-XbaI* de H42 sólo existen 5 dinucleótidos CpG, ~ 15 veces menos de lo esperado por azar. Este tipo de dinucleótido normalmente se halla infrarrepresentado en el genoma humano (International Human Genome Sequencing Consortium, 2001), aunque existen regiones de pocos cientos de pares de bases localizadas muy frecuentemente en las regiones 5' de los genes donde su frecuencia es mayor (islas CpG).

Msh42 se localizó en la banda 25.1 del brazo largo del cromosoma 15, dentro del intrón 5 del gen Q9ULM1 (Boán et al., 2002). En la Figura 8 se muestra la ubicación de la región Msh42 (minisatélite y regiones flanqueantes comprendidas entre los dos puntos *PstI* más próximos al mismo) en el genoma humano. El minisatélite Msh42 se caracteriza por su bajo grado de

polimorfismo en la población humana. Durante los primeros análisis de muestras de DNA genómico de población caucasoide digerido con *Bgl*II e hibridado con la sonda pRep42, se detectaron dos fragmentos de 1.3 y 1.2 kb (González, 1993). La posterior amplificación por PCR de Msh42 empleando los cebadores P1 y P2 (Boán et al, 1997) generó dos productos de 585 y 480 pb, cuya diferencia en tamaño se correspondía con la observada por Southern blot. A estos alelos, de 36 y 25 unidades repetitivas respectivamente y que diferían entre sí en el bloque de 11

repeticiones (105 pb) que se encuentra duplicado en el mayor de ellos, se les denominó alelo largo y alelo corto (Figura 9). Posteriormente, durante un sondeo de población caucasoide a mayor escala, se encontró un individuo portador de un alelo de Msh42 que daba lugar a un fragmento de amplificación de tamaño intermedio. Para descartar la posibilidad de que este nuevo alelo fuese consecuencia de un artefacto de la PCR, se realizó un estudio genealógico de su portador, confirmando la herencia del alelo por vía paterna (Boán et al., 2002). Este nuevo alelo, de 30 unidades



**Figura 9.** Comparación de los alelos conocidos de Msh42. **A.** Representación de los tres alelos de Msh42. Las flechas negras y azules marcan las duplicaciones de 11 y 6 repeticiones del alelo largo y que lo diferencia del alelo corto y mediano, respectivamente. **B.** Alineamiento de los tres alelos, mostrando las deleciones en el mediano y en el corto con respecto al largo. **C.** Código de las unidades repetitivas de Msh42.

repetitivas, se nombró alelo mediano y su diferencia con el largo radica en un bloque de seis repeticiones (8-13), que se encuentra duplicado en dicho alelo largo, aunque no de forma contigua. En la Figura 9 se muestra la organización de los tres alelos de MsH42, así como un código de repeticiones donde se visualizan claramente las diferencias entre estos tres alelos. Este código de colores/letras fue diseñado para facilitar la visualización de la estructura interna del minisatélite (Boán et al., 1997) y en él las unidades repetitivas se hallan agrupadas en tres conjuntos. El grupo "A" (rojo) está constituido por el decámero TGGGAGAGGC (A) y otras cuatro repeticiones de 10 pb derivadas de él (A1-A4), así como por un nonámero muy similar a A4, (A5). El grupo "B" lo constituyen los nonámeros de secuencia TGDGATTGC (B1-B3). El nonámero TGGGAGAGC es el único componente del grupo "C". Los tres alelos se heredan de forma mendeliana simple, como se comprobó mediante el estudio de diversas familias de referencia del CEPH, siendo sus frecuencias poblacionales de  $0.834 \pm 0.019$  (largo),  $0.005 \pm 0.0035$  (mediano) y  $0.161 \pm 0.018$  (corto), con una distribución genotípica en equilibrio Hardy-Weinberg ( $\chi^2 = 0.76$ ,  $P > 0.05$ ).

En nuestro laboratorio se ha demostrado mediante ensayos de *band-shifting* y *footprinting*, que la región MsH42 es específicamente reconocida por proteínas tanto de extractos de testículo, que contienen células germinales, como de tejidos somáticos -hígado y cerebro- (Boán et al., 1997; Rodríguez, 2002). Además, MsH42 es capaz de llevar a cabo procesos de *slippage* durante la PCR, lo que favorece la aparición de bucles de cadena sencilla (Boán et al., 1997). Este deslizamiento de cadena sencilla es más pronunciado cuando se emplea el alelo largo como DNA molde,

mientras que en las amplificaciones de los alelos corto o mediano apenas sí se observan bandas fruto de *slippage*. La secuenciación de los productos de *slippage* derivados del alelo largo durante la PCR demostró que uno de ellos es idéntico al alelo corto, sugiriendo que este alelo pudo generarse *in vitro* a partir del alelo largo y, por lo tanto, un posible mecanismo para su origen evolutivo. Estos datos, el reconocimiento de la región MsH42 por proteínas nucleares, la formación de lazos de cadena sencilla que podrían favorecer la interacción con el dúplex homólogo y la presencia de diversos motivos asociados con recombinación, sugería la posible implicación de esta región en procesos recombinatorios. Para comprobar *in vitro* esta hipótesis, se diseñaron una serie de constructos plasmídicos que portaban dos copias de la región MsH42 (863 pb), dos copias de una región no repetitiva 5'MsH42 (850 pb) que actuaría como secuencia control, dos copias del minisatélite sin secuencias flanqueantes (344 pb) o dos copias de la secuencia 5'MsH42.1 (356 pb), que también se utilizó como control. Diferentes combinaciones de estos plásmidos se incubaron con un extracto proteico nuclear obtenido a partir de diversos tejidos (Boán et al., 1998, 2002). La localización de las secuencias empleadas en la construcción de estos plásmidos, así como una representación de los mismos se muestra en la Figura 10.

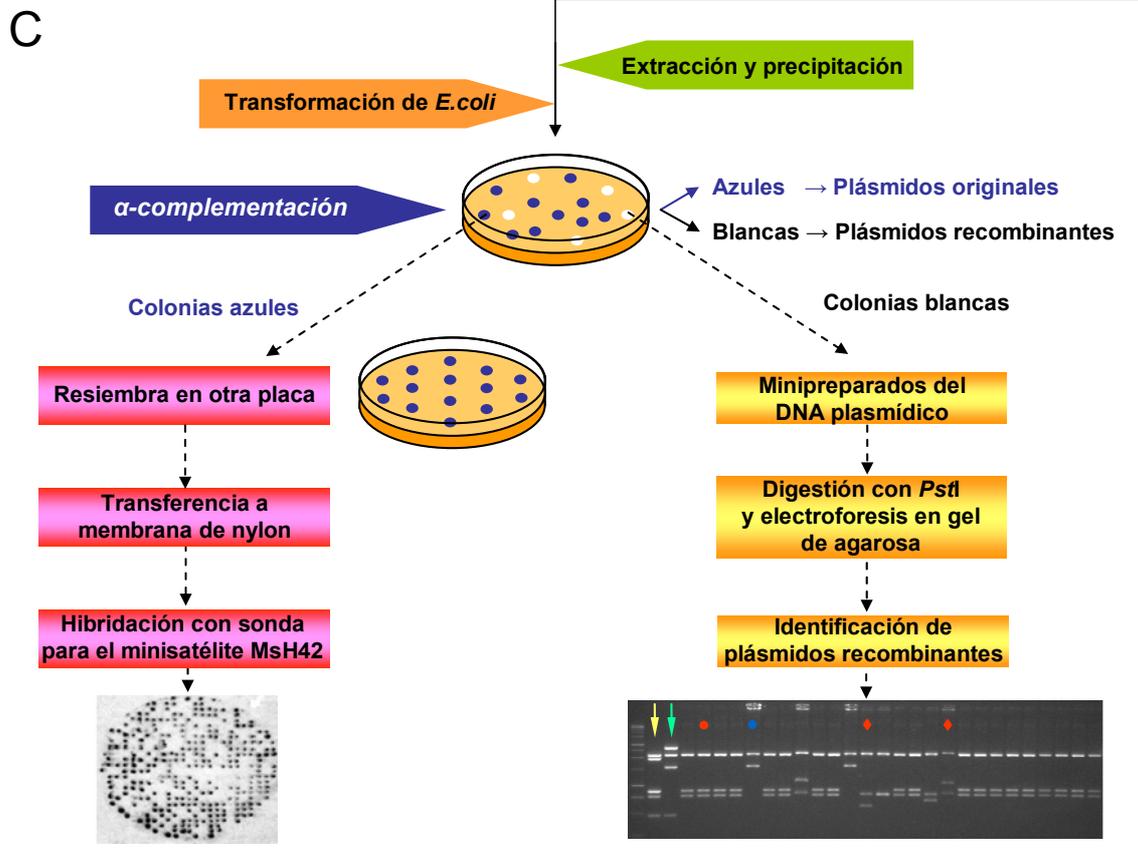
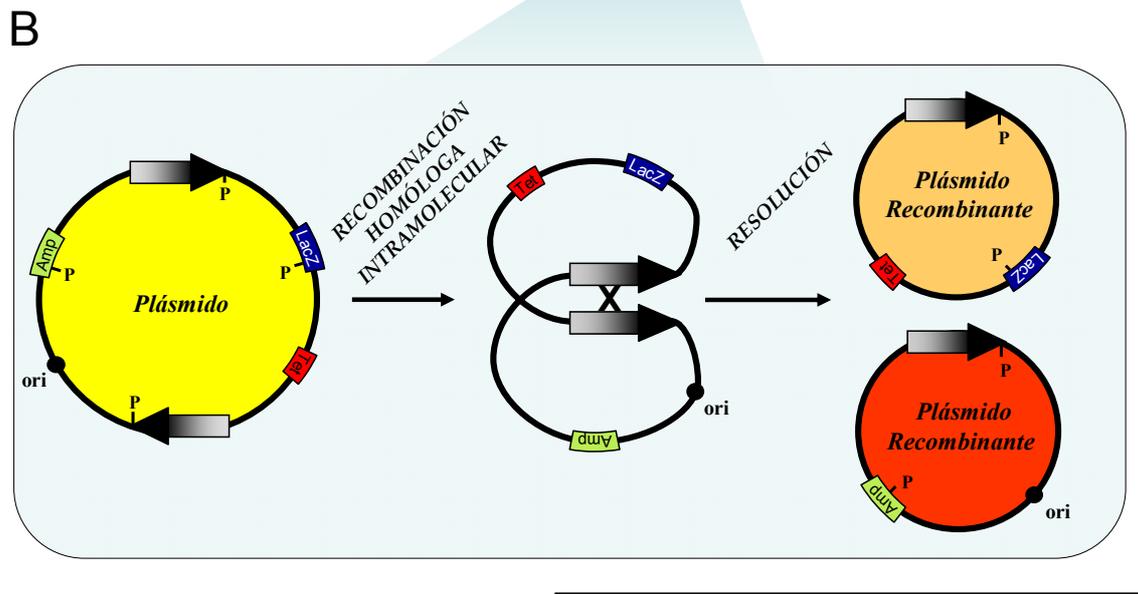
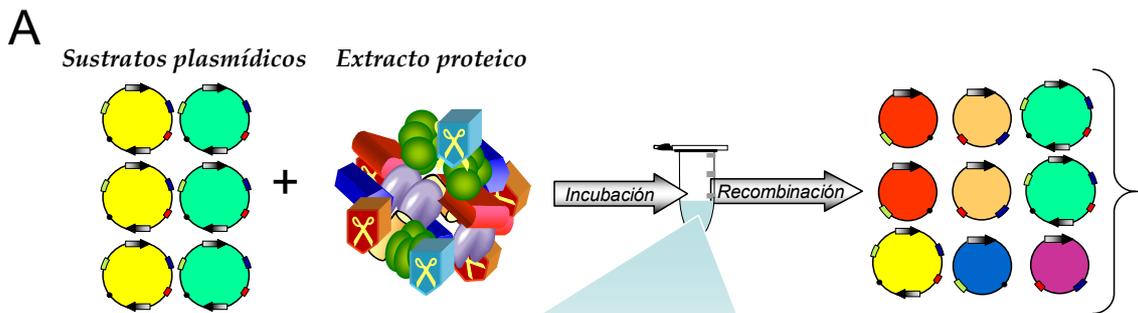
Los ensayos de recombinación estaban diseñados de tal manera que se podían detectar los eventos de recombinación intramolecular entre las dos secuencias homólogas clonadas en el plásmido (Boán et al., 1998), siguiendo la estrategia que se describe en la Figura 11. Los plásmidos se incubaban (cantidades equimoleculares de un plásmido portador de dos copias del



**Figura 11.** Estrategia seguida en los experimentos de recombinación llevados a cabo con los sustratos plasmídicos. **A.** Esquema de un ensayo donde se incuban conjuntamente un plásmido portador de dos copias de uno de los alelos de la región MsH42 (plásmidos con relleno amarillo) y un plásmido que lleva dos copias de una secuencia no repetitiva de 850 pb empleada como control (plásmidos con relleno verde). El mismo número de moléculas de cada uno de estos plásmidos se incuban conjuntamente con un extracto proteico nuclear de testículo de rata que aporta la maquinaria recombinatoria. Los plásmidos recombinantes que portan la secuencia MsH42 se muestran con relleno rojo (viable) y naranja (no viable) y los plásmidos recombinantes que portan la secuencia no repetitiva se muestran con relleno azul (viable) y morado (no viable) **B.** El proceso de recombinación intramolecular entre las dos secuencias homólogas de cualquiera de los plásmidos (se muestra el que contiene dos copias de la secuencia MsH42) concluye con la generación de dos nuevos plásmidos. Uno de ellos permitirá la formación de colonias resistentes a ampicilina, pero incapaces de realizar  $\alpha$ -complementación (rojo), mientras que el otro no puede propagarse en las bacterias al carecer de origen de replicación (naranja). **C.** Tras transformar bacterias competentes con los productos purificados del ensayo, los plásmidos originales que no han sufrido alteración alguna darán lugar a colonias azules al poseer tanto resistencia a ampicilina como capacidad de  $\alpha$ -complementación, mientras que los plásmidos recombinantes viables darán lugar a colonias blancas. Para saber si las colonias azules proceden del plásmido portador de la secuencia minisatélite o del plásmido control, éstas se transfieren a una membrana de nylon y se realiza una hibridación con una sonda para el minisatélite. Por su parte, para determinar el origen de los plásmidos recombinantes, estos se analizaron mediante digestión con enzimas de restricción. En el gel de la parte inferior se muestran los patrones obtenidos con *Pst*I a partir del plásmido pMsH42*Lac* (flecha amarilla), que contiene dos copias del alelo largo de la región MsH42, y de algunos recombinantes obtenidos a partir de él (recombinante igual, punto rojo; recombinantes no iguales, rombos rojos), así como del plásmido p5 MsH42*Lac* (flecha verde), con dos copias de la secuencia control 5' MsH42 y de un recombinante originado a partir de él (punto azul).

otra copia del minisatélite, el gen de resistencia a ampicilina, el origen de replicación *oriC*, pero no *lacZ*, por lo que generarán colonias bacterianas blancas. Esto permite distinguirlas de las colonias transformadas con los plásmidos originales, que al poseer *lacZ*, podrán realizar la  $\alpha$ -complementación y formarán colonias azules en presencia de IPTG y X-Gal. Las colonias blancas se analizan mediante la digestión del DNA plasmídico con el enzima de restricción adecuado. Si el entrecruzamiento se realiza entre secuencias perfectamente alineadas, el fragmento de restricción que contiene al minisatélite mantendrá su tamaño original, mientras que si el alineamiento fue inexacto, dicho fragmento tendrá un tamaño diferente al esperado. Las colonias azules se transfieren a una placa nueva y se hibridan con una sonda para el minisatélite MsH42 lo que permite distinguir entre aquellas producidas por los plásmidos que contienen a MsH42 de las producidas por los plásmidos control.

Gracias a esta metodología pudo comprobarse que la región MsH42 es una secuencia que potencia fuertemente la recombinación homóloga intramolecular *in vitro* cuando se compara con una secuencia control no repetitiva (Boán et al., 1998). El alelo largo se mostró como el más recombinogénico de los tres detectados, con una tasa de potenciación media de aproximadamente 12 veces, frente a las 3.6 veces obtenidas con los alelos mediano y corto, que parecen comportarse de forma similar en este sentido (Boán et al., 2002). En presencia de las regiones flanqueantes, los tres alelos promueven una elevada tasa de entrecruzamientos iguales (73-78%), mientras que el alelo largo sin sus regiones flanqueantes presenta valores de sólo un 31%, por lo que dichas secuencias influyen en la precisión con la que ocurre el emparejamiento de las cadenas de DNA. Además, el alelo largo del minisatélite MsH42 en presencia de sus regiones flanqueantes produce tasas

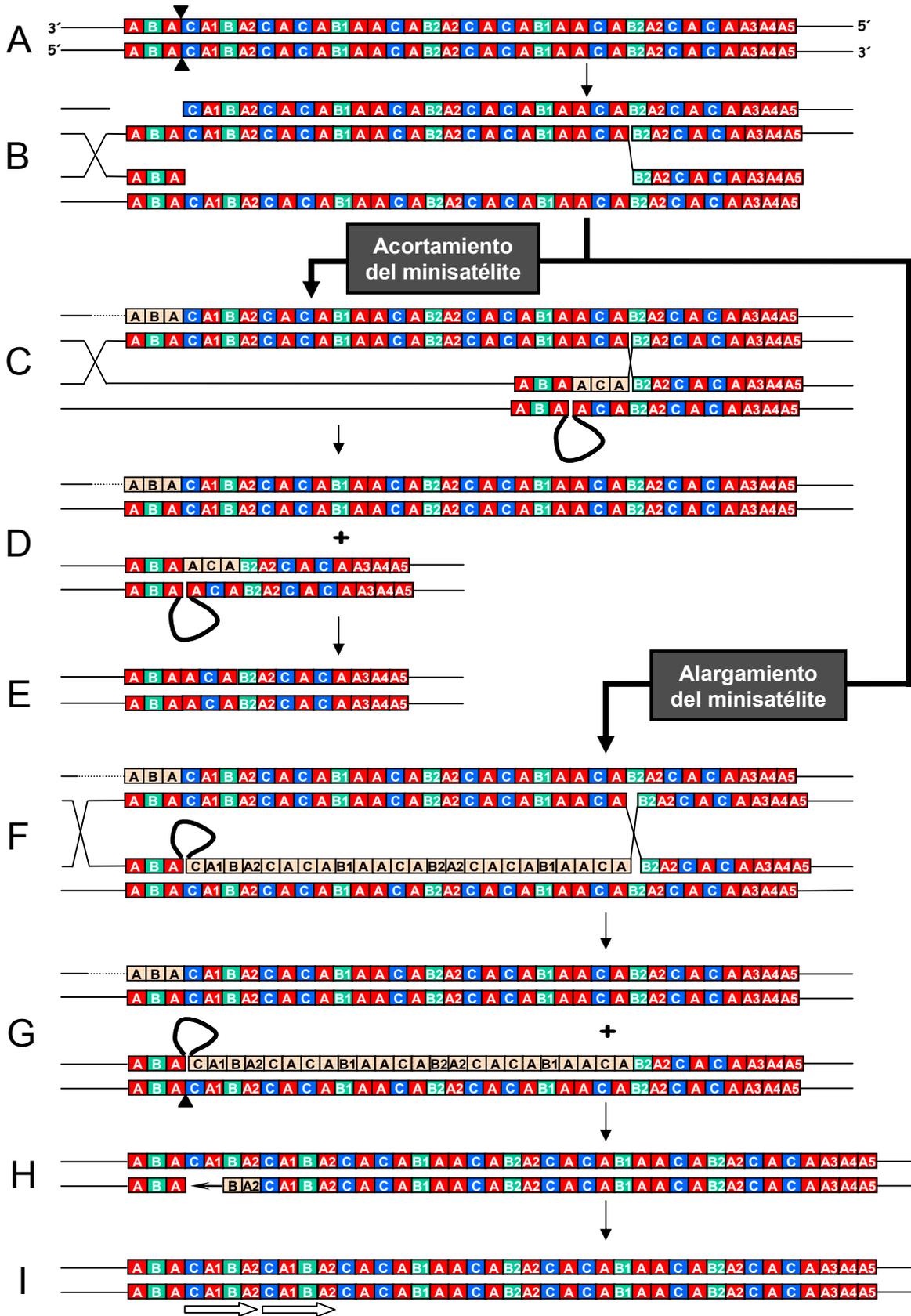


de recombinación totales 4.3 veces más altas que en ausencia de ellas, mientras que experimentos con las regiones control no repetitivas que presentan diferencias de tamaño similares sólo muestran variaciones de 1.8 veces en la tasa de recombinación. Por tanto, las regiones flanqueantes también influyen en la capacidad recombinatoria global de la región MsH42, hecho que no puede explicarse solamente por la diferencia de tamaño entre las secuencias homólogas dentro del plásmido. Es interesante señalar que el descenso en la recombinación global detectado en ausencia de dichas secuencias flanqueantes se produce principalmente a expensas de la caída en el número de recombinantes iguales, lo que refuerza su importancia en el emparejamiento exacto de las secuencias homólogas (Boán et al., 2002).

Este conjunto de datos nos ha permitido proponer una hipótesis según la cual los eventos de recombinación observados en la región MsH42 obedecen a una combinación de diferentes rutas bioquímicas (Boán et al., 2002). Por un

lado, existiría una ruta iniciada por un corte de doble cadena (DSB) en el DNA flanqueante de MsH42, seguido de la formación de una unión de Holliday que se resolvería dentro de la zona repetitiva de MsH42, generando recombinantes iguales. Este modelo es similar a otros ya propuestos anteriormente (Wahls, 1998; Vergnaud & Denoëud, 2000) y permitiría explicar aproximadamente el 75% de los sucesos de recombinación detectados. Para el 25% restante, sugerimos una segunda vía de recombinación basada en la hibridación de cadena sencilla dependiente de síntesis (SDSA). Este mecanismo se iniciaría esencialmente por un DSB dentro de MsH42. La resección de los extremos 5' generados tras el corte conlleva la generación de extremos 3' de cadena sencilla con capacidad invasiva sobre el dúplex homólogo intacto. La naturaleza repetitiva del minisatélite favorecería entonces emparejamientos erróneos de la hebra invasiva o la generación de sucesos de *slippage* durante el proceso de síntesis reparadora. Este modelo se representa gráficamente en la Figura 12.

**Figura 12.** Modelo de recombinación que explica la generación de los recombinantes con variaciones en el minisatélite MsH42, partiendo de un alineamiento perfecto de ambas copias homólogas. **A.** Se produce un corte de doble cadena en uno de los dúplex, indicado con puntas de flecha de color negro, que hemos situado entre las repeticiones 3ª y 4ª. **B.** La degradación de los extremos 5' deja extremos 3' de cadena sencilla y uno de ellos invade el dúplex homólogo, dando lugar a una unión de Holliday. **C-E.** *Slippage* en la cadena molde que producirá una delección en la hebra de nueva síntesis. El bucle de cadena sencilla (color negro) será eliminado por los sistemas de reparación, capaces de corregir bucles grandes de DNA (Kirkpatrick & Petes, 1997). **F-I.** *Slippage* en la cadena de nueva síntesis, dando lugar a una duplicación en el minisatélite. El corte de la hebra molde (punta de flecha negra) permite la extensión del bucle que será utilizado para completar la otra hebra, dando lugar a una duplicación directa dentro del minisatélite (flechas huecas). En ambos casos la resolución de las dos uniones de Holliday generará productos de recombinación derivados de un entrecruzamiento o de una conversión génica. Nuestro sistema de recombinación *in vitro* solamente detecta los originados por entrecruzamiento.







# *Objetivos*



Los anteriores estudios realizados en nuestro laboratorio dejaron abiertos dos aspectos de la biología de Msh42 que consideramos de interés y que han sido el objetivo fundamental de esta Tesis. El primero de ellos es el establecimiento del origen evolutivo de la secuencia Msh42 y de su polimorfismo en humanos. En esta dirección, son todavía muy pocos los trabajos realizados sobre la evolución de minisatélites y todos ellos tratan sobre minisatélites hipervariables o altamente polimórficos (Gray & Jeffreys, 1991; Charlesworth et al., 1994; Haber & Louis, 1998; Taylor & Breden, 2000). Sin duda, el estudio de un minisatélite no hipervariable como Msh42 podría aportar nuevos datos sobre la evolución de este tipo de secuencias.

El segundo punto que nos propusimos investigar fue la generación de cortes de doble cadena dentro de la región Msh42. Puesto que en los experimentos de recombinación con el alelo largo se habían obtenido tasas de formación de recombinantes próximas al 1%, era plausible suponer que al menos un 1% de los plásmidos substrato estarían sufriendo DSBs. En este trabajo hemos intentado localizar estos puntos de corte a lo largo de la región Msh42 y determinar si existe algún tipo de secuencia que pueda estar promoviendo la aparición de estos cortes.

Los objetivos fundamentales de esta Tesis Doctoral son:

**1.-** Estudiar el origen (nacimiento) y la historia evolutiva de la región Msh42 mediante su búsqueda y análisis en primates y otras especies disponibles en las bases de datos genómicas.

**2.-** Estudiar la generación de DSBs en la región Msh42, su distribución y las posibles moléculas implicadas en este proceso. Para ello, se desarrollará un sistema para la detección de DSBs en el DNA y su posterior análisis.



*Procedimientos  
Experimentales*



### **Metodología general de clonación**

Toda la metodología general de clonación empleada en esta Tesis Doctoral se encuentra descrita en el manual de laboratorio de Sambrook & Russell (2001). Los protocolos para la preparación de DNA plasmídico a pequeña escala (minipreparado), a gran escala (maxipreparado) y de minipreparado especial para secuenciación se realizaron siguiendo métodos basados en la lisis alcalina. La preparación y transformación de bacterias competentes DH5 $\alpha$  (recA<sup>-</sup>, endA<sup>-</sup>,  $\phi$ 80 LacZ  $\Delta$ M15) se llevó a cabo mediante el protocolo de Inoue et al., 1990.

### **Obtención del DNA genómico**

En el presente estudio se han empleado muestras de DNA genómico de 22 individuos africanos no emparentados procedentes de diferentes poblaciones subsaharianas para analizar el polimorfismo de Msh42 en población negroide. Estas muestras fueron cedidas amablemente por el Dr. J. L. B. Caeiro del Departamento de Antropología de la Facultad de Biología de la Universidad de Santiago de Compostela. Para el estudio filogenético de Msh42 hemos utilizado muestras de DNA genómico de 11 especies de primates: 7 chimpancés (*Pan troglodytes*), 6 gorilas (*Gorilla gorilla*), 4 orangutanes (*Pongo pygmaeus*), 1 gibón (*Hylobates lar*), 3 macacos (1 *Macaca mulatta* y 2 *Macaca nigra*), 1 mandril (*Mandrillus sphinx*), 1 tamarindo (*Saguinus oedipus*) y 3 lemúridos (*Lemur catta*, *Eulemur fulvus* y *Varecia variegata*). Algunas de las muestras de DNA genómico fueron generosamente cedidas por los Drs. J. Bertranpetit, A. Blancher, J. Bullerdiek, S. Pääbo, P. Rogalla, W. Schempp y R. Wimmer. También hemos contado con la amabilidad de la Dra. M. Delclaux y de sus colaboradores del Zoo de Madrid

a la hora de proporcionarnos muestras de sangre completa de algunos de sus primates. En este caso, el procesado de las muestras para la extracción del DNA genómico se realizó con el QIAamp DNA kit (QIAGEN) siguiendo las especificaciones del fabricante.

### **Preparación de los sustratos de DNA para los ensayos DSB**

Para la búsqueda de los lugares donde se generan los DSBs en la región Msh42 se puso a punto un ensayo (ensayo DSB) que consistía en la incubación de un determinado sustrato de DNA junto con un extracto proteico que aporta la maquinaria para producir tales cortes. Como sustratos para los ensayos DSB, se han utilizado tres tipos de fragmentos de DNA: i) un fragmento de ~ 0.87 kb con distintos extremos y que contiene la región Msh42, ii) dos fragmentos de ~ 0.8 y 1.0 kb que no contienen secuencias repetitivas y iii) oligonucleótidos de 50 pb con el motivo TGGGG y variantes del mismo u otros motivos diferentes (Tabla 1).

El sustrato que contiene la región Msh42 (minisatélite y secuencias adyacentes), requirió primeramente la construcción del plásmido pSSma. Este plásmido es un derivado de pGEM-T Easy (Promega), un vector preparado para clonar productos de PCR con un residuo de adenina en el extremo 3' del fragmento amplificado, producido por la *Taq* DNA polimerasa. Para formar pSSma, se clonó en pGEM-T Easy el alelo largo de la región Msh42, obtenido a partir del plásmido pRep42 (González, 1993) con los cebadores PS1Sma y PS2Sma (Tabla 3), empleando las condiciones de reacción que se describen más adelante (ver *Amplificaciones por PCR*). Este inserto puede ser liberado de pSSma mediante una digestión enzimática con *EcoRI*, *SmaI* o *PstI*, generándose los fragmentos f42E

**Tabla 1.** Secuencia de los oligonucleótidos de 50 pb empleados en los ensayos DSB

| Nombre     | Secuencia  |
|------------|--|
| <b>F1</b>  | ACATCAGTGAATTCAAGATCAAAAAAAAA <b>TGGGG</b> ATTCTGCAGCGACGTAA |
| <b>F2</b>  | ----- <b>TGGGG</b> AAAA-----                                 |
| <b>M1</b>  | ----- <b>CGGGG</b> -----                                     |
| <b>M2</b>  | ----- <b>TCGGG</b> -----                                     |
| <b>M3</b>  | ----- <b>TGCGG</b> -----                                     |
| <b>M4</b>  | ----- <b>TGGCG</b> -----                                     |
| <b>M5</b>  | ----- <b>TGGGC</b> -----                                     |
| <b>M6</b>  | ----- <b>TGGGGC</b> -----                                    |
| <b>M7</b>  | ----- <b>TG<b>T</b>GG</b> -----                              |
| <b>M8</b>  | ----- <b>TGGGA</b> -----                                     |
| <b>M9</b>  | ----- <b>TGGGGG</b> -----                                    |
| <b>M10</b> | ----- <b>TGGGGGG</b> -----                                   |
| <b>M11</b> | ----- <b>T<b>CC</b>GG</b> -----                              |
| <b>M12</b> | ----- <b>TG<b>CC</b>G</b> -----                              |
| <b>M13</b> | ----- <b>TGG<b>CC</b></b> -----                              |
| <b>M14</b> | ----- <b>T<b>TT</b>GG</b> -----                              |
| <b>M15</b> | ----- <b>TG<b>TT</b>G</b> -----                              |
| <b>M16</b> | ----- <b><b>CCC</b>GGG</b> -----                             |
| <b>M17</b> | ----- <b>CCC</b> GGG-----                                    |
| <b>S1</b>  | CTCCGGCCTGGGTCTCAGGGATGCCAGCTGCAA <b>ACT</b> GGGAGGGCCATTGT  |
| <b>S2</b>  | ----- <b>TTT</b> -----                                       |
| <b>S3</b>  | ACATGATCGAATTCAAGATAAACAGAAATGGGGATTAGGTACCGTCGTAA           |
| <b>S4</b>  | ----- <b>CAGCAGA</b> -----                                   |
| <b>S5</b>  | ----- <b>CAGCTGA</b> -----                                   |
| <b>S6</b>  | ----- <b>AAACAGAAAAACAGAAA</b> -----                         |
| <b>S7</b>  | ----- <b>AAACAGAAAAACTGAAA</b> -----                         |
| <b>S8</b>  | ----- <b>AAACAGAAAGCCTAAACAGAAA</b> -----                    |
| <b>S9</b>  | ----- <b>AAACAGAAAGCCTAAACTGAAA</b> -----                    |
| <b>S10</b> | ACTTGATTGTGTTCAGATAGAGATAAAATGATGATTGAGTATCGTGCTTG           |
| <b>S11</b> | ----- <b>GAATTC</b> -----                                    |
| <b>S12</b> | ----- <b>GGTACC</b> -----                                    |
| <b>S13</b> | ----- <b>CAGCTG</b> -----                                    |
| <b>S14</b> | ----- <b>GAATTC</b> -----                                    |
| <b>S15</b> | ----- <b>GGTACC</b> -----                                    |
| <b>S16</b> | ----- <b>GAATTT</b> -----                                    |
| <b>S17</b> | ----- <b>AAATTC</b> -----                                    |
| <b>S18</b> | ----- <b>AAATTT</b> -----                                    |
| <b>S19</b> | ----- <b>GAGTTC</b> -----                                    |
| <b>S20</b> | ----- <b>GATC</b> -----                                      |

Se muestra una de las cadenas, en orientación 5'→3', de cada oligonucleótido de 50 pb. El motivo TGGGG en F1 y F2 está sombreado. Los guiones en las secuencias indican identidad con respecto a la misma posición en el primer oligonucleótido de cada tanda (marcadas alternativamente en blanco o sombreado gris). Las mutaciones con respecto al motivo TGGGG de F1 o su entorno se indican en negrita en los oligonucleótidos M1-M16.

(*EcoRI*, 880 pb, con extremos 5'-protuberantes), f42S (*SmaI*, 869 pb, con extremos romos) y f42P (*PstI*, 863 pb, con extremos 3'-protuberantes), respectivamente. En la Figura 13A se resume el proceso de obtención de los fragmentos f42.

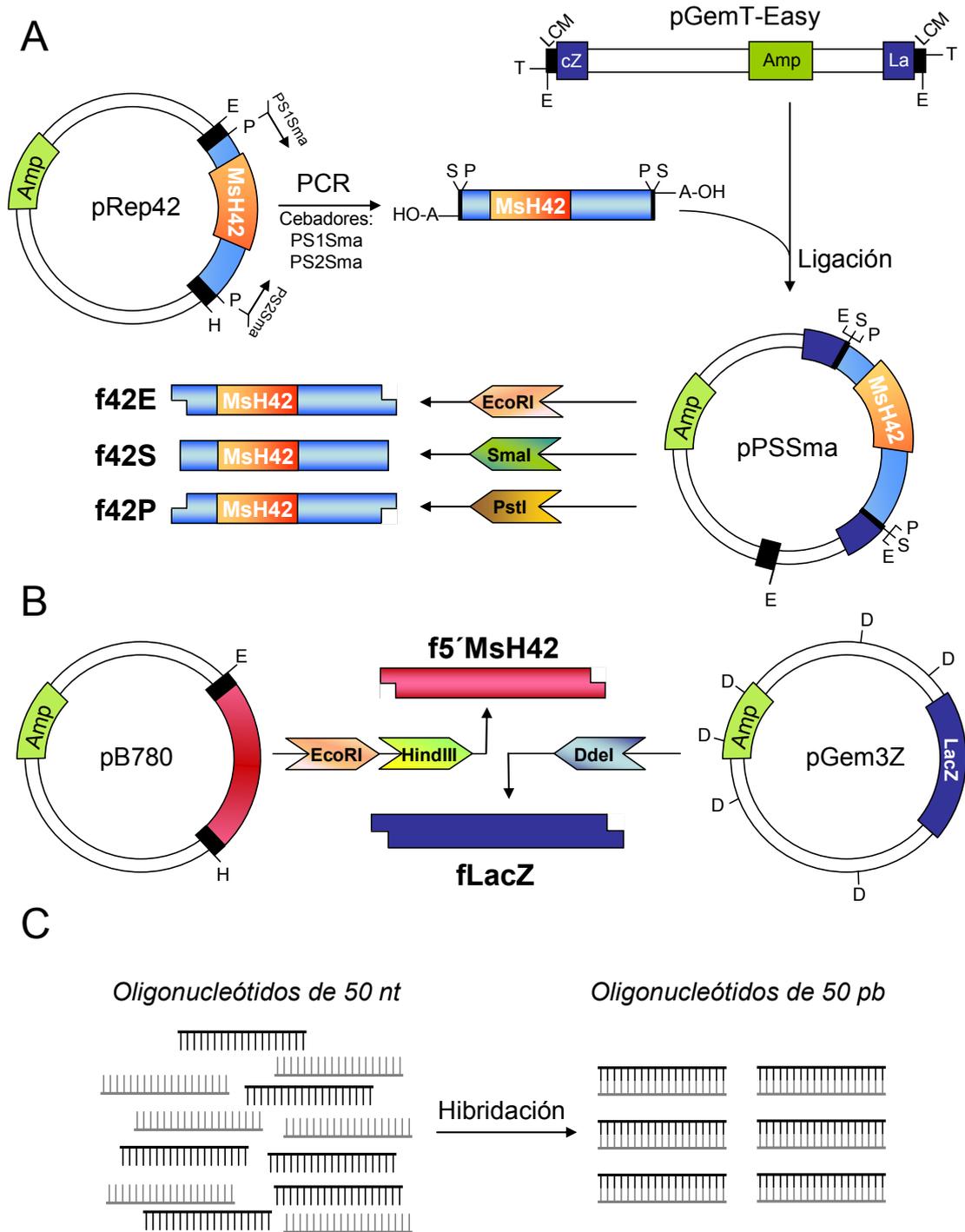
Para la obtención de cualquiera de estos tres fragmentos, 500 µg de pSSma se digirieron con 250 U del enzima correspondiente en un volumen total de 500 µl del tampón adecuado durante 8-12 h a 37°C (25°C para *SmaI*). Tras su digestión, el DNA fue extraído una vez con fenol saturado con TE pH 8.0 (Tris-HCl 10 mM, 1 mM EDTA): cloroformo: alcohol isoamílico 25:24:1 y otra con cloroformo: alcohol isoamílico 24:1, precipitado con 0.1 vol de acetato sódico 3M pH 5.2 y 2 vol de etanol 96% y finalmente se resuspendió en TE pH 8.0. Para la purificación de los fragmentos resultantes de la digestión de pSSma, el DNA se cargó en geles preparativos de agarosa al 1% en 1X TAE (40 mM Tris-acetato, 1 mM EDTA, pH 8.0). Seguidamente, la porción de agarosa conteniendo el inserto de ~ 0.87 kb se cortó del gel e introdujo en una tripa de diálisis con capacidad de retención de moléculas >12 kD. Seguidamente, el DNA fue electroeluido siguiendo el método descrito en Sambrook & Russell, 2001. La eliminación del bromuro de etidio del DNA se realizó mediante una extracción con fenol, otra con fenol: cloroformo: alcohol isoamílico 25:24:1 y otra con cloroformo: alcohol isoamílico 24:1. El DNA se precipitó como se describe previamente, resuspendiéndose en 10 mM Tris-HCl pH 8.0. Finalmente, el DNA fue cuantificado por espectrofotometría y analizado por electroforesis en geles de agarosa al 1% en 1X TAE frente a estándares de tamaño y concentración conocida. Los fragmentos que no contienen secuencias repetitivas fueron preparados a partir de los plásmidos pB780 (Boán et al., 1997)

y pGEM-3Z (Promega) de forma análoga a los fragmentos que contienen la región MsH42 (Figura 13B). El primero de ellos es un fragmento de ~ 850 pb sin secuencias repetitivas y situado en el extremo 5' de la secuencia H42 (f5'MsH42). Este fragmento se obtuvo mediante la digestión del plásmido pB780 con *EcoRI* y *HindIII*. El segundo fragmento (fLacZ), de 967 pb, se preparó por digestión del plásmido pGEM-3Z con *DdeI* y comprende el fragmento N-terminal del gen de la β-galactosidasa, implicado en la α-complementación (Ullmann et al., 1967).

Para generar los oligonucleótidos de doble cadena de 50 pb que contienen el motivo TGGGG u otros y sus variantes, se diseñaron previamente una serie de parejas de oligonucleótidos de cadena sencilla (50 nt) y secuencia complementaria que fueron sintetizados por Sigma-Genosys. La hibridación de las parejas de oligonucleótidos de cadena sencilla se realizó mezclando cantidades equimoleculares de ambos (típicamente 25 µl de un stock a 100 µM) en 10 mM Tris-HCl pH 8.0, 15 mM NaCl en un volumen total de 150 µl. La mezcla de oligonucleótidos de cadena sencilla se calentó a 95°C durante 5 min, 66°C durante 3 min y se dejó enfriar a temperatura ambiente (~ 15 min) para permitir su hibridación (Figura 13C). Los oligonucleótidos de doble cadena así sintetizados tienen una concentración de 16.2 µM, lo que equivale aproximadamente a 0.5 µg/µl.

### ***Preparación de extractos proteicos***

Los extractos proteicos nucleares y citosólicos se prepararon mediante una adaptación del protocolo de Dignam et al., 1983 (Ausubel et al., 1987). Los tejidos empleados fueron testículo, hígado, cerebro, bazo y riñón de ratas Sprague-Dawley adultas (~300 g de



peso, 60-65 días de edad) y cerebro de pollo (~20 días de edad). Brevemente, los animales se sacrificaron mediante asfixia con cloroformo o CO<sub>2</sub>, (siguiendo la normativa de la UE referente a la Experimentación y Protección Animal 86/609/CEE), tras lo cual se realizó la extracción rápida de los órganos. A partir de aquí todos los pasos se realizaron en frío y los tampones que

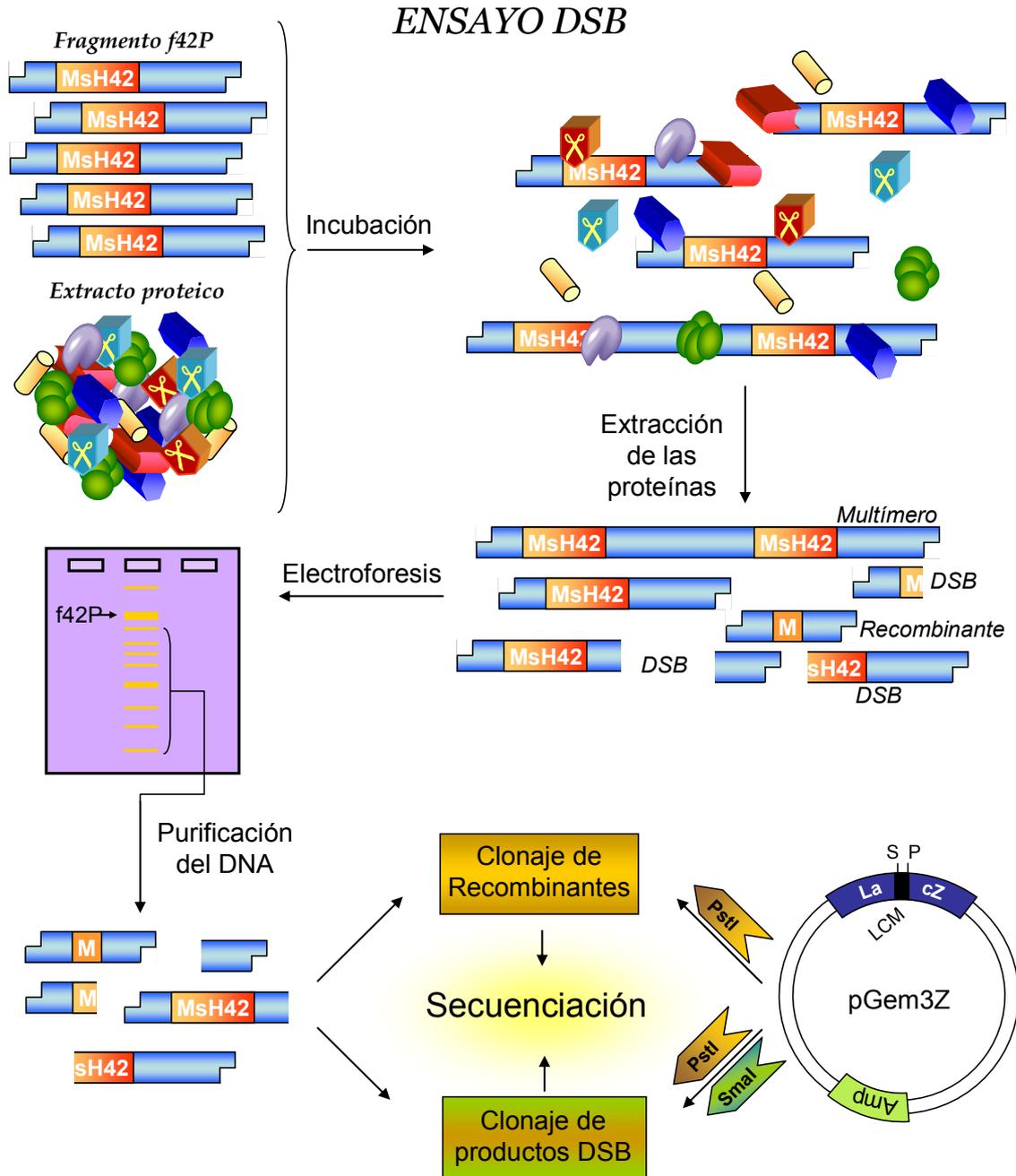
se añadieron fueron previamente enfriados en hielo. Los órganos extraídos se lavaron con tampón PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 2 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.4), eliminándose los restos de sangre, tejido graso y conjuntivo. El tejido se homogenizó en PBS con un homogenizador manual de vidrio esmerilado y a continuación las células se recogieron mediante una

**Figura 13.** Esquema de la obtención de los sustratos empleados en los ensayos DSB. **A.** Generación de los fragmentos f42E, f42P y f42S. El producto de PCR resultante de la amplificación de pRep42 con los cebadores PS1Sma y PS2Sma (Tabla 3) se clonó en pGemT-Easy, vector que permite el clonaje de productos de PCR con una adenina adicional en el extremo 3'. El plásmido resultante, pSSma, libera la región MsH42 mediante digestiones independientes con *EcoRI*, *SmaI* o *PstI*, dando lugar a los fragmentos f42E, f42S y f42P, respectivamente. **B.** Generación de los fragmentos f5'MsH42 y fLacZ. El fragmento f5'MsH42 se obtuvo a partir del plásmido pB780 (Boán et al. 1998) mediante digestión con *EcoRI* y *HindIII*. Análogamente, el fragmento fLacZ se obtuvo mediante la digestión de pGem3Z con *DdeI*. **C.** Generación de los oligonucleótidos de doble cadena de 50 pb. Inicialmente, se diseñaron y sintetizaron oligonucleótidos de cadena sencilla de 50 nt con secuencias complementarias entre sí (cadenas negras y grises). Posteriormente, cada pareja de oligonucleótidos se mezcló en cantidades equimoleculares y se hibridó, consiguiendo así los dúplex de 50 pb. D, *DdeI*; E, *EcoRI*; H, *HindIII*; P, *PstI*; S, *SmaI*; LCM, lugar de clonaje múltiple. Amp y Lac Z, igual que en Figura 6.

centrifugación a 1850g, 15 min a 4°C, midiéndose el volumen de las mismas. Este valor constituye el llamado *pcv* (*packed cell volume*). Las células se resuspendieron en un volumen de PBS igual a 5 *pcv* y se centrifugaron en las mismas condiciones que en el paso anterior. El exceso de PBS se lavó rápidamente con 5 *pcv* de tampón hipotónico (10 mM HEPES pH 7.9, 10 mM KCl, 1.5 mM MgCl<sub>2</sub>, 0.2 mM PMSF, 0.5 mM DTT), recuperando las células por centrifugación a 1850g, 5 min, 4°C. A continuación, las células se resuspendieron en 3 *pcv* de tampón hipotónico y se dejaron hinchar durante 10 min en hielo. Inmediatamente, las células se lisaron con el homogenizador manual para liberar los núcleos, que se recogieron por centrifugación a 3300g, 15 min a 4°C. En algunos casos, el sobrenadante resultante de esta última centrifugación se almacenó para obtener el extracto citoplásmico S-100 (ver más abajo). El volumen de los núcleos se midió para obtener el valor del *pnv* (*packed nuclei volume*) y se resuspendieron en 0.5 *pnv* de tampón de baja salinidad (20 mM HEPES pH 7.9, 20 mM KCl, 1.5 mM MgCl<sub>2</sub>, 0.2 mM EDTA, 0.2 mM PMSF, 0.5 mM DTT, 25% glicerol). Para realizar la extracción salina de las proteínas nucleares, se añadió a la suspensión de núcleos 0.5 *pnv* de un tampón de alta salinidad (20 mM HEPES pH 7.9, 1.2 M KCl, 1.5 mM MgCl<sub>2</sub>, 0.2 mM EDTA, 0.2 mM PMSF,

0.5 mM DTT, 25% glicerol) manteniendo la mezcla en agitación durante 30 min en hielo. Seguidamente, los núcleos se centrifugaron a 25000g, 30 min a 4°C y el sobrenadante se dializó contra 50 vol de tampón de diálisis (20 mM HEPES pH 7.9, 0.1 M KCl, 0.2 mM EDTA, 0.2 mM PMSF, 0.5 mM DTT, 20% glicerol) durante 3 h. Los restos de ácido nucleico y proteínas que precipitan durante la diálisis se eliminaron mediante una centrifugación a 25000g, 20 min a 4°C. El sobrenadante, que constituye el extracto proteico nuclear, se alicuoteó y congeló en nitrógeno líquido, almacenándose a -70°C. La actividad del extracto bajo estas condiciones permanece aparentemente inalterada durante al menos 6 meses. La concentración proteica de los extractos nucleares fue determinada mediante el método de Bradford (Bio-Rad), siendo las siguientes: testículo de rata, 4.50 - 5.75 µg/µl; hígado de rata, 8.75 - 10.50 µg/µl; cerebro de rata, 1.75 - 3.00 µg/µl; bazo de rata, 7.50 - 8.75 µg/µl; riñón de rata, 8.00 - 9.00 µg/µl; cerebro de pollo, 2.75 - 3.50 µg/µl.

Para la preparación del extracto citoplásmico, el sobrenadante obtenido tras la sedimentación de los núcleos, se mezcló con 0.11 vol de 10X tampón de extracción citoplásmica (0.3 M HEPES pH 7.9, 1.4 M KCl, 30 mM MgCl<sub>2</sub>) y se centrifugó durante 60 min a 100000g. Posteriormente, el sobrenadante se



**Figura 14.** Esquema del ensayo DSB. Un fragmento de DNA empleado como sustrato (se muestra el f42P) se mezcla con un extracto proteico nuclear y se incuba a 37°C. Los productos de la reacción (multímeros del fragmento f42P, productos de degradación –DSBs– y recombinantes) se purifican y posteriormente se analizan mediante electroforesis en geles de agarosa. El DNA de menor peso molecular que el del sustrato inicial fue extraído del gel de agarosa y clonado en el vector pGem3Z. El clonaje de las moléculas recombinantes se efectuó usando este plásmido linealizado con *PstI*, mientras que para el de los fragmentos generados por un DSB se empleó el plásmido digerido con *PstI* y *SmaI*. Finalmente, los plásmidos recombinantes se analizaron mediante digestión con enzimas de restricción y se secuenciaron. P, *PstI*; S, *SmaI*; LCM, lugar de clonaje múltiple; Amp y Lac Z, igual que en Figura 6.

dializó durante 3 horas frente a 50 vol de tampón de diálisis, tras lo cual fue alicuoteado, congelado en nitrógeno líquido y almacenado a -70°C. La

concentración de los extractos citoplásmicos fue de 1.90 µg/µl para testículo de rata y de 5.50 µg/µl para hígado de rata.

**Ensayos de corte de doble cadena (DSB) del DNA**

Los ensayos DSB, esquematizados en la Figura 14, se realizaron en un volumen final de 50 µl de tampón DSB (20 mM Tris-HCl pH 7.5, 10 mM MgCl<sub>2</sub>, 1 mM ATP, 0.1 mM de cada dNTP), conteniendo 12 µg de extracto proteico y como sustrato 1.5 µg de DNA (~ 2.8 pmol para los fragmentos f42E, f42P y f42S y ~ 48.6 pmol para los oligonucleótidos de 50 pb). Estas

condiciones fueron consideradas como estándar, con un tiempo típico de incubación de 30 min a 37°C, aunque en función del tipo de extracto o de sustrato los tiempos de incubación oscilaron entre 2 min y 4 h. A continuación, las reacciones se pararon mediante un enfriamiento rápido en hielo y posterior congelación a -70°C, seguida de una extracción con fenol: cloroformo: alcohol isoamílico 25:24:1, una con cloroformo: alcohol isoamílico 24:1 y precipitación del DNA con 0.1 vol de

**Tabla 2.** Condiciones de reacción en los ensayos DSB.

| Reacción                 | MgSO <sub>4</sub> (mM) | dNTPs (mM) | ATP (mM) | NaCl (mM) | KCl (mM) | ddNTPs (mM) | EDTA (mM) | EGTA (mM) | NEM (mM) | Preincubación extracto (5 min)* |
|--------------------------|------------------------|------------|----------|-----------|----------|-------------|-----------|-----------|----------|---------------------------------|
| Estándar                 | 10                     | 0.1        | 1        | -         | -        | -           | -         | -         | -        | -                               |
| Sin Mg                   | -                      | 0.1        | 1        | -         | -        | -           | -         | -         | -        | -                               |
| Mg variable              | 0.5-7.5                | 0.1        | 1        | -         | -        | -           | -         | -         | -        | -                               |
| Sin ATP                  | 10                     | 0.1        | 0        | -         | -        | -           | -         | -         | -        | -                               |
| Sin dNTPs                | 10                     | -          | 1        | -         | -        | -           | -         | -         | -        | -                               |
| Sin dNTPs/<br>con ddNTPs | 10                     | -          | 1        | -         | -        | 0.1         | -         | -         | -        | -                               |
| Con KCl/<br>con Mg       | 10                     | 0.1        | 1        | -         | 10-40    | -           | -         | -         | -        | -                               |
| Con KCl/<br>sin Mg       | -                      | 0.1        | 1        | -         | 10-40    | -           | -         | -         | -        | -                               |
| Con NaCl/<br>con Mg      | 10                     | 0.1        | 1        | 10-40     | -        | -           | -         | -         | -        | -                               |
| Con NaCl/<br>sin Mg      | -                      | 0.1        | 1        | 10-40     | -        | -           | -         | -         | -        | -                               |
| Con EGTA/<br>con Mg      | 10                     | 0.1        | 1        | -         | -        | -           | -         | 1-50      | -        | -                               |
| Con EGTA/<br>sin Mg      | -                      | 0.1        | 1        | -         | -        | -           | -         | 1-50      | -        | -                               |
| Con EGTA/<br>con Mg      | 10                     | 0.1        | 1        | -         | -        | -           | -         | -         | 0.5-50   | -                               |
| Con EGTA/<br>sin Mg      | -                      | 0.1        | 1        | -         | -        | -           | -         | -         | 0.5-50   | -                               |
| Preincubación/<br>con Mg | 10                     | 0.1        | 1        | -         | -        | -           | -         | -         | -        | 50 ó 60°C                       |
| Preincubación/<br>sin Mg | -                      | 0.1        | 1        | -         | -        | -           | -         | -         | -        | 50 ó 60°C                       |

El tiempo de incubación estándar fue en todos los casos de 30 min, a menos que se indique lo contrario. En todas las reacciones el tampón empleado fue Tris-HCl pH 7.5 a una concentración de 20 mM, y se añaden 1.5 µg de DNA (habitualmente el fragmento f42P) y 12 µg de extracto proteico nuclear.

\* La preincubación se realizó antes de la adición del sustrato a la reacción.

acetato sódico 3M pH 5.2 y 2 vol de etanol 96%. Las muestras se analizaron mediante electroforesis en geles de agarosa al 2% en tampón 1X TAE, para el caso de las reacciones que utilizaban los fragmentos f42E, f42S, f42P, f5' MsH42 y fLacZ, o en geles de poliacrilamida (29:1) al 18% en 1X TBE (90 mM Tris-borato, 1 mM EDTA, pH 8.3) en el caso de los experimentos realizados con los oligonucleótidos de 50 pb. Los geles se tiñeron con bromuro de etidio (0.5 µg/ml) y se visualizaron empleando un transiluminador GelDoc2000 (BioRad). Hay que tener en cuenta que a igual cantidad de oligonucleótido de 50 pb que de sustrato f42P (1.5 µg) el número de moléculas del primero es aproximadamente 17 veces mayor, por lo que será necesario que un mayor número de moléculas de sustrato sean cortadas para poder apreciar los productos DSB con una intensidad equivalente a la observada en los ensayos estándar. Debido a esto, los experimentos usando oligonucleótidos de 50 pb tuvieron que incubarse durante tiempos considerablemente más largos de lo habitual (2-4 h).

En otros experimentos se añadió [ $\alpha$ -<sup>32</sup>P]-dCTP (3000 Ci/mmol) (Amersham Biosciences) para verificar si existía incorporación de este isótopo en el DNA. En estos casos, el experimento, tras ser analizado mediante electroforesis en gel de agarosa al 2%, fue transferido mediante capilaridad a una membrana de nylon Hybond-N (Amersham Biosciences) durante 12 horas usando 20X SSC (3M NaCl, 0.3M citrato trisódico) como tampón para la transferencia. A continuación, la membrana se lavó suavemente en 2X SSC y se secó en un horno a 80°C durante 2 h para fijar el DNA, tras lo cual se realizó su autorradiografía con películas Biomax MS (Kodak) en cassettes con pantallas intensificadoras Hyperscreen (Amersham Biosciences).

En algunas reacciones se modificaron una o varias de las condiciones estándar del ensayo DSB, cambios que se encuentran resumidos en la Tabla 2. Una de estas variantes fue la supresión del MgSO<sub>4</sub> del tampón DSB (reacciones sin Mg<sup>2+</sup>) o su reducción a valores inferiores al 10 mM estándar (entre 0.5 y 7.5 mM). Otras variaciones consistieron en no añadir dNTPs o ATP al tampón DSB, añadir 2',3'-dideoxirribonucleótidos (ddNTPs) a una concentración final de 0.1 mM, adición de NaCl o KCl (10 y 40 mM), adición de EDTA o EGTA (1, 10 y 50 mM), preincubación del extracto con N-etilmaleimida (NEM) a distintas concentraciones (0.5-50 mM) o preincubación del extracto a 50°C o 60°C.

### ***Ensayos DSB / end-joining***

En esta serie de experimentos se pretendía establecer las condiciones que determinaban si el ensayo sería de tipo DSB (generación de moléculas de menor peso molecular que el sustrato inicial) o de tipo *end-joining* (predominio de eventos de ligación de extremos del DNA, generándose multímeros o moléculas circulares a partir del sustrato). Para ello, se añadieron al ensayo DSB distintas concentraciones (10-1000 ng) de alguna de las siguientes proteínas: histona H1 (H1), histona H3 (H3), una mezcla de histonas (H1, H2A, H2B, H3 y H4) o poli-arginina, así como protimosina  $\alpha$  (ProT $\alpha$ ) recombinante (Vega et al., 1998). Esta última proteína fue amablemente cedida por el Dr. F. Domínguez de la Facultad de Medicina de la Universidad de Santiago de Compostela. Todas las histonas y la poli-arginina se obtuvieron de Roche Applied Science. Cuando se realizaron los experimentos donde se incorporaron H1 y ProT $\alpha$  conjuntamente al ensayo DSB, ambas proteínas se preincubaron juntas en el tampón DSB durante 15 min a 37°C, antes de la adición del extracto

nuclear y del sustrato. En estos ensayos se emplearon como sustrato tanto el fragmento f42P como f42S y los productos fueron procesados y analizados de forma idéntica a la de los ensayos DSB estándar.

En otro grupo de experimentos se varió la concentración inicial del sustrato de DNA, manteniendo constante la cantidad de extracto proteico, para observar el efecto que tiene sobre el ensayo DSB la alteración de la relación DNA/proteína. En estos casos, el fragmento f42E se marcó terminalmente con [ $\gamma$ - $^{32}$ P]-ATP (3000 Ci/mmol) (Amersham Biosciences) mediante la reacción de intercambio de la T4 polinucleótido kinasa (Amersham Biosciences), siguiendo las instrucciones del fabricante. En estas reacciones el sustrato de DNA fue una mezcla de 3 ng (5.6 fmol) del fragmento f42E marcado radiactivamente ( $\sim$  84000 cpm) y cantidades variables (0-2.8 pmol) de fragmento f42E no marcado. El análisis de los productos de esta variante del ensayo se realizó por electroforesis en geles de agarosa al 2%, seguida de transferencia a una membrana de nylon Hybond-N y autorradiografía, de manera idéntica a la descrita en el apartado anterior.

### **Clonaje de los productos del ensayo DSB**

Los fragmentos de DNA producidos en algunos ensayos DSB fueron purificados de los geles de agarosa utilizando el QIAEXII Gel Extraction Kit (QIAGEN), siguiendo las instrucciones del fabricante. En algunos casos, estos fragmentos se ligaron en el vector pGEM-3Z, digerido con *Pst*I para el aislamiento de los recombinantes o con *Pst*I y *Sma*I para el clonaje de los productos DSB (Figura 14). Antes de la ligación, el vector fue defosforilado con fosfatasa alcalina intestinal de ternera

(Promega), siguiendo las instrucciones del fabricante, siendo purificado en gel para aislarlo de moléculas no cortadas, así como de los fragmentos liberados en la doble digestión *Pst*I/*Sma*I. Las ligaciones de los productos del ensayo DSB en pGEM-3Z se realizaron a 4°C durante toda la noche con T4 DNA ligasa (Promega) y empleando relaciones vector: inserto variables (desde  $\sim$  5:1 a  $\sim$  1:5 en volumen) puesto que la cantidad de moléculas potencialmente clonables era desconocida. La cantidad típica de vector plasmídico en estas ligaciones fue de 50 ng. Con los productos de la ligación se transformaron bacterias *E. coli* DH5 $\alpha$  competentes y se sembraron en placas de 90 mm de diámetro de LB con agar (1.5% w/v), ampicilina (50  $\mu$ g/ml) y sobre cuya superficie se habían extendido anteriormente 100  $\mu$ l de 100 mM IPTG (Promega) y 20  $\mu$ l de 50 mg/ml X-Gal (Promega) para detectar las colonias bacterianas capaces de llevar a cabo la  $\alpha$ -complementación. Las placas se incubaron 12-16 h a 37°C y se almacenaron a 4°C. Las colonias blancas recombinantes se inocularon en 3 ml de medio LB con 50  $\mu$ g/ml de ampicilina y se dejaron crecer toda la noche con agitación (250 rpm) a 37°C. El DNA plasmídico fue purificado mediante un protocolo de lisis alcalina y analizado mediante digestión con enzimas de restricción (*Pst*I para recombinantes y *Eco*RI/*Hind*III para moléculas surgidas de un DSB) para verificar el tamaño de los insertos. El DNA plasmídico de los clones positivos se purificó mediante el protocolo de maxipreparado o de minipreparado especial para secuenciación.

### **Amplificaciones por PCR**

Los distintos experimentos de PCR se llevaron a cabo sistemáticamente en un volumen de 50  $\mu$ l conteniendo 10 mM Tris-HCl pH 8.3, 50 mM KCl, 1.5 mM MgCl<sub>2</sub>, 0.01% v/v Tween-20, 0.3  $\mu$ M de

cada cebador, 0.2 mM de cada dNTP, 0.02 U/μl de *Taq* DNA polimerasa (Roche Applied Science) y una cantidad variable de DNA molde. Las amplificaciones se realizaron en un termociclador PTC-200 (MJ Research). Para el análisis de los productos de PCR se emplearon alícuotas de 10 μl de la mezcla de reacción y se separaron en geles de agarosa del 1-2% en 1X TAE, en función del tamaño de los fragmentos de amplificación. Los oligonucleótidos empleados como cebadores en los diferentes experimentos de PCR se muestran en la Tabla 3.

La generación de moléculas recombinantes en el ensayo DSB se verificó mediante PCR. Para ello, los fragmentos de DNA originados en

dichos ensayos y purificados a partir de las bandas extraídas del gel de agarosa (~ 1 ng) se amplificaron con los oligonucleótidos PS1 y PS2 (Tabla 3), empleando las siguientes condiciones: 29 ciclos de 95°C (30 s), 60°C (40 s), 72°C (1 min), y un ciclo final con una extensión de 5 min. Este ciclo adicional con extensión de 5 min se ha aplicado en todos los programas de PCR, por lo que en adelante no se hará mención a él. La amplificación de la región MsH42 a partir de DNA genómico de humanos y primates se efectuó con las parejas de oligonucleótidos PS1/PS2, P1/P2 o PC1/PC2 (Tabla 3) en función de la especie. Las condiciones de amplificación para estas tres parejas fueron idénticas a las ya descritas para PS1/PS2, pero con una temperatura de

**Tabla 3.** Secuencia de los oligonucleótidos empleados en PCR

| Nombre        | Secuencia                  |
|---------------|----------------------------|
| <b>PS1</b>    | CTGCAGCAATGGACTCAAAA       |
| <b>PS2</b>    | CTGCAGACTCCAAATCCTAA       |
| <b>P1</b>     | CTTGGGCACTCTAGGACACC       |
| <b>P2</b>     | CACAGCTCTGGCTACAAGAG       |
| <b>PC1</b>    | GGGCAGTGTTGAGAGTGAGC       |
| <b>PC2</b>    | TATCTTCATGAACTCACACT       |
| <b>E5.1</b>   | TTTGCTCTGGGATTTAAGGC       |
| <b>E5.2</b>   | CAACAAGCCATTGGAGCCAT       |
| <b>E6.1</b>   | ATCAAGGACGTTGTGGGCTA       |
| <b>E6.2</b>   | TTGCAGTCTTGCCTGGGCTT       |
| <b>Ps1Sma</b> | CCCGGGCTGCAGCAATGGACTCAAAA |
| <b>Ps2Sma</b> | CCCGGGCTGCAGACTCCAAATCCTAA |
| <b>PF1</b>    | GCCTCTCCCAGCTCTCCCAGCCTT   |

Todos los oligonucleótidos se muestran en dirección 5'-3'.

hibridación de 62°C para P1/P2 y de 56°C para PC1/PC2. Además, con esta última pareja el tiempo de extensión se redujo a 40 s. Los exones 13 y 14 (anteriormente 5 y 6) del gen Q8WUJ3, que flanquean al intrón donde se encuentra la región MsH42, se amplificaron empleando los cebadores E5.1/E5.2 y E6.1/E6.2, respectivamente (Tabla 3). Las condiciones de amplificación del exón 13 fueron 95°C (30 s), 58°C (15 s), 72°C (20 s) durante 29 ciclos, siendo iguales para el exón 14, con la salvedad de la temperatura de hibridación, que fue de 60°C. Finalmente, las amplificaciones con los cebadores E5.1/PS2 se realizaron a 95°C (1 min), 54°C (40 s), 72°C (2 min) durante 29 ciclos.

Las condiciones de amplificación de la región MsH42 clonada en pRep42 con los cebadores PS1Sma y PS2Sma, fueron idénticas a las empleadas para la pareja de cebadores PS1 y PS2. Los productos de PCR de la región MsH42 en los distintos primates se clonaron en el vector pGEM-T Easy (Promega) para su posterior secuenciación, de forma análoga a lo descrito para la generación del plásmido pPSSma.

### ***Secuenciación del DNA***

La secuenciación de los fragmentos de DNA clonados en el plásmido pGEM-3Z fue realizada de manera manual o automática empleando los cebadores universales de M13 (Sambrook et al., 2001). El DNA se preparó por el protocolo de maxipreparado o de minipreparado especial para secuenciación. Para la secuenciación manual se empleó el T7 *Sequenase* Quick-Denature Plasmid Sequencing Kit v2.0 (Amersham Biosciences), siguiendo las indicaciones del fabricante. Los productos de la secuenciación se separaron en geles desnaturizantes (7M urea) de poliacrilamida (19:1) al 8% en

1X TBE. La electroforesis se llevó a cabo a 2000 V durante 90 min a 55°C, tras lo cual el gel se lavó y fijó con dos baños de 15 min en una solución de 10% ácido acético y 10% etanol. Posteriormente el gel se secó en un horno a 80°C durante 45 min y se autorradiografió sobre películas Hyperfilm MP (Amersham Biosciences).

Para la secuenciación automática cíclica se utilizó el dGTP BigDye Terminator v3.0 Sequencing Kit (Applied Biosystems), siguiendo las especificaciones del fabricante, pero realizando las reacciones a la mitad del volumen final indicado. El DNA plasmídico se preparó del mismo modo que para la secuenciación manual, empleándose los cebadores universales de M13 en las reacciones. La secuenciación automática se realizó además directamente con productos de PCR. Para ello, las reacciones de PCR se trataron enzimáticamente con exonucleasa I y fosfatasa alcalina de camarón (Exo/Sap-It, USB) para eliminar el resto de dNTPs y cebadores. Las condiciones de secuenciación fueron 96°C (10 s), 68°C (2 min) durante 25 ciclos, en todos los casos, realizándose la incubación en un termociclador PTC-200 (MJ Research). A continuación, los productos de la reacción fueron purificados mediante precipitación con 2 vol de etanol 96% y 0.1 vol de acetato sódico 3M pH 4.6 y analizados en un secuenciador ABI PRISM 377 DNA Sequencer (Perkin Elmer). El análisis de los datos de secuenciación se llevó a cabo empleando el paquete de software BioEdit (Hall, 1999), disponible en [www.mbio.ncsu.edu/BioEdit/bioedit.html](http://www.mbio.ncsu.edu/BioEdit/bioedit.html)

### ***Ensayos de DNasaI footprinting***

El *footprinting* empleando DNasaI es una técnica que permite identificar la secuencia de un fragmento de DNA, marcado terminalmente en una de sus

hebras, que es reconocida por una proteína de modo específico (Galas & Schmitz, 1978). Se fundamenta en que la unión de una proteína a un fragmento de DNA bicatenario protege a la secuencia a la cual se une de la acción de la DNasaI. Por el contrario, en ausencia de dicha protección, el enzima introducirá mellas aleatoriamente por todo el fragmento de DNA. La desnaturalización y posterior separación mediante electroforesis de los fragmentos de diferentes tamaños originados por la acción de la DNasaI, permite comparar el patrón de fragmentos obtenido en ausencia y en presencia de proteína. Cada una de las bandas que se obtiene representa una población de fragmentos de DNA monocatenario generados por un corte en un nucleótido determinado. Por lo tanto, la desaparición de un subconjunto de bandas en la autorradiografía implica que una proteína se ha unido a una secuencia determinada de la sonda, la cual queda protegida de la acción de la DNasaI, evitando así la generación de los fragmentos correspondientes.

En estos ensayos se utilizaron como sondas dos fragmentos de la región del alelo largo del locus MsH42. Las sondas FP1a y FP1b se obtuvieron partiendo de la amplificación por PCR de la región MsH42 del alelo largo con los cebadores PS1 y P2, en las mismas condiciones que para la pareja P1/P2 (ver *Amplificaciones por PCR*). El producto de PCR fue digerido con el enzima *DdeI* para obtener un fragmento de 283 pb que contiene la región flanqueante 5' y las 16 primeras repeticiones del minisatélite. Este fragmento fue purificado y empleado como DNA molde para una amplificación por PCR con los cebadores PS1 y PF1, realizada en las mismas condiciones que para la pareja PS1/PS2 (ver *Amplificaciones por PCR*). El producto obtenido, un fragmento de 214 pb, se clonó en ambas direcciones

en el vector pGEM-T (Promega). Los plásmidos generados, que denominamos pFP1a (orientación 5'→3') y pFP1b (orientación 3'→5'), fueron digeridos con *SacII* y *NotI*, para producir los fragmentos FP1a y FP1b, respectivamente. Ambos fragmentos fueron marcados utilizando Klenow con [ $\alpha$ -<sup>32</sup>P]-dCTP y [ $\alpha$ -<sup>32</sup>P]-dGTP (6000 Ci/mmol). El protocolo seguido fue el siguiente: 1  $\mu$ g de DNA fue incubado con 5 U de enzima Klenow en 20  $\mu$ l de volumen de reacción (50 mM Tris-HCl pH 7.2, 10 mM MgSO<sub>4</sub>, 0.1 mM DTT, 20  $\mu$ g/ml BSA), que contenía 20  $\mu$ Ci de [ $\alpha$ -<sup>32</sup>P]-dCTP de 6000 Ci/mmol (Amersham Biosciences), 20  $\mu$ Ci de [ $\alpha$ -<sup>32</sup>P]-dGTP de 6000 Ci/mmol (Amersham Biosciences) y 50  $\mu$ M de cada uno de los dNTPs no marcados. La mezcla fue incubada durante 15 min a 30°C y, una vez finalizada la reacción de marcaje, se añadieron 2  $\mu$ l de una mezcla 5 mM de todos los dNTPs, prosiguiendo la incubación durante 5 min a 30°C para asegurar que los extremos fuesen rellenados completamente. La reacción se detuvo mediante una incubación de 10 min a 75°C. Las sondas fueron purificadas con el QIAquick Nucleotide Removal Kit (QIAGEN), según las indicaciones proporcionadas por el fabricante, presentando una actividad específica de  $3 \times 10^8$  cpm/ $\mu$ g. Como consecuencia del marcaje realizado, en la sonda FP1a quedaría marcada la cadena 5'→3', mientras que en la sonda FP1b quedaría marcada la 3'→5'. Por su parte, las sondas FP2a y FP2b fueron generadas a partir del fragmento de 260 pb, que comprende las 9 repeticiones del extremo 3' del alelo largo del minisatélite y la secuencia flanqueante 3'. Este fragmento fue obtenido partiendo de la digestión con *DdeI* del producto de amplificación con los cebadores P1/P2 de la región MsH42 del alelo largo. Posteriormente, se incubó 1  $\mu$ g del mismo con *Taq* DNA polimerasa en las mismas condiciones de una

reacción de PCR, pero incubando durante 30 min a 72°C. Dado que la *Taq* DNA polimerasa añade una adenina en los extremos 3' de los fragmentos de DNA producidos, pudimos clonar el fragmento en el vector pGemT, en ambas orientaciones. El resto del proceso para producir las sondas FP2a y FP2b fue similar al descrito para la preparación de las sondas FP1a y FP1b. Estas sondas fueron también purificadas con el QIAquick Nucleotide Removal Kit (QIAGEN), según las indicaciones proporcionadas por el fabricante, presentando una actividad específica de  $2 \times 10^8$  cpm/ $\mu$ g.

Las reacciones de footprinting fueron llevadas a cabo siguiendo el protocolo del SureTrack Footprinting Kit (Amersham Biosciences). Las reacciones tuvieron lugar en 50  $\mu$ l de tampón de unión (10mM Tris-HCl pH 7.5, 4% glicerol, 50 mM NaCl, 25 mM MgCl<sub>2</sub>, 0.5 mM DTT), conteniendo 50  $\mu$ g de extracto proteico nuclear, 3  $\mu$ g de poli(dI-dC)•poli(dI-dC) y 250 ng de DNA de timo de ternera sonificado. La reacción se incubó durante 7 min a temperatura ambiente y después se le añadieron 0.25 ng (~50000 cpm) de sonda, manteniéndola durante otros 30 min a temperatura ambiente. A continuación, se añadió DNasaI y se incubó a temperatura ambiente. La cantidad de enzima y el tiempo de tratamiento con DNasaI varían dependiendo del extracto proteico nuclear empleado (hígado de rata, 0.1 U, 1 min; cerebro de rata, 0.1 U, 1min; testículo de rata, 1.5 U, 1.5 min; sin extracto, 0.1 U, 1.5 min). Las reacciones fueron detenidas añadiendo 140  $\mu$ l de solución de parada para la DNasaI (192 mM acetato sódico, 32 mM EDTA, 0.14% SDS, 64  $\mu$ g/ml de RNA de levadura). A continuación, añadimos a cada reacción 2  $\mu$ g de proteinasa K y las incubamos durante 30 min a 42°C. Finalmente, fueron extraídas y

precipitadas, para ser resuspendidas en 3  $\mu$ l de tampón de carga. En cada experimento se incluyeron dos controles: uno con la sonda pero sin extracto y otro con la sonda incubada con BSA. Con cada sonda ensayada preparamos una reacción de secuenciación química de Maxam y Gilbert (Maxam & Gilbert, 1977) para las bases púricas (G+A), que al ser cargada en el mismo gel que las reacciones de footprinting de esa sonda, nos permite determinar la zona de la sonda que está siendo protegida de la digestión con DNasaI. El análisis de los experimentos se realizó mediante electroforesis, bajo las mismas condiciones descritas para los geles de secuenciación manual.

### Análisis filogenéticos

Las secuencias de DNA empleadas en los análisis filogenéticos se alinearon usando el programa Clustal W, empleando los alelos HUM(S), CHI1 y GOR1 como representantes de la región MsH42 en la especie humana, chimpancé y gorila, respectivamente. Estos análisis están basados en los métodos de *neighbour-joining* (NJ), máxima parsimonia (MP) y máxima verosimilitud (ML), disponibles en el programa PAUP\* versión 4 (Swofford, 1998). La homogeneidad de los conjuntos de datos de secuencia se evaluó con el test de incongruencia de diferencias de longitud. El modelo que mejor se ajusta a la topología estándar del árbol de los primates fue seleccionado mediante el test de la ratio de verosimilitud disponible en Modeltest versión 3.0. Todos estos métodos analíticos se encuentran revisados por Thornton & DeSalle (2000). Los valores de verosimilitud para la topología estándar de primates y de las topologías observadas se compararon con el test paramétrico de Kishino-Hasegawa (Kishino & Hasegawa, 1989).



# *Resultados y Discusión*



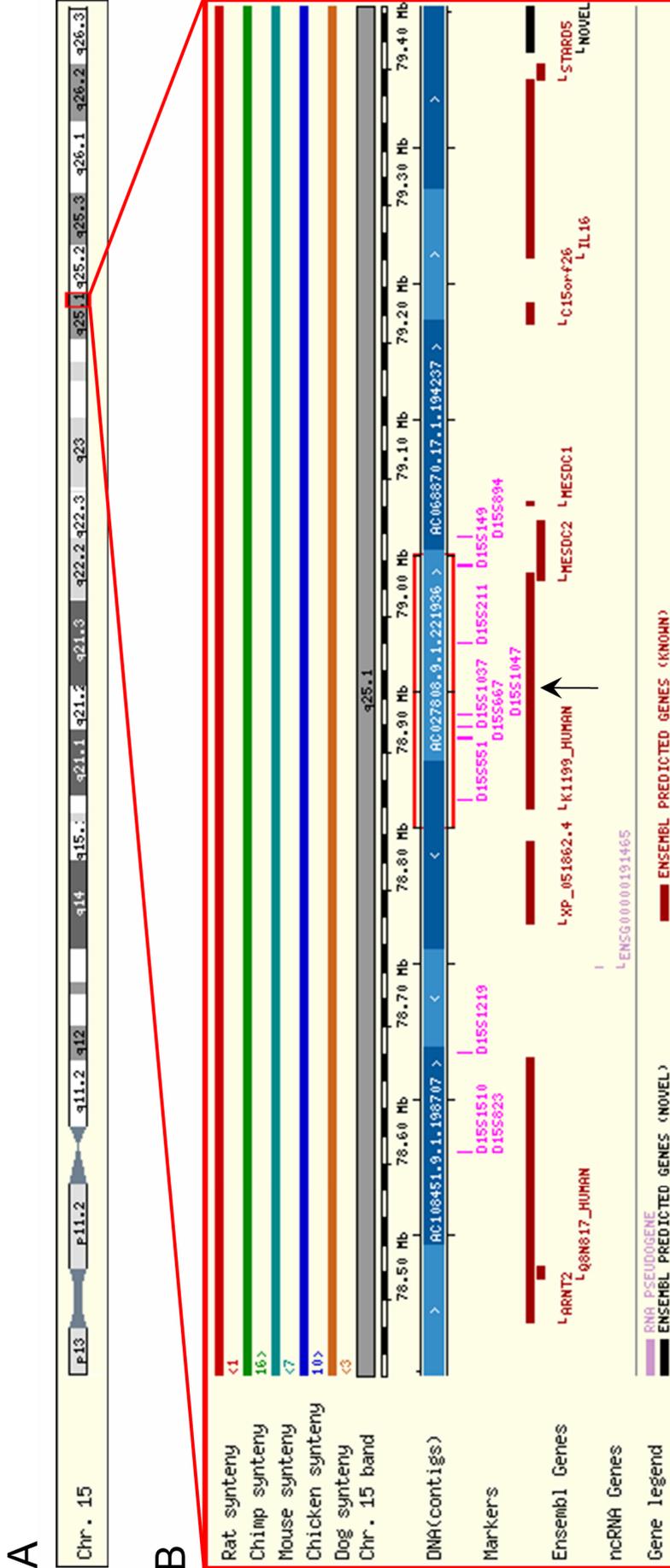
## Localización del minisatélite MsH42 en el genoma humano

La secuenciación de los genomas de diversas especies modelo y, en especial, la publicación del primer borrador de la secuencia del genoma humano en el año 2001 (International Human Genome Sequencing Consortium, 2001; Venter et al., 2001) han provocado tal aluvión de datos que el conocimiento derivado de su análisis está marcando el comienzo de una nueva era en la Biología Molecular. Además, toda esta información ha abierto las puertas al estudio de los genomas a gran escala, gracias al desarrollo de poderosos programas informáticos de búsqueda y análisis de secuencias (revisado en Ureta-Vidal et al., 2003). Desde la publicación del primer borrador del genoma humano, se ha ido completando y perfeccionando la secuencia de cada uno de los cromosomas (Stein, 2004). Por este motivo, las bases de datos donde se recoge toda esta información son actualizadas periódicamente respecto a versiones anteriores. Así, determinadas regiones especialmente difíciles de secuenciar o de interpretar, como pueden ser las duplicaciones de segmentos grandes, las repeticiones en tándem o las regiones heterocromatínicas (She et al., 2004), han sufrido cambios relativamente importantes a lo largo de las distintas actualizaciones del genoma humano (International Human Genome Sequencing Consortium, 2004).

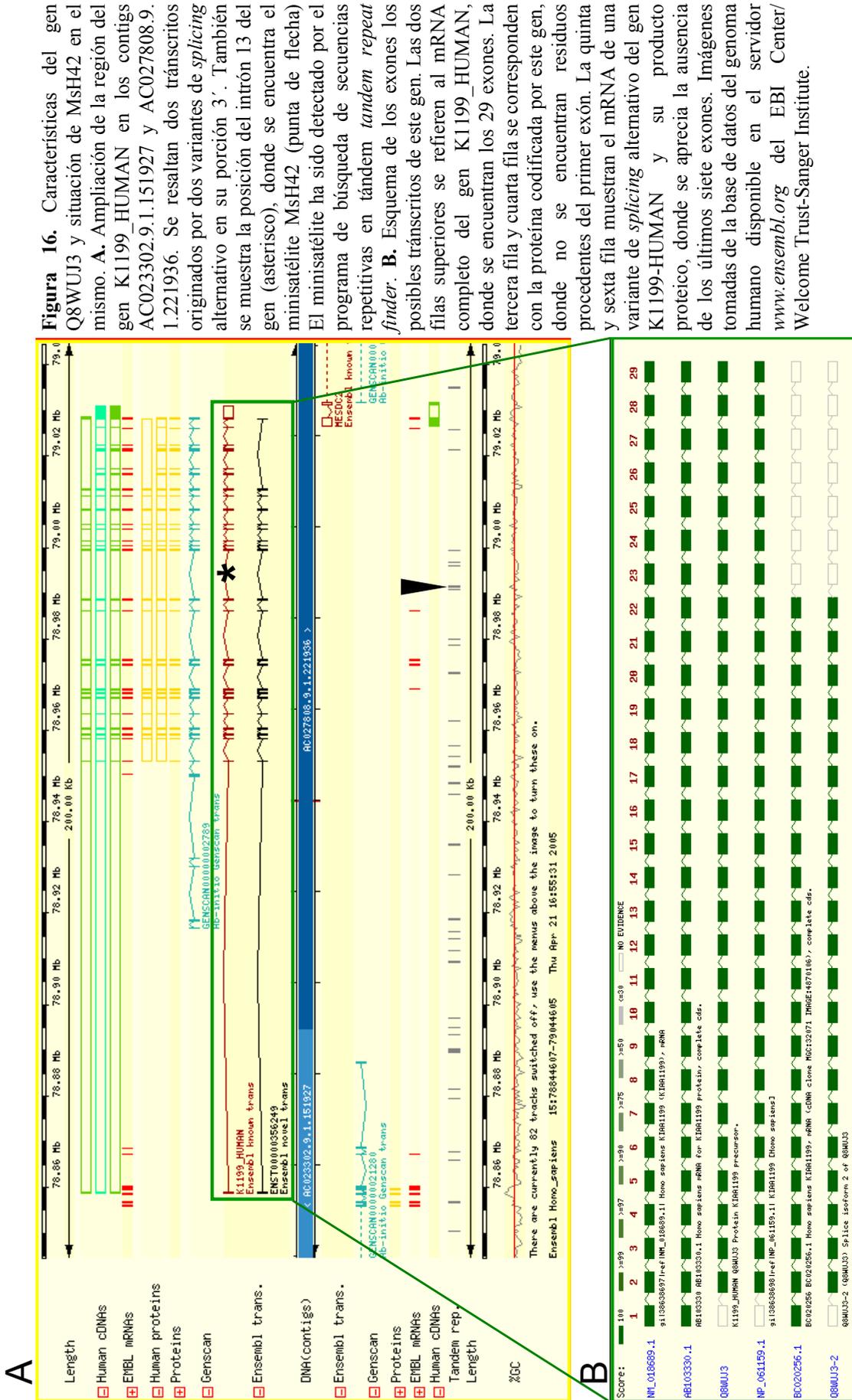
Cuando se publicó el primer borrador de la secuencia del genoma humano en 2001, realizamos la localización cromosómica del minisatélite MsH42 a través del servidor [www.ensembl.org](http://www.ensembl.org) del EBI Center/Welcome Trust-Sanger Institute. El resultado reveló que MsH42 se hallaba localizado en el intrón 5 del gen Q9ULM1 (Boán et al., 2002). Este gen, cuya función era desconocida,

estaba situado en la banda 25.1 correspondiente a la región subtelomérica del brazo largo del cromosoma 15 (contig AC027808.5.1.207303) y estaba constituido por 21 exones.

Durante el desarrollo de esta Tesis, hemos comprobado de forma periódica la situación de MsH42 en el genoma humano. La localización cromosómica de MsH42 no mostró cambio alguno hasta el año 2004. Sin embargo, en la última versión del genoma humano comprobamos que el gen Q9ULM1 había sido redefinido, aunque el entorno del minisatélite MsH42 permanecía inalterado (Figura 15). Este gen ahora se denomina Q8WUJ3 (K1199\_HUMAN) y consta de 29 exones, habiéndose detectado el cDNA completo del mismo (Figura 16). El aumento del número de exones ha provocado una reenumeración de los previamente descritos, estando MsH42 situado ahora en el intrón 13, que se corresponde con el intrón 5 del borrador inicial del genoma humano. Los exones que flanquean al intrón donde se encuentra el minisatélite son también los mismos, pero adecuando su numeración y pasando a ser los exones 13 y 14 respectivamente. El gen Q8WUJ3 presenta un tamaño considerable (~ 172 kb), estando su secuencia repartida entre los contigs AC023302.9.1.151927 y AC027808.9.1.221936, encontrándose MsH42 en el segundo de ellos. La diferencia observada con las versiones anteriores del genoma humano parece deberse a que inicialmente sólo se había detectado un cDNA incompleto de este gen. Este hecho no resulta extraño si tenemos en cuenta que el producto de este gen, KIAA1199, posee 1331 aminoácidos (~ 150 kDa), originados a partir de un



**Figura 15.** Localización del gen Q8WUJ3 en el contexto del cromosoma 15. **A.** Situación del gen Q8WUJ3 (K1199\_HUMAN) en el cromosoma 15. **B.** Localización de este gen (flecha) en la banda cromosómica 15q25.1. Se muestra la sintenia con algunos de los organismos modelos. Este gen se encuentra en todos los mamíferos analizados, pero no hemos podido localizarlo en el genoma del pollo. Imágenes tomadas de la base de datos del genoma humano disponible en el servidor [www.ensembl.org](http://www.ensembl.org) del EBI Center/Wellcome Trust-Sanger Institute.



mRNA de más de 7 kb. Como puede observarse en la Figura 16, el primer exón dista considerablemente del segundo, dando lugar a un intrón de más de 94 kb, lo que representa casi el 55% del transcrito primario. Curiosamente, este primer exón no contiene secuencia codificante, ya que el codón de iniciación está en el segundo. Existen dos formas diferentes de mRNA para este gen, originadas por *splicing* alternativo entre los intrones más próximos al extremo 3' del mismo (Figura 16).

El análisis de las bases de datos del genoma de distintas especies reveló la existencia de ortólogos del gen Q8WUJ3

en chimpancé, perro, rata y ratón, lo que apunta a una importante conservación del mismo, al menos en mamíferos ([www.ensembl.org](http://www.ensembl.org)). Actualmente se piensa que su función esté relacionada con la audición, ya que el gen Q8WUJ3 parece expresarse específicamente en el oído interno, concretamente en la cóclea y el vestíbulo (Abe et al., 2003a). Además, algunas mutaciones en este gen han sido relacionadas con sordera no sindrómica (Abe et al., 2003b).

Esta última búsqueda en el genoma humano confirma que la localización de MsH42 es intrónica, pero todavía podrían existir cambios concernientes al contexto cromosómico donde se ubica.

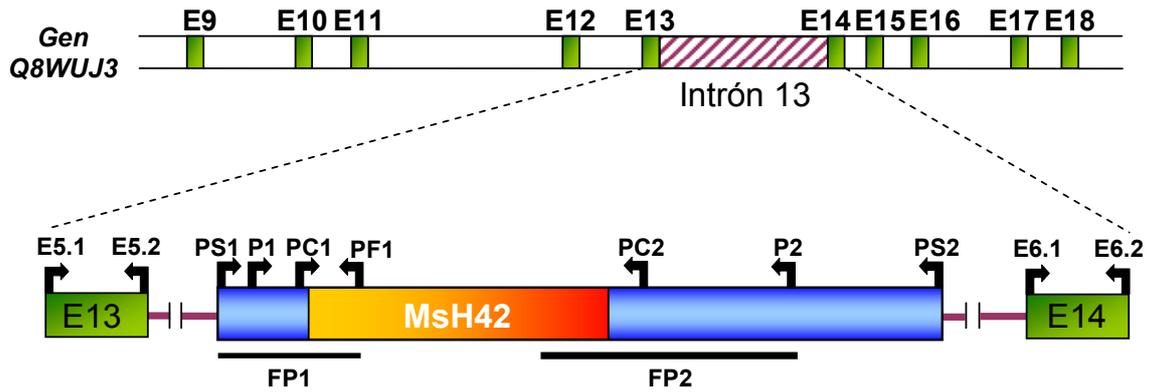
## Nacimiento e historia evolutiva de MsH42

Uno de los desafíos más excitantes de la Biología Humana es comprender de qué manera se ha ido construyendo nuestro genoma a lo largo de la evolución. Entre las principales maneras de alcanzar este objetivo, se encuentra la comparación del genoma humano con el de otras especies. Hasta hace poco, la escasa disponibilidad de genomas secuenciados completamente limitaba este tipo de comparaciones a especies como el ratón, el pez cebra, la mosca del vinagre o el nematodo *Caenorhabditis elegans*, todas ellas con una relación relativamente lejana con respecto al ser humano. Estas comparaciones, no obstante, son útiles para el estudio de secuencias conservadas durante largos períodos de tiempo, comunes a todos los mamíferos, todos los vertebrados o incluso todos los animales. Sin embargo, poco nos pueden contar acerca de la reciente evolución de los primates y, consecuentemente, de las diferencias del ser humano con el resto de este grupo (Enard & Pääbo, 2004). Con la finalización del proyecto de secuenciación del genoma del chimpancé, nuestro pariente más cercano entre los primates, se abre una nueva vía para establecer en qué nos parecemos y en qué nos diferenciamos esencialmente del resto de los primates.

Además de las grandes comparaciones a nivel de genoma completo, el estudio de la evolución de pequeñas partes del genoma también pueden ser de gran utilidad a la hora de estudiar la dinámica de los mismos (Pääbo, 2003). En este sentido, muchas secuencias de DNA repetitivo, y entre ellas los minisatélites, poseen unos mecanismos evolutivos característicos que difieren de los de las secuencias de copia única (Charlesworth et al., 1994). Puesto que en muchos organismos el DNA repetitivo constituye una fracción sustancial del genoma, la comprensión de los procesos que

gobiernan su evolución puede ayudarnos a entender con mayor claridad cómo se ha construido y desarrollado nuestro genoma hasta hoy.

En lo que respecta a los minisatélites, se han propuesto varios mecanismos para explicar la generación de cambios en este tipo de secuencias, tales como el entrecruzamiento desigual, el deslizamiento de cadena durante la replicación (*slippage*), las mutaciones puntuales, la generación de inserciones/delecciones o la conversión génica (Olsen, 1999; Vergnaud & Denoeud, 2000). No obstante, su origen filogenético y dinámica evolutiva siguen siendo relativamente desconocidos. Existe un estudio evolutivo en primates sobre los minisatélites hipervariables MS1 y MS32 donde se muestra que su grado de polimorfismo ha fluctuado notablemente en un período relativamente corto de tiempo (Gray & Jeffreys, 1991). Otros trabajos han versado sobre los procesos que dan lugar al nacimiento de minisatélites, como el realizado por Taylor & Breden (2000) en el teleosteo *Poecilia reticulata* y que sugiere que son sucesos de tipo *slippage* entre unidades similares no contiguas los que permiten la aparición de este tipo de secuencias y cuyos remanentes son detectables como unidades de repetición truncadas o imperfectas que flanquean al minisatélite. Una conclusión similar a esta había sido previamente obtenida por Haber & Louis (1998), aunque dichos autores proponían el entrecruzamiento desigual como una alternativa adicional al *slippage* en la generación de los minisatélites. En este sentido, se ha descrito una secuencia de tipo microsatélite en el gen de la  $\eta$ -globina cuya evolución obedece claramente a procesos de mutación y *slippage*, como principales causantes de su expansión (Messier et al., 1996) y que, aunque



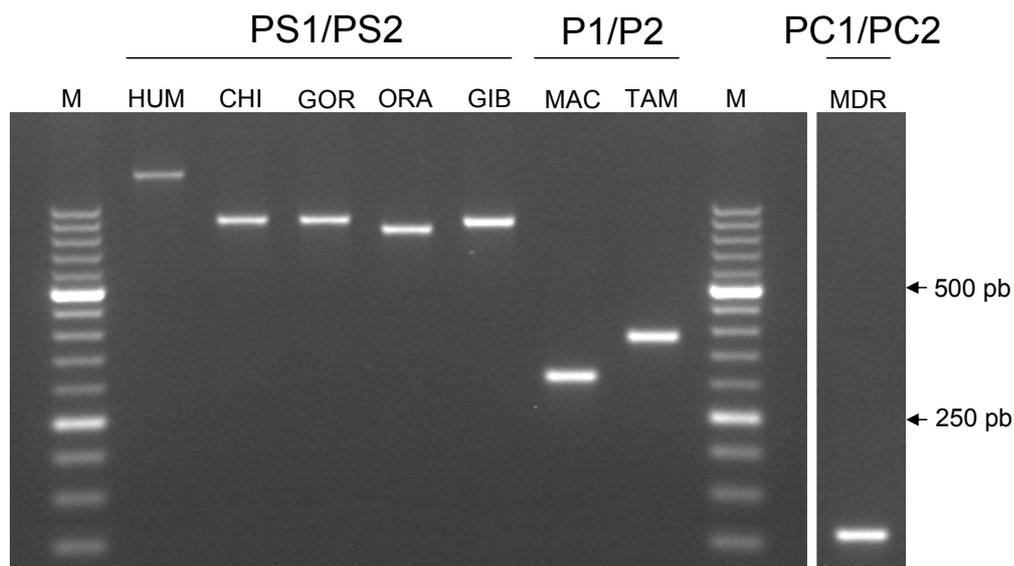
**Figura 17.** Localización de los cebadores de PCR empleados en el análisis de MsH42. Se muestra la ubicación de estos cebadores (flechas) en el gen Q8WUJ3. En la parte superior se destaca la zona del gen donde se encuentra el intrón 13 que incluye a la región MsH42. E9-E18 indican los exones del gen. La caja azul señala la región MsH42, que comprende al minisatélite MsH42 (caja naranja). Se indican además las regiones que ocupan los fragmentos FP1 y FP2, empleados en los experimentos de DNasal *footyping*.

pertenece a un grupo de secuencias diferentes de los minisatélites, puede ilustrar los procesos responsables de la variación de las secuencias repetidas en tándem. Sin embargo, se sabe poco acerca de la evolución de los minisatélites poco variables, como es el caso de MsH42. Existe un estudio que emplea una secuencia relativamente parecida a la nuestra, el minisatélite del gen de la tiopurina metiltransfera, con una unidad repetitiva de 17-18 pb (Alves et al., 2002). Este trabajo hace un análisis poblacional detallado en gorila, chimpancé y humanos, concluyendo que dentro del minisatélite las distintas unidades repetitivas pueden ser más o menos tendentes a sufrir cambios y condicionando así la evolución de todo el conjunto. Sin embargo, este trabajo no profundiza en el origen del minisatélite en cuestión y su origen evolutivo permanece incierto. En esta Tesis, nos propusimos investigar la existencia del minisatélite MsH42 en otros primates y en las especies modelo cuyo genoma ha sido secuenciado para, de este modo, poder establecer su posible origen (nacimiento) e historia evolutiva.

### Origen evolutivo de MsH42

La búsqueda de la región MsH42 (minisatélite y secuencias adyacentes) en

otras especies de primates se realizó mediante amplificación de muestras de DNA genómico por PCR, utilizando las parejas de cebadores PS1/PS2, P1/P2 y PC1/PC2 (Tabla 3), que se encuentran localizados en las secuencias flanqueantes de MsH42 (Figura 17). Los cebadores PS1/PS2 produjeron productos de amplificación en chimpancé, gorila, orangután y gibón (Figura 18). En el caso de gorila, los productos de la PCR fueron muy débiles, probablemente debido cambios en la secuencia de alguno de los cebadores, por lo que debió efectuarse una segunda ronda de amplificación. En esta reamplificación, se empleó como sustrato el DNA purificado de la banda del gel de agarosa que habíamos obtenido inicialmente. Tanto en chimpancés como en gorilas se detectó la presencia de bandas de tamaños diferentes, indicando que la región MsH42 presenta variabilidad en estas especies. Los cebadores PS1 y PS2 no produjeron amplificación cuando se empleó DNA genómico de monos del Viejo Mundo (macaco y mandril), del Nuevo Mundo (tamarindo) o de prosimios (*Eulemur*, *Lemur* y *Varecia*). Con estas especies se probó la pareja de oligonucleótidos P1/P2, localizados más cerca del minisatélite, pero que todavía permiten la amplificación de una porción



**Figura 18.** Amplificación de la región MsH42 en primates. Se muestra el análisis electroforético en gel de agarosa al 2% de los productos de amplificación de la región MsH42 en diversos primates (HUM, humano; CHI, chimpancé; GOR, gorila; ORA, orangután; GIB, gibón; MAC, macaco; TAM, tamarindo; MDR, mandril). En la parte superior se especifican las distintas parejas de cebadores empleadas en la PCR. Hay que señalar que las diferencias en tamaño se deben parcialmente al uso de distintas parejas de cebadores. La imagen de la amplificación de mandril procede de un experimento distinto a los demás. En el caso de humano, chimpancé y gorila, las bandas mostradas se corresponden con los alelos HUM(L), CHI1 y GOR1, respectivamente. La amplificación de macaco fue obtenida de un individuo de la especie *Macaca mulatta* y fue idéntica a la obtenida con individuos de la especie *Macaca nigra*. M, marcador XIII, con bandas cada 50 pb (Roche Applied Science).

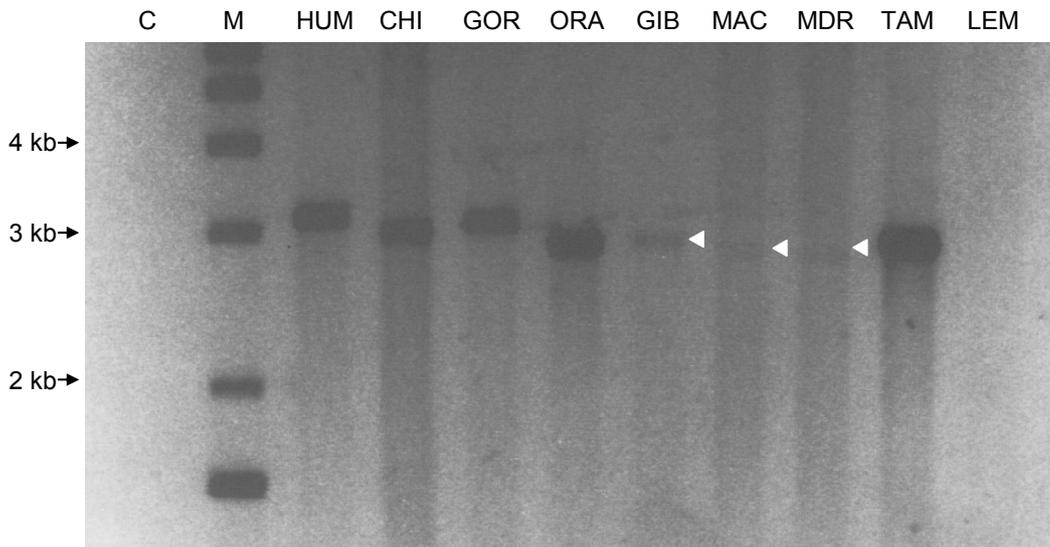
importante de las secuencias flanqueantes del mismo. Con estos cebadores obtuvimos amplificaciones para macaco y tamarindo (Figura 18), pero no fueron capaces de producir amplificación a partir del DNA de mandril y lemures. Los tres macacos analizados (dos individuos de la especie *Macaca nigra* y uno de *Macaca mulatta*) generaron bandas del mismo tamaño, por lo que aparentemente este locus puede ser monomórfico en este género. El DNA de mandril pudo ser amplificado con los cebadores PC1/PC2 (Figura 18), diseñados a partir de dos zonas especialmente conservadas de la región MsH42 y que comentaremos más adelante. Tampoco con estos cebadores se obtuvo amplificación a partir del DNA genómico de lemures. Todos los productos resultantes de esta serie de experimentos de PCR se clonaron en pGEM-T Easy para su posterior secuenciación. Para asegurarnos de que

las amplificaciones que estábamos observando en los diferentes primates se correspondían con una secuencia homóloga al minisatélite humano, es decir que comparte el mismo origen evolutivo, se realizaron una serie de experimentos de PCR en los que usaron los cebadores E5.1/PC2 para amplificar las muestras de DNA genómico. Dado que E5.1 se localiza en el extremo 5' del exón 13 del gen Q8WUJ3 y PC2 en la secuencia flanqueante 3' del minisatélite, una amplificación positiva del tamaño esperado indicaría que el minisatélite que hemos detectado en estos primates se encuentra ubicado en una posición equivalente a la que ocupa en humanos. Los resultados de estos experimentos confirmaron la localización de MsH42 en el intrón equivalente del gen Q8WUJ3 (Figura 19). En la Tabla 4 se resumen los tamaños de los fragmentos de PCR obtenidos para cada especie.

La imposibilidad para amplificar secuencias relacionadas con la región MsH42 en lemures indicaba que MsH42 podría ser exclusivo de primates. Entonces, para averiguar si existía una secuencia similar a MsH42 en los genomas de especies filogenéticamente más distantes, realizamos una búsqueda *in silico* en las bases de datos de los genomas de las siguientes especies: ratón (*Mus musculus*), rata (*Rattus norvegicus*), pollo (*Gallus gallus*), sapo (*Xenopus laevis*), pez cebra (*Danio rerio*) y la mosca del vinagre (*Drosophila melanogaster*), así como en los borradores de los genomas de la vaca (*Bos taurus*) y la zarigüeya (*Monodelphis domestica*). No pudimos encontrar ninguna secuencia ortóloga a MsH42 en ninguno de estos genomas. No obstante, sí hallamos evidencia de la

existencia de un gen ortólogo a Q8WUJ3, aunque solamente en mamíferos (datos no mostrados).

Este conjunto de datos, i) la presencia en mamíferos de los dos exones del gen Q8WUJ3 que flanquean a la región MsH42 en humanos, ii) la ausencia del minisatélite en los genomas de mamíferos no primates analizados, iii) la incapacidad para obtener una amplificación de la región MsH42 en lemures con cualquiera de las combinaciones de cebadores empleadas y iv) la existencia del minisatélite en todos los platirrininos y catarrinos ensayados, sugiere que el origen de esta secuencia tuvo lugar durante una fase temprana de la evolución de los primates, antes de la divergencia entre monos del Nuevo y del Viejo Mundo.



**Figura 19.** Demostración de la localización de MsH42 y sus ortólogos para diversos primates en el gen Q8WUJ3. Se muestra el análisis electroforético en gel de agarosa al 1% de los productos de amplificación obtenidos con el DNA genómico de diversos primates empleando los cebadores E5.1 y PC2. La imagen muestra el negativo de una fotografía del gel para facilitar la observación de los productos menos intensos. En el caso de gibón, macaco y mandril, los productos de amplificación fueron muy débiles (puntas de flecha blancas), probablemente debido a la existencia de mutaciones en la secuencia de alguno de los cebadores. C, reacción control sin DNA molde; M, marcador de 1kb (Invitrogen). Los acrónimos para las diferentes especies son los mismos que en la Figura 18, con excepción de LEM, lemur (*Varecia variegata*).

**Tabla 4.** Alelos de MsH42 en primates

| Organismo | Alelos | Tamaño (pb) | Unidades repetitivas | Cebadores | Variantes de las unidades repetitivas <sup>a</sup> |    |    |    |    |    |   |    |    |    |    |   |    |    |    |    |    |    |   |
|-----------|--------|-------------|----------------------|-----------|--|----|----|----|----|----|---|----|----|----|----|---|----|----|----|----|----|----|---|
|           |        |             |                      |           | A  | A1 | A2 | A3 | A4 | A5 | B | B1 | B2 | B3 | B4 | C | C1 | C2 | C3 | C4 | C5 | C6 | D |
| Humano    | HUM(L) | 863         | 36                   | PS1/PS2   | +  | +  | +  | +  | +  | +  | + | +  | +  | +  | +  | + | +  | +  | +  | +  | +  | +  | + |
|           | HUM(M) | 806         | 30                   | "         | +  | +  | +  | +  | +  | +  | + | +  | +  | +  | +  | + | +  | +  | +  | +  | +  | +  | + |
|           | HUM(S) | 758         | 25                   | "         | +  | +  | +  | +  | +  | +  | + | +  | +  | +  | +  | + | +  | +  | +  | +  | +  | +  | + |
| Chimpancé | CHI1   | 728         | 22                   | "         | +  | +  | +  | +  | +  | +  | + | +  | +  | +  | +  | + | +  | +  | +  | +  | +  | +  | + |
|           | CHI2   | 690         | 18                   | "         | +  | +  | +  | +  | +  | +  | + | +  | +  | +  | +  | + | +  | +  | +  | +  | +  | +  | + |
| Gorila    | GOR1   | 742         | 23                   | "         | +  | +  | +  | +  | +  | +  | + | +  | +  | +  | +  | + | +  | +  | +  | +  | +  | +  | + |
|           | GOR2   | 800         | 29                   | "         | +  | +  | +  | +  | +  | +  | + | +  | +  | +  | +  | + | +  | +  | +  | +  | +  | +  | + |
|           | GOR3   | 1010        | 51                   | "         | +  | +  | +  | +  | +  | +  | + | +  | +  | +  | +  | + | +  | +  | +  | +  | +  | +  | + |
| Orangután | ORA    | 657         | 13                   | "         | +  | +  | +  | +  | +  | +  | + | +  | +  | +  | +  | + | +  | +  | +  | +  | +  | +  |   |
| Gibón     | GIB    | 680         | 16                   | "         | +  | +  | +  | +  | +  | +  | + | +  | +  | +  | +  | + | +  | +  | +  | +  | +  | +  | + |
| Macaco    | MAC    | 308         | 7                    | P1/P2     | +  | +  | +  | +  | +  | +  | + | +  | +  | +  | +  | + | +  | +  | +  | +  | +  | +  | + |
| Mandrill  | MDR    | 145         | 7                    | PC1/PC2   | +  | +  | +  | +  | +  | +  | + | +  | +  | +  | +  | + | +  | +  | +  | +  | +  | +  | + |
| Tamarindo | TAM    | 386         | 19                   | P1/P2     | +  | +  | +  | +  | +  | +  | + | +  | +  | +  | +  | + | +  | +  | +  | +  | +  | +  | + |

NOTA.- <sup>a</sup>Código de las variantes de las unidades repetitivas. **A:** TGGGAGAGGC, **A1:** TGACAGAGGC, **A2:** TGGGAAAGGC, **A3:** TGGGAGAGAC, **A4:** TGGGAAAGAC, **A5:** TGGGAAAGGT, **B:** TGGGATTGC, **B1:** TGTGATTGC, **B2:** TGAGATTGC, **B3:** TGGAATTGC, **B4:** CGAGATTGC, **C:** TGGGAGAGC, **C1:** TGGGAAAGA, **C2:** TGTGAGAGC, **C3:** TGGGAAAGC, **C4:** TGGGAGAC, **C5:** TAGGAGAGC, **C6:** TGGGAGGGC, **D:** TCAGAGAC, **A6:** TGGGAGAAGG, **A7:** TGGGAGATGG, **A8:** TAGGGGAGGC, **A9:** TGGGAGAGGG, **A10:** TGGGGGAGGC, **C7:** TGGGAGAGC, **C8:** TGGGAGAAC, **C9:** TGGGGAAGC, **C10:** TCGGAGAGA, **C11:** TAGGAGACC, **C12:** TGGAAAGAC, **C13:** TGAGAGAGA. Este código ha sido modificado con respecto al empleado en trabajos previos (Boán et al. 1998, 2002). HUM(L), HUM(M) y HUM(S) son los alelos largo, mediano y corto de MsH42, respectivamente, CHI1, CHI2, GOR1, GOR2, y GOR3 representan distintos alelos encontrados en las especies de chimpancé y gorila.

### Estructura de MsH42 en primates

Para estudiar la organización de la región MsH42 en los primates, los productos de PCR se secuenciaron y compararon entre sí. La Figura 20 muestra el alineamiento de un alelo representativo de cada especie con respecto a la secuencia del alelo corto humano de MsH42. Las diferencias de tamaño de las secuencias mostradas se deben en parte a la pareja de cebadores con las que fueron amplificadas. En la Tabla 4 se muestran el número y tipo de unidades repetitivas que componen el minisatélite en cada una de las especies y también utiliza el código de letras que emplearemos en adelante para representar cada unidad repetitiva de MsH42. Este código se modificó con respecto al anteriormente descrito por Boán et al. (1997, 1998, 2002) (Figura 9). La razón de esta modificación fue para acomodar mejor los nuevos datos evolutivos y agrupar de forma más coherente algunas de las variantes repetitivas. Un análisis detallado de la

organización del minisatélite reveló una serie de características interesantes. Así, el minisatélite de tamarindo, cuyo origen evolutivo se remonta a ~ 40 millones de años, presenta dos variantes repetitivas comunes al minisatélite humano (A y C), intercaladas con otras variantes similares exclusivas de esta especie, constituyendo una secuencia que aunque guarda cierta semejanza con el minisatélite MsH42 humano, muestra una organización interna claramente diferente. Esta variante del minisatélite pudo originarse a partir de una secuencia progenitora que contenía las unidades repetitivas primigenias (A y C) o unas muy similares a ellas, que comenzaron a sufrir mutaciones y a expandirse en una fase temprana de la evolución de los primates. El minisatélite que encontramos en los monos de la familia Cercopithecidae (divergencia con la familia Hominidae estimada en unos 25 millones de años), macaco y mandril, consta de siete unidades repetitivas con la misma organización en ambas especies, exceptuando una transición

G→A en la sexta repetición. Esto concuerda con la estrecha relación filogenética que existe entre estas ambas especies. Además, en estos primates el minisatélite MsH42 presenta por primera vez una característica que se mantendrá constante en todas las demás especies: las unidades repetitivas A y C1 marcan el principio (5') y el final (3') del minisatélite (tomando como orientación 5'-3' la misma que la del gen donde se encuentra incluido). En el gibón (divergencia estimada hace ~ 18 millones de años), el minisatélite adquiere una organización relativamente peculiar, constando de 16 unidades repetitivas, siendo cuatro de ellas distintas a las que se encuentran en humanos. Además, una de estas repeticiones (D) es exclusiva de esta especie. Otra peculiaridad del gibón es que aparecen por primera vez las repeticiones tipo B, que se mantendrán en adelante a lo largo de la evolución de MsH42. En el análisis de los póngidos, los primates más próximos a la especie humana, el orangután (12-16 millones de años), ya presenta una composición en repeticiones muy similar a la de los humanos, aunque todavía su organización interna es sustancialmente diferente. De entre las trece repeticiones que lo componen sólo una (B4) no está presente en el minisatélite MsH42, siendo aparentemente exclusiva de esta especie.

En los póngidos africanos, chimpancé (4-6 millones de años) y gorila (6-8 millones de años), el minisatélite ya se

encuentra organizado de manera similar a la del locus humano. El análisis de las muestras de gorila reveló la existencia de al menos tres alelos de 23 (GOR1), 29 (GOR2) y 51 (GOR3) unidades repetitivas, con una única variante (C2) ausente en los alelos humanos. Estos tres alelos del gorila comparten entre sí un grupo de 9 repeticiones (A-B-A-C-A1-B-A2-C-A) en el extremo 5' del minisatélite y otro grupo de 5 repeticiones en el extremo 3' (A-C-A-A3-C1), por lo que el polimorfismo observado se debe a variaciones en su porción central. En el chimpancé los dos alelos detectados CHI1 y CHI2, de 22 y 18 repeticiones respectivamente, también presentan una organización homóloga a la del alelo corto humano, con la salvedad de dos repeticiones (B3 y C2), que no se encuentran en humanos. La unidad C2 está presente tanto en gorilas como en chimpancés, mientras que B3 parece una variante infrecuente originada por un SNP. En este sentido, nuestros resultados concuerdan con la hipótesis de que los póngidos africanos están más relacionados con los humanos que con los orangutanes asiáticos (Miyamoto et al., 1987; Pääbo, 2003).

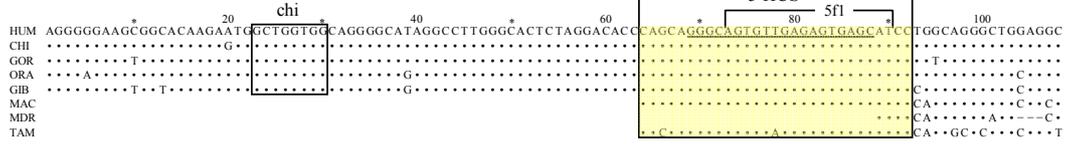
Gracias a las comparaciones entre los alelos de los distintos primates se puede afirmar que: i) las repeticiones A4 y B1 son exclusivas de los humanos, ii) las repeticiones A y C deben de ser las más antiguas, porque ya se encuentran en tamarindo, iii) las repeticiones A y C1 marcan los extremos del minisatélite desde hace al menos 25 millones de años

**Figura 20.** Alineamiento de las secuencias obtenidas a partir de los diferentes primates para la región MsH42 (minisatélite MsH42 y secuencias adyacentes 5' y 3') y los exones 13 y 14. Las secuencias humana, de chimpancé y de gorila están representadas por los alelos HUM(S), CHI1 y GOR1 debido a su mayor similitud en tamaño. En el resto de primates se muestra el único alelo conocido. Para facilitar las comparaciones entre alelos, el alineamiento del minisatélite se ha esquematizado con el código de una letra empleado en la Tabla 4. Las cajas 5'HCS y 3'HCS, resaltadas en amarillo, indican las secuencias altamente conservadas donde se encuentran los cebadores PC1 y PC2 (subrayados) y los *footprints* de DNaseI 5f1 y 3f1. Las cajas chi contienen el motivo de recombinación de procariontas *Chi* (GCTGGTGG). En el alineamiento del exón 13, las cajas delimitan dos codones que codifican para metionina en casi todos los primates no humanos y para treonina y valina en humanos.

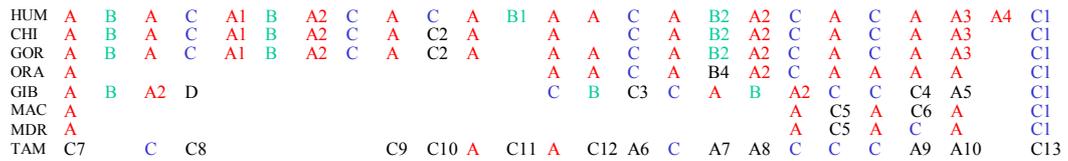
y iv) existe un bloque de repeticiones muy bien conservado en todas las especies de la familia Homínidae (C-A-B/B2/B4-A2-C) que sugieren que su presencia podría haber sido necesaria

para la construcción del minisatélite. La comparación de secuencias genómicas entre especies distantes es una herramienta de gran utilidad para detectar la presencia de elementos

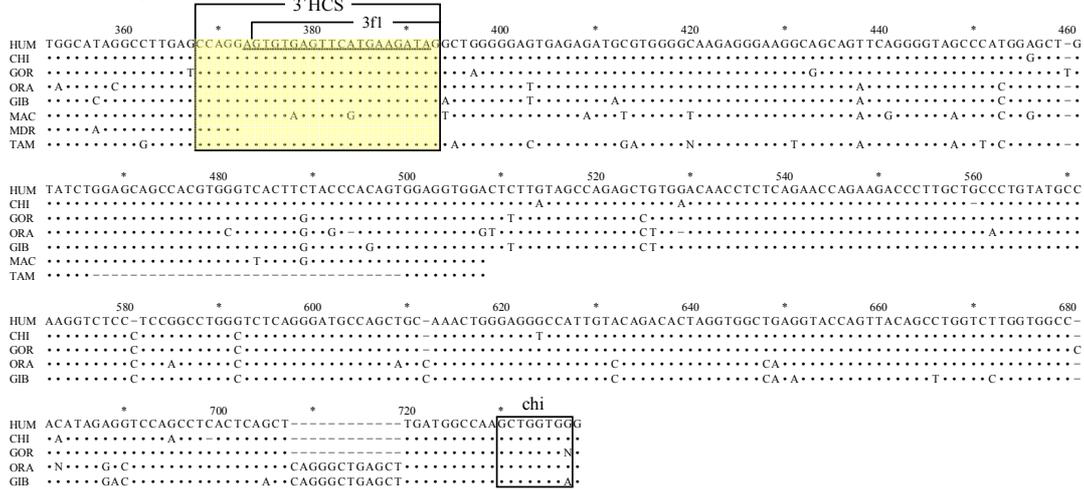
**Secuencia flanqueante 5'**



**Minisatélite Msh42**

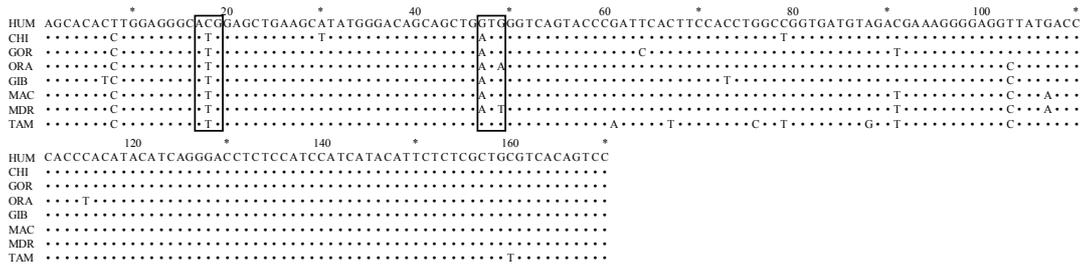


**Secuencia flanqueante 3'**

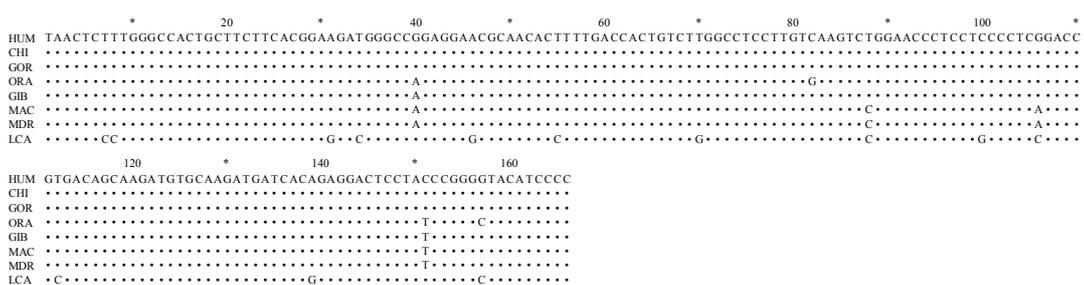


Región minisatélite (Intrón 13)

**Exón 13**



**Exón 14**



Exones

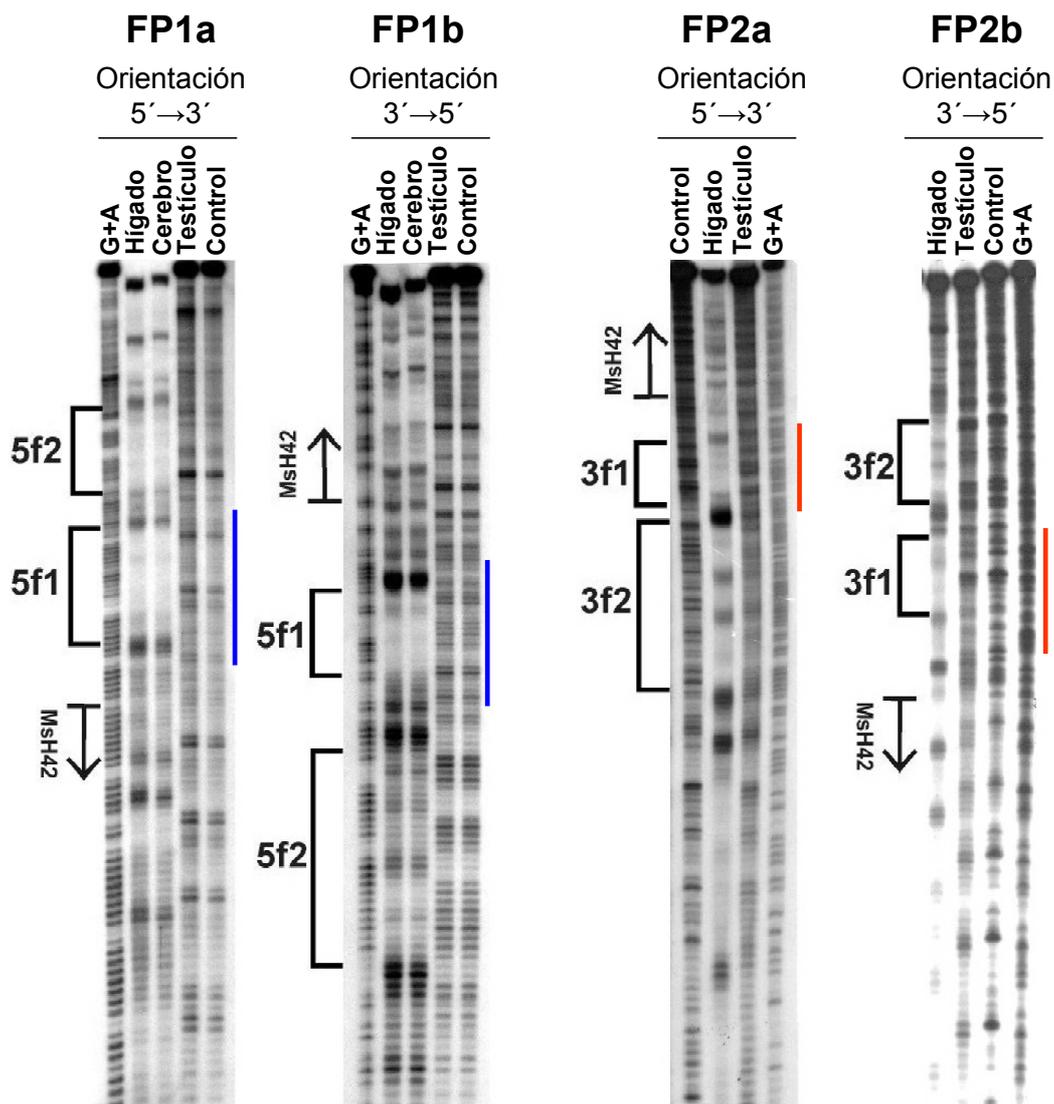
reguladores en las regiones no codificadoras del genoma (Boffelli et al., 2003). Por esta razón, llevamos a cabo el análisis de las secuencias flanqueantes 5' y 3' de MsH42. Los resultados obtenidos demostraron que hay una importante conservación entre dichas secuencias en los distintos primates (Figura 20). Concretamente, detectamos a ambos lados del minisatélite la presencia de dos secuencias altamente conservadas que denominamos 5'HCS y 3'HCS (*highly conserved sequences*). La secuencia 5'HCS comprende 29 pb y sólo presenta dos cambios nucleotídicos en el tamarindo. Por su parte, la secuencia 3'HCS, de 26 pb, sólo muestra dos transiciones A→G en el macaco. Aunque no pudimos analizar directamente las HCS en el mandril, como los oligonucleótidos PC1/PC2 con los que se obtuvo su amplificación están diseñados sobre estas secuencias, el hecho de obtener una amplificación con estos cebadores indica que en esta especie también deben hallarse bien conservadas. Asimismo, cabe destacar la alta conservación de las dos secuencias *Chi* de recombinación en procariontes presentes a ambos lados del minisatélite.

### ***Interacción de las HCS con proteínas***

La conservación específica de las secuencias no codificadoras puede implicar una determinada función, con la consiguiente presión selectiva y ralentización de las tasas de mutación de las mismas (Ludwig, 2002). Estrategias como la huella filogenética (Tagle et al., 1988; Gumucio et al., 1992) y el sombreado filogenético (Boffelli et al., 2003) han sido utilizadas para identificar potenciales elementos reguladores altamente conservados, mediante el alineamiento de secuencias procedentes de numerosas especies poco o muy emparentadas filogenéticamente. Esta

alta conservación evolutiva puede estar reflejando una función de estas secuencias probablemente dentro del gen Q8WUJ3. Para profundizar más en el análisis de esta región del genoma, nos planteamos estudiar si estas secuencias HCS serían reconocidas por proteínas nucleares, lo que estaría manifestando una evidencia más a favor de una posible función de las mismas. En nuestro laboratorio, se habían detectado previamente interacciones DNA-proteína en esta región mediante experimentos de *band shifting* (Boán et al., 1997), lo que refuerza la hipótesis de que las regiones conservadas interactúan de forma preferencial con proteínas de unión al DNA.

Para intentar averiguar qué secuencias concretas eran las que estaban siendo específicamente reconocidas por proteínas, se realizaron experimentos de *footprinting* con DnasaI. En estos experimentos se emplearon las sondas FP1 y FP2. FP1 comprende las 127 pb de la región flanqueante 5' contiguas al minisatélite, incluyendo el 5'HCS, y las nueve primeras repeticiones de MsH42. FP2, por su parte, posee las últimas nueve repeticiones de MsH42 y las siguientes 175 pb de la región flanqueante 3', incluyendo por lo tanto el 3'HCS. Los resultados obtenidos con ambas orientaciones de la sonda FP1 (FP1a y FP1b) y extractos nucleares de tejidos somáticos (cerebro e hígado) demostraron la presencia de dos zonas protegidas, 5f1 y 5f2, localizadas en la región flanqueante 5' del minisatélite MsH42 (Figura 21). Análogamente, los experimentos realizados con las dos orientaciones de la sonda FP2 (FP2a y FP2b) pusieron de manifiesto la existencia de otras dos zonas protegidas, 3f1 y 3f2, en la región flanqueante 3' cuando se empleaba el extracto de hígado de rata. Hay que destacar que las regiones 5f1 y 3f1 incluyen respectivamente a las secuencias 5'HCS



5f1 5'-AGTGTTGAGAGTGAGCAT-3'

5f2 5'-CATAGGCCTTGGGCACATCTAGGACAC-3'

3f1 5'-GTGTGAGTTCATGAAGATAG-3'

3f2 5'-GGAGTGAGAGATGCGTGGGGCAAGAGGGAAGGCAGCAGTTCAGGG-3'

**Figura 21.** Análisis de *footprinting* de DNasal con las sondas FP1 y FP2. Se muestran las autorradiografías correspondientes a los experimentos realizados con las dos orientaciones posibles de cada sonda (FP1a y FP2a, 5'→3'; FP1b y FP2b, 3'→5'). FP1 comprende parte de la región flanqueante 5' y el principio de Msh42, mientras que FP2 abarca el final del minsatélite y parte de la región flanqueante 3' contigua. En los dos casos se detectaron regiones de *footprint* (5f1, 5f2, 3f1 y 3f2), cuya secuencia se muestra en la parte inferior de la figura. 5f1 y 3f1 se incluyen dentro de las regiones conservadas 5'HCS (línea vertical azul) y 3'HCS (línea vertical roja), respectivamente (Figura 19). G+A, reacciones de secuenciación G+A de Maxam & Gilbert; Control, reacciones con BSA; Hígado, Cerebro y Testículo indican los distintos extractos proteicos nucleares empleados en las reacciones.

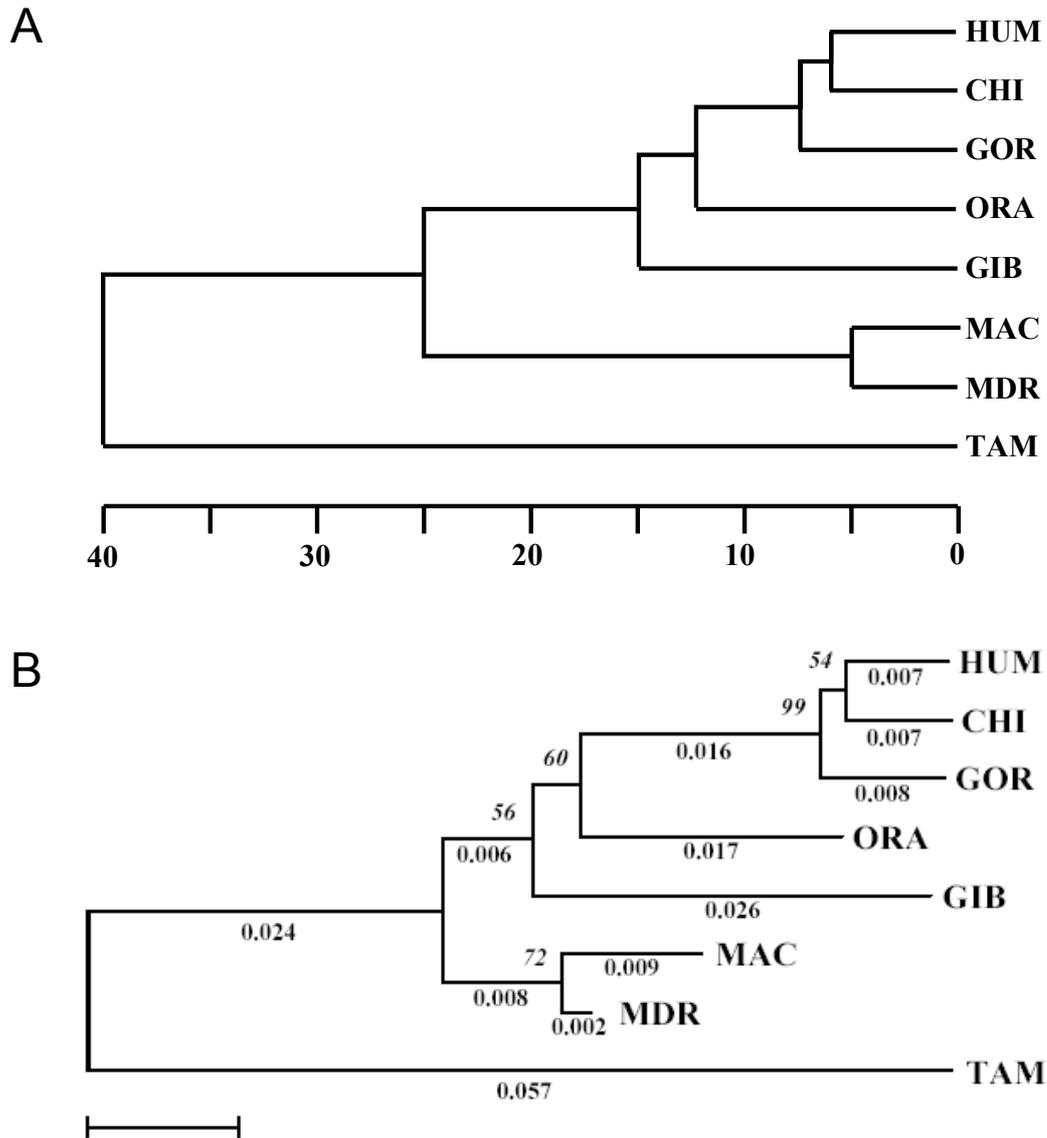
y al 3'HCS (Figura 20), lo que demuestra la existencia de una o varias proteínas que interactúan con estas secuencias conservadas, por lo que dicha conservación estaría verdaderamente marcando *footprints* filogenéticos.

Teniendo en cuenta que el locus Msh42 se encuentra situado dentro de un intrón, se podría especular con que estas HCS pudiesen desempeñar una posible función reguladora dentro del gen Q8WUJ3.

**Análisis filogenético de los datos de secuencia**

Para comparar la dinámica evolutiva de la región MsH42 localizada dentro del intrón 13 del gen Q8WUJ3 con respecto a las secuencias codificantes más próximas, amplificamos y secuenciamos

los exones vecinos (13 y 14) utilizando para ello los cebadores E5.1/E5.2 y E6.1/E6.2, respectivamente (Figura 17). El exón 13 se amplificó en todos los primates estudiados, excepto en los lemúridos, siendo la mayoría de los cambios con respecto a la secuencia humana se daban en la tercera posición



**Figura 22.** Comparación entre el árbol filogenético estándar de los primates (A) y el obtenido en nuestro estudio (B). En la parte inferior de la topología estándar de primates se muestra la escala temporal aproximada en millones de años (tomada de Goodman, 1999 y Pääbo, 2003). Nuestro árbol filogenético fue obtenido por *neighbour-joining* a partir del conjunto total de datos (región minisatélite y exones). Las distancias se estimaron bajo criterio de máxima verosimilitud, generándose mil versiones a partir de combinaciones de los datos originales mediante *bootstrapping* para evaluar el soporte de los nodos. Los números en cursiva indican los valores de *bootstrapping*, mientras que los números normales indican los valores de distancia de Tamura-Nei. La secuencia de tamarindo se usó como grupo externo para enraizar el árbol. La barra de escala representa 0.01 unidades de distancia.

de los codones y eran sinónimos (Figura 20). Existe, sin embargo, un codón que codifica para metionina en todos los primates, excepto en humanos donde codifica treonina, suprimiéndose así un codón de iniciación potencial dentro de la pauta de lectura. Quizás este hipotético codón de iniciación podría servir para producir más de una proteína a partir del mismo mRNA mediante un mecanismo que implicase la iniciación de la traducción en lugares internos de unión del ribosoma (Oh & Sarnow, 1993). El exón 14 pudo amplificarse en todos los primates con la excepción del tamarindo. En este caso todos los cambios con respecto al exón humano resultaron ser sinónimos (Figura 20).

Para estudiar si los datos experimentales se adecuaban a la filogenia estándar de primates (Miyamoto et al., 1987; Goodman et al., 1998), se realizó un análisis filogenético empleando las secuencias de la región minisatélite y de los exones 13 y 14. Los resultados obtenidos al aplicar el test de diferencias en la incongruencia de longitud indicaron que había una congruencia

significativa de las secuencias ( $P=1$ ), lo cual permitió la reconstrucción de una topología por *neighbour-joining* a partir del conjunto total de datos (Figura 22). El árbol filogenético obtenido de esta forma reproduce las relaciones filogenéticas entre los primates generalmente aceptadas (Miyamoto et al., 1987). Bajo criterios de parsimonia, los datos de todas las secuencias tomadas en su conjunto, así como los de las secuencias flanqueantes de MsH42 por separado, también reflejan el árbol de primates aceptado generalmente (Tabla 5). Sin embargo, se obtuvieron topologías alternativas a partir de los datos de secuencia de los exones y del minisatélite (datos no mostrados), aunque los valores del índice de consistencia, así como del test de Kishino-Hasegawa indican que dichos datos de secuencia pueden ser consistentes con la topología estándar de los primates (Tabla 5). La concordancia entre el árbol de los primates y el estimado a partir del minisatélite MsH42 sugiere que los eventos mutacionales observados en MsH42 son similares a los acaecidos en otras secuencias no

**Tabla 5.** Evaluación de los árboles filogenéticos generados a partir de las secuencias obtenidas de los distintos primates con respecto a la topología estándar establecida para este grupo.

|  | Exones<br>13 y 14    | Secuencias<br>flanqueantes<br>de MsH42 | Minisatélite<br>MsH42 | Conjunto de<br>todas las<br>secuencias |
|--|----------------------|--|-----------------------|--|
| Señal Filogenética (-g <sub>i</sub> )                  | -0.890               | -1.019                                 | -1.075                | -0.997                                 |
| Número de Caracteres Incluidos                         | 336                  | 495                                    | 243                   | 1074                                   |
| Caracteres Informativos (Parsimonia)                   | 15                   | 51                                     | 164                   | 207                                    |
| Longitud del árbol: Topología MP Observada             | 28 <sup>a</sup>      | 135 <sup>b</sup>                       | 293 <sup>c</sup>      | 456 <sup>b</sup>                       |
| Índice de Consistencia                                 | 0.893                | 0.963                                  | 0.761                 | 0.829                                  |
| Longitud del árbol: Topología Estándar <sup>d</sup>    | 28                   |  | 293                   |  |
| Índice de Consistencia                                 | 0.893                |  | 0.761                 |  |
| Diferencia de Longitud                                 | 0                    |  | 0                     |  |
| -lnL de la Topología ML Observada                      | 631.216 <sup>a</sup> | 1109.097 <sup>b</sup>                  | 473.966 <sup>c</sup>  | 2289.185 <sup>b</sup>                  |
| -lnL de la Topología estándar de primates <sup>d</sup> | 631.216              |  | 475.945               |  |
| $\Delta$ lnL   | 0                    |  | 1.979                 |  |
| P (Test Kishino-Hasegawa)                              | 1.0                  |  | 0.304                 |  |

<sup>a</sup> Topología observada: (TAM, ((MAC, MDR),(ORA, GIB, (CHI, GOR, HUM)))). <sup>b</sup> Topología observada idéntica a la topología estándar de primates. <sup>c</sup> Topología observada: (TAM, (MAC, MDR), (ORA, (GIB, (HUM, (CHIM, GOR))))). <sup>d</sup> Topología estándar de primates: (TAM, ((MAC, MDR), (GIB, (ORA, (GOR,(CHIM, HUM)))))). <sup>e</sup> Topología observada: (TAM, (GIB, ((HUM, CHI, GOR), (ORA, (MAC, MDR))))). MP, máxima parsimonia; ML, máxima verosimilitud.

codificadoras durante la historia evolutiva de los primates (Saitou & Ueda, 1994; Apoil & Blancher, 2000). Además, la estimación del parámetro  $\theta$  ( $\theta=0.256$ ) obtenido de los datos poblacionales sugiere una tasa de mutación que se encuentra dentro del rango de las estimadas para intrones en el genoma humano (Huang, 1998).

### ***Historia evolutiva del minisatélite MsH42***

Los resultados obtenidos en nuestro laboratorio indican que la secuencia MsH42 se originó dentro de un intrón, probablemente a partir de una secuencia progenitora sin una estructura de minisatélite bien definida (Figura 23). Esta secuencia inicial pudo haberse originado mediante procesos de *slippage* y/o entrecruzamiento desigual entre unidades repetitivas no contiguas formadas al azar (Levinson & Gutman, 1987; Haber & Louis, 1998; Taylor & Breden, 2000) o por otro tipo de cambios que condujesen a la formación de las primeras unidades repetitivas del precursor de MsH42. La existencia de una región parecida a MsH42 en el tamarindo denota la generación de esta secuencia durante la fase temprana de la evolución del linaje de los primates, antes de la divergencia entre los monos del Nuevo y Viejo Mundo, aproximadamente hace 40 millones de años (Goodman, 1999). Esta deducción se apoya en la ausencia de amplificación al emplear el DNA genómico de tres especies de prosimios diferentes y porque no pudo encontrarse ninguna secuencia ortóloga a MsH42 en ninguna de las bases de datos genómicas de organismos filogenéticamente más distantes (vaca, zarigüeya, ratón, rata, pez cebra y mosca del vinagre). En el macaco y el mandril, la región ortóloga a MsH42 ya presenta una organización en tándem más parecida a la del minisatélite humano, mostrando las unidades

características inicial y final (A y C1). El minisatélite de gibón presenta una unidad repetitiva relativamente diferente a las demás (D), así como otras variantes exclusivas de esta especie (A5, C3 y C4). Sin embargo, ya encontramos un grupo de repeticiones C-A-B-A2-C-C muy similar al bloque de repeticiones C-A-B2-A2-C-A de la secuencia humana, y que pudo constituir el ordenamiento primordial del minisatélite en los póngidos. La similitud con la organización del minisatélite humano se hace aún más evidente en el orangután, donde el bloque A-A-C-A-B4-A2-C-A es idéntico al presente en el alelo corto humano, con la excepción de una transición entre las repeticiones B4 (exclusiva de orangután) y B2.

La homología entre MsH42 en humanos y los grandes monos africanos, chimpancé y gorila, es muy elevada, tanto en secuencia como en organización de las repeticiones. La comparación de los alelos HUM(S), CHI1 y GOR1 reveló una identidad casi completa entre las secuencias de las tres especies. Cabe resaltar la existencia de dos bloques exactamente iguales entre los tres alelos: A-B-A-C-A1-B-A2-C-A y A-A-C-A-B2-A2-C-A-C-A-A3 (Figura 23). Este hecho sugiere que debió de existir de un alelo común ancestral a partir del cual se generaron los alelos HUM(S), CHI1 y GOR1, pero que ha desaparecido de las poblaciones humanas actuales y de las de chimpancé y gorila. No obstante, es destacable el gran parecido que existiría entre tal alelo ancestral y los alelos mencionados. Desde este punto de vista es posible que el polimorfismo actual existente en la región minisatélite de dichas especies haya sido originado a partir de estos alelos altamente homólogos. Sorprendentemente GOR1 mantiene más diferencias con GOR2 y GOR3 que con los alelos CHI1 y HUM(S). Estas variaciones en el gorila no parecen fácilmente explicables

mediante entrecruzamiento desigual, *slippage* o mutaciones puntuales, al contrario que entre los alelos presentes en humano y chimpancé. Por lo tanto, es plausible que diversas formas alélicas de la región MsH42 coexistiesen en el acervo genético ancestral y que GOR2 y GOR3 evolucionasen a partir de alguna de estas variantes, mientras que GOR1, CHI1 y HUM(S) se originaron a partir de otra forma alélica distinta. Al mismo tiempo, la propia existencia de estos tres alelos altamente homólogos en las tres especies sugiere que el alelo común ancestral debió haber sido el más abundante. De hecho, la identificación de alelos compartidos entre especies, más que las propias frecuencias relativas en cada especie, son un método muy indicativo para establecer los posibles alelos ancestrales de un locus polimórfico (Iyengar et al., 1998).

Los datos previos de que disponíamos acerca de las frecuencias alélicas de MsH42 se referían a población caucasoide (Boán et al., 2002). Por esta razón, hemos realizado un estudio poblacional en individuos negroides. En estas poblaciones el alelo corto HUM(S) presentó una frecuencia de  $0.48 \pm 0.07$  lo que constituye un valor tres veces más alto que en la población caucasoide ( $0.16 \pm 0.02$ ) (Boán et al., 2002). El hecho de que este alelo tenga mayor frecuencia en negroides, junto con su gran homología con los alelos CHI1 y GOR1 en los grandes monos africanos, concuerdan con la hipótesis del “*out of Africa*” (Cann et al., 1989; Armour et al., 1996). En otras palabras, los datos sobre las frecuencias alélicas diferentes entre las poblaciones europeas y africanas y concretamente el hecho de que en estas últimas el alelo HUM(S), el más semejante a los alelos de chimpancé y gorila, sea más abundante que en europoides parecen concordar con la teoría del origen africano de los humanos modernos. Esta teoría se apoya

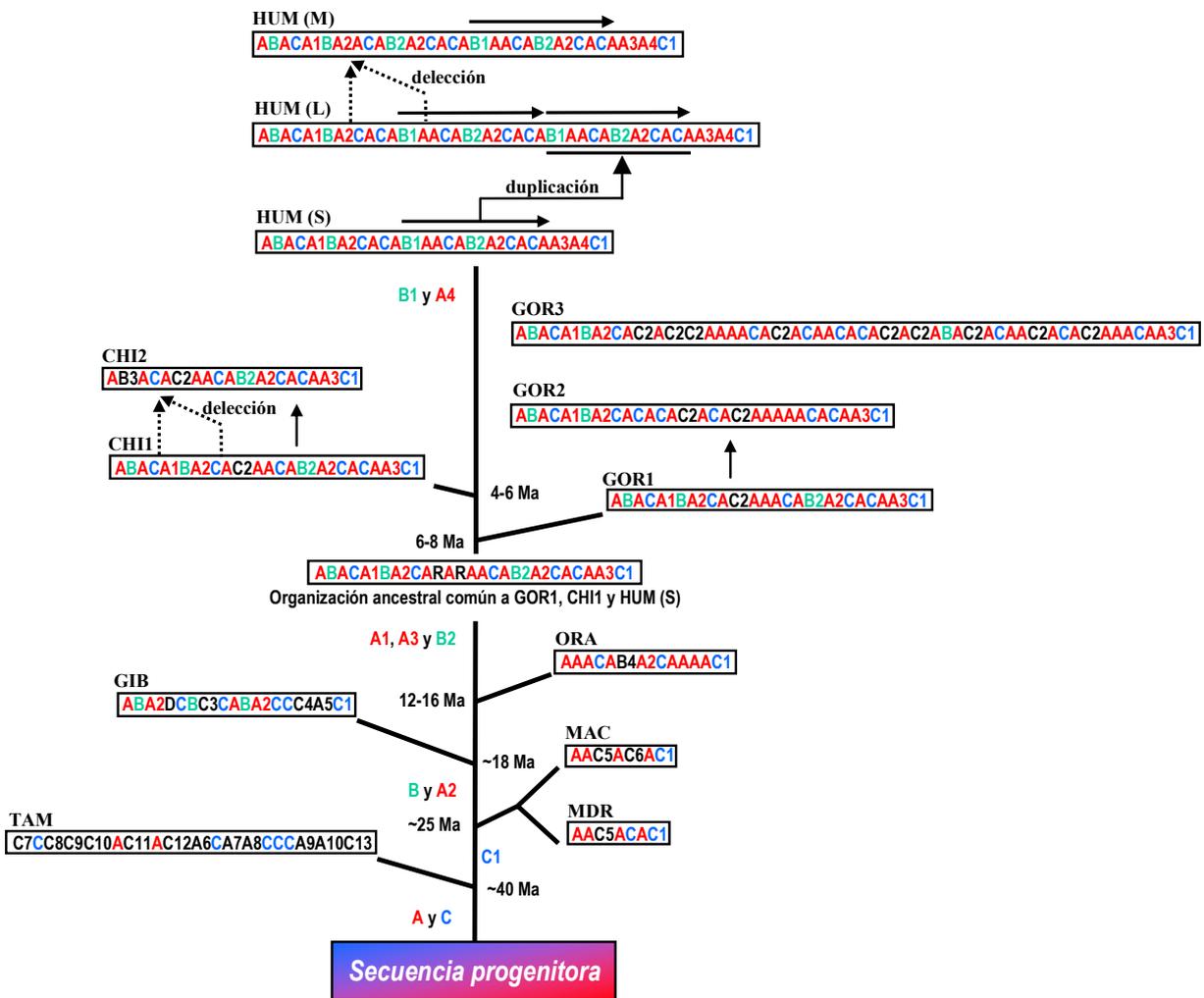
fuertemente tanto en estudios de variación genética en poblaciones actuales como en el registro fósil y es la más ampliamente aceptada en la actualidad (Stringer, 2003).

A la vista de los resultados obtenidos al analizar la historia evolutiva de MsH42 en primates, nos planteamos la cuestión de cómo se podría haber originado el polimorfismo de este locus en la especie humana. Así, una hipótesis sobre la sucesión de acontecimientos que expliquen este polimorfismo podría ser la siguiente: inicialmente se formó el alelo HUM(S) a partir del ancestro común con chimpancé y gorila; después una duplicación directa del bloque de 11 repeticiones B1-A-A-C-A-B2-A2-C-A-C-A dio lugar al alelo largo, del cual surgió el alelo mediano mediante una delección del bloque C-A-C-A-B1-A (Figura 23). El mantenimiento de la organización interna de las repeticiones en los tres alelos de MsH42 apunta hacia el emparejamiento incorrecto de bloques de repeticiones como el principal evento mutacional en este locus, mecanismo que ya ha sido propuesto para explicar la variabilidad en otros minisatélites (Charlesworth et al., 1994).

Una aproximación útil al estudio de la evolución humana a nivel molecular es considerar a nuestro genoma como un mosaico en el cual los diferentes segmentos de DNA tienen su propia historia evolutiva (Pääbo, 2003). Aunque dispersas a lo largo del genoma, distintos tipos de secuencias de tipo repetitivo, especialmente los retroelementos, pueden haber influenciado notablemente en la expansión general que ha sufrido el genoma humano con respecto al de otros primates (Liu et al., 2003). No obstante, la contribución del DNA repetido en tándem a este crecimiento del genoma puede haber sido también considerable (Comeron, 2001). En este sentido, existen evidencias que apuntan hacia una

tendencia de los minisatélites a expandirse hasta un determinado momento en el cual expansión y contracción se equilibran, estado en el cual parecen hallarse ya algunos minisatélites en determinadas poblaciones humanas (Buard et al., 2002). Bajo esta perspectiva, nuestros resultados nos permiten imaginar cuando inició MsH42 su andadura evolutiva en los primates y cómo este locus evolucionó hasta convertirse en la secuencia actual que conocemos en los primates. Así, en el modelo propuesto en

la Figura 23, podemos observar como MsH42 es una secuencia que se originó antes de la divergencia entre los monos del Nuevo y del Viejo Mundo, que probablemente evolucionó mediante cambios relativamente sencillos del tipo *slippage*, entrecruzamiento desigual, mutaciones puntuales y/o duplicaciones/delecciones, y que el polimorfismo actual en las poblaciones humanas, de chimpancé y de gorila, las especies con minisatélites más largos, podría haberse generado a partir de un único alelo ancestral de la región MsH42.



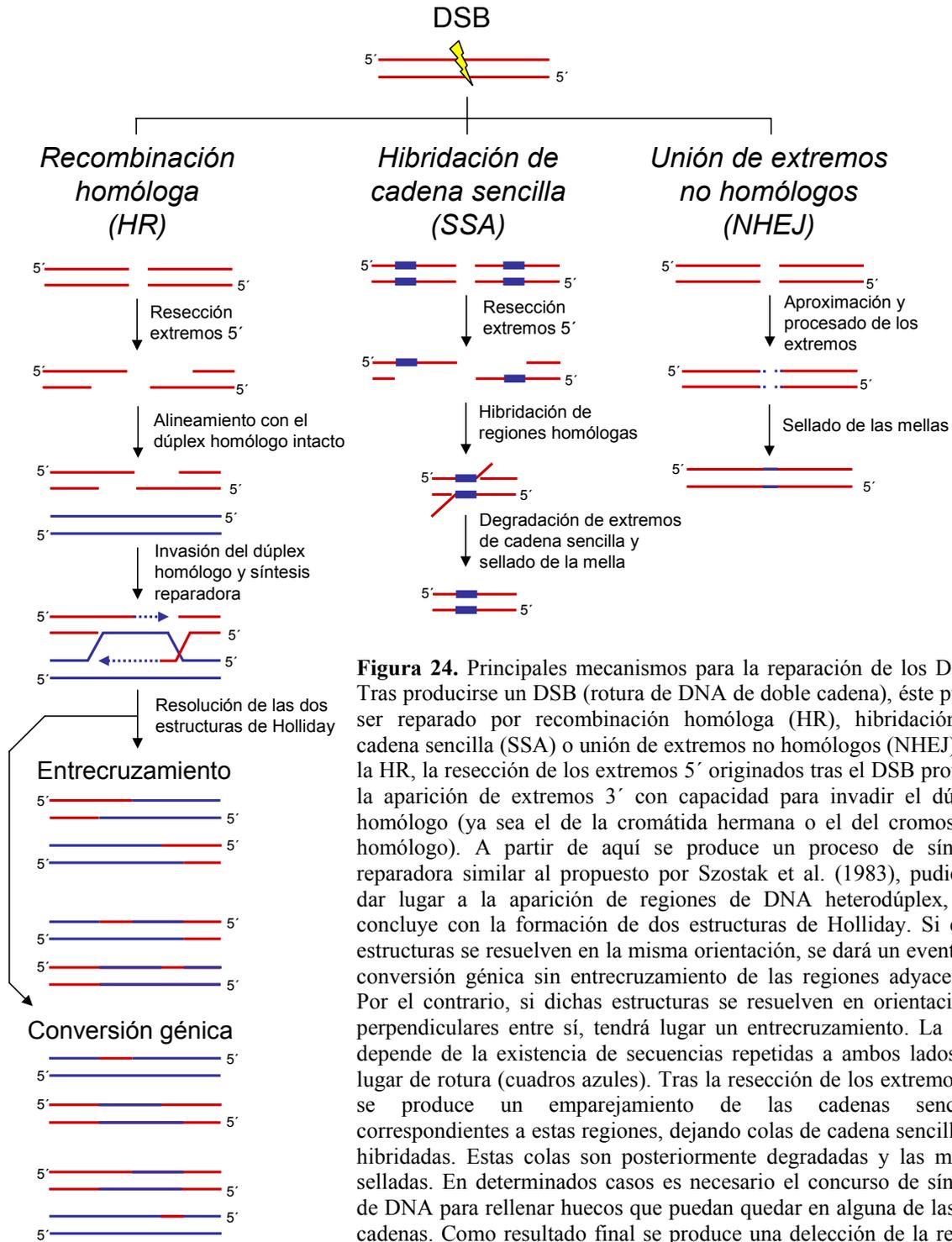
**Figura 23.** Historia evolutiva hipotética del minisatélite MsH42. La organización del minisatélite para cada uno de los alelos de las distintas especies se indica con el código de una letra (Tabla 4). Las repeticiones presentes en el minisatélite humano se hallan coloreadas y su posible momento de aparición en el linaje de los primates se muestra en los internodos a lo largo del árbol. R en la organización de las unidades repetitivas del hipotético ancestro común denota repeticiones de secuencia desconocida. Los tiempos de divergencia se tomaron de Goodman (1999) y Pääbo (2003). Ma, millones de años.

## Generación y localización de cortes de doble cadena en la región *MsH42*

Muchos procesos recombinatorios se inician tras la generación de un corte de doble cadena (DSB, *double-strand break*) en un dúplex de DNA (Szostak, 1983; Pâques & Haber, 1999). Además, estos DSBs también pueden ser causados por la exposición de las células a agentes lesivos para el DNA, como determinadas sustancias químicas genotóxicas o radiaciones ionizantes, o bien como consecuencia de horquillas de replicación bloqueadas o dañadas (Chu, 1997; West et al., 2000). La persistencia de estas rupturas del DNA podría conducir a la generación de mutaciones, translocaciones genéticas y finalmente cáncer (Sancar et al., 2004; Esashi et al., 2005; Gorgoulis et al., 2005). Debido a esto, el buen funcionamiento de los mecanismos de reparación de los DSBs es crucial para el mantenimiento de la estabilidad de los genomas (Kanaar et al., 1998; Pastink et al., 2001).

Los DSBs que aparecen en células germinales en meiosis son habitualmente procesados mediante recombinación homóloga (HR, *homologous recombination*), mientras que en las células somáticas son reparados tanto por HR como por unión de extremos no homólogos (NHEJ, *non-homologous end-joining*) (Takata et al., 1998; Singleton & Jeggo, 1999). La HR es un mecanismo conservativo que implica el uso de una molécula de DNA doble cadena intacta para la reparación precisa del corte por medio de eventos de intercambio de cadena sencilla y pasos de síntesis de DNA (Figura 24). Este mecanismo es el principal responsable de la eliminación de DSBs en levaduras (Pâques & Haber, 1999). La hibridación de cadena sencilla (SSA, *single-strand annealing*) puede considerarse como una forma no conservativa de HR que

requiere la existencia de regiones de homología en las zonas que flanquean el corte y conlleva la degradación de los extremos 5' de las moléculas de DNA implicadas, facilitando la asociación entre dichas regiones (Figura 24). Seguidamente, se produce una eliminación de los extremos no homólogos, rellenándose los posibles huecos que se generen y la mella en cada cadena se sella con el establecimiento de un enlace fosfodiéster mediante una ligasa (Göttlich et al., 1998). Por otro lado, la recombinación tipo NHEJ es la ruta de reparación no conservativa predominante en los vertebrados, además de estar directamente implicada en la recombinación V(D)J y en el cambio de clase de los genes de las inmunoglobulinas en los mamíferos (Singleton & Jeggo, 1999). En este caso, la reparación del DSB se logra sin la necesidad de regiones de homología, o en cualquier caso muy pequeñas, entre los extremos que van a ser ligados. Aunque se han descrito más proteínas que pueden intervenir en este proceso, el NHEJ (Figura 24) implica básicamente la unión de las proteínas Ku70 y Ku80 a cada extremo de DNA, lo que activa la subunidad catalítica de la proteína quinasa dependiente de DNA (DNA-PK) mediante la estabilización de su interacción con dichos extremos del DNA (revisado en Critchlow & Jackson, 1998; Doherty & Jackson, 2001). Esto promueve la ligación del corte por el heterodímero DNA ligasa IV-XRCC4 (revisado en Karran, 2000; Lieber et al., 2003; Weterings & van Gent, 2004). Las evidencias más recientes obtenidas en este campo hacen pensar en una cooperación entre HR y NHEJ para el mantenimiento de la integridad y de la estabilidad genómica en los mamíferos (Richardson & Jasin, 2000).



**Figura 24.** Principales mecanismos para la reparación de los DSBs. Tras producirse un DSB (rotura de DNA de doble cadena), éste puede ser reparado por recombinación homóloga (HR), hibridación de cadena sencilla (SSA) o unión de extremos no homólogos (NHEJ). En la HR, la resección de los extremos 5' originados tras el DSB provoca la aparición de extremos 3' con capacidad para invadir el dúplex homólogo (ya sea el de la cromátida hermana o el del cromosoma homólogo). A partir de aquí se produce un proceso de síntesis reparadora similar al propuesto por Szostak et al. (1983), pudiendo dar lugar a la aparición de regiones de DNA heterodúplex, que concluye con la formación de dos estructuras de Holliday. Si estas estructuras se resuelven en la misma orientación, se dará un evento de conversión génica sin entrecruzamiento de las regiones adyacentes. Por el contrario, si dichas estructuras se resuelven en orientaciones perpendiculares entre sí, tendrá lugar un entrecruzamiento. La SSA depende de la existencia de secuencias repetidas a ambos lados del lugar de rotura (cuadros azules). Tras la resección de los extremos 5', se produce un emparejamiento de las cadenas sencillas correspondientes a estas regiones, dejando colas de cadena sencilla no hibridadas. Estas colas son posteriormente degradadas y las mellas selladas. En determinados casos es necesario el concurso de síntesis de DNA para rellenar huecos que puedan quedar en alguna de las dos cadenas. Como resultado final se produce una delección de la región entre las secuencias repetidas. En el NHEJ, se produce un reclutamiento de proteínas hacia la zona de la ruptura que procesa los extremos, con la ganancia o pérdida de algunos nucleótidos, en caso de que no sean ligables. Estas mismas proteínas mantienen próximos estos extremos para facilitar su posterior ligación.

La región MsH42 es una secuencia que es capaz de promover fuertemente la recombinación intramolecular *in vitro* (Boán et al., 1998, 2002). En esta Tesis Doctoral, nos hemos planteado si el

potencial recombinogénico de esta región se debe a que contiene motivos que son reconocidos como dianas por la actividad que origina los DSBs. En nuestros experimentos previos con el

minisatélite, hemos podido detectar valores de recombinación intramolecular *in vitro* cercanos al 1% (Boán et al., 2002). Además, si tenemos en cuenta que la estrategia empleada sólo permitía el análisis de los eventos intramoleculares, muy probablemente el nivel de recombinación total en la región Msh42 sería aún mayor. Esto nos llevó a pensar que si los sucesos recombinatorios eran tan frecuentes y estaban precedidos por la formación de un DSB, quizá pudiésemos detectar la generación de estos cortes en un ensayo *in vitro*.

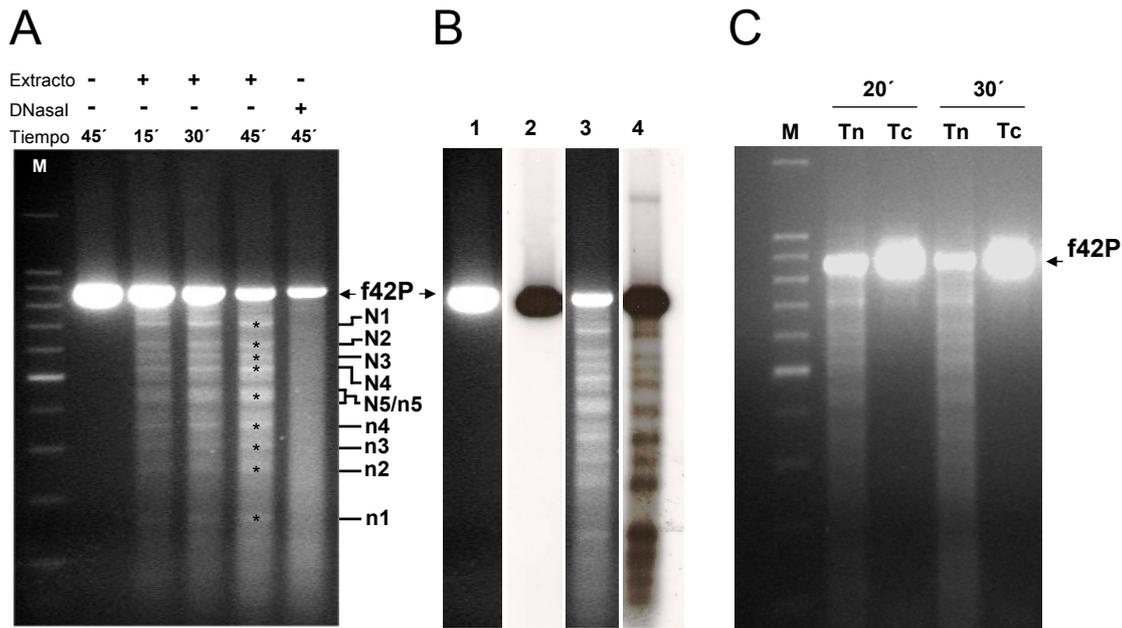
### ***El ensayo DSB in vitro***

En nuestro laboratorio, decidimos desarrollar un ensayo *in vitro* que nos permitiese detectar tanto la generación de DSBs en la región Msh42 como los productos de recombinación originados a partir de las moléculas cortadas. En estos ensayos, que en adelante denominaremos genéricamente ensayos DSB, se incubó un fragmento de DNA que contenía a la región Msh42 junto con un extracto proteico nuclear que aportaría las actividades nucleásicas y de reparación del DNA. Rutinariamente, se ha empleado un extracto proteico preparado a partir de testículo de rata, ya que fue en este tejido donde se detectó una mayor capacidad recombinogénica para la región Msh42 (Boán et al., 2002) y por tanto donde teóricamente podríamos encontrar una mayor tasa de generación de DSBs. Esta incubación del DNA sustrato y el extracto proteico se realizó además utilizando un tampón que contenía los elementos necesarios para llevar a cabo la recombinación / reparación, es decir,  $Mg^{2+}$ , dNTPs y ATP.

En un primer ensayo DSB, el fragmento de DNA f42P, de 863 pb y extremos *Pst*I, que comprende la región Msh42 (Figura 13A), se incubó a 37°C con un

extracto proteico nuclear de testículo de rata, en condiciones estándar de 10 mM  $Mg^{2+}$  (Tabla 2) y a diferentes tiempos (15, 30 y 45 min). La Figura 25A muestra los resultados de este experimento, observándose que el fragmento f42P es degradado progresivamente por la acción nucleásica del extracto. Esta actividad genera un patrón de bandas definido y que siempre se mantiene constante entre los experimentos realizados, consecuencia de la acumulación de moléculas originadas por cortes en puntos concretos del sustrato. A este patrón de bandas nos referiremos en adelante como *patrón DSB*. La naturaleza proteica de dicha actividad se demostró mediante la desnaturalización previa del extracto por calor, tratamiento que anulaba la degradación del fragmento f42P (datos no mostrados).

Las bandas de este patrón DSB pueden agruparse en parejas cuyo tamaño suma aproximadamente el del sustrato original f42P. De esta manera, se definieron las parejas N1/n1, N2/n2, N3/n3, N4/n4 y N5/n5 (Figura 25A). Esta complementariedad en los tamaños de las bandas del patrón DSB apoya que su origen sea debido a un único corte sobre el sustrato inicial. Para verificar la especificidad de los cortes, se realizó un control en el cual se sustituía el extracto de testículo por DNasaI. El resultado de este experimento fue la formación de una serie continua de productos de degradación sin la formación de bandas aparentes (Figura 25A), lo que apunta a que las bandas del patrón DSB no se deben a lugares hipersensibles a la acción de la DNasaI. Por otra parte, para descartar que la formación del patrón DSB pudiese ser debida a la acción de exonucleasas que degradasen el sustrato desde sus extremos hasta puntos concretos con una secuencia o estructura secundaria determinada, se llevó a cabo un ensayo DSB empleando un sustrato



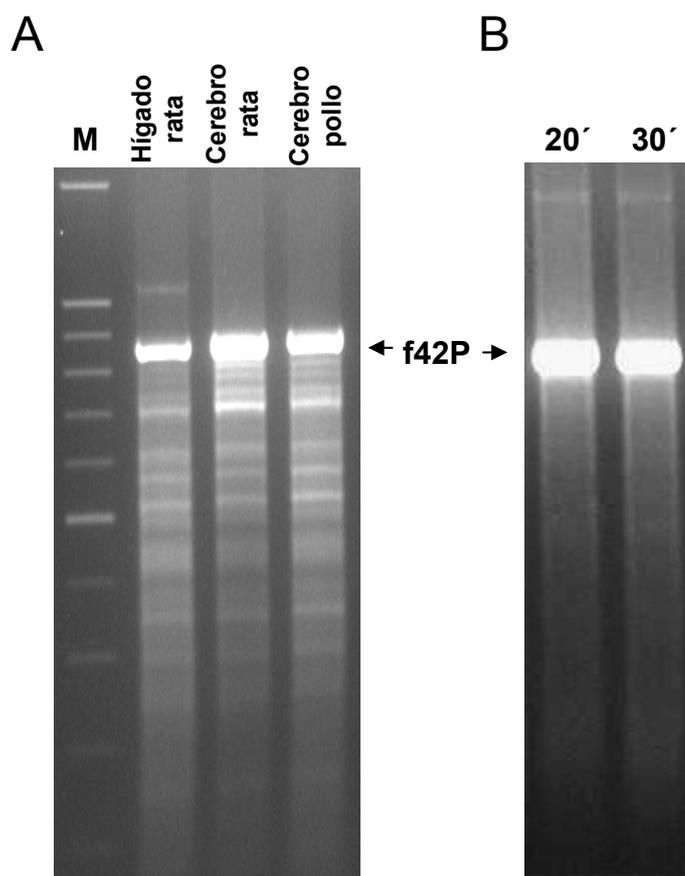
**Figura 25.** Ensayo DSB en condiciones estándar (10 mM  $Mg^{2+}$ ). **A.** Ensayo DSB empleando extracto nuclear de testículo de rata y fragmento f42P a distintos tiempos (15, 30 y 45 min) y de una digestión de f42P con DNasaI. Se muestran las bandas N1-N5 y n1-n5 (asteriscos) que forman parejas cuyo tamaño suma aproximadamente el del sustrato original. **B.** Ensayo DSB con fragmento f42E marcado radiactivamente. Carriles 1 y 2, fragmento sin extracto; 3 y 4, reacciones con extracto de testículo. Se muestra el gel de agarosa teñido con bromuro de etidio (1 y 3) y el autorradiograma obtenido a partir de él (2 y 4). **C.** Ensayo DSB con extracto citosólico (Tc) o nuclear (Tn) de testículo de rata 20 y 30 min de incubación. El análisis electroforético de los productos del ensayo DSB se realizó en gels de agarosa al 2%. M es el marcador *100 bp ladder* de Promega.

cuyos extremos estaban marcados radiactivamente. Puede observarse como tras la incubación con el extracto, las moléculas que forman el patrón DSB permanecen marcadas, lo que descartaría la actividad exonucleásica como la causante de la aparición de estas bandas (Figura 25B). En este experimento el sustrato empleado fue f42E, fragmento equivalente al f42P, pero con extremos *EcoRI* (Figura 13A). La elección de este sustrato se debió a que el marcaje terminal mediante T4 polinucleótido kinasa es más efectivo sobre extremos 5' protuberantes. Por otra parte, no observamos diferencia alguna en los ensayos DSB al emplear los fragmentos f42E, f42S y f42P, lo que indica que la generación del patrón DSB no se ve influido por el tipo de extremos (protuberantes 5', romos y protuberantes 3', respectivamente) del sustrato (datos no mostrados). En algunos ensayos, se empleó el extracto citosólico S-100 de

testículo de rata en lugar del extracto proteico nuclear. En estos experimentos, el DNA no sufrió alteración alguna sugiriendo la probable localización nuclear de la actividad detectada (Figura 25C). Cuando se emplearon sustratos alternativos a la región minisatélite, f5'MsH42 o fLacZ, se formó como era de esperar un patrón DSB diferente (datos no mostrados), aunque con la misma característica de la existencia de parejas de bandas que suman aproximadamente el tamaño del sustrato original. Esto apunta a que cualquier DNA puede ser sustrato de la actividad endonucleásica del ensayo DSB, siempre y cuando contenga las dianas adecuadas para esta enzima. Esta serie de resultados demuestran que existe una actividad endonucleolítica en nuestro ensayo DSB capaz de generar cortes de manera específica en secuencias concretas a lo largo de la región MsH42.

Al emplear extractos proteicos nucleares procedentes de distintas fuentes (testículo, hígado, cerebro y riñón de rata, y cerebro de pollo) no encontramos diferencias cualitativas en la formación del patrón DSB, aunque sí observamos variaciones en el tiempo de incubación necesario para su generación. En concreto, los extractos de hígado de rata y cerebro de pollo resultaron ser especialmente rápidos en la generación del patrón DSB (Figura 26A), necesitando tan sólo 10-15 min frente a

los 30 min requeridos por el extracto de testículo. Sin embargo, extractos como los de cerebro (Figura 26A) y riñón de rata (no mostrado) exhibieron una cinética de degradación del sustrato similar a la del testículo. Por el contrario, no fuimos capaces de detectar la formación del patrón DSB característico en extractos proteicos nucleares obtenidos a partir de bazo de rata (Figura 26B). Esto puede indicar que la proteína responsable de la formación de los DSBs se expresa poco o nada en este tejido.



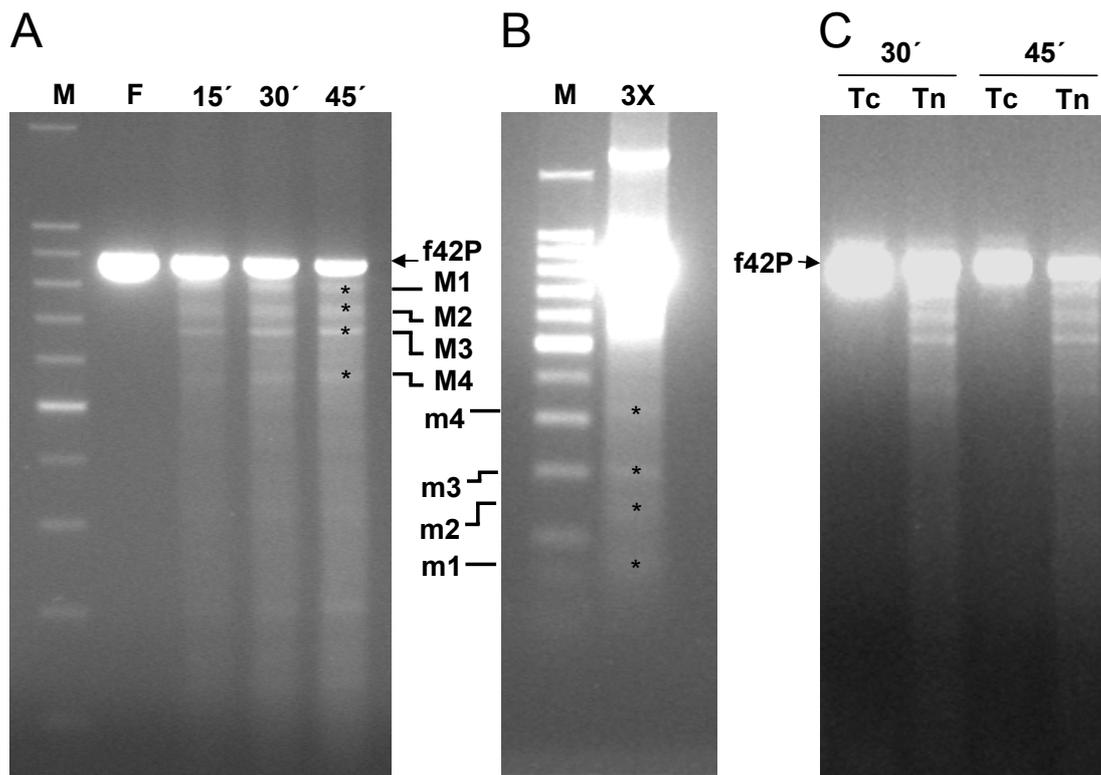
**Figura 26.** Ensayo DSB variando el extracto proteico nuclear o el sustrato empleado. **A.** Ensayo DSB con 12  $\mu$ g de extracto nuclear procedente de distintos tejidos: hígado de rata (incubado 15 min), cerebro de rata (incubado 30 min) y cerebro de pollo (incubado 10 min). **B.** Ensayo DSB con extracto nuclear de bazo a 20 y 30 min de incubación. El análisis de los productos de los distintos ensayos DSB se realizó en geles de agarosa al 2%. F, fragmento sin extracto; T, reacciones con extracto nuclear de testículo. M, marcador 100 bp ladder de Promega.

### ***Influencia del $Mg^{2+}$ en la actividad endonucleásica del ensayo DSB***

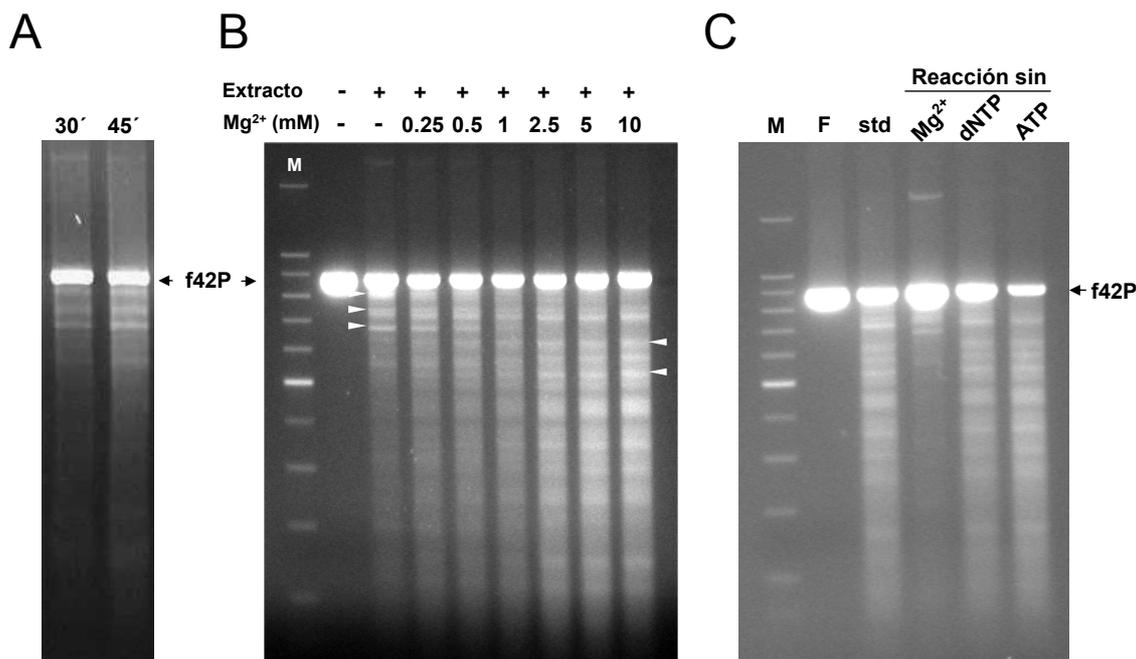
Para estudiar la posible influencia del  $Mg^{2+}$  sobre la actividad nucleásica responsable de la generación del patrón DSB, se realizaron una serie de ensayos DSB con un tampón de reacción que no contenía dicho catión (Tabla 2). La ausencia de  $Mg^{2+}$  provocó una alteración notable respecto a los resultados obtenidos en condiciones estándar (10 mM  $Mg^{2+}$ ), apareciendo un patrón de bandas completamente diferente (Figura 27A). También aquí pudimos comprobar, que las bandas de este patrón podían agruparse en parejas (M1/m1, M2/m2, M3/m3, M4/m4) cuyos tamaños sumaban aproximadamente el del fragmento de DNA f42P empleado como

sustrato. Esto apuntaba a que su generación se debía a la acumulación de cortes en puntos concretos de las moléculas del sustrato como ocurría en presencia de  $Mg^{2+}$ . Las bandas m1, m2, m3 y m4 son difíciles de apreciar dada su escasa intensidad, por lo que para visualizarlas fue necesario cargar varias reacciones en un mismo pocillo (Figura 27B). Al igual que en el caso del ensayo DSB en condiciones estándar, la actividad nucleásica sólo se detectó en los extractos nucleares, no pudiéndose apreciar productos de degradación cuando se emplearon extractos proteicos citosólicos (Figura 27C).

La actividad independiente de  $Mg^{2+}$  fue observada en todos los extractos nucleares de rata analizados (datos no mostrados), incluido el extracto de bazo



**Figura 27.** Ensayo DSB sin  $Mg^{2+}$ . **A.** Ensayos DSB sin  $Mg^{2+}$  con extracto nuclear de testículo a distintos tiempos de incubación, mostrando las bandas DSB de mayor tamaño, M1-M4 (marcadas por asteriscos). F, fragmento f42P sin extracto. **B.** Análisis de los productos de tres ensayos DSB sin  $Mg^{2+}$  con extracto nuclear de testículo de rata, cargados juntos en un mismo pocillo (3X) para poner de manifiesto las bandas DSB de menor tamaño, m1-m4 (marcadas por asteriscos). **C.** Ensayo DSB sin  $Mg^{2+}$  con extractos citosólicos (Tc) o nucleares (Tn) de testículo de rata, a tiempos de incubación de 30 y 45 min. M, marcador *100 bp ladder* de Promega.



**Figura 28.** Ensayos DSB sin  $Mg^{2+}$  y con otras variaciones del ensayo DSB con respecto a la reacción estándar. **A.** Ensayo DSB sin  $Mg^{2+}$  empleando extracto nuclear de bazo a 30 y 45 min de incubación. **B.** Ensayos DSB con extracto nuclear de testículo de rata y a diferentes concentraciones de  $Mg^{2+}$ . Las flechas blancas señalan algunas de las bandas de DNA de los patrones DSB en ausencia como en presencia de 10 mM  $Mg^{2+}$  y que desaparecen de forma más evidente al variar la concentración de este catión. **C.** Ensayo DSB estándar (std) con extracto de testículo de rata y otros realizados con un tampón DSB que carece de  $Mg^{2+}$ , de dNTPs o de ATP. F, fragmento sin extracto. El análisis electroforético de los productos de esta serie de ensayos se realizó en geles de agarosa al 2%. M, marcador 100 bp ladder de Promega.

que no mostraba la actividad dependiente de  $Mg^{2+}$  (Figura 28A), aunque tal actividad  $Mg^{2+}$ -independiente no pudo detectarse en extractos de cerebro de pollo.

Dado que la ausencia de  $Mg^{2+}$  en el tampón del ensayo DSB conducía a la generación de un patrón DSB diferente del obtenido bajo condiciones estándar, decidimos hacer una serie de experimentos en los que se incrementaba progresivamente la concentración de  $Mg^{2+}$  en el tampón. El incremento gradual de la concentración de  $Mg^{2+}$  desde 0 hasta 10 mM provocó una transición paulatina entre los dos patrones de bandas DSB descritos (Figura 28B). Puede observarse como determinadas bandas del patrón DSB en condiciones estándar (10 mM  $Mg^{2+}$ ) van desapareciendo a medida que

disminuimos la concentración de  $Mg^{2+}$ . De forma similar, las bandas que conforman el patrón DSB obtenido sin  $Mg^{2+}$  se difuminan a medida que aumenta la concentración de este catión. A concentraciones de 0.5 y 1 mM  $Mg^{2+}$  se aprecia la coexistencia de bandas características de ambos patrones.

Después de descubrir el efecto de la ausencia de  $Mg^{2+}$ , decidimos investigar si otros componentes del tampón DSB podían también influir en el resultado del ensayo DSB. Para ello, se realizaron ensayos DSB con un tampón que carecía de dNTPs o de ATP. Como se muestra en la Figura 28C, los productos obtenidos en un ensayo DSB sin dNTPs son idénticos a los procedentes de ensayos estándar, mientras que los originados en ensayos DSB sin ATP muestran un grado ligeramente mayor

de degradación, muy probablemente causado por la pérdida de la actividad ligasa que se reconstruye parte de las moléculas cortadas por la actividad endonucleásica. Sin embargo, de todos los componentes del tampón DSB, el  $Mg^{2+}$  fue el único cuya supresión tuvo como resultado un cambio cualitativo tan importante en la formación de productos de degradación.

La generación de patrones DSB distintos provocados por la presencia o ausencia de  $Mg^{2+}$  podría explicarse mediante dos posibles hipótesis: i) la existencia de dos actividades enzimáticas diferentes (una  $Mg^{2+}$ -dependiente y otra  $Mg^{2+}$ -independiente) o la acción de una enzima cuya especificidad cambia en función de la concentración de  $Mg^{2+}$ . En principio, la distinta distribución en

tejidos y en vertebrados (murinos y pollo) de ambas actividades endonucleásicas apoyaría la existencia de dos actividades enzimáticas independientes.

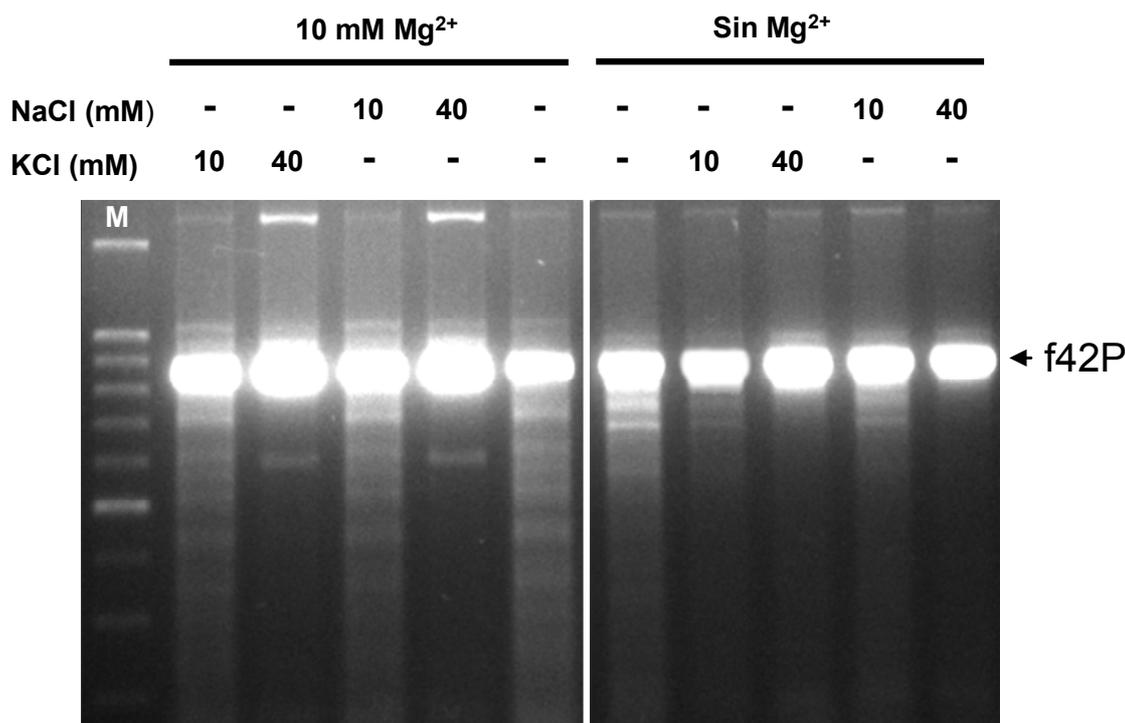
**Caracterización de las actividades endonucleásicas en el ensayo DSB**

El ensayo DSB estaba mostrando la presencia de dos actividades independientes generadoras de DSBs en los extractos nucleares analizados, por lo cual nos propusimos caracterizar con mayor detalle tales actividades. Para ello, se llevaron a cabo una serie de ensayos DSB en los que se modificaron las condiciones experimentales respecto a la reacción estándar (Tabla 2). Uno de

**Tabla 6.** Comparación de las dos actividades generadoras de DSBs. El efecto de las concentraciones de EGTA, NEM, NaCl y KCl y el calentamiento de los extractos nucleares antes del ensayo DSB fue determinado bajo condiciones estándar con  $Mg^{2+}$  o en ausencia de  $Mg^{2+}$ .

|                            |      | Ensayo con $Mg^{2+}$<br>(10 mM) | Ensayo sin $Mg^{2+}$ |
|----------------------------|------|---------------------------------|----------------------|
| EGTA (mM)                  | 1    | N                               | L                    |
|                            | 10   | N                               | I                    |
|                            | 50   | I                               | I                    |
| NEM (mM)                   | 0.5  | I                               | N                    |
|                            | 5    | I                               | N                    |
|                            | 50   | I                               | N                    |
| NaCl (mM)                  | 10   | N                               | L                    |
|                            | 40   | I                               | I                    |
| KCl (mM)                   | 10   | N                               | L                    |
|                            | 40   | I                               | I                    |
| Calentamiento del extracto | 50°C | L                               | N                    |
|                            | 60°C | I                               | L                    |

N, actividad normal; L, actividad disminuida; I, actividad inhibida.

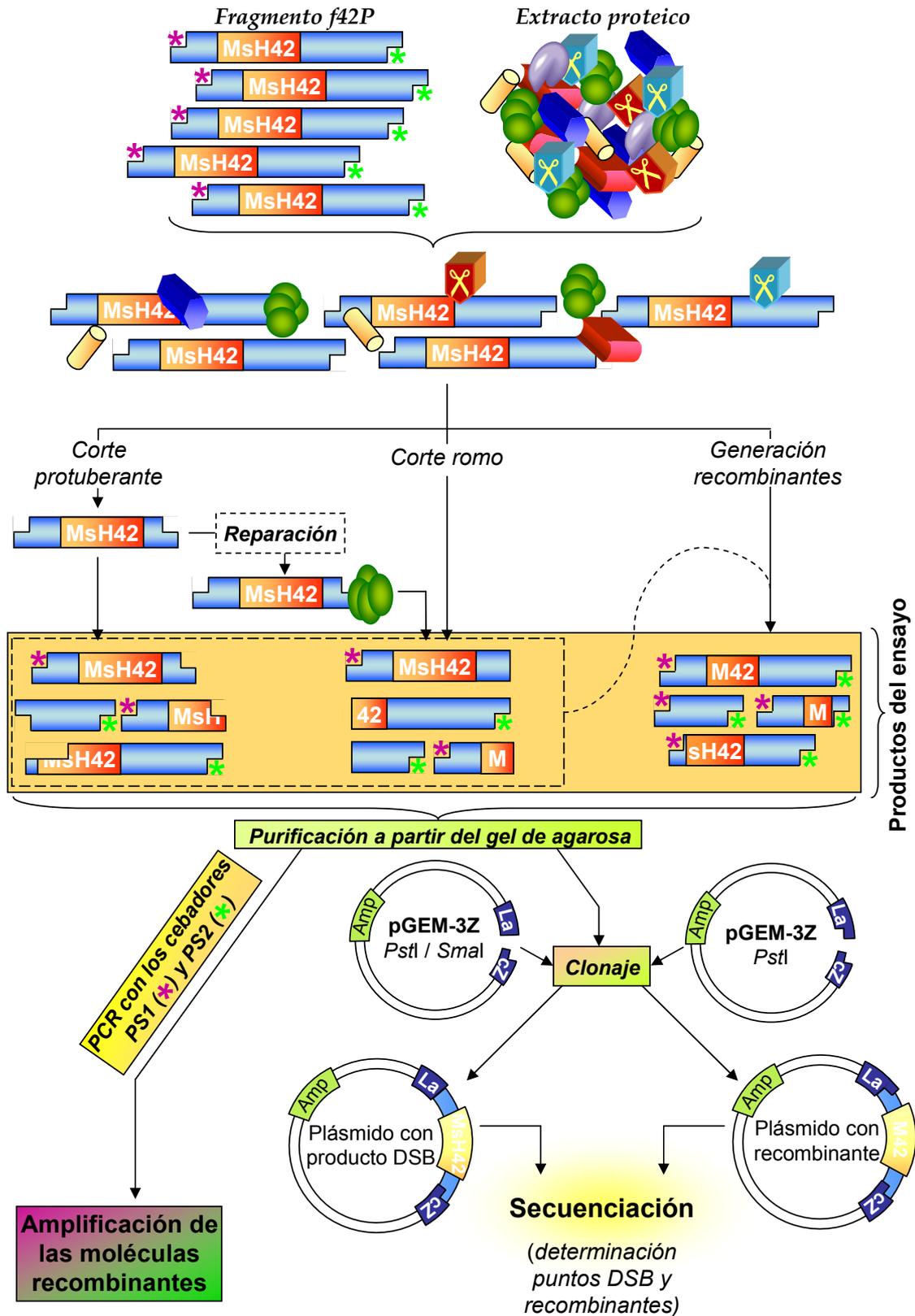


**Figura 29.** Efecto de la fuerza iónica sobre las actividades generadoras de DSBs. Se muestra el análisis electroforético en geles del 2% de una serie de ensayos DSB en presencia de NaCl y KCl. En el panel de la izquierda se muestra el resultado de la adición de NaCl o KCl a las concentraciones indicadas en ensayos DSB estándar (10 mM Mg<sup>2+</sup>). En el de la derecha, se muestra el resultado de la incorporación de estas mismas sales a ensayos DSB sin Mg<sup>2+</sup>. M, marcador 100 bp ladder de Promega.

los ensayos consistió en la adición a la reacción de EGTA, un quelante específico de Ca<sup>2+</sup>, a concentraciones finales de 1, 10 y 50 mM. Este compuesto inhibió la actividad independiente de Mg<sup>2+</sup>, sugiriendo que esta actividad pueda requerir Ca<sup>2+</sup>, aunque su presencia apenas afectó a la Mg<sup>2+</sup>-dependiente. Otra diferencia destacable entre ambas actividades DSB se reveló mediante el tratamiento de los extractos nucleares con NEM, un agente alquilante de radicales sulfhidrilo capaz de modificar los residuos de cisteína de las proteínas. Así, la adición de NEM a una concentración de 0.5 mM eliminó la actividad formadora de DSBs dependiente de Mg<sup>2+</sup>, mientras que la actividad independiente de Mg<sup>2+</sup> permaneció inalterada incluso a 50 mM. Otra característica que permitió distinguir ambas actividades fue su resistencia al calor. Así, los extractos se sometieron a una preincubación de 5 min

a 50 ó 60°C, antes de su adición al ensayo DSB. De esta forma, la actividad dependiente de Mg<sup>2+</sup> resultó inactivada completamente por la preincubación a 60°C, mientras que la Mg<sup>2+</sup>-independiente todavía se mantuvo activa. Los resultados de esta serie de ensayos DSB (resumidos en la Tabla 6) tomados en su conjunto reforzaron la existencia de dos actividades enzimáticas diferentes.

La ausencia del patrón DSB correspondiente a la reacción independiente de Mg<sup>2+</sup> en aquellos ensayos que sí contenían este catión era un tanto sorprendente si pensamos en dos actividades enzimáticas independientes. En este sentido, en las reacciones con 10 mM Mg<sup>2+</sup> cabría esperar que se formase un patrón DSB compuesto por bandas procedentes de la actividad de ambas enzimas, mientras que en ausencia de este catión debería



observarse la desaparición de aquellas bandas correspondientes a la actividad dependiente de  $Mg^{2+}$ . Sin embargo, como se mostró en la Figura 28B, los

patrones DSB cambian completamente en presencia (10 mM) o ausencia de  $Mg^{2+}$ . ¿Podría ocurrir que los cambios que estábamos observando al retirar el

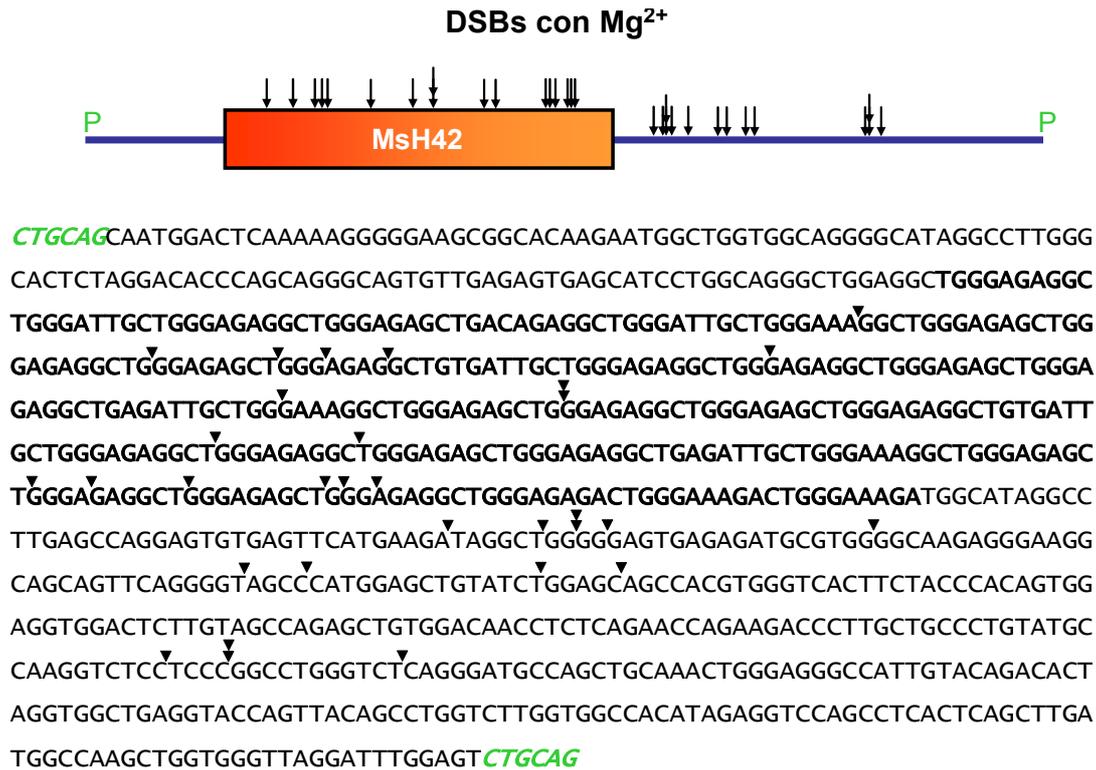
**Figura 30.** Estrategia seguida para la clonación de productos originados en el ensayo DSB con el fragmento f42P. Se asume que al menos una fracción de las moléculas originadas tras un DSB mantendrá uno de sus extremos *PstI* inalterado, mientras que el extremo generado por la actividad endonucleásica podría ser roto debido a la propia actividad endonucleásica o por el procesado de extremos protuberantes. Estas moléculas serían ligables en un vector con extremos *PstI* y *SmaI*. El caso de los recombinantes, se asume que algunos de ellos mantendrán sus extremos originales intactos. Esto permitirá su detección por dos métodos diferentes: i) Amplificación mediante los cebadores PS1 (asterisco morado) y PS2 (asterisco verde), cuya secuencia se halla en los extremos de la región Msh42 y ii) clonaje en un vector linearizado con *PstI*. El inserto de los plásmidos recombinantes se secuenció con los cebadores universales M13 para determinar tanto los puntos de corte en los productos DSB como la secuencia de los recombinantes.

$Mg^{2+}$  del tampón DSB se debiesen a que ambas actividades enzimáticas tuviesen una sensibilidad diferente al descenso de la fuerza iónica en la reacción?

Para analizar esta posibilidad, realizamos un conjunto de ensayos DSB en los cuales se sustituía el 10 mM de  $Mg_2SO_4$  del tampón de reacción por NaCl o KCl bien a la misma molaridad (10 mM) o a una concentración que se correspondiese con la misma fuerza iónica (40 mM). Estas mismas concentraciones de NaCl o KCl se añadieron a reacciones DSB estándar para ver su efecto sobre la reacción  $Mg^{2+}$ -dependiente. Como se muestra en la Figura 29, la actividad independiente de  $Mg^{2+}$  se inhibía fuertemente por  $Na^+$  o  $K^+$  a concentraciones de 10 mM, siendo totalmente eliminada a 40 mM. Por el contrario, la actividad  $Mg^{2+}$ -dependiente no mostraba alteraciones aparentes a una concentración 10 mM, aunque a 40 mM se encontraba muy inhibida. Estos resultados permitieron extraer las siguientes conclusiones acerca de estas dos actividades generadoras de DSBs: i) la actividad independiente de  $Mg^{2+}$  se inhibe fuertemente por un aumento en la fuerza iónica del tampón de reacción, lo que explica la no existencia de bandas generadas a partir de esta actividad en el ensayo estándar DSB (10 mM  $Mg^{2+}$ ); ii) el aumento de la fuerza iónica en una reacción sin  $Mg^{2+}$  no restaura el patrón DSB de la actividad dependiente de

$Mg^{2+}$ , lo que verifica el requerimiento específico de este ión; iii) la actividad dependiente de  $Mg^{2+}$  es notablemente más tolerante a variaciones en la fuerza iónica del tampón de reacción que la actividad independiente de  $Mg^{2+}$ . Hay que tener en cuenta que la actividad dependiente de  $Mg^{2+}$  soporta tanto la adición de concentraciones moderadas de NaCl o KCl (10 mM) en un tampón que ya de por sí contiene 10 mM de  $Mg_2SO_4$ , lo que equivaldría a una fuerza iónica un 25% mayor con respecto a la reacción estándar, como un descenso del  $Mg_2SO_4$  a 2.5 mM, lo que implica una fuerza iónica un 75% menor que en las condiciones estándar (Figura 28A).

Recapitulando acerca de las propiedades que hemos estudiado para las actividades generadoras de DSB observadas en nuestro ensayo ( $Mg^{2+}$ -dependiente y  $Mg^{2+}$ -independiente), hemos constatado que, además de en su dependencia por el  $Mg^{2+}$ , se diferencian en las siguientes características: i) la distribución en tejidos y especies; ii) la susceptibilidad a NEM (existencia de grupos sulfhidrilo accesibles); iii) la susceptibilidad a EGTA (dependencia de  $Ca^{2+}$ ); iv) inhibición por NaCl y KCl y v) inactivación por calor. Este conjunto de datos apoya la hipótesis de que los patrones DSB observados en presencia y ausencia de  $Mg^{2+}$  son consecuencia de la acción de dos enzimas independientes entre sí.



**Figura 31.** Distribución de los DSBs a lo largo de la región Msh42 generados en el ensayo DSB estándar. La caja anaranjada representa al minisatélite Msh42, cuya secuencia se muestra en negrita en la parte inferior. Los puntos de corte se representan como flechas sobre el esquema y como puntas de flecha sobre la secuencia. Los dianas *PstI* se indican en verde.

## ***Análisis de los productos DSB***

### **Actividad Mg<sup>2+</sup>-dependiente**

Una vez analizadas algunas de las propiedades de las dos actividades endonucleásicas detectadas en el ensayo DSB, nos planteamos la caracterización de los productos de digestión derivados de estas actividades enzimáticas para averiguar las secuencias que estaban siendo reconocidas específicamente. Para abordar este estudio, asumimos que al menos algunas de las moléculas originadas a partir de un único DSB sobre el sustrato f42P (en adelante, *productos DSB*) mantendrían su extremo *PstI* intacto, mientras que el punto de corte producido por la actividad endonucleásica sería romo, bien porque la propia actividad DSB genera este tipo de extremos o porque otras proteínas del extracto puedan procesarlo, eliminando extremos protuberantes (Figura 30).

Los productos del ensayo DSB se purificaron a partir de las bandas correspondientes del gel de agarosa y se clonaron en el vector pGEM-3Z linearizado con *PstI* y *SmaI* (esta enzima genera extremos romos). Los clones recombinantes se secuenciaron para determinar el lugar exacto de corte. En la Figura 31, se muestra un mapa de la región Msh42 donde se indican los puntos de corte correspondientes a los clones recombinantes aislados a partir de ensayos DSB estándar y que, por lo tanto reflejan la actividad de la nucleasa Mg<sup>2+</sup>-dependiente. De los 28 puntos DSB analizados, 16 están localizados dentro del minisatélite y 12 en la región flanqueante 3' de Msh42. Un número significativo de puntos DSB (~ 60%) estaban situados dentro o flanqueando motivos TGGG (Figura 31), especialmente abundantes en la secuencia minisatélite. Cabe destacar que varios DSBs localizados fuera del

minisatélite también estaban asociados a motivos similares al TGGG. Concretamente, 4 de ellos se dieron en dos secuencias con más residuos de guanina (TGGGG y TGGGGG), que se encuentran situadas en la región flanqueante 3' de MsH42 (Figura 31).

El análisis de las secuencias de los puntos de corte detectados a lo largo de la región MsH42 para la actividad generadora de DSBs dependiente de  $Mg^{2+}$  indicaba que su diana estaba relacionada con motivos ricos en guanina. Para confirmar esta preferencia por secuencias ricas en guanina diseñamos una batería de oligonucleótidos de 50 pb que contienen el motivo TGGGG o variaciones de éste y de sus secuencias flanqueantes (Tabla 1). Estos oligonucleótidos se emplearon como sustrato en ensayos DSB para averiguar si dicho motivo era reconocido como diana por la actividad endonucleásica  $Mg^{2+}$ -dependiente. En estos ensayos se empleó el extracto de cerebro de pollo por ser uno de los que produce un menor ruido de fondo y, por lo tanto, un patrón de bandas más definido. Cuando se realizó el ensayo DSB empleando el oligonucleótido F1, que contiene el motivo TGGGG en las posiciones 29-33 a partir de su extremo 5', se generaron dos bandas prominentes (Figura 32, carril F1). Estas bandas presentaban la misma movilidad electroforética que los fragmentos de 19 y 31 pb producidos por la digestión con *SmaI* del oligonucleótido M16, que contiene un lugar de restricción para esta enzima en la posición 29-34 (carril M). Este resultado demuestra que el DSB está teniendo lugar principalmente entre los residuos segundo y tercero del motivo TGGGG en el oligonucleótido F1. Si se altera la posición de este motivo, como en el caso del oligonucleótido F2 donde el motivo TGGGG ocupa las posiciones 25-29 pb, el patrón de bandas generado se

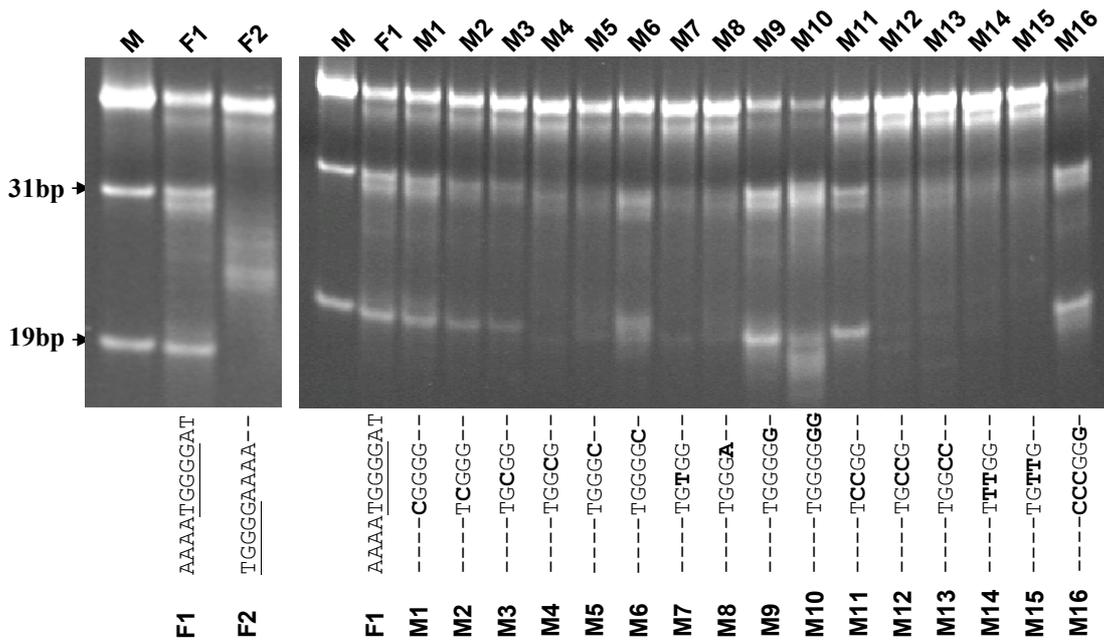
modifica no sólo en el tamaño de sus bandas, sino también en su definición (Figura 32, carril F2). Estos resultados ponen de manifiesto que el motivo TGGGG forma parte de la secuencia de reconocimiento de la actividad generadora de DSBs  $Mg^{2+}$ -dependiente. El hecho de que la precisión en el corte sea menor cuando se desplaza la diana TGGGG en el oligonucleótido F2 sugiere que la actividad enzimática puede estar siendo influenciada por la composición del DNA adyacente a dicho motivo.

Para analizar de qué manera influyen las variaciones en el TGGGG en la generación de los DSBs, se emplearon una serie de sustratos en los que se introdujeron determinadas mutaciones en su secuencia (Figura 32 y Tabla 1). Los resultados obtenidos en los ensayos con los oligonucleótidos M1-M5 mostraron que mutaciones en el extremo 3' de este motivo reducían la actividad DSB de forma más eficaz que aquellas localizadas en su extremo 5', lo que denota la existencia de una polaridad en el reconocimiento del motivo por parte de la actividad enzimática. Esta polaridad también se observó en los experimentos con los oligonucleótidos M11-M13, en los cuales se introdujeron dos mutaciones en el motivo TGGGG. Dos transversiones G→C en las dos primeras guaninas apenas redujeron la tasa de corte, mientras que si estos cambios se realizaban en la segunda y tercera o bien en la tercera y cuarta guanina, la actividad cae bruscamente (carriles M11-M13). También hay que destacar que el tipo de mutación en una posición determinada influye en la actividad formadora de DSBs. Así, los cambios G→C son más tolerados por esta actividad que los cambios G→T (comparar carriles M3 con M7 y M11 con M14), indicando que esta actividad puede verse favorecida por la existencia de residuos de guanina en ambas

cadena. Esta hipótesis se ve reforzada por el hecho de que las secuencias palindrómicas con residuos de guanina a 3' son buenos sustratos para la actividad que produce los DSBs (ver M11 y M16). La adición de un residuo de guanina en el extremo 3' del motivo TGGGG (M9), provoca un incremento en la tasa de corte, confirmando la preferencia de este enzima por tramos de guaninas. Cuando se incorpora otro residuo más de guanina (M10) se potencia más la tasa de generación de DSBs, aunque la precisión del enzima disminuye, ya que se observa una difuminación del principal producto de corte.

DSB dependiente de  $Mg^{2+}$  en la región MsH42 junto con los obtenidos mediante la mutagénesis dirigida del motivo TGGGG en los oligonucleótidos de 50 pb revelaron las siguientes características de dicha actividad: i) el corte se produce mayoritariamente entre la segunda y la tercera guanina del motivo TGGGG; ii) la tasa de corte aumenta al incrementarse el número de residuos de guanina en el motivo, aunque la precisión del corte disminuye, sugiriendo que la diana de reconocimiento de esta enzima es relativamente pequeña, probablemente menor de 6 pb, con lo que tramos largos de guaninas contendrían múltiples dianas potenciales; iii) las mutaciones en la zona 3' del motivo afectan más a la tasa

Los resultados acerca de la localización de los puntos de corte de la actividad



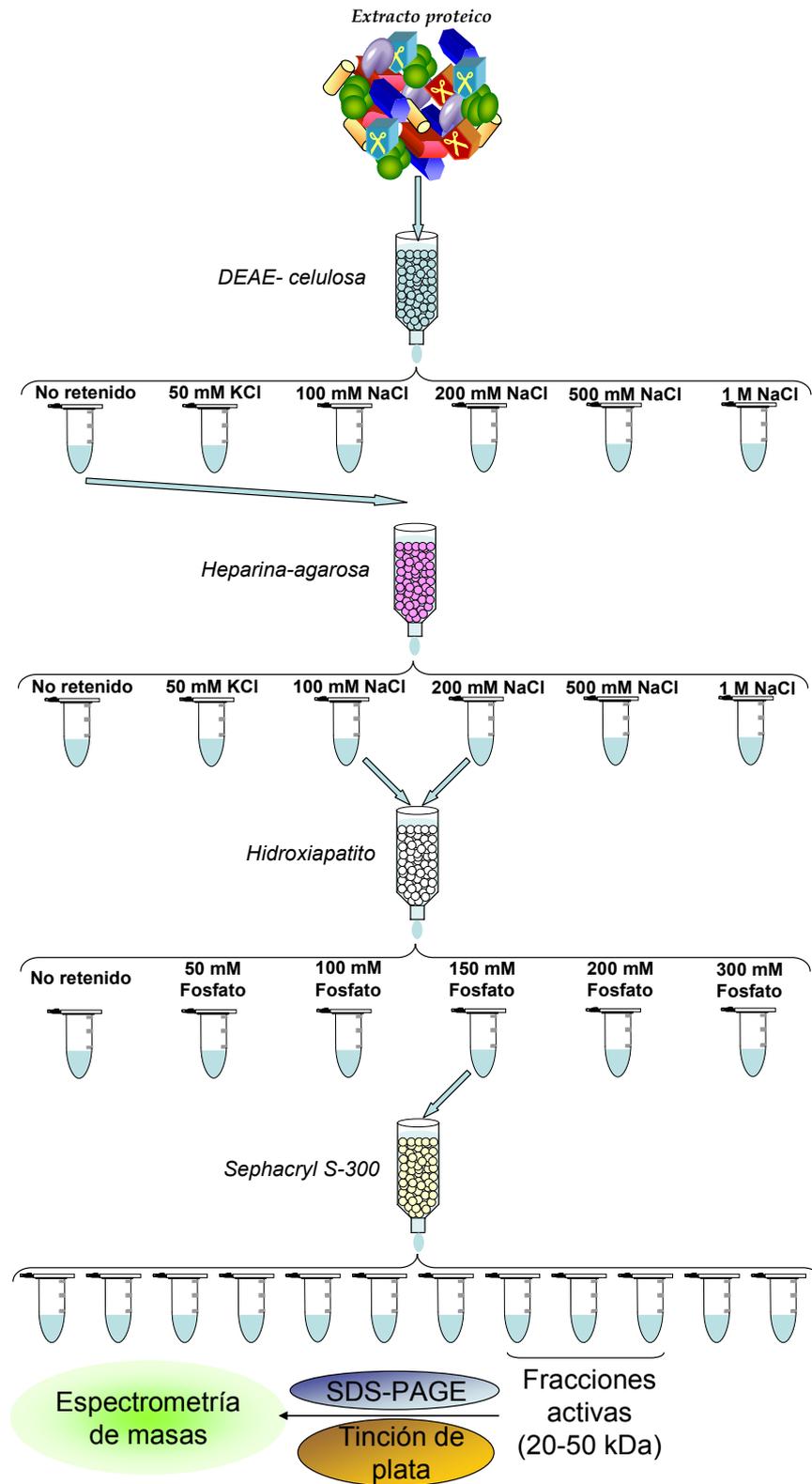
**Figura 32.** Caracterización de la secuencia de reconocimiento de la actividad  $Mg^{2+}$ -dependiente. Las imágenes muestran el análisis electroforético en gel de acrilamida al 18% de los productos obtenidos en ensayos DSB en condiciones estándar y empleando como sustratos los oligonucleótidos de 50 pb mostrados en la Tabla 1. En el panel de la izquierda se muestra el efecto del desplazamiento del motivo TGGGG desde su posición original en el oligonucleótido F1 a una ubicación más central y con un cambio en la secuencia flanqueante (F2). En el panel de la derecha se muestra una serie de experimentos utilizando como sustrato el oligonucleótido F1 y un conjunto de sustratos derivados de él que contienen una o más modificaciones en el motivo TGGGG. En la parte inferior de la figura se muestra la secuencia comprendida entre las posiciones 25-35 de cada oligonucleótido. El resto de la secuencia de estos oligonucleótidos es idéntica entre todos ellos se encuentra descrita en la Tabla 1. El motivo TGGGG se haya subrayado en los oligonucleótidos F1 y F2. En los oligos M1-M16, se indican en negrita las mutaciones con respecto al oligonucleótido F1. M es un marcador de peso molecular generado mediante la digestión de F16 con *SmaI*, produciendo dos fragmentos de 19 y 31 pb.

de corte que las localizadas en su extremo 5', sugiriendo que existe una polaridad en el reconocimiento de la secuencia; iv) existe un aumento de la actividad de corte cuando hay presentes secuencias palindrómicas compuestas por GC que mantengan al menos dos residuos de guanina en el extremo 3'. Dicho incremento en la formación de DSBs podría deberse al propio reconocimiento de la estructura del palíndromo por parte de la actividad o a que el palíndromo genera dianas en ambas cadenas del dúplex con un lugar de corte común.

Uno de los objetivos que nos propusimos tras observar los resultados del ensayo DSB estándar fue la purificación de la actividad endonucleásica dependiente de  $Mg^{2+}$ . Para ello, optamos por una serie de procesos de separación cromatográfica empleando como fuente proteica los extractos nucleares de testículo de rata (Figura 33). Debido a que los resultados obtenidos no fueron concluyentes, describiremos brevemente los pasos seguidos en este intento de purificación. El primer paso consistió en pasar el extracto por una columna de DEAE-celulosa equilibrada en el tampón de diálisis empleado durante la preparación del extracto, pero con una concentración de 50 mM KCl. La actividad nucleásica dependiente de  $Mg^{2+}$  no fue retenida por esta columna y se produjo una gran pérdida de actividad (> 90%). A continuación, esta fracción no retenida se introdujo en una columna de heparina-agarosa equilibrada en el mismo tampón. En este caso, la actividad fue eluida mediante un lavado de la columna con tampón de diálisis con un gradiente de 100-200 mM KCl. Esta segunda fracción fue posteriormente diluida 1:1 con 10 mM tampón fosfato potásico pH 6.8 y cargada en una columna de hidroxapatito equilibrada en 10 mM tampón fosfato potásico pH 6.8 y 10% glicerol. La actividad se recuperó a

~ 100 mM tampón fosfato con 10% glicerol. Finalmente esta última fracción obtenida se separó en una columna de Sephacryl S300 para estimar el tamaño aproximado de la actividad nucleásica. Esta apareció repartida en diversas fracciones en torno a un tamaño relativamente pequeño de 20 a 50 kDa. Además, eventualmente se detectaron fracciones activas de mayor tamaño, lo que sugería que la proteína responsable de la generación de DSBs podría existir en un estado dimérico o multimérico. Llegados a este punto y tras haber fraccionado el extracto mediante cuatro columnas diferentes (DEAE-celulosa, heparina-agarosa, hidroxapatito y Sephacryl S-300), decidimos analizar el grado de pureza de las fracciones con más actividad, para lo cual dichas fracciones se concentraron, se separaron mediante SDS-PAGE al 14% y se tiñeron con plata. Las fracciones más activas coincidían con las que contenían mayor cantidad de proteína y mostraban todavía un gran número de bandas. De todas formas, decidimos analizar mediante espectrometría de masas aquellas bandas que parecían seguir mejor el perfil de la actividad nucleásica. Entre las proteínas detectadas se hallaban determinadas proteínas de unión a RNA, la endonucleasa apurínica/apirimidínica, una proteína de ratón homóloga a Piwi, etc. Ninguna de ellas parecía una buena candidata para ser la responsable de la generación de DSBs observada en nuestro ensayo. No pudimos avanzar más en la purificación puesto que las pérdidas en actividad durante los procesos cromatográficos resultaron tan grandes que dejamos de detectarla rápidamente.

Para intentar identificar la actividad  $Mg^{2+}$ -dependiente, realizamos una revisión bibliográfica de las endonucleasas descritas en la literatura comparando sus propiedades con las de otras endonucleasas de mamíferos que



**Figura 33.** Esquema de los procesos cromatográficos empleados en el intento de purificación de la endonucleasa dependiente de  $Mg^{2+}$ .

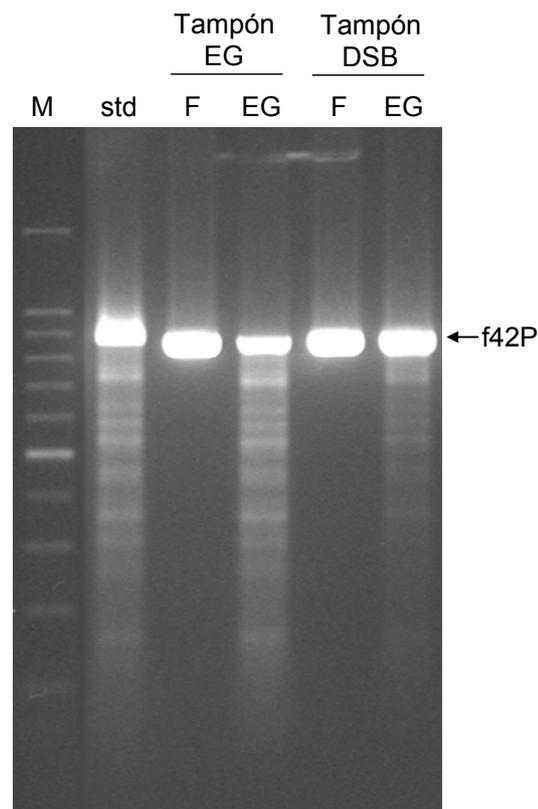
cortan específicamente secuencias ricas en guaninas. Se ha descrito una actividad endonucleásica en tejidos de mono y toro con cierta especificidad por residuos

de A y G. Esta proteína exhibe un pH óptimo neutro, siendo fuertemente inhibida por concentraciones  $>25$  mM de NaCl (Brown et al., 1978; McKenna et

al., 1981). Tanto el tamaño de esta proteína (dímero de 130 kDa) como su requerimiento de DTT o  $\beta$ -mercaptoetanol para una actividad óptima, la excluyen como la proteína que estábamos observando. Existe otra endonucleasa específica de residuos de guanina (endonucleasa mitocondrial bovina), purificada a partir de mitocondrias de corazón de vaca, que presenta un tamaño de 29 kDa, su forma activa es un dímero y con un doble pH óptimo a 5.4 y 7.5 (Low et al., 1987; Cummings et al., 1987). Aunque las propiedades de esta enzima eran relativamente parecidas a las que nosotros conocíamos para la actividad presente en nuestro extracto, tanto su origen mitocondrial, como su doble pH óptimo parecían descartarla. Endo R es otra endonucleasa que presenta notables similitudes con nuestra actividad, aunque era inactivada durante procesos cromatográficos en hidroxipatito o heparina-agarosa (Gottlieb & Muzyczka, 1988, 1990), dos matrices que eran las que mejores resultados daban en nuestro caso. Por esta razón, tampoco esta enzima parecía una firme aspirante para ser la actividad generadora de DSBs que estábamos detectando. Otras endonucleasas de mamíferos ampliamente conocidas, tal como DNasaII (Evans & Aguilera, 2003), las proteínas RAG (Gellert, 2002) o Artemis (Ma et al., 2002) mostraban ya grandes diferencias con nuestra actividad, tanto por sus requerimientos de cofactores como por el tipo de sustrato que procesan.

Una de las proteínas nucleares con actividad endonucleásica más abundantes en las células es la denominada Endo G (Ruiz-Carrillo & Renaud, 1987; Cote et al., 1989). Este enzima reconoce preferentemente secuencias ricas en guanina, presenta un tamaño de 26 kDa, dentro rango que nosotros habíamos estimado y se

comporta de forma análoga durante los procesos cromatográficos empleados. Puesto que las características y comportamiento de dicha endonucleasa encajaban bien con nuestras observaciones experimentales, digerimos nuestro fragmento f42P con esta enzima para ver si el patrón de bandas DSB se correspondía con el que ya conocíamos. Para ello empleamos Endo G



**Figura 34.** Actividad de Endo G sobre el fragmento f42P. En la imagen se muestra el análisis electroforético en gel de agarosa al 2% de los productos de digestión del fragmento f42P después de la incubación con Endo G. Tampón EG indica el tampón descrito para Endo G por Schäfer et al. (2004): 10 mM acetato sódico, 10 mM MES, 20 mM Tris, 2.5 mM MgCl<sub>2</sub>, pH ajustado a 6.0 con KOH. Tampón DSB es el empleado en las condiciones estándar del ensayo DSB. F, fragmento f42P sin extracto ni Endo G; EG reacciones con 5  $\mu$ l de la fracción purificada de Endo-G recombinante; std, reacción estándar del ensayo DSB; M, marcador 100 pb ladder de Promega.

recombinante, amablemente cedida por el Dr. Gregor Meiss (Schäfer et al., 2004), que incorporamos al ensayo DSB en lugar del extracto proteico nuclear, tanto con el tampón DSB habitual como con el especificado por estos autores para esta proteína. El resultado de estos experimentos mostró claramente que los patrones de corte generados por el extracto y Endo G eran el mismo (Figura 34), lo cual es una prueba muy importante a favor de que sea esta endonucleasa la responsable de la mayor parte de los DSBs que se producen en nuestro sistema *in vitro* bajo condiciones estándar. Ruiz-Carrillo & Renaud (1987) han sugerido la posible implicación de Endo G en procesos recombinatorios, basándose en las pruebas experimentales que ligaban la presencia de secuencias tipo minisatélite, de las que muchas son ricas en guanina, con algunos puntos calientes de recombinación. Ahora, casi 20 años después, hemos visto como esta nucleasa es capaz de digerir eficazmente la secuencia de la región minisatélite Msh42, relacionada con procesos recombinatorios *in vitro* (Boán et al, 1998, 2002). Por tanto, es posible que la actividad de corte de Endo G sea la que promueva la generación de DSBs en el entorno de la región Msh42 que provoquen los eventos de recombinación descritos anteriormente.

#### Actividad $Mg^{2+}$ -independiente

El análisis de los productos DSB originados en el ensayo sin  $Mg^{2+}$  se realizó siguiendo la misma estrategia que en el caso del ensayo estándar. Contrastando con los resultados fruto de la actividad  $Mg^{2+}$ -dependiente, la mayoría de los DSBs generados por la actividad  $Mg^{2+}$ -independiente (15 de 16) se localizaron en las regiones flanqueantes de Msh42, fuera de la región repetitiva (Figura 35). El análisis de la secuencia de dichos puntos de corte no reveló ningún consenso bien definido para esta actividad endonucleolítica.

Algunos DSBs (11 de 16) aparecieron en motivos con 2 ó mas guaninas consecutivas, pero en contextos de secuencia diferentes. También pudo observarse una cierta correlación de los cortes con motivos CAG, pero esta asociación podría explicarse simplemente por la abundancia de este motivo en las secuencias flanqueantes de Msh42.

En un intento de dilucidar cuáles podrían ser los motivos que estaban siendo reconocidos por la actividad independiente de  $Mg^{2+}$  para la generación de los DSBs, hicimos una serie de ensayos con oligonucleótidos de 50 pb de manera análoga a la empleada en el estudio de la actividad  $Mg^{2+}$ -dependiente. Debido a que el mapa de los DSBs no nos permitió descifrar un motivo claro que sirviese de referencia inicial para comenzar el análisis, empleamos una secuencia de 50 pb similar a una porción de la región flanqueante 3' de Msh42 en la cual detectábamos varios puntos de corte (oligonucleótido S1, ver Tabla 1 y Figura 35). Se realizó un ensayo DSB empleando este oligonucleótido y una variante del mismo en la cual se cambiaba una secuencia CAG por TTT (S2, ver Tabla 1). Este experimento resultó en un patrón de bandas con tamaños que podían explicarse por cortes en motivos ricos en guanina. Además, este cambio entre los sustratos S1 y S2 se traducía en una diferencia muy evidente en el tamaño de una de estas bandas (18-19 pb, ver Figura 36A). Estos dos hechos parecían apuntar en la misma línea que el mapeo que habíamos realizado mediante el clonaje de los productos DSB.

En vista de estos resultados, diseñamos una serie de oligonucleótidos (S3-S9) en los que se incluían motivos CAG y CTG a distancias determinadas (1 CAG en S3, CAGCAG en S4, CAGCTG en S5,

CAG-CAG separados por 6 pb en S6, CAG-CTG separados por 6 pb en S7, CAG-CAG separados por 10 pb en S8 y CAG-CTG separados por 10 pb en S9; Tabla 1). Además, S3-S5 portaban un motivo TGGGG centrado en la posición 31 del oligonucleótido. El resultado del ensayo DSB con esta serie de sustratos mostró que el patrón de bandas era el mismo al emplear los sustratos S3-S5, presentando una pareja de bandas prominentes a tamaños de ~ 20 y 30 pb, que podrían estar siendo generadas por cortes en el motivo TGGGG. En este sentido, los sustratos S6-S9, que no poseían este motivo, no generaban este par de bandas. Sin embargo, S6 y S7 sustratos que sólo se diferencian en el cambio de un motivo CAG por CTG, de nuevo presentaban patrones de bandas ligeramente diferentes (Figura 36A), indicando que estos motivos podrían

estar implicados en la generación de al menos parte de los cortes que estábamos observando.

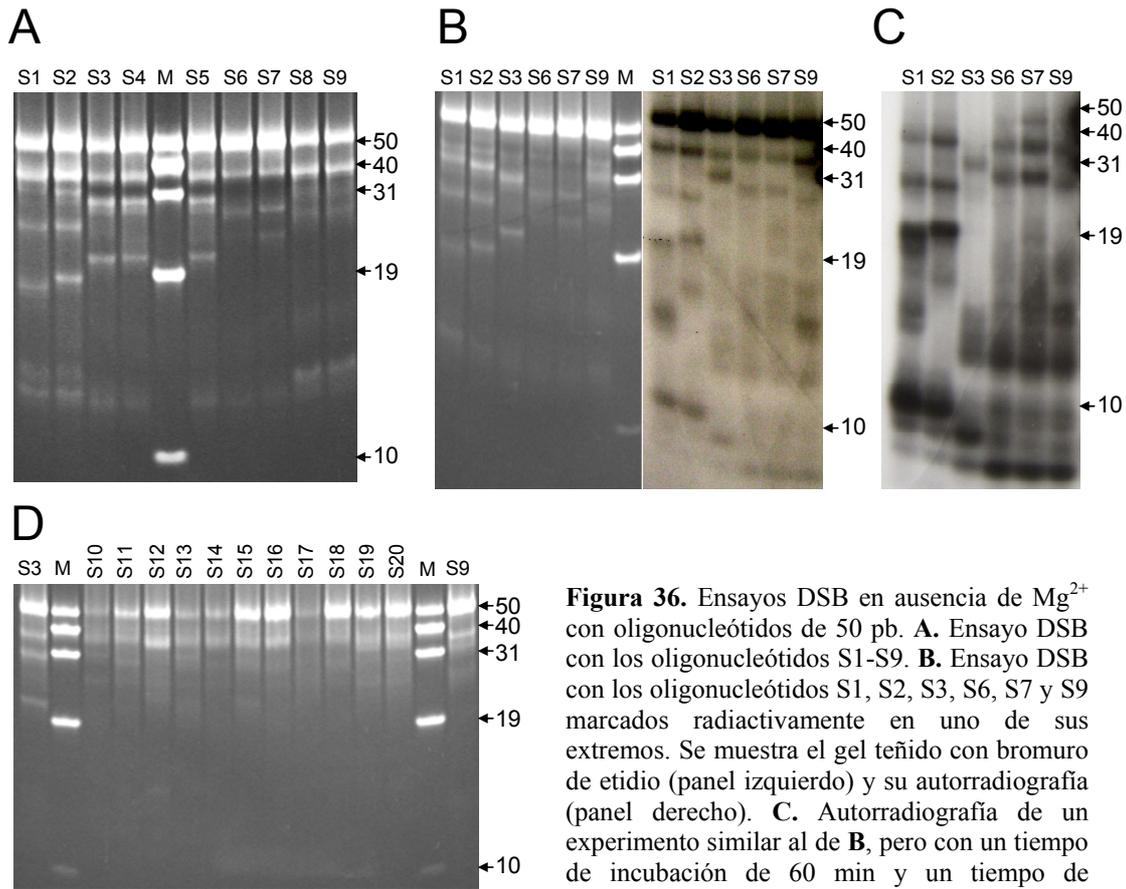
Para determinar la orientación con la que se estaban produciendo los DSBs en los sustratos, marcamos terminalmente uno de los oligonucleótidos de cadena sencilla que componen cada uno de los sustratos de 50 pb (concretamente el que se muestra en la Tabla 1) y construimos el sustrato de doble cadena. Se incorporaron 5 ng de este sustrato marcado en uno solo de sus extremos a los 1.5 µg de sustrato no marcado habitual y se realizó el ensayo DSB en condiciones de ausencia de Mg<sup>2+</sup>. El resultado de estos experimentos con los oligonucleótidos S1, S2, S3, S6, S7 y S9 se muestra en las Figuras 36B y 36C. Como se puede comprobar, S1 y S2 son cortados idénticamente en la posición

### DSBs sin Mg<sup>2+</sup>



**CTGCAG**CAATGGACTCAAAAAGGGGAAGCGGCACAAGAATGGCTGGTGGCAGGGGCATAGGCCTTGGG  
 CACTCTAGGACACCCAGCAGGGCAGTGTGAGAGTGAGCATCCTGGCAGGGCTGGAGGCTGGGAGAGGC  
 TGGGATTGCTGGGAGAGGCTGGGAGAGCTGACAGAGGCTGGGATTGCTGGGAAAGGCTGGGAGAGCTGG  
 GAGAGGCTGGGAGAGCTGGGAGAGGCTGTGATTGCTGGGAGAGGCTGGGAGAGGCTGGGAGAGCTGGGA  
 GAGGCTGAGATTGCTGGGAAAGGCTGGGAGAGCTGGGAGAGGCTGGGAGAGCTGGGAGAGGCTGTGATT  
 GCTGGGAGAGGCTGGGAGAGGCTGGGAGAGCTGGGAGAGGCTGAGATTGCTGGGAAAGGCTGGGAGAGC  
 TGGGAGAGGCTGGGAGAGCTGGGAGAGGCTGGGAGAGACTGGGAAAGACTGGGAAAGATGGCATAGGCC  
 TTGAGCCAGGAGTGTGAGTTCATGAAGATAGGCTGGGGGAGTGAGAGATGCGTGGGGCAAGAGGGAAGG  
 CAGCAGTTCAGGGGTAGCCCATGGAGCTGTATCTGGAGCAGCCACGTGGGTCACTTCTACCCACAGTGG  
 AGGTGGACTCTTGTAGCCAGAGCTGTGGACAACCTCTCAGAACCAGAAGACCCTTGCTGCCCTGTATGC  
 CAAGGTCTCCTCCCGGCTGGGTCTCAGGGATGCCAGCTGCAAACTGGGAGGGCCATTGTACAGACACT  
 AGGTGGCTGAGGTACAGTTACAGCCTGGTCTTGGTGGCCACATAGAGGTCCAGCCTCACTCAGCTTGA  
 TGGCCAAGCTGGTGGGTTAGGATTTGGAGT **CTGCAG**

**Figura 35.** Distribución de los DSBs generados en el ensayo DSB sin Mg<sup>2+</sup> a lo largo de la región MsH42. La caja anaranjada representa al minisatélite MsH42, cuya secuencia se muestra en negrita en la parte inferior. Los puntos de corte se representan como flechas sobre el esquema y como puntas de flecha bajo la secuencia. Los dianas *PstI* se indican en verde. En azul se remarca la región que se empleó para diseñar el oligonucleótido S1.



**Figura 36.** Ensayos DSB en ausencia de  $Mg^{2+}$  con oligonucleótidos de 50 pb. **A.** Ensayo DSB con los oligonucleótidos S1-S9. **B.** Ensayo DSB con los oligonucleótidos S1, S2, S3, S6, S7 y S9 marcados radiactivamente en uno de sus extremos. Se muestra el gel teñido con bromuro de etidio (panel izquierdo) y su autorradiografía (panel derecho). **C.** Autorradiografía de un experimento similar al de **B**, pero con un tiempo de incubación de 60 min y un tiempo de exposición mayor para poner de manifiesto las bandas menos intensas. **D.** Ensayo DSB con los

oligonucleótidos S10-S20, incluyendo a S3 y S9 como referencia. Todos los ensayos se han realizado con extracto de bazo de rata y tiempo de incubación de 30 min, excepto en **C**, 45 min. Las muestras se analizaron en geles de poli-acrilamida al 18% y en **B** y **C** fueron autorradiografiados. La secuencia de todos estos oligonucleótidos se encuentra en la Tabla 1. M, marcador de peso molecular generado mediante la digestión de los oligonucleótidos M16 y M17 con *Sma*I.

10, donde existe un motivo TGGG, y de forma diferente en la zona de las 19 y 20 pb, donde la sustitución de CAG por TTT deshace un tramo de tres guaninas consecutivas. Otro prominente corte en la zona de las 36-38 pb también recae en un motivo rico en guaninas. En el caso de los oligonucleótidos S3, S6, S7 y S9, cabe destacar el corte que aparece en la zona de las 9 pb en S3, que se corresponde con un palíndromo *Eco*RI y otros más pequeños, y que se ve desplazado a 8 pb por una sustitución C→A que deshace este motivo en los otros tres. S3 muestra también un claro corte hacia las 31 pb, donde se encuentra el motivo TGGGG, corte que desaparece en S6, S7 y S9. Existe un punto de corte muy evidente en estos tres últimos

sustratos en la zona de las 15 pb, que podría estar relacionado con secuencias ricas en adenina, pero no hemos observado este tipo de cortes en otros tramos de secuencia con esta misma composición. Finalmente, en todos los carriles hay una banda muy clara en la zona de 36-38 pb. Esta banda aparece tanto en S1 y S2 como en S3-S9. Puesto que la secuencia en dicha zona del sustrato es diferente para estos dos grupos de oligonucleótidos, puede que esta banda se trate de un producto inespecífico, lo cual no descarta que pueda estar enmascarando DSBs reales en esa misma porción del sustrato.

Llegados a este punto, prácticamente habíamos descartado el que una

actividad enzimática estuviese asociada específicamente a los motivos CAG, y por otra parte nos parecía evidente que existía una actividad que tenía afinidad por tramos de guaninas. Además, algunos resultados apuntaban hacia una enzima con afinidad por zonas de palíndromos. Por esta razón, diseñamos una nueva tanda de sustratos de 50 pb donde se colocaron secuencias palindrómicas únicas y variantes de las mismas, evitando la presencia de tramos de guaninas (S10-S20), con el fin de evitar la aparición de bandas debidas a la otra actividad. Los resultados de ensayos DSB con esta tanda de oligonucleótidos se muestran en la Figura 36C y, como puede verificarse en la imagen, nos aportaron muy poca información con respecto a esta posible actividad. El sustrato S10, que actuaba como control (no contenía secuencias palindrómicas) es degradado muy rápidamente por el extracto sin generar apenas bandas. Otros oligonucleótidos de esta serie muestran bandas débiles, pero que no se corresponden con los tamaños esperados.

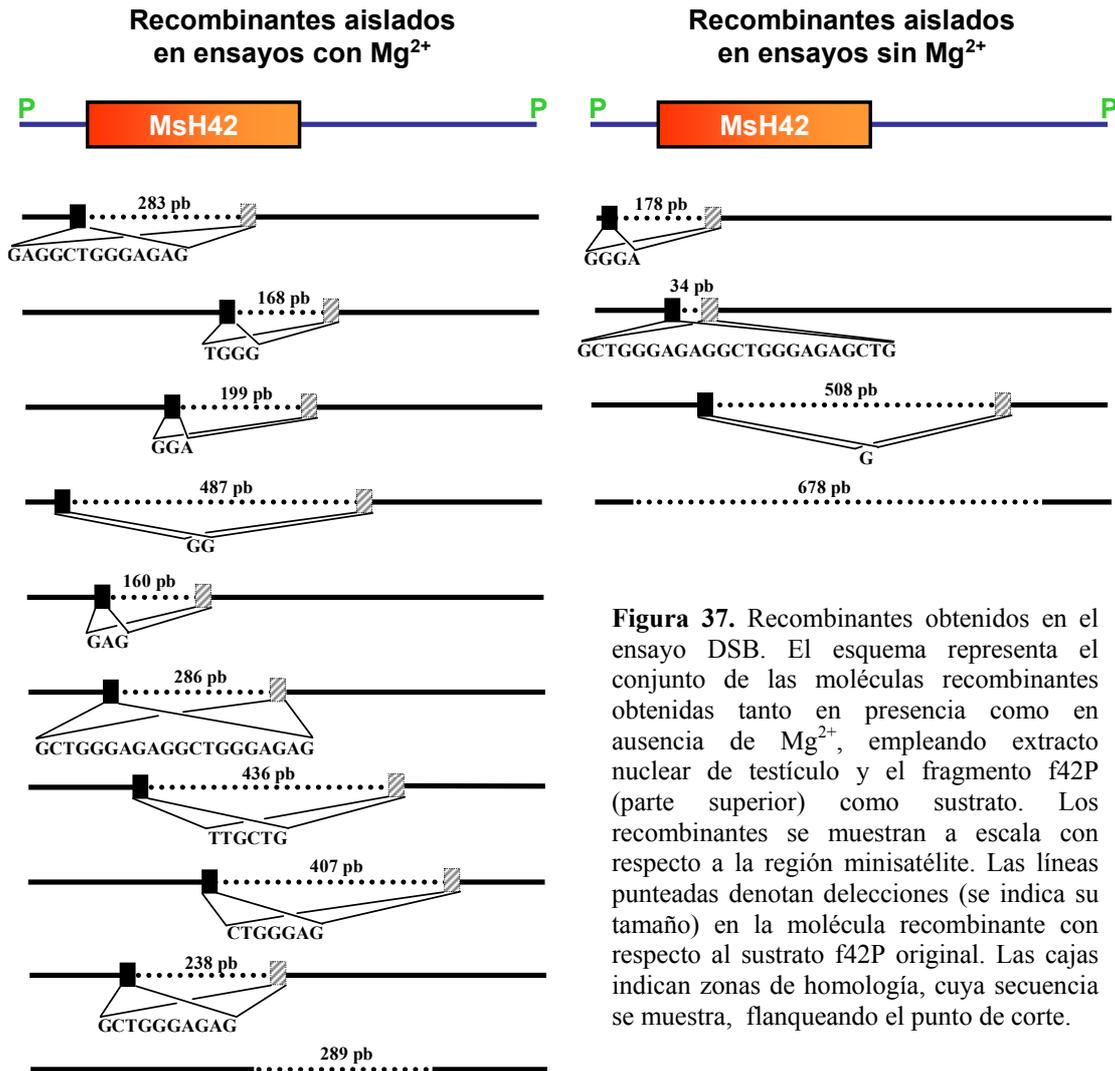
Hay que señalar que durante el proceso de purificación de la actividad dependiente de  $Mg^{2+}$  se intentó seguir paralelamente a esta otra actividad durante los fraccionamientos, pero la perdimos en los primeros pasos. Ensayos DSB con las fracciones más purificadas de la actividad  $Mg^{2+}$ -dependiente realizados en ausencia del catión no produjeron degradación del fragmento (datos no mostrados), reforzando que se trate de proteínas diferentes.

Como conclusión general con respecto a la actividad independiente de  $Mg^{2+}$ , podemos decir que una buena parte de los DSBs que se están detectando están siendo producidos por una actividad que tiene afinidad por residuos de guanina. Esta actividad parece distinta de Endo-G, pues ensayos con esta enzima en ausencia de  $Mg^{2+}$  no generan banda

alguna (datos no mostrados). Ante la hipótesis de una posible actividad residual de Endo-G asociada a pequeñas cantidades de este catión en el extracto, cabría esperar una coincidencia del patrón obtenido tanto en presencia como en ausencia de  $Mg^{2+}$ , hecho que no se da. Se han descrito otras actividades como Endo-SR (Evans & Aguilera, 2003), que es independiente de  $Mg^{2+}$  y muestra preferencia por tramos de guaninas o guaninas y adeninas. No tenemos datos hasta el momento que puedan confirmar o desmentir que sea esta la proteína implicada.

### ***Análisis de los productos recombinantes generados en los ensayos DSB***

Después de la identificación y caracterización de los productos generados por DSBs sobre la región MsH42, nos propusimos buscar moléculas cuyo origen fuese los procesos recombinatorios catalizados por la maquinaria de recombinación presente en el extracto nuclear. Para detectar estas moléculas tanto en el ensayo DSB estándar como en el realizado sin  $Mg^{2+}$ , purificamos los fragmentos de DNA resultantes del ensayo DSB obtenidos del de agarosa (Figura 30). Este DNA fue utilizado como molde en amplificaciones con los cebadores PS1 y PS2, que se encuentran en los extremos del sustrato f42P. Si obteníamos productos de PCR que conservaran las secuencias del extremo del fragmento f42P al amplificarlo con PS1/PS2 y si eran de menor peso molecular que el fragmento original, entonces deberían ser moléculas recombinantes. En estas amplificaciones, el DNA procedente de las bandas N1, N2, N4 y N5 generó productos con el tamaño correspondiente a cada banda, indicando que además de moléculas DSB, en esas bandas de DNA existen moléculas recombinantes. Sin embargo, este procedimiento únicamente



**Figura 37.** Recombinantes obtenidos en el ensayo DSB. El esquema representa el conjunto de las moléculas recombinantes obtenidas tanto en presencia como en ausencia de  $Mg^{2+}$ , empleando extracto nuclear de testículo y el fragmento f42P (parte superior) como sustrato. Los recombinantes se muestran a escala con respecto a la región minisatélite. Las líneas punteadas denotan deleciones (se indica su tamaño) en la molécula recombinante con respecto al sustrato f42P original. Las cajas indican zonas de homología, cuya secuencia se muestra, flanqueando el punto de corte.

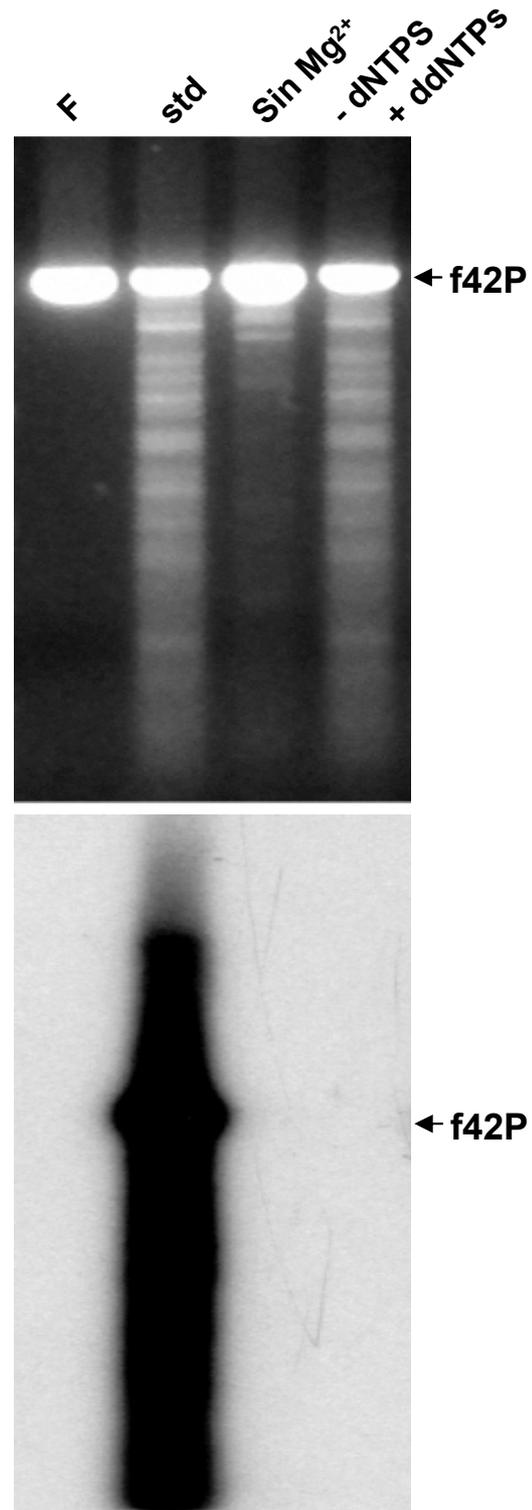
se usó como método de detección de estas moléculas recombinantes, y sus productos no fueron clonados ni secuenciados puesto que podrían generarse artefactos durante la PCR por uniones de extremos imperfectas o *jumping PCR* (Pääbo et al., 1990). Para determinar la secuencia de los recombinantes, se clonaron directamente los productos de los ensayos DSB, estándar y sin  $Mg^{2+}$ , en el vector pGEM-3Z linealizado con *Pst*I. De esta forma, sólo aquellos productos que mantuviesen intactos sus extremos *Pst*I podrían ligarse en el vector y, puesto que el DNA purificado es de menor tamaño que el sustrato f42P original, deberían ser recombinantes. Con esta estrategia se obtuvieron varios clones positivos cuya secuenciación reveló la presencia de

deleciones que oscilan entre 34 y 678 pb con respecto al fragmento original (Figura 37). De forma llamativa, 12 de las moléculas recombinantes presentaban tramos cortos de homología a ambos lados del punto de corte (Figura 37), sugiriendo que su formación estuvo guiada por microhomologías.

Nos planteamos analizar si el mecanismo mediante el cual se estaban originando estos recombinantes precisaba de nueva síntesis de DNA, lo que indicaría que podrían estar formando por mecanismos como HR o SSA. Para ello añadimos dCTP marcado radiactivamente a ensayos DSB estándar y sin  $Mg^{2+}$ , con el fin de determinar si el sistema era capaz de realizar dicha síntesis *de novo*. El análisis de los

productos de estos ensayos mostró que los fragmentos DSB aparecían marcados en la reacción estándar (Figura 38), demostrando que efectivamente el nucleótido marcado se estaba incorporando al DNA en nuestro sistema *in vitro*. Por el contrario, no se observó marcaje en las reacciones sin  $Mg^{2+}$ , lo que se puede explicar por la dependencia de este catión por parte de las DNA polimerasas. Aunque el extracto era capaz de realizar síntesis de DNA, todavía restaba por establecer si ésta era necesaria para la formación de las moléculas recombinantes detectadas en nuestro ensayo. Para ello realizamos experimentos en los que se eliminaron los dNTPs del tampón DSB, a la vez que se adicionaban ddNTPs, que bloquean la síntesis de DNA. La ausencia de incorporación de dCTP en el ensayo DSB puso de manifiesto que, bajo estas condiciones, la síntesis *de novo* está bloqueada (Figura 38), pero el patrón de bandas DSB fue igual al obtenido en condiciones estándar. El hecho de que se pudiesen obtener recombinantes a partir de los productos de los ensayos con y sin  $Mg^{2+}$  en condiciones en que la síntesis de DNA está fuertemente inhibida, es una sólida evidencia de que los procesos de síntesis de DNA no están implicados en la generación de, al menos, parte de las moléculas recombinantes originadas durante el ensayo DSB. El marcaje que detectamos en la reacción estándar era probablemente debido al relleno de extremos 3'-recesivos producidos por exonucleasas (Mason et al., 1996) o por procesos de *nick translation*.

Si consideramos conjuntamente que para la formación de recombinantes no se requiere síntesis de DNA, la presencia de deleciones en la gran mayoría de las moléculas recombinantes y la existencia de secuencias homólogas cortas a ambos lados del punto de corte, podemos concluir que la HR no parece ser el mecanismo de formación de los



**Figura 38.** Síntesis de DNA en el ensayo DSB. Se realizaron una serie de experimentos donde se incorporó dCTP en ensayos DSB en condiciones estándar (std), en reacciones sin  $Mg^{2+}$  (- $Mg^{2+}$ ) o en reacciones sin dNTPs y con ddNTPs (-dNTPs / +ddNTPs). F, fragmento f42P sin extracto. El panel superior muestra el gel teñido con bromuro de etidio y el inferior la autorradiografía obtenida a partir del mismo gel.

recombinantes detectados en nuestro ensayo, ya que ésta se caracteriza por la implicación de síntesis *de novo* de DNA y por dar lugar a productos recombinantes que son normalmente iguales a los originales. La actividad que da lugar a las moléculas recombinantes sería diferente a las descritas anteriormente en otros ensayos *in vitro*, donde sí se ha constatado el concurso de eventos de síntesis de DNA para la generación de moléculas recombinantes (Singleton & Jeggo, 1999; Pospiech et al., 2001). Las deleciones y la presencia de secuencias duplicadas a ambos lados del DSB podrían hacer pensar en mecanismos del tipo SSA (Göttlich et al., 1998) o NHEJ guiado por microhomologías (Boulton & Jackson, 1996) como responsables de la generación de los recombinantes detectados, pero la ausencia de requerimiento de síntesis de DNA descartaría en principio al primero de estos procesos. Por lo tanto, el modelo que mejor se ajusta a nuestras observaciones es el NHEJ, que se encontraría facilitado por pequeñas regiones de homología en las secuencias adyacentes al punto de generación del DSB. La naturaleza repetitiva de nuestra secuencia haría especialmente efectiva esta forma de recombinación, ya que la similitud entre las unidades repetitivas de MsH42 e incluso determinados motivos ricos en guanina de las secuencias adyacentes permite múltiples opciones de emparejamiento homólogo. No podemos descartar, sin embargo, que puedan estar ocurriendo otro tipo de

procesos recombinatorios que hayan pasado inadvertidos en nuestro análisis. El sistema que acabamos de describir nos ha permitido el estudio de la generación de DSBs en la región MsH42 y de la formación de moléculas recombinantes por NHEJ. Hemos visto como los extractos nucleares poseen actividades endonucleásicas capaces de producir DSBs en un fragmento que comprende a la región MsH42 en condiciones tanto de presencia como de ausencia del catión  $Mg^{2+}$ . La principal actividad endonucleásica en condiciones estándar muestra una gran preferencia por secuencias ricas en guanina y todos los indicios de que disponemos apuntan a que tal enzima es la endonucleasa G. La actividad independiente de  $Mg^{2+}$  parece deberse a un enzima distinto de la endonucleasa G, pero que no hemos sido capaces de identificar. Además, hemos detectado eventos recombinatorios en este mismo ensayo DSB, caracterizados esencialmente por la no necesidad de síntesis de DNA, por la aparición de deleciones en los recombinantes y por la presencia de regiones cortas de homología flanqueando a las deleciones en la mayoría de ellos. Estas propiedades apoyan la premisa de un mecanismo de tipo NHEJ favorecido por microhomologías como el responsable de la formación de las moléculas recombinantes observadas. Todo este conjunto de datos puede permitirnos explicar los resultados obtenidos anteriormente en nuestro laboratorio acerca de la capacidad recombinogénica de MsH42.

## Una paradoja en los ensayos DSB

### **Generación de DSBs vs. Ligación de extremos de DNA (*end-joining*)**

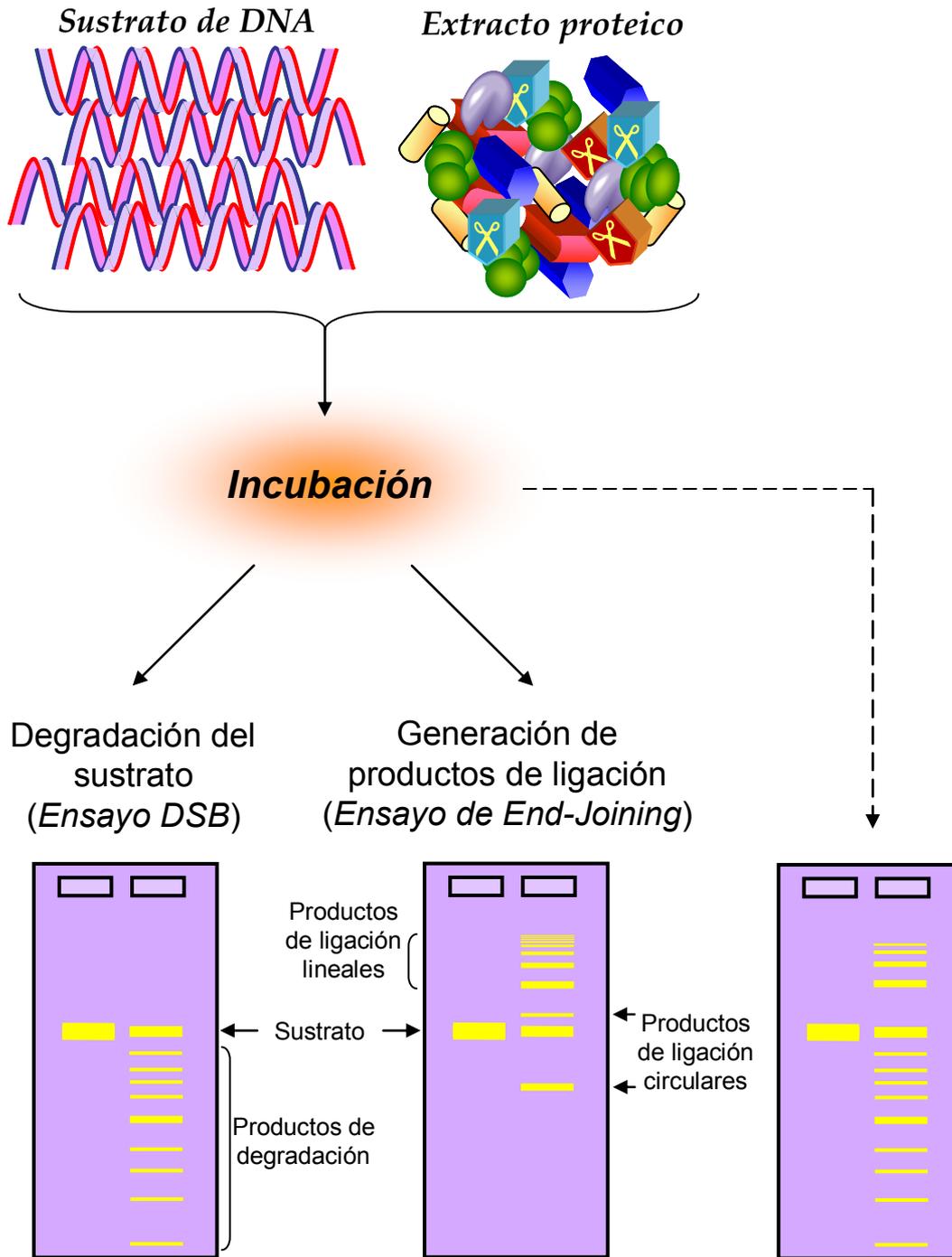
Desde hace años, el estudio de muchos de los procesos de reparación de DNA se ha venido realizando, entre otras metodologías, mediante ensayos de ligación de extremos de DNA (DNA *end-joining*) *in vitro*. Este tipo de experimentos consisten en la incubación de un determinado DNA lineal, empleado como sustrato, junto con proteínas que catalizan la ligación de sus extremos. Estas proteínas pueden añadirse ya purificadas a la reacción (Ramsden & Gellert, 1998; Van Dyck et al., 1999) o bien en forma de un extracto proteico crudo o nuclear (North et al., 1990; Nicolas & Young, 1994). En la Figura 39 se muestran esquemáticamente los resultados que se pueden obtener al incubar un sustrato de DNA con un extracto proteico nuclear. En un caso, que denominamos ensayo DSB, se produce mayoritariamente la degradación del DNA mientras que en el ensayo de *end-joining* tiene lugar la multimerización o circularización del sustrato inicial.

Existen todo un conjunto de investigaciones que se basan en la adición de extractos poco o nada fraccionados a un DNA sustrato, y cuya aportación al conocimiento de los sistemas de reparación de DNA mediante NHEJ ha sido notoria. Un buen ejemplo es el trabajo de Baumann & West (1998), quienes desarrollaron el primer sistema de ligación de extremos de DNA *in vitro* donde se demostraba que la capacidad del ensayo para realizar NHEJ era dependiente de proteínas como Ku70, Ku80, la subunidad catalítica de la proteína-quinasa DNA dependiente (DNA-PK<sub>cs</sub>), la DNA ligasa IV y XRCC4. Estos factores son

imprescindibles para el funcionamiento normal del NHEJ *in vivo* y la implicación de algunos de ellos en esta vía de reparación ya se había deducido por estudios genéticos previos (Blunt et al., 1995; Grawunder et al., 1997; Singleton et al., 1997). Posteriormente, el uso de una metodología similar permitió relacionar otras proteínas como BRCA-1 (Zhong et al., 2002), la DNA polimerasa  $\alpha$  (Pospiech et al., 2001) o la polinucleótido kinasa humana (Chappell et al., 2002) con el NHEJ.

Durante la puesta a punto del ensayo DSB, llamaron nuestra atención las profundas diferencias entre el tipo de resultados obtenidos en nuestro ensayo con respecto a los presentados por otros autores en trabajos basados en sistemas de *end-joining in vitro*, máxime teniendo en cuenta que las metodologías empleadas en ambos experimentos son esencialmente las mismas. En el ensayo DSB realizado en nuestro laboratorio, el DNA es degradado de una forma relativamente rápida y se detecta una baja formación de multímeros lineales o moléculas circulares (Figura 25). Por el contrario, en la mayoría de los ensayos de *end-joining*, las moléculas de DNA son ligadas de forma eficiente tanto intra como intermolecularmente, sin apenas degradación del DNA (Mason et al., 1996; Baumann & West, 1998; Sathees & Raman, 1999; Feldmann et al., 2000; Pospiech et al., 2001; Chappel et al., 2002; Odersky et al., 2002; Zhong et al., 2002; Diggle et al., 2003).

En la Tabla 7 se hace un análisis comparativo entre las características de los ensayos de *end-joining* realizados por varios autores y las de nuestro ensayo DSB. A primera vista, parece extraño que en ninguno de los ensayos arriba mencionados apenas se pudiese apreciar la actividad nucleásica endógena que con



**Figura 39.** Comparación entre el ensayo DSB y el ensayo de *end-joining*. Se muestra una representación de estos dos tipos de experimentos. Ambos implican la incubación de un determinado sustrato de DNA con un extracto proteico. Sin embargo, los procesos que suceden en cada uno de ellos son diferentes, ya que los productos finales son totalmente opuestos. En el caso del ensayo DSB, el DNA es degradado generando un patrón de bandas característico. En los ensayos de *end-joining*, el DNA sufre procesos de ligación de extremos que conducen a su multimerización o la formación de moléculas circulares. En determinadas situaciones, pueden ocurrir simultáneamente la degradación y la multimerización del fragmento (panel derecho), indicando que ambos procesos no son excluyentes.

toda seguridad deben poseer los extractos empleados y que nosotros observábamos con tanta claridad. Este hecho resultó aún más llamativo al

constatar que en la mayoría de estos trabajos se incubaban cantidades en el rango de pocos ng de DNA con decenas de  $\mu\text{g}$  de proteína. Se nos planteaba de

**Tabla 7.** Características de diversos ensayos de *end-joining* comparados con nuestro ensayo DSB.

| Fuente de los datos  | DNA sustrato                    |             |                      | Extracto proteico                          |               | Relación DNA/proteína (fmol/μg) | Principales productos obtenidos <sup>†</sup> | Proporción relativa en nuestro ensayo <sup>‡</sup> |
|--|---------------------------------|-------------|----------------------|--|---------------|---------------------------------|--|--|
|  | Tipo                            | Tamaño (pb) | Cantidad (ng) (fmol) | Tipo* y origen                             | Cantidad (μg) |                                 |  |  |
| <i>Baumann &amp; West, 1998</i>                              | pFB585 linearizado              | 7700        | 100.00<br>20.00      | Crudo, línea linfoblastoide humana GM00558 | 68            | 0,29                            | Mult.  | 804.62   |
| <i>Sathees &amp; Raman, 1999; Raghavan &amp; Raman, 2004</i> | pUC12 linearizado               | 2680        | 600.00<br>344.46     | Crudo, testículo de ratón                  | 60            | 5,74                            | Mult.  | 40.65  |
| <i>Feldmann et al., 2000</i>                                 | Derivado de pSP65 linearizado   | 3000        | 15.00<br>7.70        | Crudo, línea CHO                           | 70            | 0,11                            | OC, CCC                                      | 2121.27  |
| <i>Pospiech et al., 2001</i>                                 | Derivado de M13mp18 linearizado | 7500        | 100.00<br>20.25      | Crudo, HeLa                                | 75            | 0,27                            | Mult.  | 864.22   |
| <i>Chappel et al., 2002</i>                                  | pACSG2 linearizado              | 5500        | 40.00<br>11.20       | Crudo, línea linfoblastoide humana GM00558 | 80            | 0,14                            | Mult.  | 1666.71  |
| <i>Odersky et al., 2002</i>                                  | Derivado de pUC19 linearizado   | 2727        | 10.00<br>5.60        | Crudo, línea CHO-K1                        | 70            | 0,08                            | OC, CCC, Dim.                                | 2916.75  |
| <i>Zhong et al., 2002</i>                                    | pBSK(+) linearizado             | 2960        | 9.62<br>5.00         | Crudo, fibroblastos embrionarios de ratón  | 20            | 0,75                            | Mult.  | 311.12   |
| <i>Diggle et al., 2003</i>                                   | pGEM3zf+ linearizado            | 3200        | 20.00<br>9.60        | Crudo, varias líneas y tejidos humanos     | 40            | 0,24                            | Mult.  | 972.25   |
| <i>Nuestro ensayo DSB estándar</i>                           | Fragmento lineal f42P           | 863         | 1500.00<br>2800.08   | Nuclear, testículo de rata                 | 12            | 233,34                          | Degradación                                  | 1.00   |

\* Por crudo, se entienden extractos proteicos obtenidos a partir de la lisis de células completas sin fraccionamiento de ningún tipo.

† Mult., multímeros lineales; OC, moléculas circulares relajadas; CCC, moléculas circulares covalentemente ligadas; Dim., dímeros lineales.

‡ Esta proporción se calcula como el cociente entre la relación DNA/proteína en nuestro ensayo DSB dividida por la relación DNA/proteína del ensayo en cuestión.

esta forma una paradoja para la que no teníamos una explicación evidente: ¿cómo era posible que cuando se incubaban cantidades relativamente elevadas de DNA con un extracto proteico, caso de nuestro ensayo DSB, el DNA se degradase mientras que si la cantidad de DNA era proporcionalmente

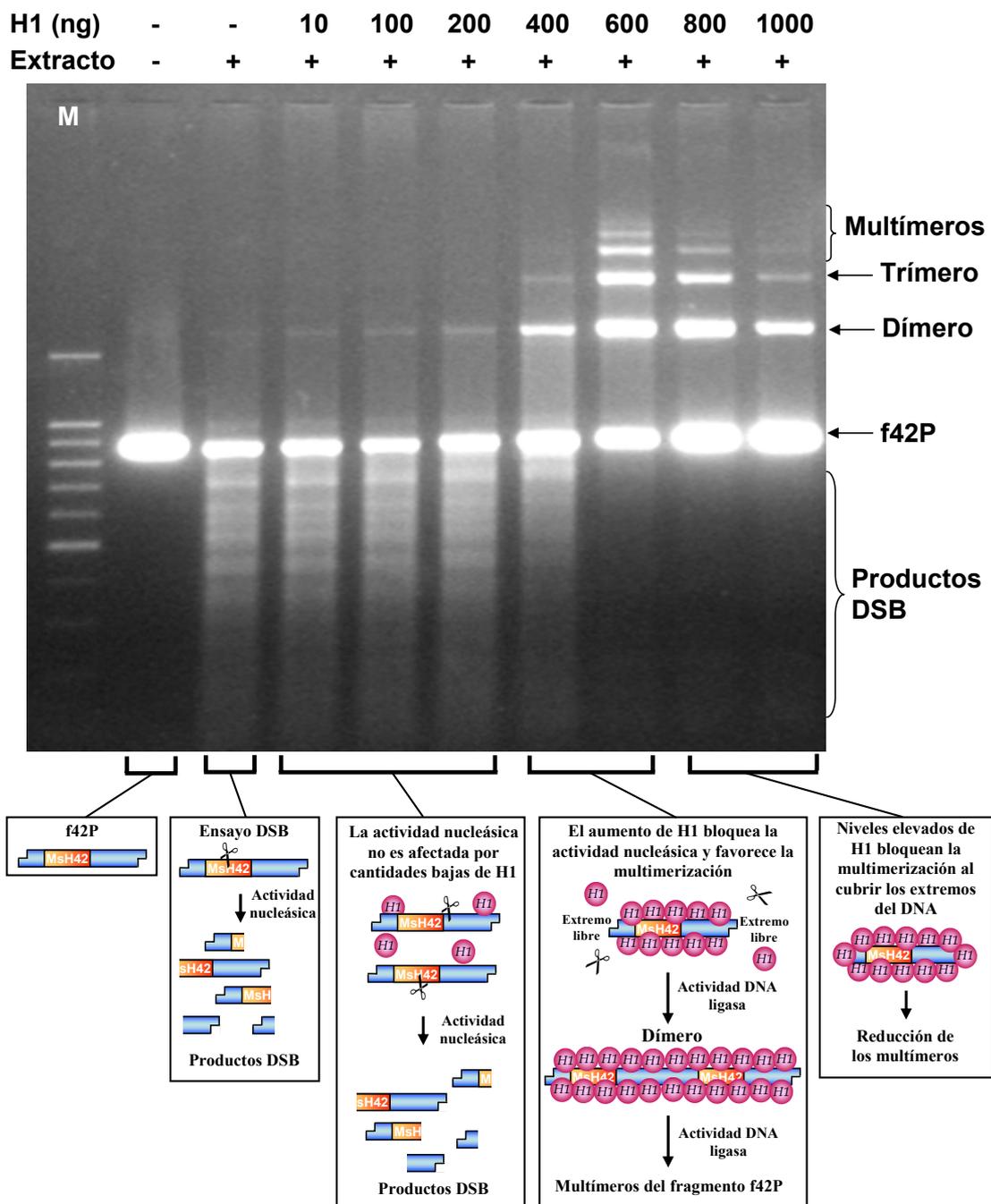
mucho menor este DNA resistía ante la actividad nucleásica y era además ligado por sus extremos? Para dar respuesta a esta cuestión, decidimos investigar qué era lo que podría estar influenciando en la prevalencia de la actividad ligasa o nucleasa en cada caso.

### ***Efecto de la adición de histona H1 al ensayo DSB***

Uno de los primeros aspectos que necesitábamos dilucidar era si la escasa actividad de ligación que detectábamos en nuestros experimentos se debía a una carencia en DNA ligasas de nuestro extracto derivada del protocolo empleado para su preparación o bien si era consecuencia del propio método de ensayo. Para comprobar este punto, decidimos incorporar al ensayo DSB estándar cantidades crecientes (0.01-1 µg) de la histona H1 (H1), ya que se había descrito su capacidad para estimular la actividad de ligación intermolecular de la T4 DNA ligasa (Sobczak & Duguet, 1988) y de las DNA ligasas III y IV de mamíferos (Yamanaka et al., 2002). Como se observa en la Figura 40, la adición de H1 tuvo dos efectos notables en el resultado de nuestro ensayo DSB: i) se favoreció la multimerización del sustrato f42P, existiendo un óptimo de cantidad de H1 (600 ng) a partir del cual se produce una disminución de la tasa de ligación intermolecular; ii) se detectó una fuerte inhibición de la actividad nucleásica, que provocó una desaparición progresiva del patrón de bandas correspondientes a los productos originados por DSBs. Ambos hechos pusieron de manifiesto que la adición de H1 no sólo potenciaba la multimerización del sustrato sino que además impedía su degradación. Resultados previos han puesto de manifiesto la capacidad de H1 para formar complejos nucleoproteicos largos que reúnen a las moléculas de DNA (Yamanaka et al., 2002), de modo que la resistencia a la degradación podría deberse a que el ácido nucleico es inaccesible para las nucleasas cuando se halla incluido en dichos complejos. En nuestro experimento puede apreciarse como la protección del DNA se va adquiriendo de forma gradual,

desapareciendo progresivamente el patrón de bandas originadas por DSBs, hasta llegar a un punto donde la degradación es mínima y la multimerización máxima. A este nivel, las moléculas de DNA probablemente estarían recubiertas casi en su totalidad por H1, aunque existiendo suficiente margen en sus extremos como para facilitar la actividad de las ligasas. Si se continúa incrementando la cantidad de H1, la protección alcanza los extremos de los sustratos, con lo que se produce un descenso de la actividad ligasa y la consiguiente reducción en la multimerización del fragmento f42P. Estos resultados dan respuesta a la pregunta inicial, al demostrar la presencia de una actividad ligasa en el extracto que es plenamente funcional bajo las condiciones estándar del ensayo DSB.

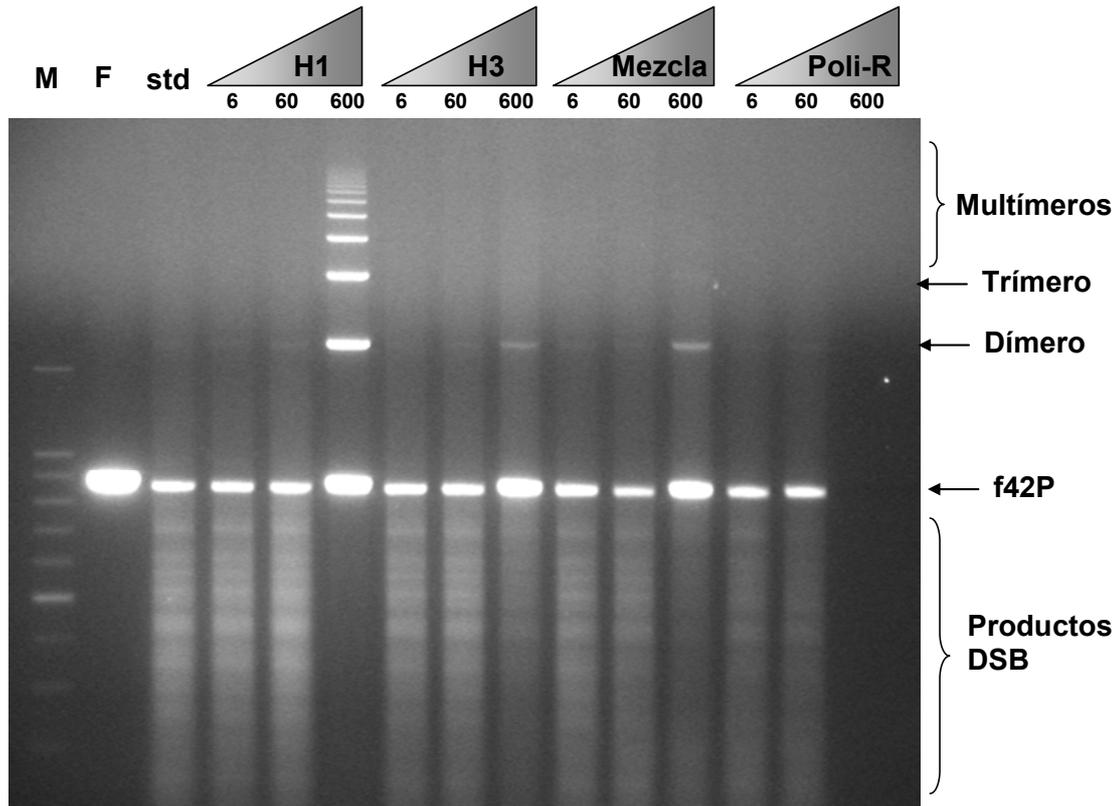
Para verificar que el efecto de H1 era específico de esta histona y no debido a una neutralización de las cargas de los fosfatos del DNA, decidimos comparar su efecto con el de otras histonas y el de una poliamina. Para ello, se realizaron una serie de reacciones estándar del ensayo DSB a las que se incorporaron separadamente distintas cantidades (6, 60 ó 600 ng) de H1, histona H3 (H3), poliarginina o de una mezcla equimolecular de las histonas H1, H2A, H2B, H3 y H4. Como se muestra en la Figura 41, la ligación intermolecular es mucho más elevada en el ensayo con 0.6 µg de H1 que con cualquiera de las otras proteínas ensayadas, coincidiendo con el mayor descenso en la degradación del fragmento f42P. La misma cantidad de H3 o de la mezcla de histonas apenas produjo un leve incremento en la cantidad de dímeros formados a partir de f42P con respecto a la reacción estándar. En estas mismas reacciones pudo apreciarse una cierta caída en la formación de productos generados por DSBs, aunque la actividad nucleásica no



**Figura 40.** Efecto de la adición de histona H1 al ensayo DSB. En la parte superior se muestra el resultado del análisis por electroforesis en gel de agarosa al 2% de los productos obtenidos en una serie de ensayos DSB a los cuales se han añadido las cantidades indicadas de H1. En la imagen pueden apreciarse las moléculas generadas a partir de DSBs (productos DSB), así como los dímeros, trímeros y multímeros lineales del sustrato f42P. En la parte inferior de la figura se presentan esquemas interpretativos de los resultados. Las tijeras representan proteínas con actividad endonucleásica y las esferas moradas representan moléculas de H1.

resultó totalmente bloqueada. El efecto de 0.6  $\mu\text{g}$  de poliarginina no pudo detectarse mediante este sistema ya que probablemente esta poliamina arrastrase consigo al DNA a la fase orgánica durante la extracción de los productos del experimento antes de la

electroforesis. Como control adicional, también se utilizaron otras proteínas como BSA o IgG de ratón en un rango desde 0.01 hasta 10  $\mu\text{g}$  por reacción, sin observarse alteración alguna del nivel de degradación del fragmento o de su multimerización (datos no mostrados).



**Figura 41.** Ensayo DSB utilizando distintos polipéptidos básicos. Se muestra un gel de agarosa al 2% donde se han separado los productos de distintas reacciones DSB en las que se han añadido cantidades crecientes (6, 60 y 600 ng) de H1, de H3, de una mezcla equimolecular de las histonas H1, H2A, H2B, H3 y H4 (Mezcla) y de un polímero de arginina (Poli-R). Pueden apreciarse tanto el patrón de bandas generado por la acción nucleásica (Productos DSB) como la aparición de dímeros, trímeros y multímeros lineales del fragmento f42P. M, marcador de 100 pb (Promega); F, fragmento sin extracto; std, reacción estándar.

Por tanto, parece que el efecto de inhibición de la degradación del DNA y de potenciación de su ligación intermolecular es específico de H1, o al menos con esta proteína es mucho más acentuado que con otras histonas.

Una forma alternativa para comprobar el efecto de H1 en el ensayo DSB fue contrarrestar su actividad mediante la adición de protimosina  $\alpha$  (ProT $\alpha$ ), una proteína capaz de unirse específicamente a H1 (Gómez-Márquez & Rodríguez, 1998; Karetsov et al., 1998; Segade & Gómez-Márquez, 1999). Se realizaron ensayos DSB en condiciones estándar a los cuales se adicionaron independientemente H1, H3 o ProT $\alpha$ . Otras reacciones se prepararon añadiendo una mezcla de estas proteínas, con o sin preincubación de las mismas

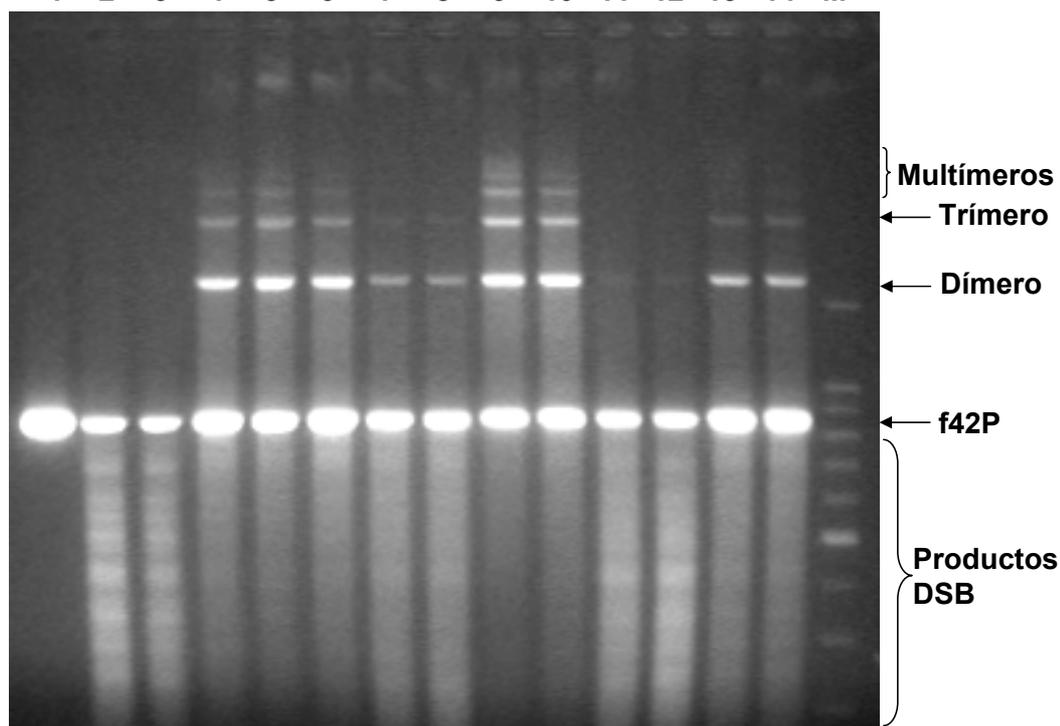
en el tampón DSB antes de la adición del sustrato f42P. Los datos derivados de estos experimentos mostraron que la adición de ProT $\alpha$  no altera el resultado final del ensayo, pero sí que es capaz de atenuar la acción de H1 (Figura 42). Sin embargo, hay que destacar que dicha atenuación se detecta solamente cuando H1 y ProT $\alpha$  se preincuban conjuntamente antes de la adición del DNA (carriles 7 y 8), mientras que si la adición de estas proteínas es simultánea a la del DNA, la potenciación de la ligación es similar a la observada en ausencia de ProT $\alpha$  (carriles 5 y 6). Además, en la misma figura se aprecia otro aspecto importante de la interacción entre H1, ProT $\alpha$  y el DNA: una vez que H1 se ha unido al DNA, la ProT $\alpha$  es prácticamente incapaz de retirar moléculas de H1 del complejo

nucleoproteico una vez que éste se ha formado (comparar carriles 4, 9 y 10), una característica de ProTα que se ha descrito anteriormente (Karetsou et al., 1998). Esto explicaría el hecho de que la disminución que ProTα provoca sobre el efecto de H1 sea tan leve: ProTα solamente interacciona a moléculas de H1 libre, siendo incapaz de liberar todas aquellas moléculas que consiguen unirse al DNA y que, por tanto, permanecen establemente asociadas a él. Por tanto, parece claro que la actuación de ProTα en nuestro ensayo consiste en la unión de moléculas de H1 libres, disminuyendo

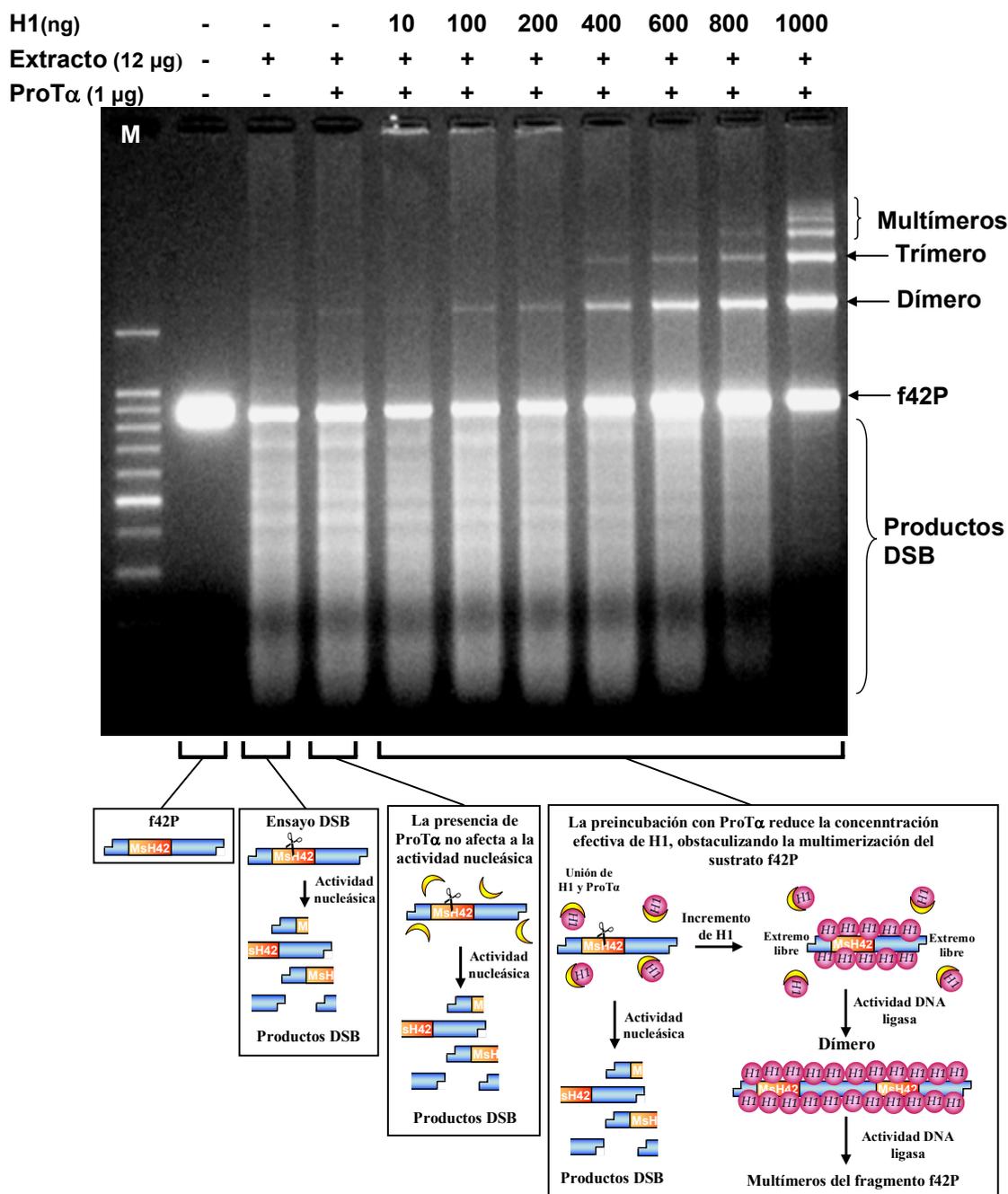
de esta forma su concentración efectiva y, por tanto, el número de moléculas disponibles para interactuar con el DNA. Además, hemos visto que la adición de H3 apenas fue capaz de desplazar a ProTα de su unión con H1, ya que su presencia no favoreció sustancialmente la formación de multímeros en presencia de H1 y ProTα (comparar carriles 7 y 8 con 13 y 14).

A continuación realizamos una serie de ensayos DSB en los que se preincubaba una cantidad constante de ProTα (1 µg) con cantidades crecientes de H1

|                 |          |          |          |          |          |          |          |          |          |           |           |           |           |           |          |
|-----------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|-----------|-----------|-----------|-----------|----------|
| <b>H1</b>       | -        | -        | -        | +        | +        | +        | +        | +        | +        | -         | -         | +         | +         |           |          |
| <b>H3</b>       | -        | -        | -        | -        | -        | -        | -        | -        | -        | +         | +         | +         | +         |           |          |
| <b>ProTα</b>    | -        | -        | ++       | -        | +        | ++       | +        | ++       | -        | ++        | -         | ++        | +         | ++        |          |
| <b>Extracto</b> | -        | +        | +        | +        | +        | +        | +        | +        | +        | +         | +         | +         | +         | +         |          |
|                 | <b>1</b> | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b> | <b>6</b> | <b>7</b> | <b>8</b> | <b>9</b> | <b>10</b> | <b>11</b> | <b>12</b> | <b>13</b> | <b>14</b> | <b>M</b> |



**Figura 42.** Ensayos DSB con la adición de H1, H3 y ProTα. Se muestra un gel de agarosa al 2% con los productos de una serie de reacciones DSB en condiciones estándar, pero donde se han incorporado 1 µg de H1 (+), 0.73 µg de H3 y 0.5 µg (+) o 1 µg (++) de ProTα. Estos valores representan cantidades aproximadamente equimoleculares de H1, H3 y ProTα (0.5 µg ProTα), o una relación 1:1:2 (1 µg ProTα). En las reacciones 7, 8, 12-14, se hizo una preincubación de estas proteínas en tampón DSB (5 min, 37°C) antes de la adición del sustrato f42P. Todas las reacciones se incubaron durante 30 min a 37°C, excepto 9 y 10, en las que la incubación fue de 60 min. En la reacción 10, ProTα se añadió a los 30 min de incubación, mientras que en la 9 se añadió 1 µl de agua. En la figura se indican los productos generados por la acción nucleásica (Productos DSB) y los dímeros, trímeros y multímeros lineales del fragmento f42P. M, marcador 100 bp ladder (Promega).



**Figura 43.** Efecto de la adición de H1 y ProTα al ensayo DSB. En la parte superior se muestra el resultado del análisis por electroforesis en gel de agarosa al 2% de los productos de una serie de ensayos DSB en los que se han añadido diversas cantidades de H1 junto con 1 µg de ProTα. Ambas proteínas se preincubaron juntas antes de la adición del sustrato f42P. Se señalan las moléculas generadas a partir de DSBs (productos DSB), así como los dímeros, trímeros y multímeros lineales del sustrato f42P. En la parte inferior de la figura se presenta un esquema interpretativo de los resultados. Las tijeras representan proteínas con actividad endonucleásica, las esferas moradas a H1 y las curvas amarillas a ProTα. M, marcador *100 bp ladder* (Promega).

idénticas a las utilizadas en la Figura 40 (0.01-1 µg), antes de la adición del sustrato f42P. En la Figura 43 se pone de manifiesto como en presencia de ProTα

el óptimo de ligación se desplazó hasta 1 µg de H1 (comparar con los carriles equivalentes en ausencia de ProTα, Figura 17), y que la degradación del

DNA se mantuvo hasta niveles de 0.8  $\mu\text{g}$  de H1, aproximadamente el doble que sin ProT $\alpha$ . Este resultado concuerda con una reducción en la concentración efectiva de H1 en el ensayo y confirma la especificidad de su papel en la protección del DNA frente a la actividad nucleásica y en la potenciación de su ligación.

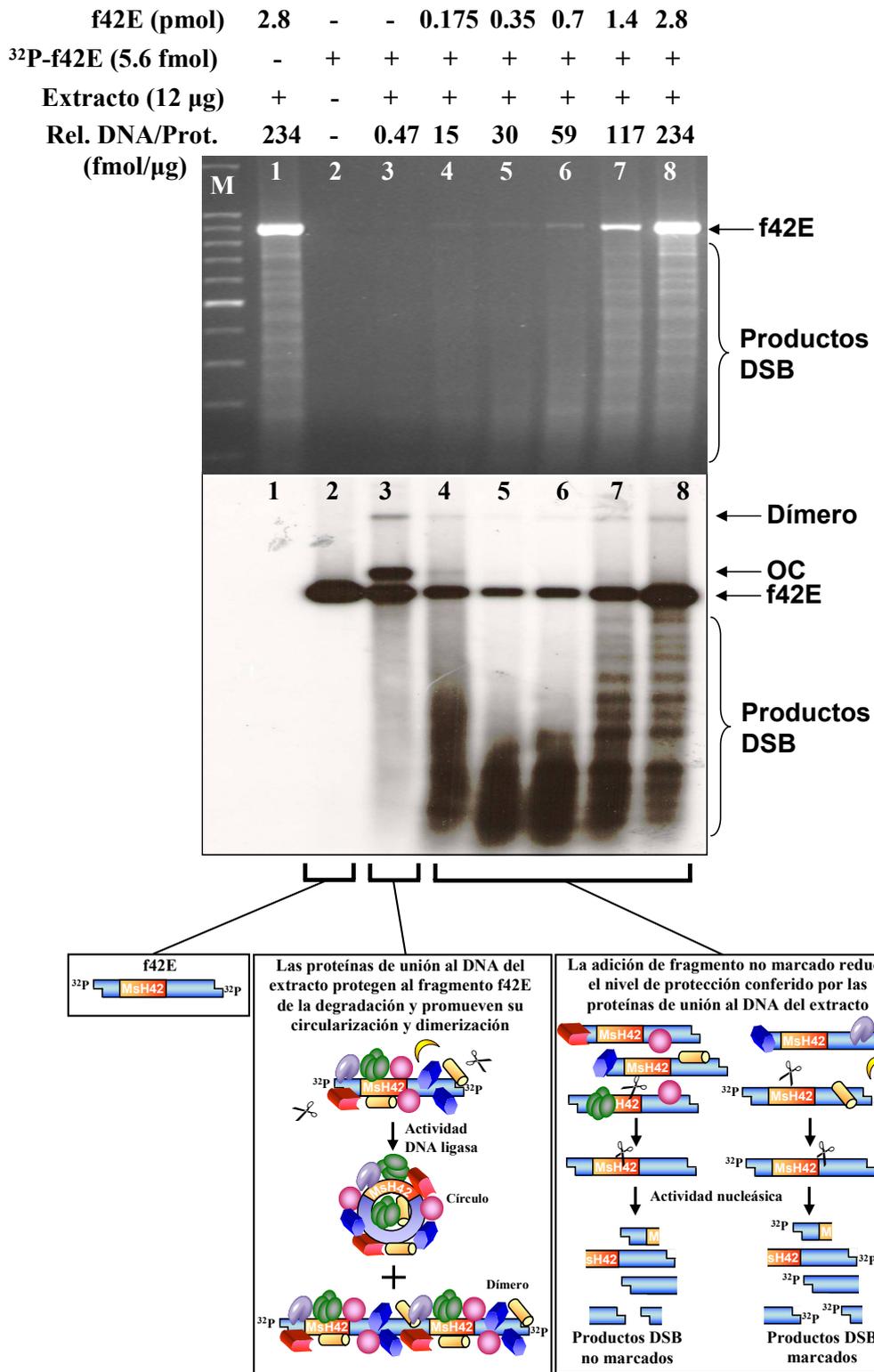
### ***Diferencias en la relación DNA/proteína en los ensayos DSB***

Una vez confirmado que tanto nuestro extracto como las condiciones del ensayo DSB que estábamos empleando tenían plena capacidad para la ligación del DNA, volvimos a la cuestión de por qué en nuestras condiciones estándar no ocurre tal ligación y, en cambio, se produce la degradación del DNA. Al realizar un análisis comparativo entre nuestras condiciones de ensayo y las descritas en la bibliografía referente a ensayos de *end-joining* (Tabla 7), descubrimos que una de las diferencias principales entre nuestro sistema y los empleados por otros autores era la proporción entre moléculas de DNA sustrato y la cantidad de proteína añadida, es decir, la relación DNA/proteína.

Así, la relación DNA/proteína en nuestro ensayo es de  $\sim 234$  fmol/ $\mu\text{g}$ , lo que constituye un valor entre  $\sim 40$  y 2900 veces más elevado que los empleados en determinados ensayos de *end-joining*, que oscilan entre los 0.08 y los 5.74 fmol/ $\mu\text{g}$  (Tabla 7). Este dato plantea la siguiente paradoja: ¿por qué cantidades proporcionalmente elevadas de DNA son degradadas fácilmente por las nucleasas de los extractos, mientras que cantidades relativamente mucho menores no solamente no se degradan, sino que además se ligan eficientemente? La clave para resolver este aparente contrasentido nos la proporcionó el

experimento de adición de H1 al ensayo (Figura 40). Este experimento, además de indicarnos que en nuestro sistema la actividad ligasa estaba intacta, también puso de manifiesto que el fragmento de DNA puede ser protegido frente a la degradación si existe suficiente cantidad de proteínas que se unan a él y lo recubran. Es decir, que la relación sustrato de DNA / cantidad de proteína del extracto nuclear podría ser la responsable de que el ensayo fuese de tipo *end-joining* o de generación de DSBs.

Para verificar la hipótesis de que la causante de las diferencias observadas pudiese ser la relación DNA/proteína, decidimos realizar un ensayo DSB empleando relaciones DNA/proteína crecientes (variando las cantidades de DNA y manteniendo la cantidad de extracto constante), partiendo desde las empleadas en los ensayos de *end-joining* realizados por otros autores hasta las de nuestro ensayo DSB. De este modo podríamos comprobar si, efectivamente, una alteración de esta variable condicionaría el comportamiento general del sistema. Se realizaron una serie de reacciones con una cantidad fija de extracto proteico nuclear de testículo de rata (12  $\mu\text{g}$ ) y, como sustrato, una cantidad constante de fragmento f42E marcado terminalmente con  $[\gamma\text{-}^{32}\text{P}]\text{-ATP}$  (5.6 fmol) suplementada con cantidades crecientes de fragmento f42E no marcado (0-2.8 pmol). Con esta estrategia se pretendía modificar ampliamente la relación DNA/proteína en cada ensayo, pero manteniendo constante la señal que obtendríamos del DNA por autorradiografía. El uso del fragmento f42E en lugar del f42P se debió a que el marcaje terminal de los grupos 5'-fosfato protuberantes con la T4 polinucleótido kinasa es mucho más eficiente que el de los extremos recesivos o romos. Los resultados de estos experimentos (Figura 44)



demonstraron que la reducción de la relación DNA/proteína a un valor de 0.47 fmol/µg (carril 3), característico de ensayos de *end-joining* y unas 500 veces menor que en el ensayo DSB, provoca la inhibición de la actividad nucleásica y

una fuerte potenciación de la ligación intramolecular, generando principalmente moléculas circulares relajadas. Este dato indica que con una relación DNA/proteína de 0.47 fmol/µg existen suficientes moléculas de proteína

**Figura 44.** Efecto de la alteración de la relación DNA/proteína en el ensayo DSB. En el panel superior se muestra el análisis en un gel de agarosa al 2% de una serie de experimentos DSB donde se ha empleado como sustrato una mezcla de fragmento f42E marcado terminalmente y sin marcar. La cantidad de f42E marcado fue de 5.6 fmol (3 ng) y la de fragmento sin marcar osciló desde 0 a 2.8 pmol. Las relaciones DNA/proteína aproximadas se indican en cada caso. En el panel inferior se muestra la autorradiografía del mismo gel, donde puede apreciarse como a relaciones DNA/proteína bajas, el fragmento apenas sufre la acción de las nucleasas y se circulariza, mientras que a relaciones más elevadas el DNA se degrada generando el típico patrón de bandas observadas en los ensayos DSBs. En la parte inferior de la figura se interpretan esquemáticamente estos resultados. OC, monómero circular relajado. M, marcador de 100 pb (Promega)

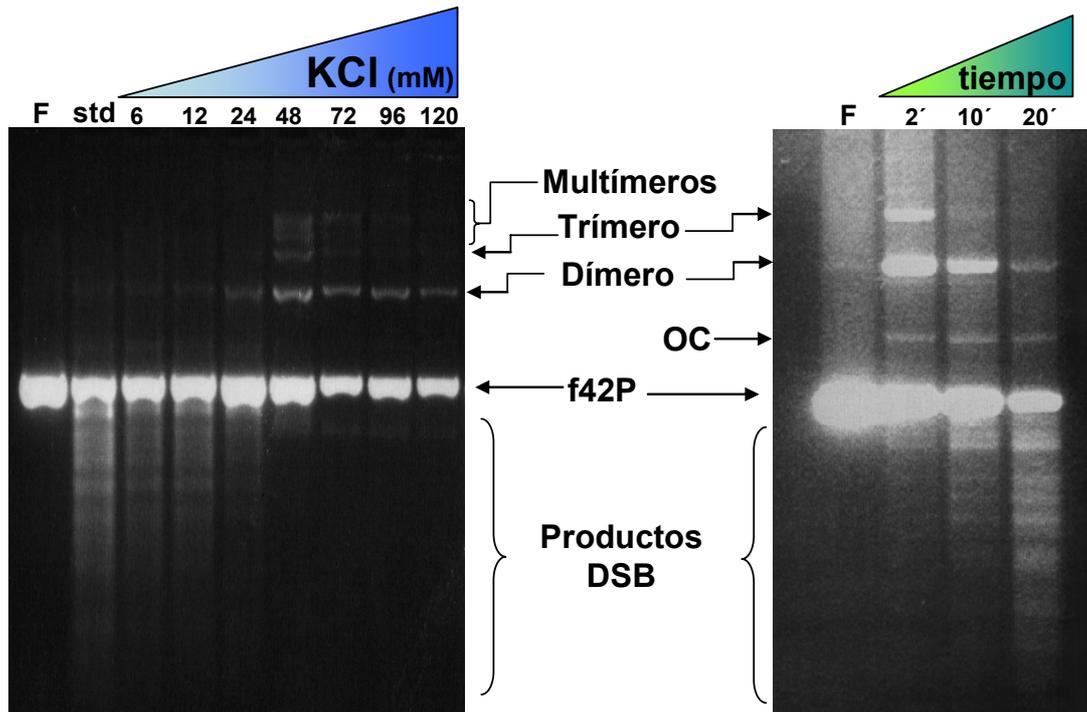
capaces de unirse al DNA como para protegerlo completamente frente a la acción nucleásica. Sin embargo, a medida que se aumenta la relación DNA/proteína (hasta llegar a los 30 fmol/ $\mu$ g), la degradación se va haciendo mayor. Esto ocurre porque al aumentar la cantidad de DNA, la cantidad de proteínas con capacidad para protegerlo se vuelve insuficiente, siendo degradado por las nucleasas. Al seguir aumentado la relación DNA/proteína hasta los 234 fmol/ $\mu$ g se observó la aparición progresiva de bandas producidas por DSBs sobre f42E. En este caso, la elevada cantidad de sustrato sobrepasa completamente la capacidad protectora de las proteínas de unión a DNA, por lo que es degradado por las nucleasas. Sin embargo, al ser esta cantidad de DNA inicial suficientemente alta, se permite el establecimiento de un equilibrio entre la formación de moléculas originadas por DSBs y su posterior degradación, lo que se traduce en una acumulación transitoria de las mismas hasta que se agota el sustrato. Esto permite la detección de dichos productos de degradación, constituyendo el patrón de bandas propio de nuestro ensayo DSB.

Un hecho que merece ser comentado es que los principales productos de ligación observados en nuestro sistema a una relación de 0.47 fmol/ $\mu$ g fueron moléculas circulares relajadas, mientras que en la mayoría de los ensayos de *end-joining* que hemos citado los productos principales eran dímeros, trímeros y multímeros lineales. En este sentido se

conoce que HMGB1 y HMGB2, unas abundantes proteínas nucleares que se unen al DNA espaciador de los nucleosomas de forma competitiva con H1 (Ogawa et al., 1995), son capaces de estimular la ligación de DSBs *in vitro* (Nagaki et al., 1998), favoreciendo preferentemente los eventos de ligación intramolecular frente a la intermolecular (Yamanaka et al., 2002). Por tanto, una abundancia de este tipo de proteínas en nuestros extractos, o de otras que se comporten de forma análoga, podría explicar la mayor proporción de fragmento circularizado frente a los multímeros lineales.

### ***Otros modos de favorecer la ligación frente a la degradación***

Hemos visto como la clave para detectar los eventos de ligación que suceden en estos ensayos depende de poder proteger al DNA frente a la actividad nucleásica. Por ello, una vez demostrado que la adición de H1 o bien el uso de relaciones DNA/proteína suficientemente bajas minimizan el efecto de esta actividad degradativa, nos planteamos buscar formas alternativas de conferir esta resistencia al DNA, dado que no siempre puede ser viable la adición de H1 o la alteración de las cantidades de DNA y extracto. En este sentido, se ha descrito que el ión  $K^+$ , así como otros cationes monovalentes, son fuertes inhibidores de diversas actividades endonucleásicas presentes en los extractos obtenidos a partir de tejidos de mamíferos (McKenna



**Figura 45.** Ensayos DSB con distintas concentraciones de KCl o tiempos de incubación cortos. Se muestran los análisis electroforéticos en geles de agarosa al 2% de distintos ensayos DSB. **A.** Efecto de la concentración de KCl. **B.** Efecto del tiempo de incubación. Las condiciones de reacción fueron las estándar, incubando durante los tiempos indicados. F, fragmento sin extracto; std, reacción estándar; OC, monómero circular relajado.

et al., 1981; Ruiz-Carrillo & Renaud, 1987; Gottlieb & Muzyczka, 1990; Evans & Aguilera, 2003). Además, estos cationes pueden potenciar la ligación de moléculas de DNA a concentraciones moderadas (~ 50 mM) (Sambrook & Russell, 2001). Para verificar si esta sería una aproximación útil, realizamos una serie de ensayos DSB con concentraciones crecientes de KCl.

Como se muestra en la Figura 45A, el incremento en la concentración de KCl hasta niveles de ~ 50 mM KCl provoca una disminución de la degradación del sustrato y la potenciación de la ligación intermolecular. Por encima de este punto, la actividad de multimerización también se ve afectada y desciende progresivamente. Este resultado es bastante similar al obtenido con H1 (Figura 40), aunque en este caso el efecto está producido por una inactivación de las principales actividades nucleásicas del extracto por

aumento de la fuerza iónica. También descubrimos que un modo de detectar los procesos de ligación frente a los de degradación en nuestro ensayo DSB era el de emplear tiempos de incubación muy cortos (2 min). En este intervalo de tiempo ya se ha producido una población considerable de dímeros y trímeros, mientras que apenas se han puesto en marcha los procesos de degradación (Figura 45B). A medida que se prolonga el tiempo de incubación, tanto el sustrato original como los sus multímeros son blanco de la actividad nucleásica que termina por degradarlos para dar lugar al patrón de bandas DSB.

Los ensayos *in vitro* de DNA *end-joining* son un método sencillo y rápido que ha sido ampliamente utilizado para estudiar procesos de reparación del DNA, especialmente el NHEJ (Labhart, 1999). Actualmente siguen apareciendo nuevas variantes y refinamientos de este tipo de experimentos que permiten

nuevas variantes y refinamientos de este tipo de experimentos que permiten representar mejor los procesos de reparación del DNA *in vivo* (Ma et al., 2004; Pfeiffer et al., 2005). En este sentido, en un artículo reciente se conjuga la metodología de este tipo de ensayos con la gran sensibilidad de la PCR cuantitativa para aumentar el nivel de detección de los eventos de *end-joining* (Budman & Chu, 2005). En nuestro estudio hemos demostrado que uno de los parámetros de reacción, la relación DNA/proteína, puede condicionar de manera crucial cuál será el proceso bioquímico predominante en el experimento, degradación de las moléculas de sustrato o ligación de los extremos del mismo, tanto intramolecularmente como de manera intermolecular. Las soluciones concretas que hemos propuesto para mejorar la actividad de ligación en los ensayos podría ser útil en aquellos casos donde sea necesario trabajar con extractos que procedan de tejidos o líneas celulares

con una fuerte actividad nucleásica endógena, como la línea linfoblastoide GM06315 (Baumman & West, 1998).

A modo de conclusión, podemos decir que un ensayo DSB y un ensayo de *end-joining* son las dos facetas de un mismo experimento y que el factor crucial que lo modula, haciendo que el sistema que se decante en uno u otro sentido, es un bloqueo de la actividad nucleásica sobre el sustrato. En este trabajo hemos mostrado cuatro formas de cómo ejecutar este bloqueo: i) protección del sustrato mediante la adición de proteínas que se unan a él; ii) protección del sustrato mediante el mantenimiento de una relación DNA/proteína suficientemente baja; iii) inhibición directa de la actividad nucleásica, mediante un incremento en la concentración de KCl; iv) limitando el tiempo de actuación de la actividad nucleásica y aprovechando que, al menos en nuestro sistema, los procesos de ligación son más rápidos.



*Conclusiones*



De los resultados presentados en esta Tesis Doctoral, las principales conclusiones que hemos obtenido son:

- 1) MsH42 es un minisatélite que se originó durante la evolución de los primates antes de la divergencia entre platirrinos y catarrinos (~ 40 millones de años) dentro del intrón 13 del gen Q8WUJ3. Este minisatélite ha evolucionado probablemente mediante eventos como inserciones/delecciones, mutaciones sencillas y procesos de expansión y contracción mediados por entrecruzamiento desigual y/o deslizamiento de cadena durante la replicación. El polimorfismo de este minisatélite en las poblaciones actuales de gorila, chimpancé y del ser humano parece haberse generado a partir de un único alelo ancestral.
  
- 2) El minisatélite MsH42 es degradado por nucleasas específicas en ensayos DSB. Estos experimentos nos han llevado a caracterizar dos rutas para la generación de cortes en el fragmento de DNA empleado como sustrato. Una de estas vías implica la actuación de una nucleasa específica  $Mg^{2+}$ -dependiente sobre los abundantes tramos de guaninas que existen en la región MsH42, tratándose muy probablemente de la endonucleasa G. La otra vía se basa en una actividad  $Mg^{2+}$ -independiente que también es capaz de degradar la región minisatélite, pero que a diferencia de la anterior parece actuar principalmente fuera del minisatélite.
  
- 3) Las diferencias observadas en el resultado de los ensayos DSB y de *end-joining*, predominancia de eventos de degradación frente a los de ligación y viceversa, están influenciadas por la relación DNA/proteína que se emplea en el ensayo.



# *Bibliografía*



- Abe, S., Katagiri, T., Saito-Hisaminato, A., Usami, S., Inoue, Y., Tsunoda, T. & Nakamura, Y.** (2003a) Identification of CRYM as a candidate responsible for nonsyndromic deafness, through cDNA microarray analysis of human cochlear and vestibular tissues. *Am. J. Hum. Genet.* **72**, 73-82.
- Abe, S., Usami, S. & Nakamura, Y.** (2003b) Mutations in the gene encoding KIAA1199 protein, an inner-ear protein expressed in Deiters' cells and the fibrocytes, as the cause of nonsyndromic hearing loss. *J. Hum. Genet.* **48**, 564-570.
- Ali, S. & Wallace, R.B.** (1988) Intrinsic polymorphism of variable number tandem repeat loci in the human genome. *Nucleic Acids Res.* **16**, 8487-8496.
- Alves, S., Amorim, A. & Prata, M. J.** (2002) Evolution of a VNTR located within the promoter region of the thiopurine methyltransferase gene: inferences from population and sequence data. *Hum. Genet.* **111**, 172-178.
- Amarger, V., Gauguier, D., Yerle, M., Apiou, F., Pinton, P., Giraudeau, F. et al.** (1998) Analysis of distribution in the human, pig, and rat genomes points towards a general subtelomeric origin of minisatellite structures. *Genomics* **52**, 62-71.
- Andersen T. H. & Nilsson-Tillgren T.** (1997) A fungal minisatellite. *Nature* **386**, 771.
- Andreassen, R., Egeland, T. & Olaisen, B.** (1996) Mutation rate in the hypervariable VNTR g3 (D7S22) is affected by allele length and a flanking DNA sequence polymorphism near the repeat array. *Am. J. Hum. Genet.* **59**, 360-367.
- Anjos, S. & Polychronakos, C.** (2004) Mechanisms of genetic susceptibility to type-1 diabetes: beyond HLA. *Mol. Genet. Metab.* **81**, 187-195.
- Apoil, P.-A. & Blancher, A.** (2000) Rh gene evolution in primates: study of intron sequences. *Mol. Biol. Evol.* **17**, 127-136.
- Armour, J. A. L., Harris, P. C. & Jeffreys, A.J.** (1993) Allelic diversity at minisatellite MS205 (D16S309): evidence for polarized variability. *Hum. Mol. Genet.* **2**, 1137-1145.
- Armour, J. A., Anttinen, T., May, C. A., Vega, E. E., Sajantila, A., Kidd, J. R. et al.** (1996) Minisatellite diversity supports a recent African origin for modern humans. *Nat. Genet.* **13**, 154-160.
- Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K.** (1987) *Current Protocols in Molecular Biology*. Wiley Interscience, NY.
- Bartel, D. P.** (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281-297.
- Baumann, P. & West, S. C.** (1998) DNA end-joining catalyzed by human cell-free extracts. *Proc. Natl. Acad. Sci.* **95**, 14066-14070.
- Bell, G. L., Selby, M. J. & Rutter, W. J.** (1982) The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences. *Nature* **295**, 31-35.
- Benjamin, J., Li, L., Patterson, C., Greenberg, B. D., Murphy, D. L. & Hammer, D. H.** (1996) Population and familial association between the D4 dopamine receptor gene and measures of novelty seeking. *Nat. Genet.* **12**, 81-84.
- Bennett, S. T., Lucassen, A. M., Gough, S. C., Powell, E. E., Undlien, D. E., Pritchard, L. E. et al.** (1995) Susceptibility to human type 1 diabetes at IDDM2 is determined by tandem repeat variation at the insulin gene minisatellite locus. *Nat. Genet.* **9**, 284-292.

- Berg, E. S. & Olaisen, B.** (1993) Characterization of the COL2A1 VNTR polymorphism. *Genomics* **16**, 350-354.
- Blunt, T., Finnie, N. J., Taccioli, G. E., Smith, G. C., Demengeot, J., Gottlieb, T. M. et al.** (1995) Defective DNA-dependent protein kinase activity is linked to V(D)J recombination and DNA repair defects associated with the murine scid mutation. *Cell* **80**, 813-823.
- Boán, F., González, A.I., Rodríguez, J. M. & Gómez-Márquez, J.** (1997) Molecular characterization of a new human minisatellite that is able to form single-stranded loops in vitro and is recognized by nuclear proteins. *FEBS Lett.* **418**, 251-257.
- Boán, F., Rodríguez, J. M. & Gómez-Márquez, J.** (1998) A non-hypervariable human minisatellite strongly stimulates in vitro intramolecular homologous recombination. *J. Mol. Biol.* **278**, 499-505.
- Boán, F., Rodríguez, J. M., Mouriño, S., Blanco, M. G., Viñas, A., Sánchez, L. & Gómez-Márquez, J.** (2002) Recombination analysis of the human minisatellite Msh42 suggests the existence of two distinct pathways for initiation and resolution of recombination at Msh42 in rat testes nuclear extracts. *Biochemistry* **41**, 2166-2176.
- Boffelli, D. J., McAuliffe, J., Ovcharenko, D., Lewis, K. D., Ovcharenko, I., Pachter, L. & Rubin, E. M.** (2003) Phylogenetic shadowing of primate sequences to find functional regions of the human genome. *Science* **299**, 1391-1394.
- Boguski, M. S., Birkenmeier, E. H., Elshourbagy, N. A., Taylor, M. T. & Gordon, J. I.** (1986) Evolution of the apolipoproteins. *J. Biol. Chem.* **261**, 6398-6407.
- Bois, P. & Jeffreys, A. J.** (1999) Minisatellite instability and germline mutation. *Cell. Mol. Life. Sci.* **55**, 1636-1648.
- Bois, P. R.** (2003) Hypermutable minisatellites, a human affair? *Genomics* **81**, 349-355.
- Bois, P., Stead, J. D. H., Bakshi, S., Williamson, J., Neumann, R., Moghadaszadeh, B. & Jeffreys, A. J.** (1998) Isolation and characterization of mouse minisatellites. *Genomics* **52**, 317-330.
- Boulton, S. J. & Jackson, S. P.** (1996) *Saccharomyces cerevisiae* Ku70 potentiates illegitimate DNA double-strand break repair and serves as a barrier to error-prone DNA repair pathways. *EMBO J.* **15**, 5093-5103.
- Britten, R. J. & Kohne, D. E.** (1968) Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. *Science* **161**, 529-540.
- Brown, F. L., Musich, P. R. & Maio, J. J.** (1978) Cae I: an endonuclease isolated from the African green monkey with properties indicating site-specific cleavage of homologous and heterologous mammalian DNA. *Nucleic Acids Res.* **5**, 1093-1107.
- Brown, T.A.** (1999) Genomes. *BIOS Scientific Publishers.* Oxford.
- Brusco, A., Saviozzi, S., Cinque, F., Bottaro, A. & DeMarchi, M.** (1999) A recurrent breakpoint in the most common deletion of the Ig heavy chain locus (del A1-GP-G2-G4-E). *J. Immunol.* **163**, 4392-4398.
- Buard, J. & Vergnaud G.** (1994) Complex recombination events at the hypermutable minisatellite CEB1 (D2S90). *EMBO J.* **13**, 3203-3210.
- Buard, J., Bourdet, A., Yardley, J., Dubrova, Y. & Jeffreys, A. J.** (1998) Influences of array size and homogeneity on minisatellite mutation. *EMBO J.* **17**, 3495-3502.

- Buard, J., Brenner, C. & Jeffreyes, A. J.** (2002) Evolutionary fate of an unstable human minisatellite deduced from sperm-mutation spectra of individual alleles. *Am. J. Hum. Genet.* **70**, 1038-1043.
- Buard, J., Collick, A., Brown, J. & Jeffreyes, A. J.** (2000) Somatic versus germline mutation processes at minisatellite CEB1 (D2S90) in humans and transgenic mice. *Genomics* **65**, 95-103.
- Budman, J. & Chu, G.** (2005) Processing of DNA for nonhomologous end-joining by cell-free extract. *EMBO J.* **24**, 849-860.
- Buitkamp, J., Ammer, H. & Geldermann, H.** (1991) DNA fingerprinting in domestic animals. *Electrophoresis* **12**, 169-174.
- Burke, T. & Bruford, M. W.** (1987) DNA fingerprinting in birds. *Nature* **327**, 149-152.
- Canestro, C., González-Duarte, R. & Albalat, R.** (2002) Minisatellite instability at the Adn locus reveals somatic polymorphism in amphioxus. *Nucleic Acids Res.* **30**, 2871-2876.
- Cann, R. L., Stoneking, M. & Wilson, A. C.** (1987) Mitochondrial DNA and human evolution. *Nature* **325**, 31-36.
- Chaillet, J. R., Bader, D. S. & Jeder, P.** (1995) Regulation of genomic imprinting by gametic and embryonic processes. *Genes Dev.* **9**, 1177-1187.
- Chappell, C., Hanahaki, L. A., Karimi-Busheri, F., Weinfeld, M. & West, S. C.** (2002) Involvement of human polynucleotide kinase in double-strand break repair by non-homologous end joining. *EMBO J.* **21**, 2827-2832.
- Charlesworth, B., Sniegowski, P. & Stephan, W.** (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371**, 215-220.
- Chu, G.** (1997) Double-strand break repair. *J. Biol. Chem.* **275**, 26196-26205.
- Collick, A. & Jeffreyes, A. J.** (1990) Detection of a novel minisatellite-specific DNA binding protein. *Nucleic Acids Res.* **18**, 625-629.
- Comeron, J. M.** (2001) What controls the length of noncoding DNA? *Curr. Opin. Genet. Dev.* **11**, 652-659.
- Cote, J., Renaud, J. & Ruiz-Carrillo, A.** (1989) Recognition of (dG)n.(dC)n sequences by endonuclease G. Characterization of the calf thymus nuclease. *J. Biol. Chem.* **264**, 3301-3310.
- Critchlow, S. E. & Jackson, S. P.** (1998) DNA end-joining: from yeast to man. *Trends Biochem. Sci.* **23**, 394-398.
- Cummings, O. W., King, T. C., Holden, J. A. & Low, R. L.** (1987) Purification and characterization of the potent endonuclease in extracts of bovine heart mitochondria. *J. Biol. Chem.* **262**, 2005-2015.
- Dallas, J. F.** (1988) Detection of DNA "fingerprints" of cultivated rice by hybridization with a human minisatellite DNA probe. *Proc. Natl. Acad. Sci. USA* **85**, 6831-6835.
- Debrauwère, H., Buard, J., Tessier, J., Aubert, D., Vergnaud, G. & Nicolas, A.** (1999) Meiotic instability of human minisatellite CEB1 in yeast requires DNA double-strand breaks. *Nat. Genet.* **23**, 367-371.
- Diggle, C. P., Bentley, J. & Kiltie, A. E.** (2003) Development of a rapid small-scale DNA repair assay for use on clinical samples. *Nucleic Acids Res.* **31**, e83.
- Dignam, J. D., Lebovitz, R. M. & Roeder, R. G.** (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **11**, 1475-1489.

- Doherty, A. J. & Jackson, S. P.** (2001) DNA repair: how Ku makes DNA ends meet. *Curr. Biol.* **11**, R920-R924.
- Doolittle, W. F. & Sapienza, C.** (1980) Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284**, 601-603.
- Dunnick, W., Hertz, G. Z., Scappino, L. & Gritzmacher, C.** (1993) DNA sequences at immunoglobulin switch region recombination sites. *Nucleic Acids Res.* **21**, 365-372.
- Ebstein, R. P., Novick, O., Umansky, R., Priel, B., Osher, Y., Blaine, D. et al.** (1996) Dopamine D4 receptor (D4DR) exon III polymorphism associated with the human personality trait of novelty seeking. *Nat. Genet.* **12**, 78-80.
- Ellsworth, D. L., Shriver, M. D. & Boerwinkle, E.** (1995) Nucleotide sequence analysis of the apolipoprotein B 3' VNTR. *Hum. Mol. Genet.* **4**, 937-944.
- Enard, W. Pääbo, S.** (2004) Comparative primate genomics. *Annu. Rev. Genomics Hum. Genet.* **5**, 351-378.
- Enokizono, Y., Matsugami, A., Uesugi, S., Fukuda, H., Tsuchiya, N., Sugimura, T., Nagao, M., Nakagama, H. & Katahira, M.** (2003) Destruction of quadruplex by proteins and its biological implications in replication and telomere maintenance. *Nucleic Acids Res. Suppl.* **3**, 231-232.
- Esashi, F., Christ, N., Gannon, J., Liu, Y., Hunt, T., Jasin, M. & West, S. C.** (2005) CDK-dependent phosphorylation of BRCA2 as a regulatory mechanism for recombination and repair. *Nature* **434**, 598-604.
- Evans, C. J. & Aguilera, R. J.** (2003) DNase II: genes, enzymes and functions. *Gene* **322**, 1-15.
- Feldmann, E., Schmiemann, V., Goedecke, W., Reichenberger, S. & Pfeiffer, P.** (2000) DNA double-strand break repair in cell-free extracts from Ku80-deficient cells: implications for Ku serving as an alignment factor in non-homologous DNA end joining. *Nucleic Acids Res.* **28**, 2585-2596.
- Fukuda, H., Takashi, S., Nagao, M. & Nakagama, H.** (2001) Detection and isolation of minisatellite Pc-1 binding proteins. *Biochim. Biophys. Acta* **1528**, 152-158.
- Galas D.J. & Schmitz A.** (1978) DNase I footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Res.* **5**, 3157-3170.
- Gellert, M.** (2002) V(D)J recombination: RAG proteins, repair factors, and regulation. *Annu. Rev. Biochem.* **71**, 101-132.
- Giraud, T., Fortini, D., Levis, C. & Brygoo, Y.** (1998) The minisatellite MSB1, in the fungus *Botrytis cinerea*, probably mutates by slippage. *Mol. Biol. Evol.* **15**, 1524-1531.
- Gómez-Márquez, J. & Rodríguez, P.** (1998) Prothymosin  $\alpha$  is a chromatin-remodelling protein in mammalian cells. *Biochem. J.* **333**, 1-3.
- Gómez-Márquez, J., Puga, A. & Notkins, A.L.** (1985) Regions of the terminal repetitions of the *Herpes simplex* virus type I genome. *J. Biol. Chem.* **260**, 3490-3495.
- González, A.I.** (1993) Homología entre el DNA humano y el fragmento SmaIA del genoma del virus Herpes simplex tipo 1: nuevos minisatélites en el genoma humano. *Tesis Doctoral*. Departamento de Bioquímica e Biología Molecular, Facultad de Biología, Universidad de Santiago de Compostela.
- Goodman, M.** (1999) The genomic record of humankind's evolutionary roots. *Am. J. Hum. Genet.* **64**, 31-39.
- Goodman, M., Porter, C. A., Czelusniak, J., Page, S. L., Schneider, H., Shoshani, J. et al.**

- (1998) Towards a phylogenetic classification of Primates based on DNA evidence complemented by fossil evidence. *Mol. Phylogenet. Evol.* **9**, 585-598.
- Gorgoulis, V. G., Vassiliou, L. F., Karakaidos, P., Zacharatos, P., Kotsinas, A., Liloglou, T. et al.** (2005) Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature* **434**, 907-913.
- Göttlich, B., Reichenberger, S., Feldmann, E. & Pfeiffer, P.** (1998) Rejoining of double-strand breaks *in vitro* by single-strand annealing. *Eur. J. Biochem.* **258**, 387-395.
- Gottlieb, J. & Muzyczka, N.** (1988) In vitro excision of adeno-associated virus DNA from recombinant plasmids: isolation of an enzyme fraction from HeLa cells that cleaves DNA at poly(G) sequences. *Mol. Cell. Biol.* **8**, 2513-2522
- Gottlieb, J. & Muzyczka, N.** (1990) Purification and characterization of HeLa endonuclease R. *J. Biol. Chem.* **265**, 10836-10841.
- Grawunder, U., Wilm, M., Wu, X., Kulesza, P., Wilson, T. E., Mann, M. & Lieber, M. R.** (1997) Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. *Nature* **388**, 492-495.
- Gray, I. C. & Jeffreys, A. J.** (1991) Evolutionary transience of hypervariable minisatellites in man and the primates. *Proc. R. Soc. London* **243**, 241-253.
- Gumucio, D. L., Heilstedt-Williamson, H., Gray, T. A., Tarlé, S. A., Shelton, D. A., Tagle, D. A. et al.** (1992) Phylogenetic footprint reveals a nuclear protein which binds to silencer sequences in the human gamma and epsilon globin genes. *Mol. Cell. Biol.* **12**, 4919-4929.
- Haber, J. E. & Louis, E. J.** (1998) Minisatellite origins in yeast and humans. *Genomics* **48**, 132-135.
- Hall, T. A.** (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids Symp. Ser.* **41**, 95-98.
- Handt, O., Sutherland, G. R., & Richards, R. I.** (2000) Fragile sites and minisatellite repeat instability. *Mol. Genet. Metab.* **70**, 99-105.
- He, Q., Cederberg, H. & Rannug, U.** (2002) The influence of sequence divergence between alleles of the human MS205 minisatellite incorporated into the yeast genome on length-mutation rates and lethal recombination events during meiosis. *J. Mol. Biol.* **319**, 315-327.
- He, Q., Cederberg, H., Armour, J. A. L., May, C. A. & Rannug U.** (1999) Cis-regulation of inter-allelic exchanges in mutation at human minisatellite MS205 in yeast. *Gene* **232**, 143-153.
- Heller, M., Henderson, A. & Kieff, E.** (1982) Repeat array in Epstein-Barr virus DNA is related to cell DNA sequences interspersed on human chromosomes. *Proc. Natl. Acad. Sci. USA* **79**, 5916-5920.
- Huang, W., Fu, Y. X., Chang, B. H., Gu, X., Jorde, L. B. & Li, W. H.** (1998) Sequence variation in ZFX introns in human populations. *Mol. Biol. Evol.* **15**, 138-142.
- Inoue, H., Nojima, H. & Okayama, H.** (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**, 23-28.
- International Human Genome Sequencing Consortium** (2001) Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921.
- International Human Genome Sequencing Consortium** (2004) Finishing the euchromatic sequencing of the human genome. *Nature* **431**, 931-945.
- Iyengar, S., Marinero, M., Deinard, C., Rosenbaum, H. C., Sirugo, G., Castiglione, C. et al.**

(1998) Analysis of cross species polymerase chain reactions products to infer the ancestral state of human polymorphisms. *DNA Seq.* **8**, 317-327.

**Jauert, P. A. & Kirkpatrick, D. T.** (2005) Length and sequence heterozygosity differentially affect HRAS1 minisatellite stability during meiosis in yeast. *Genetics*. Publicación electrónica (16-04-2005) pendiente de impresión.

**Jeffreys, A. J. & Morton, D. B.** (1987) DNA fingerprints of dogs and cats. *Anim. Genet.* **18**, 1-15.

**Jeffreys, A. J., Barber, R., Bois, P., Buard, J., Dubrova, Y. E., Grant, G. et al.** (1999) Human minisatellites, repeat DNA instability and meiotic recombination. *Electrophoresis* **20**, 1665-1675.

**Jeffreys, A. J., MacLeod, A., Tamaki, K, Neil, D. L. & Monckton, D. G.** (1991) Minisatellite repeat coding as a digital approach to DNA typing. *Nature* **354**, 204-209.

**Jeffreys, A. J., Neil, D. L. & Neumann, R.** (1998) Repeat instability at human minisatellites arising from meiotic recombination. *EMBO J.* **17**, 4147-4157.

**Jeffreys, A. J., Neumann, R. & Wilson, V.** (1990) Repeat unit sequence variation in minisatellites: a novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. *Cell* **60**, 473-485.

**Jeffreys, A. J., Royle, N. J., Wilson, V. & Wong, Z.** (1988) Spontaneous mutation rates to new length alleles at tandem repetitive hypervariable loci in human DNA. *Nature* **332**, 278-281.

**Jeffreys, A. J., Tamaki, K., MacLeod, A., Monckton, D. G., Neil, D. L. & Armour, J.A.L.** (1994) Complex gene conversion events in germline mutation at human minisatellites. *Nat. Genet.* **6**, 136-145.

**Jeffreys, A. J., Wilson, V. & Thein, S. L.** (1985a) Hypervariable minisatellite regions in human DNA. *Nature* **314**, 67-73

**Jeffreys, A. J., Wilson, V. & Thein, S. L.** (1985b) Individual-specific "fingerprints" of human DNA. *Nature* **316**, 76-79.

**Jeffreys, A. J., Wilson, V., Kelly, R., Taylor, B. A. & Bulfield, G.** (1987) Mouse DNA "fingerprints": analysis of chromosome localization and germ-line stability of hypervariable loci in recombinant inbred strains. *Nucleic Acids Res.* **15**, 2823-2836.

**Jobling, M. A., Bouzekri, N. & Taylor, P. G.** (1998) Hypervariable digital DNA codes for human parental lineages: MVR-PCR at the Y-specific minisatellite, MSY1 (DYF155S1). *Hum. Mol. Genet.* **7**, 643-653.

**Kanaar, R., Hoeijmakers, J. H. J. & van Gent, D. C.** (1998) Molecular mechanisms of DNA double-strand break repair. *Trends Cell Biol.* **8**, 483-489.

**Karetsou, Z., Sandaltzopoulos, R., Frangou-Lizaridis, M., Lai, C.-Y., Tsolas, O., Becker, P. B. & Papamarcaki, T.** (1998) Prothymosin alpha modulates the interaction of histone H1 with chromatin. *Nucleic Acids Res.* **26**, 3111-3118.

**Karran, P.** (2000) DNA double strand break repair in mammalian cells. *Curr. Opin. Genet. Dev.* **10**, 144-150.

**Kennedy, G. C., German, M. S. & Rutter, W. J.** (1995) The minisatellite in the diabetes susceptibility locus IDDM2 regulates insulin transcription. *Nat. Genet.* **9**, 293-298.

**Kirkpatrick D.T. & Petes T.D.** (1997) Repair of DNA loops involves DNA-mismatch and nucleotide-excision repair proteins. *Nature* **387**, 929-931.

**Kishino, H., & Hasegawa, M.** (1989) Evaluation of the maximum

likelihood estimate of the evolutionary tree topologies from DNA sequence data and the branching order in hominoidea. *J. Mol. Evol.*, **29**, 170-179.

**Labhart, P.** (1999) Nonhomologous DNA end joining in cell-free extracts. *Eur. J. Biochem.* **265**, 849-861.

**Lalioi, M. D., Scott, H. S., Buresi, C., Rossier, C., Bottani, A., Morris, M. A. et al.** (1997) Dodecamer repeat expansion in cystatin B gene in progressive myoclonus epilepsy. *Nature*. **386**, 847-850.

**Levinson, G. & G. A. Gutman.** (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **4**, 203-221.

**Lichter, J. B., Barr, C. L., Kennedy, J. L., Van Tol H. H. M., Kidd, K. K. & Livak, K. J.** (1993) A hypervariable segment in the human dopamine receptor D4 (D4DR) gene. *Hum. Mol. Genet.* **2**, 767-773.

**Lieber, M. R., Ma, Y., Pannicke, U. & Schwarz, K.** (2003) Mechanism and regulation of human non-homologous DNA end-joining. *Nat. Rev. Mol. Cell Biol.* **4**, 712-720.

**Low, R. L., Cummings, O. W. & King, T. C.** (1987) The bovine mitochondrial endonuclease prefers a conserved sequence in the displacement loop region of mitochondrial DNA. *J. Biol. Chem.* **262**, 16164-16170.

**Liu, G., Zhao, S., Bailey, J. A., Sahinalp, S. C., Alkan, C., Tuzun, E. et al.** (2003) Analysis of primate genomic variation reveals a repeat-driven expansion of the human genome. *Genome Res.* **13**, 358-368.

**Ludwig, M. Z.** (2002) Functional evolution of noncoding DNA. *Curr. Opin. Genet. Dev.* **12**, 634-639.

**Ma, Y., Lu, H., Tippin, B., Goodman, M. F., Shimazaki, N., Koiwai, O. et al.** (2004) A biochemically defined system for

mammalian nonhomologous DNA end-joining. *Mol. Cell* **16**, 701-13.

**Ma, Y., Lu, H., Tippin, B., Goodman, M. F., Shimazaki, N., Koiwai, O. et al.** (2004) A biochemically defined system for mammalian nonhomologous DNA end joining. *Mol. Cell* **16**, 701-713.

**Ma, Y., Pannicke, U., Schwarz, K. & Lieber, M.R.** (2002) Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell* **108**, 781-94.

**Mason, R. M., Thacker, J. & Fairman, M. P.** (1996) The joining of non-complementary DNA double-strand breaks by mammalian extracts. *Nucleic Acids Res.* **24**, 4946-4953.

**Maxam A.M. & Gilbert W.** (1977) A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA* **74**, 560-564.

**May, C. A., Jeffreys, A. J. & Armour J. A. L.** (1996) Mutation rate heterogeneity and the generation of allele diversity at the human minisatellite MS205 (D16S309). *Hum. Mol. Genet.* **5**, 1823-1833.

**McKenna, W. G., Maio, J. J. & Brown, F. L.** (1981) Purification and properties of a mammalian endonuclease showing site-specific cleavage of DNA. *J. Biol. Chem.* **256**, 6435-6443.

**Messier, W., Li, S.-H. & Stewart, C.-B.** (1996) The birth of microsatellites. *Nature* **381**, 483.

**Miyamoto, M. M., Slightom, J. L. & Goodman, M.** (1987) Phylogenetic relations of human and African apes from DNA sequences of the  $\psi\eta$ -globin region. *Science* **238**, 369-373.

**Monckton, D. G., Neumann, R., Guram, T., Fretwell, N., Tamaki, K., MacLeod, A. & Jeffreys, A. J.** (1994) Minisatellite mutation rate variation associated with a flanking DNA

sequence polymorphism. *Nat. Genet.* **8**, 162-170.

**Nagaki, S., Yamamoto, M., Yumoto, Y., Shirakawa, H., Yoshida, M. & Teraoka, H.** (1998) Non-histone chromosomal proteins HMG1 and 2 enhance ligation reaction of DNA double-strand breaks. *Biochem. Biophys. Res. Comm.* **246**, 137-141.

**Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, R., Culver, M. et al.** (1987) Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* **235**, 1616-1622.

**Neumann, B., Kubicka, P. & Bartel, D. P.** (1995) Characteristics of imprinted genes. *Nat. Genet.* **9**, 12-13.

**Nicolas, A. L. & Young, C. S. H.** (1994) Characterization of DNA end joining in a mammalian cell nuclear extract: junction formation is accompanied by nucleotide loss, which is limited and uniform but not site specific. *Mol. Cell. Biol.* **14**, 170-180

**North, P., Ganesh, A. & Thacker, J.** (1990) The rejoining of double-strand breaks in DNA by human extracts. *Nucleic Acids Res.* **18**, 6205-6210.

**O'Connell, P., Lathrop, G. M., Law, M., Leppert, M., Nakamura, Y., Hoff. et al.** (1987) A primary genetic linkage map for human chromosome 12. *Genomics* **1**, 93-102.

**Odersky, A., Panyutin, I. V., Panyutin, I. G., Schunck, C., Feldmann, E., Goedecke, W. et al.** (2002) Repair of sequence-specific <sup>125</sup>I-induced double-stranded breaks by nonhomologous end joining in mammalian cell-free extracts. *J. Biol. Chem.* **277**, 11756-11764

**Ogawa, Y., Aizawa, S., Shirakawa, H. & Yoshida, M.** (1995) Stimulation of transcription accompanying relaxation of chromatin structure in cells overexpressing high mobility group 1 protein. *J. Biol. Chem.* **270**, 9272-9280.

**Oh, S. & Sarnow, P.** (1993) Gene regulation: translational initiation by internal ribosomal binding. *Curr. Opin. Genet. Dev.* **3**, 295-300.

**Olsen, K. M.** (1999) Minisatellite variation in a single-copy nuclear gene: phylogenetic assesment of repeat length homoplasmy and mutational mechanism. *Mol. Biol. Evol.* **16**, 1406-1409.

**Orgel, L. E. & Crick, F. H. C.** (1980) Selfish DNA: the ultimate parasite. *Nature* **284**, 604-607.

**Pääbo, S.** (2003) The mosaic that is our genome. *Nature* **421**, 409-412.

**Pääbo, S., Irwin, D. M. & Wilson, A. C.** (1990) DNA damage promotes jumping between templates during enzymatic amplification. *J. Biol. Chem.* **265**, 4718-4721.

**Pâques, F. & Haber, J. E.** (1999) Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **63**, 349-404.

**Pastink, A., Eecken, J. C. & Lohman, P. H.** (2001) Genomic integrity and the repair of double-strand break. *Mutat. Res.* **480-481**, 37-50.

**Peden, K., Mounts, P. & Hayward G. S.** (1982) Homology between mammalian cell DNA sequences and human herpesvirus genomes detected by a hybridization procedure with high-complexity probe. *Cell* **31**, 71-80.

**Pfeiffer, P., Feldmann, Elke, Odersky, A., Kuhfittig-Kulle, S. & Goedecke, W.** (2005) Analysis of double-strand break repair by non-homologous DNA end joining in cell-free extracts from mammalian cells. *Methods Mol. Biol.* **291**, 351-371.

**Phelan, C. M., Rebbeck, T. R., Weber, B. L., Devilee, P., Ruttledge, M. H., Lynch, H. T. et al.** (1996) Ovarian cancer risk in BRCA1 carriers is modified by the HRAS1 variable

number tandem repeat (VNTR) locus. *Nat. Genet.* **12**, 309-311.

**Pospiech, H., Rytönen, A. K. & Syväoja, J. E.** (2001) The role of DNA polymerase activity in human non-homologous end joining. *Nucleic Acids Res.* **29**, 3277-3288.

**Puga, A., Cantin, E. M. & Notkins, A.L.** (1982) Homology between murine and human cellular DNA sequences and terminal repetition of the S component of herpes simplex virus type 1 DNA. *Cell* **31**, 81-87.

**Pugliese, A., Zeller, M., Fernandez, A. Jr., Zalberg, L. J., Bartlett, R. J., Ricordi, C. et al.** (1997) The insulin gene is transcribed in the human thymus and transcription levels correlated with allelic variation at the INS VNTR-IDD2 susceptibility locus for type-1 diabetes. *Nat. Genet.* **15**, 293-297.

**Raghavan, S. C. & Raman, M. J.** (2004) Nonhomologous end joining of complementary and noncomplementary DNA termini in mouse testicular extracts. *DNA Repair (Amst)* **3**, 1297-1310.

**Ramsden, D. A. & Gellert, M.** (1998) Ku protein stimulates DNA end-joining by mammalian DNA ligases: a direct role for Ku in repair of DNA double-strand breaks. *EMBO J.* **17**, 609-614.

**Richardson, C. & Jasin, M.** (2000) Coupled homologous and nonhomologous repair of a double-strand break preserves genomic integrity in mammalian cells. *Mol. Cell. Biol.* **20**, 9068-9075.

**Rodríguez, J. M.** (2002) Estudio *in vitro* del minisatélite Msh42: Análisis poblacional y localización en el genoma humano, comparación de la capacidad recombinogénica de los alelos e interacción con proteínas nucleares. *Tesis Doctoral*. Departamento de Bioquímica e Biología Molecular, Facultad de

Biología, Universidade de Santiago de Compostela.

**Royle, N.J., Clarkson, R.E., Wong, Z. & Jeffreys, A.J.** (1988) Clustering of hypervariable minisatellites in the proterminal regions of the human autosomes. *Genomics* **3**, 352-360.

**Ruiz-Carrillo, A. & Renaud, J.** (1987) Endonuclease G: a (dG)n X (dC)n-specific DNase from higher eukaryotes. *EMBO J.* **6**, 401-407.

**Saitou, N. & Ueda, S.** (1994) Evolutionary rates of insertion and deletion in noncoding nucleotide sequences of primates. *Mol. Biol. Evol.* **11**, 504-512.

**Sakano, H., Kurosawa, Y., Weigert, M. & Tonegawa S.** (1981) Identification and nucleotide sequence of a diversity DNA segment (D) of immunoglobulin heavy-chain genes. *Nature* **290**, 562-565.

**Sambrook, J. & Russell D.W.** (2001) Molecular cloning. A laboratory manual. *Cold Spring Harbor Laboratory Press*. New York.

**Sancar, A., Lindsey-Boltz, L. A., Unsal-Kacmaz, K. & Linn, S.** (2004) Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu. Rev. Biochem.* **73**, 39-85.

**Sathees, C. R. & Raman, M. J.** (1999) Mouse testicular extracts process DNA double-strand breaks efficiently by DNA end-to-end joining. *Mutat. Res.* **433**, 1-13.

**Schäfer, P., Scholz, S. R., Gimadutdinov, O., Cymerman, I. A., Bujnicki, J. M., Ruiz-Carrillo, A. et al.** (2004) Structural and functional characterization of mitochondrial EndoG, a sugar non-specific nuclease which plays an important role during apoptosis. *J. Mol. Biol.* **338**, 217-228.

**Segade, F. & Gómez-Márquez, J.** (1999) Molecules in Focus: prothymosin  $\alpha$ . *Int. J. Biochem. Cell Biol.* **3**, 1241-1246.

- She, X., Jiang, Z., Clark, R. A., Liu, G., Cheng, Z., Tuzun, E. et al.** (2004) Shotgun sequence assembly and recent segmental duplications within the human genome. *Nature* **431**, 927-930.
- Singleton, B. K. & Jeggo, P. A.** (1999) Double-strand break repair and V(D)J recombination. In Smith, P. J. and Jones, C. J. (eds.) *DNA Recombination and Repair*. Oxford University Press, Oxford, pp.16-37.
- Singleton, B. K., Priestley, A., Steingrimsdottir, H., Gell, D., Blunt, T., Jackson, S. P. et al.** (1997) Molecular and biochemical characterization of xrs mutants defective in Ku80. *Mol. Cell. Biol.* **17**, 1264-1273.
- Smith, G. P.** (1976) Evolution of repeated DNA sequences by unequal crossover. *Science* **191**, 528-535.
- Smith, G. R., Kunes S., Schultz, D., Taylor, A., & Triman, K. L.** (1981) Structure of *Chi* hotspots of generalized recombination. *Cell* **24**, 429-436.
- Sobczak, J. & Duguet, M.** (1988) Effect of histone H1, poly(ethyleneglycol) and DNA concentration on intermolecular and intramolecular ligation by T4 DNA ligase. *Eur. J. Biochem.* **175**, 379-385.
- Stein, L. D.** (2004) End of the beginning. *Nature* **431**, 915-916.
- Storz, G.** (2002) An expanding universe of noncoding RNAs. *Science* **296**, 1260-1263.
- Stringer, C.** (2003) Human evolution: out of Ethiopia. *Nature* **423**, 692-694.
- Swofford, D. L.** (1998) PAUP\*. Phylogenetic Analysis Using Parsimony (\* and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.
- Szostak J. W., Orr-Weaver T. L., Rothstein R. J. & Stahl F. W.** (1983) The double-strand break repair model for recombination. *Cell* **33**, 25-35.
- Tagle, D. A., Koop, B. F., Goodman, M., Slightom, J. L., Hess, D. & Jones, R.T.** (1988) Embryonic epsilon and gamma globin genes of a prosimian primate (*Galago crassicaudatus*): nucleotide and aminoacid sequences, developmental regulation and phylogenetic footprints. *J. Mol. Biol.* **203**, 439-455.
- Takata, M., Sasaki, M. S., Sonoda, E., Morrison, C., Hashimoto, M., Utsumi, H. et al.** (1998) Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *EMBO J.* **17**, 5497-5508.
- Taylor, J. S. & Breden, F.** (2000) Slipped-strand mispairing at non-contiguous repeats in *Poecilia reticulata*: A model for minisatellite birth. *Genetics* **155**, 1313-1320.
- Thornton, J. W., & DeSalle, R.** (2000) Gene family evolution and homology: genomics meets phylogenetics. *Annu. Rev. Genomics Hum. Genet.* **1**, 41-73.
- Trepicchio, W. L. & Krontiris, T. G.** (1993) IGH minisatellite suppression of USF binding site and E $\mu$ -mediated transcriptional activation of the adenovirus major late promoter. *Nucleic Acids Res.* **21**, 977-985.
- Turri, M. G., Kuin, C. A. & Porter, A. C. G.** (1995) Characterization of a novel minisatellite that provides multiple splice donor sites in an interferon-induced transcript. *Nucleic Acids Res.* **23**, 1854-1861.
- Ullman, A., Jacob, F. & Monod, J.** (1967) Characterization by *in vitro* complementation of a peptide corresponding to an operator proximal segment of the beta-galactosidase

- structural gene of *Escherichia coli*. *J. Mol. Biol.* **24**, 339-343.
- Ureta-Vidal, A., Ettwiller, L. & Birney, E.** (2003) Comparative genomics: genome-wide analysis in metazoan eukaryotes. *Nat. Rev. Genet.* **4**, 251-262.
- Vafiadis, P., Bennett, S. T., Todd, J. A., Nadeau, J., Grabs, R., Goodyer, C. G. et al.** (1997) Insulin expression in human thymus is modulated by INS VNTR alleles at the IDDM2 locus. *Nat. Genet.* **15**, 289-292.
- Van Dyck, E., Stasiak, A. Z., Stasiak, A. & West, S. C.** (1999) Binding of double-strand breaks in DNA by human Rad52 protein. *Nature* **398**, 728-731.
- Van Tol, H. H. M., Wu, C. M., Guan, H. C., Ohara, K., Bunzow, J. R., Civelli, O. et al.** (1992) Multiple dopamine D4 receptor variants in the human population. *Nature.* **358**, 149-152.
- Vega, F., Vidal, A., Hellman, U., Wernstedt, C., & Domínguez, F.** (1998) Prothymosin alpha stimulates Ca<sup>2+</sup>-dependent phosphorylation of elongation factor 2 in cellular extracts. *J. Biol. Chem.* **273**, 10147-10152.
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G. et al** (2001) The sequence of the human genome. *Science* **291**, 1304-1351.
- Vergnaud, G. & Denoeud, F.** (2000) Minisatellites: mutability and genome architecture. *Genome Res.* **10**, 899-907.
- Vogt P.** (1990) Potential genetic functions of tandem repeated DNA blocks in the human genome are based on a highly conserved "chromatin folding code". *Hum. Genet.* **84**, 301-336.
- Wahls, W. P. & Moore, P. D.** (1998) Recombination hotspot activity of hypervariable minisatellite DNA requires minisatellite DNA binding proteins. *Somat. Cell Mol. Genet.* **24**, 41-51.
- Wahls, W. P. & Smith G. R.** (1994) A heteromeric protein that binds to a meiotic homologous recombination hotspot: correlation of binding and hot spot activity. *Genes Dev.* **8**, 1693-1702.
- Wahls, W. P.** (1998) Meiotic recombination hotspots: shaping the genome and insights into hypervariable minisatellite DNA change. *Curr. Top. Dev. Biol.* **37**, 37-75.
- Wahls, W. P., Swenson, G. & Moore, P. D.** (1991) Two hypervariable minisatellite DNA binding proteins. *Nucleic Acids Res.* **19**, 3269-3274.
- Wahls, W. P., Wallace, L. J. & Moore, P. D.** (1990) Hypervariable minisatellite DNA is a hotspot for homologous recombination in human cells. *Cell* **60**, 95-103.
- West, S. C., Chappell, L., Hanakahi, L. A., Massom, J. Y., Mcilwraith, M. & Van Dyck, E.** (2000). Double-strand break repair in human cells. *Cold. Spring Harbor Symp. Quant. Biol.* **65**, 315-321.
- Weterings, E. & van Gent, D. C.** (2004) The mechanism of non-homologous end-joining: a synopsis of synapsis. *DNA Rep (Amst.)* **3**, 1425-1435.
- Wyman, A. R. & White, R.** (1980) A highly polymorphic locus in human DNA. *Proc. Natl. Acad. Sci. USA* **77**, 6754-6758.
- Yamanaka, S., Katayama, E., Yoshioka, K., Nagaki, S., Yoshida, M. & Teraoka, H.** (2002) Nucleosome linker proteins HMGB1 and histone H1 differentially enhance DNA ligation reactions. *Biochem. Biophys. Res. Comm.* **292**, 268-273.
- Yu, S., Mangelsdorf, M, Hewett, D., Hobson, L., Baker, E., Eyre, H. J. et al.** (1997) Human chromosome fragile site FRA16B is an amplified AT-rich minisatellite repeat. *Cell* **88**, 367-374.

**Zhong, Q., Boyer, T. G., Chen, P.-L. & Lee, W.-H.** (2002) Deficient nonhomologous end-joining activity in cell-free extracts from Brca1-null fibroblasts. *Cancer Res.* **62**, 3966-3970.

# *Apéndice*



Los resultados obtenidos en la presente Tesis Doctoral han sido publicados en las siguientes revistas internacionales (se adjuntan a continuación):

**Boán, F., Blanco, M.G., Quinteiro, J., Mouriño, S. & Gómez-Márquez, J.** (2004) Birth and evolutionary history of a human minisatellite. *Mol. Biol. Evol.* **21**, 228-235.

**Blanco, M. G., Boán, F. & Gómez-Márquez, J.** (2004) A paradox in the end-joining assays. *J. Biol. Chem.* **279**, 26797-26801.

Además, hay un tercer artículo que se encuentra bajo revisión para su publicación:

**Blanco, M. G., Boán, F., Barros, P. & Gómez-Márquez, J.** (2005) Analysis of DNA double-strand breaks generated by two independent enzymatic activities in nuclear extracts. *En revision para J. Mol. Biol.*