



11281

**UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO**

INSTITUTO DE FISIOLÓGIA CELULAR  
DOCTORADO EN CIENCIAS BIOMEDICAS

**IDENTIFICACION Y CARACTERIZACION DE LA  
ESTRUCTURA DE LA CROMATINA DEL LÍMITE 5' DEL  
DOMINIO  $\alpha$ -GLOBINA DE POLLO**

tesis  
para obtener el grado de

**DOCTORA EN CIENCIAS**

Biol. Viviana del Carmen Valadez Graham

Director de Tesis: Dr. Félix Recillas Targa



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

La Dirección General de Bibliotecas de la  
a difundir en formato digital el contenido de mi trabajo  
obtenido de mi trabajo.

NOMBRE: Viviana Valdez  
Graham

FECHA: 16/3/04

~~\_\_\_\_\_~~

**ESTA TESIS NO SALE  
DE LA BIBLIOTECA**

**IDENTIFICACIÓN Y CARACTERIZACIÓN DE LA ESTRUCTURA DE LA CROMATINA  
DEL LIMITE 5' DEL DOMINIO  $\alpha$ -GLOBINA DE POLLO**

<b>RESUMEN</b> .....	<b>3</b>
<b>ABSTRACT</b> .....	<b>4</b>
<b>INTRODUCCIÓN</b> .....	<b>5</b>
<b>DOMINIOS CROMOSÓMICOS</b> .....	<b>6</b>
<b>ESTRUCTURA DE LA CROMATINA</b> .....	<b>8</b>
<b>MODIFICACIONES POST-TRADUCCIONALES DE LAS HISTONAS</b> .....	<b>14</b>
<b>VARIANTES DE HISTONAS</b> .....	<b>19</b>
<i>Variantes de la histona H3</i> .....	<i>19</i>
<i>Variantes de la histona H2A</i> .....	<i>20</i>
<i>Variantes de la histona H1</i> .....	<i>22</i>
<b>SITIOS DE HIPERSENSIBILIDAD</b> .....	<b>23</b>
<b>REGIONES DE CONTROL DEL LOCUS Ó LCR'S</b> .....	<b>26</b>
<b>LÍMITES DE LOS DOMINIOS</b> .....	<b>40</b>
<b>LAS SECUENCIAS DE UNIÓN A LA MATRIZ NUCLEAR O MAR'S</b> .....	<b>54</b>
<b>TERRITORIOS NUCLEARES</b> .....	<b>56</b>
<b>LOS "INSULATORS" Y SU RELACIÓN CON LA ORGANIZACIÓN NUCLEAR Y SUB-NUCLEAR.</b> .....	<b>59</b>
<b>EL FACTOR TRANSCRIPCIONAL CTCF</b> .....	<b>61</b>
<i>Estructura peptídica y génica de CTCF</i> .....	<i>62</i>
<i>CTCF y represión</i> .....	<i>65</i>
<i>CTCF y BORIS</i> .....	<i>66</i>
<b>LOS DOMINIOS <math>\alpha</math> Y <math>\beta</math>-GLOBINA DE POLLO</b> .....	<b>68</b>
<i>El dominio <math>\alpha</math>-globina de pollo</i> .....	<i>73</i>
<b>PLANTEAMIENTO DEL PROBLEMA</b> .....	<b>76</b>
<b>HIPÓTESIS</b> .....	<b>78</b>
<b>OBJETIVOS</b> .....	<b>78</b>
<b>MATERIALES Y METODOS</b> .....	<b>79</b>
<b>RESULTADOS</b> .....	<b>85</b>
<b>"INSULATORS" DEPENDIENTES DE CTCF EN EL COSTADO 5' DEL DOMINIO <math>\alpha</math>-GLOBINA DE POLLO</b> .....	<b>85</b>
<i>I. Análisis de la actividad de bloqueo de enhancer de los fragmentos que corresponden a los tres sitios de hipersensibilidad.</i> .....	<i>87</i>
<i>II.- Participación del factor CTCF en la actividad de bloqueo de enhancer.</i> .....	<i>91</i>
<i>III. Interacción in vivo del factor CTCF.</i> .....	<i>104</i>
<b>DISCUSIÓN</b> .....	<b>109</b>
<b>PERSPECTIVAS</b> .....	<b>113</b>
<b>BIBLIOGRAFIA</b> .....	<b>126</b>



ANEXO 1.....	135
<b>MODULACIÓN DE LA ACTIVIDAD DEL ENHANCER LOCALIZADO EN EL COSTADO 3' DEL DOMINIO <math>\alpha</math>-GLOBINA DE POLLO DURANTE LA DIFERENCIACIÓN ERITROIDE (RINCÓN-ARANO <i>ET AL</i> 2004).....</b>	<b>135</b>
<b>DISCUSIÓN DE RESULTADOS .....</b>	<b>136</b>
<i>Los factores Sp1 y YY1 .....</i>	<i>136</i>
<i>Los factores EKLF y NF-E2 .....</i>	<i>137</i>
<i>El factor GATA-1.....</i>	<i>138</i>
<b>PERSPECTIVAS.....</b>	<b>138</b>
APENDICE .....	140

## RESUMEN

La necesidad funcional de compactar al genoma de una célula eucariota en cromatina para contenerlo al interior del núcleo restringe el acceso de proteínas que regulan la actividad genética a sus secuencias blanco en el ADN. Por lo tanto debe existir un remodelaje regulado de la cromatina que permita la accesibilidad de las proteínas a sus secuencias blanco en el ADN y así permitir que se lleve a cabo la expresión eficiente y ordenada de un gen. Para que la expresión génica sea regulada en un tiempo y espacio específicos durante el desarrollo de un organismo multicelular, se ha propuesto que los genes se encuentran organizados dentro del genoma en dominios cromosómicos, los cuales son unidades transcripcionales cuya expresión es regulada de manera independiente. Estos dominios pueden contener uno o varios genes y tienen límites bien definidos que ayudan a mantenerlos aislados de elementos de regulación ajenos al dominio y a llevar a cabo una correcta regulación de la expresión de los genes dentro del dominio. Estos límites llamados "insulators" correlacionan con sitios de hipersensibilidad al corte por nucleasas y tienen dos propiedades funcionales: 1) bloquean las interacciones entre enhancers y promotores, únicamente cuando se colocan entre el enhancer y el promotor, en los "insulators" de vertebrados esta propiedad depende del factor CTCF y 2) protegen a un transgen del efecto de posición causado por el entorno cromatínico. En este trabajo definimos lo que podría ser el límite 5' del dominio  $\alpha$ -globina de pollo. Hemos identificado en el extremo 5' del dominio, tres sitios de hipersensibilidad al corte por la DNasa I, que corresponden a dos nuevos "insulators" llamados  $\alpha$ EHS-1.4 y  $\alpha$ CHS-1.2, mediante un análisis de secuencia se determinó la presencia de posibles sitios de unión a la proteína CTCF. Ensayos funcionales en transfecciones transitorias permitieron demostrar que dos de estos tres sitios tienen una actividad de bloqueo de enhancer. Mediante ensayos de retardo y de inmunoprecipitación de la cromatina (ChIP) confirmamos que el factor CTCF está presente en

estas secuencias. Proponemos que esta región no-codificante pueda contribuir a la regulación del dominio a distintos niveles: estructura de la cromatina, regulación transcripcional y/o la presencia y función de transcritos intergénicos.

## ABSTRACT

In higher eukaryotes the genome is packaged into a compact and repressive chromatin structures. To allow small and specific portions of the genome to be expressed in a given cell-type, tissue or developmental stage, it is necessary a series of ordered and hierarchical chromatin remodeling events. On recent years it is more generally accepted that eukaryotic genome is organized into structural and functional discrete domains. It has been proposed that chromatin domain formation represents one of the first steps in genome remodeling influencing the control of gene expression. In recent years it has been suggested that chromatin domain boundaries could contribute to their formation and maintenance in a regulated way.

In the present work we describe the discovery of two new "insulator" elements, named  $\alpha$ EHS-1.4 and  $\alpha$ CHS-1.2 that could represent the 5' limit of the chicken  $\alpha$ -globin domain. We have identified three hypersensitive sites in the 5' side of the chicken  $\alpha$ -globin domain, we performed a sequence analysis and found putative sites for the union of the transcriptional factor CTCF. Transient transfection assay allowed us to determine that two of these three hypersensitive sites possess enhancer blocking activity. Chromatin immunoprecipitation assays and gel shifts confirmed the presence of CTCF at those sites both *in vivo* and *in vitro*. These elements might contribute to chromatin domain organization, gene regulation and/or the presence or function of intergenic transcripts. We believe these sites could represent the 5' limit of the chicken  $\alpha$ -globin domain

## INTRODUCCIÓN

La diferencia más notoria entre las células procariotas y eucariotas, es la presencia en estas últimas de un núcleo el cuál contiene al genoma. El genoma de una célula eucariota se encuentra asociado a proteínas llamadas histonas, las cuales contribuyen a empaquetar el ADN. Esta asociación del ADN con proteínas se denomina cromatina y permite que la información genética pueda ser contenida dentro del núcleo. Hasta hace algunos años se estudiaba a la estructura de la cromatina y a la activación transcripcional como dos procesos biológicos independientes, de hecho se pensaba que la cromatina no jugaba un papel esencial dentro de la regulación transcripcional y que su participación era meramente estructural. Actualmente se entiende que la regulación de la expresión de los genes de las células eucariotas, se encuentra íntimamente asociada a cambios en la arquitectura de la cromatina. El remodelaje de la estructura de la cromatina contribuye además al desarrollo de funciones esenciales de las células como la progresión del ciclo celular, condensación y descondensación de los cromosomas durante la mitosis y durante la interfase, respectivamente, así como la reparación, duplicación y recombinación del ADN.

El empaquetamiento del ADN constituye *per se* una estructura represora que impide la transcripción, pero que a su vez es capaz de remodelarse para permitir la expresión regulada de los genes. Este remodelaje en la cromatina favorece o impide de manera coordinada el ensamblaje del complejo de iniciación de la transcripción asociado a la ARN polimerasa II en las regiones promotoras de los genes y de regiones tipo enhancer, repercutiendo en la expresión de los mismos (Felsenfeld y Groudine 2003; Felsenfeld *et al.* 1996). En un nivel más elevado de complejidad, la transcripción de

los genes eucariotes es modulada además por elementos distantes, es decir, que se encuentran alejados de las regiones promotoras de los mismos. Estos elementos en *cis* son sitios de unión a factores que en cooperación con diversos complejos son capaces de regular la transcripción genética positiva o negativamente y posiblemente contribuyen también al remodelaje de la cromatina.

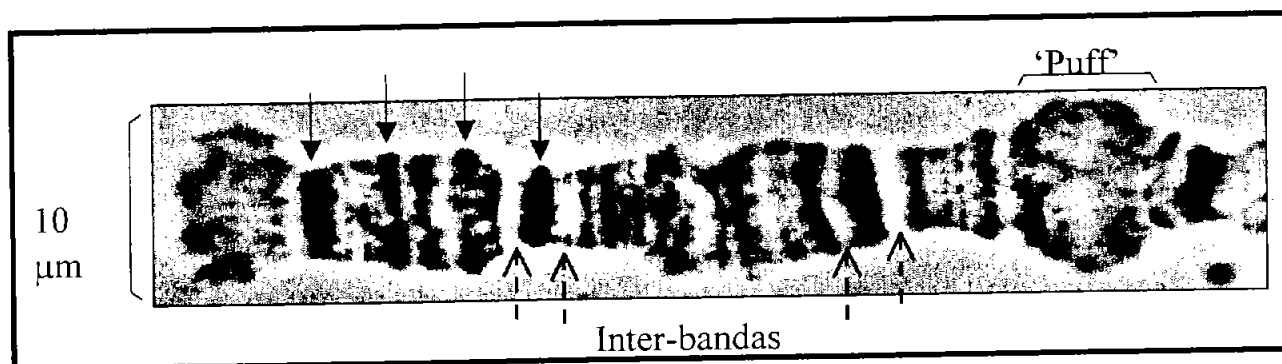
Los sitios de unión a factores, se encuentran en regiones que son hipersensibles al corte por nucleasas como por ejemplo la DNasa I; esta enzima es capaz de digerir con mayor frecuencia en regiones donde la cromatina está más laxa haciendo cortes en la doble cadena en particular en el surco menor de la hebra del ADN. La hipersensibilidad al corte por estas enzimas correlaciona con una estructura más "abierta" o accesible de la cromatina, lo que permite la unión de diferentes factores de transcripción a sus secuencias blanco en el ADN (Boyes y Felsenfeld 1996). En ocasiones, cuando los genes son activamente transcritos, esta accesibilidad al corte por endonucleasas puede extenderse por varias kilobases (kb) pudiendo incluir a uno o varios genes (Gross y Garrard 1988).

### **Dominios Cromosómicos**

A las regiones de la cromatina que se encuentran "abiertas" en un tiempo y espacio determinado del desarrollo celular se les ha denominado "dominios transcripcionalmente activos".

En años recientes se ha propuesto que estos dominios tienen límites bien definidos y la transcripción del gen o genes contenidos en ellos es regulada diferencialmente a lo largo del desarrollo por elementos situados dentro de estos límites. Existen además, evidencias citológicas que apoyan el concepto de dominios, siendo la más evidente lo que ocurre durante la expresión genética en los cromosomas politénicos de las glándulas salivales de larvas de *Drosophila*. Al observar al microscopio los cromosomas politénicos, presentan un patrón de bandas e interbandas bien definido (Fig. 1). En las bandas se encuentran los genes, los cuales al ser expresados generan una estructura conocida como "puff", la cual se observa como un ensanchamiento de la banda representativo de un relajamiento en la

estructura de la cromatina. Cabe señalar que estas estructuras o "puffs" están acotados, es decir, que no invaden bandas adyacentes y que recientemente se ha propuesto que estos límites contribuyen a la formación y mantenimiento de un dominio transcripcionalmente activo. Así pues, se ha comprobado que en los "puffs" existe un dominio que está siendo regulado y transcrito por elementos inherentes a él y que no invaden otros dominios gracias a la presencia de sus límites (Kellum y Scheld 1992).



**Fig. 1. Evidencias citológicas de la formación de los dominios transcripcionalmente activos.** Cromosoma politénico de las glándulas salivales de una larva de *Drosophila*, en la figura puede observarse el patrón de bandas de cromatina condensada donde se encuentran los genes y la formación de la estructura conocida como "Puff", en la cual se observa un relajamiento de la cromatina lo que permite la transcripción de los genes.

Además de las evidencias bioquímicas y citológicas mencionadas, existe actualmente un cúmulo de datos experimentales recientes que apoyan el concepto de una organización del genoma eucariote en dominios transcripcionalmente activos (Recillas-Targa y Razin 2001). La formación de un dominio representa probablemente uno de los niveles iniciales dentro de la regulación de la expresión genética, y el hecho de que existan límites de esos dominios nos habla del alto grado de especificidad y regulación que debe ocurrir al expresarse un gen en un tiempo y espacio definidos (Dillon y Grosveld 1994).

El estudio a nivel molecular sobre los aspectos relacionados con las proteínas involucradas en la compactación de la información genética y los elementos en *cis* que participan en la regulación transcripcional de los genes ha ayudado a entender como cooperan todos estos elementos de manera coordinada para formar, mantener y regular la expresión de los genes contenidos en un dominio.

### **Estructura de la Cromatina**

Para comprender como se forma un dominio transcripcionalmente activo, es necesario entender el contexto en el cual el ADN está organizado en cromatina y su relación con el núcleo celular. Si extendemos de manera lineal la molécula de ADN contenida en el interior del núcleo de una célula eucariota, su longitud varía entre 2 y 4 metros; tomando en cuenta que el diámetro del núcleo oscila entre 10 y 20 micras, se hace evidente que debe existir un sistema de empaquetamiento que compacte el genoma celular de manera que pueda ser contenido dentro del núcleo (Felsenfeld y Groudine 2003). Para empaquetar el ADN, la célula se vale de unas proteínas llamadas histonas las cuales forman un octámero cilíndrico sobre el que el ADN es enrollado para formar lo que se conoce como nucleosoma. El nucleosoma tiene un diámetro de 6.5 nanómetros (nm), está compuesto de un tetrámero central formado por dos copias de las histonas H3 y H4, flanqueado por dos heterodímeros de las histonas H2A y H2B (Widom 1998). Las histonas contactan con los residuos de argininas al surco menor del ADN (Luger *et al* 1997; Workman y Kingston 1998). Esta unión se estabiliza gracias a los extremos amino-terminales de las histonas, los cuales quedan expuestos por fuera del octámero de histonas y son ricos en residuos de lisina. La abundancia en residuos de lisina en los extremos amino-terminales genera una carga neta positiva, la cual interacciona con la carga negativa del ADN estabilizando la

unión del ADN al octámero de histonas, de esta manera el ADN queda enrollado alrededor del octámero.

En los extremos amino-terminales de las histonas ocurren diversas modificaciones post-traduccionales, como acetilación, fosforilación, metilación y otras, las cuáles contribuyen al remodelaje de la cromatina y juegan un papel importante en la regulación de la transcripción génica, estas modificaciones serán discutidas más adelante (Iizuka y Smith 2003; Fischle *et al* 2003)

La estructura nucleosomal es considerada como el primer nivel estructural de empaquetamiento, donde 146 pares de bases (pb) de ADN se enrollan en 1.65 vueltas alrededor del octámero de histonas. Los cilindros de histonas van posicionándose uno tras otro regularmente a lo largo del ADN, de manera independiente de secuencias de nucleótidos específicas formando una cadena de polinucleosomas conocida como "collar de perlas" o fibra de 11 nm (Fig. 2).

En el siguiente nivel de estructuración de la cromatina, la histona H1 (cuyo homólogo en pollo es la histona H5), juega un papel tanto estructural como de regulación, al menos en eucariotes superiores (Thomas 1984). El homólogo de la histona H1 en levadura ha sido recientemente descrito, y se ha demostrado que no contribuye de manera vital en la organización de la cromatina, además juega un papel importante en mecanismos de reparación del ADN, aunque este papel en eucariotes superiores no ha sido observado (Conconi y Wellinger, 2003). La histona H1 contribuye a la formación de lo que se conoce como cromatosoma que corresponde al siguiente nivel de estructuración del genoma, el cual contiene al octámero de histonas, a la histona H1 y abarca 168 pb de ADN (20 pb más por la interacción con la histona H1; Widom 1998). La histona H1 se localiza en la región internucleosomal e interacciona con el ADN que entra y sale del nucleosoma (ADN de unión), tiene un papel estructural fundamental, ya que contribuye a que la fibra de 11 nm escale al siguiente nivel de compactación del ADN en cromatina: la fibra de 30 nm o solenoide (Thomas 1984; Fig 2). La relevancia de la histona H1 en estos niveles de estructuración es el efecto que tiene la



fosforilación de H1 en 5 residuos de serina específicos, contribuyendo a la formación de los cromosomas durante la mitosis. En este nivel, la compactación llega hasta 1400 nm (Herrera *et al*, 2000; Fig 2). En cromatina *de novo*, la histona H1 se encuentra moderadamente fosforilada, a medida que la cromatina madura, la histona H1 se hiperfosforila. La fosforilación de H1 es mayor en células que se dividen rápidamente y menor en células que no están proliferando. Los niveles de fosforilación de la histona H1 son menores en G1 y mayores en la fase S del ciclo celular y en la mitosis. En esta última fase del ciclo, la fosforilación aumenta específicamente durante la metafase, que es cuando los cromosomas se encuentran condensados. Así mismo, la histona H1 está presente en una menor densidad y con un grado menor de fosforilación en dominios transcripcionalmente activos (Kamakaka y Thomas 1990; Ericsson *et al* 1990). Estos resultados han llevado a pensar que existe una relación directa entre la fosforilación de la histona H1 y la compactación de los cromosomas (Chadee *et al* 1997; Mizzen *et al* 1999).

Además de su importante papel estructural se ha propuesto que la histona H1 puede actuar en la regulación transcripcional. La regulación mediada por la histona H1, se lleva a cabo de diferentes maneras:

- 1) Un estudio de los años 90's indica que en genes activos la presencia de la histona H1 es menor que en genes inactivos (Kamakaka y Thomas 1990), sin embargo estudios más recientes en sistemas experimentales *in vitro* han demostrado que la histona H1 se encuentra en intercambio activo, es decir, que no permanece siempre unida al ADN. Además la fosforilación en sitios específicos de esta histona, correlaciona con un nivel de empaquetamiento más elevado que el del solenoide durante la mitosis. Se sugiere que la histona H1 podría actuar modulando la exposición de un sitio en el ADN a proteínas reguladoras. Al estar ausente la histona H1, no existen contactos estables entre los nucleosomas adyacentes, por lo tanto se impide el siguiente nivel de empaquetamiento de la cromatina

manteniéndose una estructura más laxa lo que permite una mayor accesibilidad al ADN (Workman y Kingston 1998).

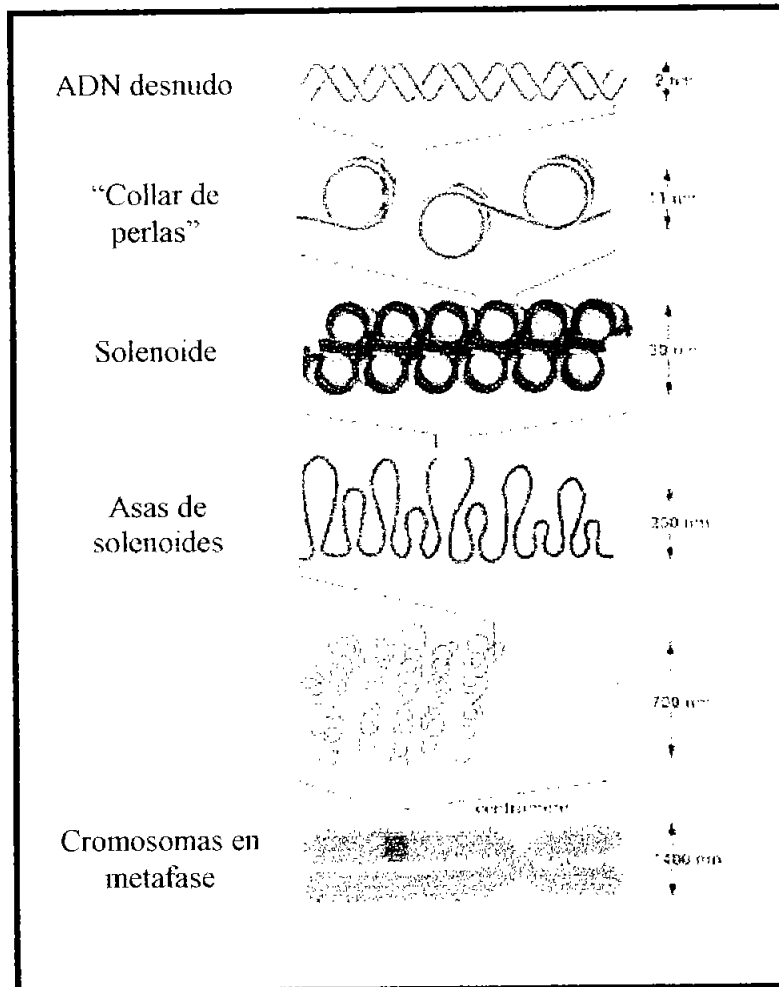
2) Se han hecho estudios *in vitro* en los cuales se ha demostrado una interferencia estérica entre los extremos amino-terminales de la histona H1 y otras histonas como la H3 impidiendo la acetilación de sus extremos amino-terminales y por lo tanto que la cromatina se descompacte interfiriendo con la unión de factores reguladores de la estructura de la cromatina y de la transcripción genética (Widom 1998).

La histona H1 estabiliza la fibra de 30 nm al fomentar contactos entre nucleosomas adyacentes (Kornberg y Lorch, 1999). Se han propuesto varios modelos para explicar la formación de la fibra de 30 nm y hasta la fecha el más aceptado es el que sugiere que la histona H1 coopera para formar una estructura de hélice con 6 nucleosomas por vuelta (Thomas 1984). Recientemente se ha comprobado por métodos físico-químicos que una región de cromatina condensada y altamente metilada de 15.5 kb localizada río arriba del dominio  $\beta$ -globina de pollo está formada por una fibra de 30 nm (Ghirlando *et al* 2004). Cabe señalar que durante la mayor parte del ciclo celular, la cromatina permanece empaquetada en fibras de 30 nm de diámetro, aunque existen niveles más complejos de compactación siendo el más alto el de los cromosomas, pero a la fecha se desconocen los mecanismos por los cuales se alcanzan los niveles más elevados de estructuración y compactación de la cromatina.

La fibra de 30 nm es una estructura altamente compacta y consecuentemente represora para la transcripción. Anteriormente se pensaba que la fibra de 11 nm era el molde natural de la transcripción, ya que esta constituía una estructura lo suficientemente "abierta" y accesible de la cromatina sobre la cual el complejo transcripcional podría realizar su función (Bulger y Groudine 1999). Sin embargo, las siguientes evidencias sugieren que el molde real donde ocurre el remodelaje de la cromatina para una subsiguiente activación de la transcripción es la fibra de 30 nm:

- Las condiciones en las que se hicieron los estudios *in vitro* utilizando la fibra de 11 nm como molde de la transcripción, no son las mismas que existen en una célula *in vivo*. De hecho en condiciones fisiológicas la fibra de cromatina tiende de manera natural a compactarse al siguiente nivel, en el cual se forman los llamados "loops" o asas de solenoides, los que se forman cuando las proteínas que se unen al ADN anclan la fibra de 30 nm a la matriz nuclear, formando de esta manera el cuello de las asas. Cada asa de solenoides contiene de 20,000 a 80,000 pb y alcanza un nivel superior de compactación de hasta 400 nm. Diversos cálculos teóricos han estimado que cada cromosoma humano contiene aproximadamente 2600 asas (Wanner y Formanek 2000).
- Si se realiza una estimación teórica del tamaño que abarca el ADN descompactado es prácticamente imposible considerar a la fibra de 11 nm como el molde real de la transcripción debido a que en un momento dado de diferenciación de la célula, los genes activos generarían una descompactación de la cromatina que impediría su inclusión dentro del núcleo (Wanner y Formanek 2000). Además, en estudios hechos por microscopía electrónica se ha observado que una descondensación por debajo del nivel de la fibra de 30 nm, ocurre cuando hay una gran densidad de polimerasas que se encuentran en la etapa de extensión. Cuando las polimerasas terminan de extender y se encuentran lo suficientemente alejadas de la región que estaba siendo transcrita, la estructura de la cromatina regresa rápidamente a la fibra de 30 nm (Orphanides y Reinberg 2000). Esta última observación explicaría la presencia de la histona H1 en dominios que están siendo transcritos contribuyendo a un pronto empaquetamiento de los genes recién transcritos al nivel de la fibra de 30 nm (Fig. 2).

En resumen este conjunto de evidencias apoya la hipótesis de que el molde real de la transcripción es la fibra de 30 nm o solenoide y que la histona H1 juega un papel importante en la regulación de la transcripción de un dominio.



**Fig. 2. Niveles de empaquetamiento del ADN.** El ADN "desnudo" (fibra de 2 nm) se enrolla en 1.6 vueltas alrededor del octámero de histonas para formar un nucleosoma. Los nucleosomas se van ordenando periódicamente a lo largo de la fibra de ADN dando lugar a lo que se conoce como "collar de perlas" ó fibra de 11 nm, al participar la histona H1, los contactos entre los nucleosomas se estabilizan y se constituye el solenoide ó fibra de 30 nm. El siguiente nivel corresponde a las asas de solenoides, las cuales se forman cuando la histona H1 se fosforila, en este nivel la compactación llega hasta 300 nm, abarcando de 20 a 80 kb, después de este nivel todavía se conoce muy poco acerca de como se llega a la compactación de los siguientes niveles hasta llegar a los cromosomas, el nivel más elevado de compactación. Tomado de Felsenfeld y Groudine 2003.

## Modificaciones post-traduccionales de las histonas

Como se ha hecho evidente, la transcripción de los genes en células eucariotas no ocurre sobre la hebra de ADN desnudo, sino en el contexto de la cromatina. Esto trae como consecuencia el que modificaciones en la cromatina y en particular en las histonas tenga repercusiones directas sobre la expresión genética. Modificaciones post-traduccionales en los extremos amino-terminales de las histonas y el remodelaje de la cromatina representan un punto crítico para dar lugar a procesos como la represión o activación de la expresión génica. Las modificaciones post-traduccionales de las histonas que han sido mayormente estudiadas incluyen acetilación, fosforilación, ubiquitinación y metilación. Estas modificaciones son hechas por enzimas especializadas conocidas como acetil-transferasas de histonas (HAT's), desacetilasas de histonas (HDAC's) y metiltransferasas de histonas (HMT's), la fosforilación está dada por distintas cinasas (Struhl 1998; Rice y Allis, 2001; Izuka y Smith 2003). Las modificaciones pueden ocurrir al mismo tiempo en distintos residuos del extremo amino-terminal de las histonas, esta combinación de modificaciones da como resultado un estado transcripcionalmente activo o inactivo. Strahl y Allis propusieron que estas diferentes modificaciones representan un "código" el cuál se traduce en señales para eventos que ocurren río abajo en la cascada y que pueden dar lugar a la activación o represión transcripcional (Strahl y Allis 2000). Las modificaciones de las histonas podrían interferir con la integridad y la estabilidad de un solo nucleosoma o de un arreglo nucleosomal. Por ejemplo, se ha visto que la acetilación tiene varios efectos: alterar la estructura secundaria del extremo terminal amino de la histona, debilitar interacciones entre el ADN y los extremos amino de las histonas y reducir interacciones internucleosomales afectando la estructuración de la cromatina. Generalmente la acetilación en los residuos de lisinas de las histonas H3 y H4 correlaciona con un estado activo mientras que la desacetilación se asocia a un estado inactivo. La fosforilación de la serina 10 de la histona H3 es característica de un estado activo, esta fosforilación

facilita y es necesaria para la acetilación de la lisina 14 y la metilación de la lisina 4 (Fischle *et al* 2003).

Mientras que la metilación de la lisina 4 correlaciona con una expresión génica, la metilación de la lisina 9 es una señal para la condensación de la cromatina; de manera que las modificaciones que existen en el extremo amino-terminal de la histona H3 son: desacetilación de las lisinas 9 y 14, desfosforilación de la serina 10 y metilación de la lisina 9, siendo esta última modificación y la acetilación de la lisina 14 eventos excluyentes (Fischle *et al* 2003).

Actualmente se sabe que diferentes marcas son capaces de estar presentes al mismo tiempo en el extremo terminal amino de una histona en particular. Los estudios vanguardistas se centran en ver como estos patrones de modificaciones actúan al nivel de una histona, del nucleosoma y de dominios nucleosomales (Fischle *et al* 2003).

Estas modificaciones y combinaciones de las histonas, forman un código el cuál, dependiendo del gen, puede ser reconocido por distintas proteínas activadoras o represoras. Un ejemplo de ello es la proteína HP1 (del inglés "heterochromatin protein 1") la cuál es capaz de unirse a la histona H3 cuando se encuentra metilada en la lisina 9. HP1 es un componente principal de la heterocromatina y contribuye al mantenimiento del estado silenciado (Lachner y Jenuwein 2002).

Muchas de estas proteínas que se unen a las histonas modificadas tienen ciertos dominios particulares, por ejemplo se ha visto que proteínas con Bromodominios (Hassan *et al* 2002), son capaces de reconocer lisinas acetiladas específicamente y de unirse a ellas en el contexto de la cromatina. Adicionalmente existen los Cromodominios los cuáles se ha visto que se reconocen lisinas metiladas como es el caso de la proteína HP1 (Lachner y Jenuwein 2001). Sería interesante saber si otras proteínas que contengan estos dominios también pueden unirse a sitios de acetilación y metilación de las histonas y si reconocen combinaciones de diferentes marcas post-traduccionales en las histonas.

Un nivel adicional de complejidad sucede cuando estas modificaciones ocurren por duplicado o triplicado. El estado exacto (mono-metilado, di-metilado o tri-metilado) de una sola lisina tiene un gran impacto en procesos fisiológicos, por ejemplo, se ha determinado que la lisina 4 de la histona H3 puede ser mono, di o tri metilada, la di-metilación de esta lisina correlaciona con genes activos e inactivos, mientras que la tri-metilación está presente exclusivamente en genes activos (Santos-Rosa *et al* 2002; Jenuwein y Allis 2001).

Allis y Jenuwein han propuesto en base a los datos experimentales acumulados por un número considerable de grupos de investigación que el “código de histonas” tiene implicaciones biológicas a tres niveles: 1) a nivel local, es decir en regiones de regulación como promotores o enhancers; 2) a nivel de un dominio compuesto por un gen o grupo de genes o 3) a nivel de un cromosoma completo como por ejemplo el caso del silenciamiento del cromosoma X (Fischle *et al* 2003). Si la hipótesis del “código de histonas” es correcta, entonces descifrar cómo se traduce el código en respuestas biológicas es muy relevante, aunque representa un reto difícil metodológicamente (Jenuwein y Allis, 2001).

Muchas de las alteraciones en la estructura de la cromatina se derivan de cambios en la carga neta de los extremos de las histonas hechas por la presencia o ausencia de las modificaciones. Recientemente se ha descrito que la lisina 79 de la histona H3 es capaz de metilarse, esta lisina se encuentra en el dominio globular de la histona, en la levadura *S. cerevisiae* adyacente a esta lisina existen dos secuencias específicas llamadas “parches” que son necesarias para el silenciamiento (Park *et al* 2002). Estos parches son grupos de residuos que forman una zona con cierta carga la cuál afecta sus interacciones con el ADN. Hasta ahora no se han descrito más residuos que puedan ser modificados en el dominio globular de las histonas, sin embargo se ha identificado una cisteína altamente conservada de la histona H3, la C110, la cuál está altamente conservada entre todos los organismos y que le da la capacidad a la histona H3 de dimerizarse al formar un puente disulfuro,

además la formación de este puente ha sido relacionada con el silenciamiento transcripcional (Fischle *et al* 2003).

Actualmente los estudios acerca de cómo son controlados los eventos de modificaciones de las histonas para que sean localizados a nivel de un dominio se centran en encontrar elementos como los “insulators” o delimitadores (ver capítulo sobre sitios de hipersensibilidad y límites de los dominios) los cuáles juegan un papel predominante en la formación de dominios cromosómicos y en establecer zonas de transición de diferentes modificaciones de las histonas. Como se discutirá mas adelante, hasta la fecha se desconoce como actúan estos elementos, así que el interés principal de las investigaciones actuales está enfocado a encontrar su mecanismo de acción (Burgess-Beusse *et al.* 2002). En capítulos posteriores se hablará más a fondo de estos elementos.



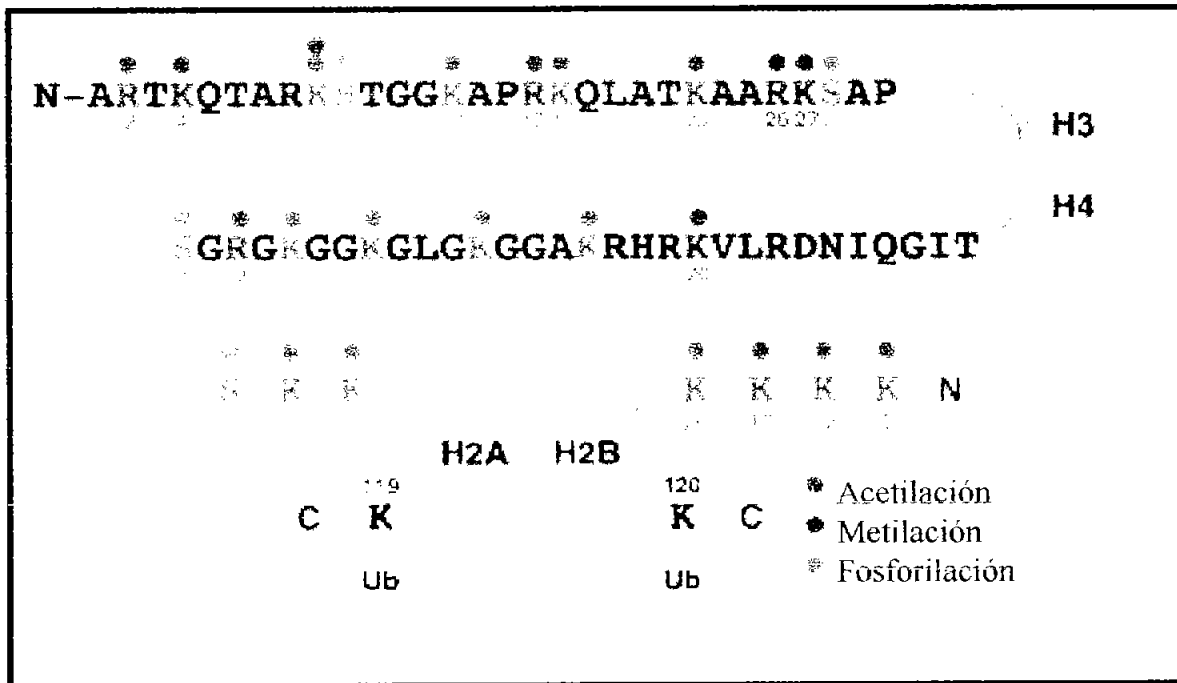


Fig. 3 Se representan las histonas H2A, H2B, H3 y H4 y sus extremos terminales amino, con los residuos y las modificaciones encontradas hasta la fecha. Tomado de Zhang y Reinberg 2001

	Estado Activo/ Accesible	Estado Inactivo/ Condensado
H3	Ac - lis 9 y 14 Me - lis 4 P - ser 10	Des - lis 9 y 14 Me - lis 9
H4	Ac - lis 5 Me - arg 3	Ac - lis 12

Tabla 1 Ejemplos de modificaciones post-traduccionales de las histonas H3 y H4. En la Tabla I se muestran algunas combinatorias de las modificaciones en ciertos residuos de los extremos amino-terminales de las histonas y el significado a nivel de la estructura de la cromatina. Ac: acetilación, Me: metilación, P: fosforilación, Lis: lisina, Ser: serina, Arg: arginina

## Variantes de Histonas

Muchas de estas variantes fueron descritas durante la década de los 70's y 80's su importancia ha sido reconocida en los últimos años, se sabe que están involucradas en distintos procesos como reparación del ADN, transcripción y silenciamiento. Muchas veces estas variantes sustituyen a las histonas en regiones determinadas del genoma, lo cuál tiene consecuencias importantes para distintos eventos, por esto es importante el hacer un resumen de las actividades de estas variantes de histonas y de algunos de los procesos en los que se han visto involucradas.

## Variantes de la histona H3

Generalmente las histonas se sintetizan *de novo* durante la fase S del ciclo celular, estas histonas se depositan durante la duplicación del ADN, sin embargo hay histonas que se sintetizan a partir de genes "huérfanos" fuera de la fase S del ciclo celular. Un ejemplo son las variantes de la histona H3 de *Drosophila*. Existen dos variantes de esta histona, la histona H3.3, la cuál se localiza principalmente en regiones de cromatina activa y Cid la cual es un componente estructural de la cromatina centromérica (McKittrick *et al.* 2004).

La histona H3.3 es casi idéntica a la histona H3, sólo cambian cuatro aminoácidos, mientras que Cid conserva únicamente el dominio de plegado de la histona H3. Una característica particular de estas variantes es que se depositan fuera de la fase S del ciclo celular (Ahmad y Henikoff 2002)

El ensamblaje de la cromatina durante la replicación o cromatina *de novo* tiene dos procesos, el primero, es la incorporación de un tetrámero de  $(H3-H4)_2$  seguido por los dímeros de H2A y H2B. Estas histonas se van incorporando a la cromatina post-duplicativa con la ayuda de la proteína CAF-1 ("chromatin assembly factor 1"), la cuál a su vez es atraída por PCNA ("proliferating cell nuclear antigen"). El hecho de que las variantes de histonas H3.3 y Cid sean depositadas en una fase diferente

a la fase S, quiere decir que debe existir otro sistema de ensamblaje de la cromatina para estas variantes de histonas. Se han hecho diversos intentos para determinar qué factores están involucrados en este proceso, poniendo particular atención en complejos remodeladores y componentes de la maquinaria basal de la transcripción, sin embargo hasta la fecha no se ha encontrado ningún factor cuya delección afecte el posicionamiento de estas variantes. Un dato importante es que la variante Cid carece del extremo amino-terminal de la histona H3. Se sabe que modificaciones específicas en el extremo amino-terminal de la histona H3 son la señal para el reclutamiento de HP1 (“heterochromatin protein 1”), y esta proteína no está asociada a la cromatina centromérica donde se encuentra Cid. Este dato ha llevado a proponer que una conformación “abierta” de la cromatina es la base común para el posicionamiento nucleosomal independiente de la duplicación. En *Drosophila* el centrómero está constituido por secuencias repetidas que están formando heterocromatina, esto forma una compartimentalización que puede observarse en células en interfase, la cuál correlaciona con la presencia de Cid únicamente en regiones centroméricas, por lo tanto parece que la presencia de esas secuencias repetidas está restringiendo el posicionamiento de Cid al centrómero (Blower y Karpen 2001; Smith 2002).

### **Variantes de la histona H2A**

Las variantes de la histona H2A fueron identificadas en los años 80's. La variante H2AX difiere en su extremo carboxilo-terminal, tiene un motivo de 4 serinas llamado SQ, invariable entre los miembros de esta familia, y su posición con respecto al carboxilo también es invariable. En mamíferos, *Xenopus*, *Drosophila* y *Saccharomyces*, el dominio SQ puede ser fosforilado en respuesta a radiación ionizante o a otros agentes que introducen cortes de doble cadena en el ADN. La fosforilación de estos residuos puede ser visualizada como agregados llamados *foci* en las células, el número de estos agregados de H2AX fosforilada es directamente proporcional al número de los cortes de doble cadena

producidos (Bassing *et al* 2003). El motivo SQ puede ser fosforilado por diferentes cinasas como la ATM (“ataxia telangiectasia mutated”) durante la radiación ionizante, y ATR durante estrés metabólico y otros agentes como luz ultravioleta o hidroximetiluracil. También se presentan estos *foci* de H2AX fosforilada durante los cortes de doble cadena programados, un ejemplo de esto ocurre durante el rearrreglo de los genes de las inmunoglobulinas. También es un ejemplo de especificidad, ya que son capaces de reconocer un solo corte de doble cadena en el genoma por ejemplo durante la recombinación del locus V(D)J (Reina-San-Martín *et al.* 2003). En base a estos datos, es evidente que esta variante tiene una función importante durante la reparación del ADN.

Otra variante de la histona H2A es la histona H2AZ, las regiones de cromatina que contienen a esta variante tienen una menor estabilidad (medida por propiedades bioquímicas como por ejemplo extracciones a diferentes concentraciones de sal), esto puede traducirse en la pérdida de los dímeros H2AZ-H2B, permitiendo una mayor accesibilidad de la maquinaria transcripcional al ADN (Ausio y Abbott 2002). Sin embargo, ésta variante juega un papel importante en el silenciamiento en *S. cerevisiae*. La delección de H2AZ tiene un efecto parcial de desrepresión en el locus HMR, pero un efecto mayor en la pérdida del silenciamiento telomérico. El sobre-expresar H2AZ puede complementar a una mutante en Sir (familia de proteínas importante para el silenciamiento telomérico) que puede ser reclutada pero no silenciar el locus HMR (Redon *et al.* 2002).

Recientemente el grupo de Carl Wu ha demostrado que un complejo remodelador de la cromatina es responsable del intercambio de la histona H2A por la variante de histona H2AZ de manera ATP dependiente (Mizuguchi *et al.* 2004).

La variante Macro H2A es 64% idéntica a H2A, y el 57% restante no está relacionada a ninguna histona conocida, contiene una posible cremallera de leucinas y es similar a proteínas involucradas en la replicación de ARN viral. Esta variante está asociada al cromosoma X inactivo de las hembras de mamíferos, pero aún se desconoce su función (Redon *et al.* 2002).

## Variantes de la histona H1

Estas histonas tienen una considerable heterogeneidad funcional, el número de variantes y el grado de divergencia es mayor que el de las histonas del core nucleosomal. En el ratón existen al menos ocho variantes de la histona H1, algunas tienen un papel potencial en la regulación de la expresión génica durante la oogénesis y embriogénesis temprana como H1t y H1oo. Además otras están involucradas en ciclo celular (Conconi y Wellinger, 2003).

En conclusión es evidente que las variantes de histonas participan y regulan diversos procesos biológicos, y que la presencia de estas variantes en ciertas regiones es importante para una correcta estructuración de la cromatina en la célula. Si bien se han encontrado involucradas en la reparación del ADN y como componentes estructurales de cromatina especializada, como lo es la cromatina centromérica, aún se desconoce si existe la presencia de estas histonas en los dominios cromosómicos y si tienen un papel predominante en la regulación de la transcripción genética. Es importante el hecho de que muchas de estas variantes difieran en sus extremos amino-terminales, ya que esto indica que no son sujetas a las mismas modificaciones post-traduccionales que las demás histonas. Sin embargo en algunas de estas variantes existen motivos de aminoácidos repetidos que de acuerdo a la carga que presentan hacen que la estructura nucleosomal sea más estable, estos dominios son conocidos como "parches de carga" y se ha propuesto que estos representan otro nivel de regulación de la estructura de la cromatina a nivel de las proteínas (Fischle *et al.* 2003). La incorporación de variantes de histonas como una estrategia para delimitar y definir un dominio cromosómico resulta ser una idea muy atractiva que merece tomarse en consideración (McKittrick *et al.* 2004).

A continuación se analizará la presencia de regiones en la cromatina, las cuales son sitios accesibles para la unión de elementos que participan en la transcripción y en su regulación.

## Sitios de Hipersensibilidad

En 1976 Weintraub y Groudine se preguntaron si la estructura de la cromatina en genes activos sería la misma que en genes inactivos. Para contestar a su pregunta, aislaron núcleos de células eritroides de pollo que expresan los genes de las globinas y núcleos de células que no expresan estos genes y los incubaron con concentraciones crecientes de la nucleasa pancreática DNasa I, la cual reconoce la doble hebra del ADN y corta inespecíficamente en zonas libres de nucleosomas (Weintraub y Groudine 1976). Posteriormente mediante una técnica de hibridación indirecta, la cual actualmente ha sido reemplazada por el "Southern blot", determinaron la magnitud del corte por parte de la Dnasa I. Observaron que en el ADN genómico de las células que expresaban los genes de las globinas, la cromatina del locus tenía un mayor grado de digestión, mientras que el locus globina de las células que no expresaban los genes era resistente al corte. Esto hizo suponer que la estructura de la cromatina se alteraba cuando un gen es transcripcionalmente activo. Actualmente se sabe que la sensibilidad al corte por la DNasa I correlaciona con un relajamiento en la estructura de la cromatina del locus de genes activos, aunque el locus continúe ensamblado en nucleosomas. En 1979 Wu y Elgin realizaron digestiones con la DNasa I en cromatina de núcleos de células de *Drosophila* y encontraron que a bajas concentraciones de la enzima había en el Southern blot bandas más pequeñas que los fragmentos originales. A estas bandas se les denominó sitios de hipersensibilidad al corte por la DNasa I. Se propuso que estos sitios correspondían a regiones con una estructura alterada de la cromatina, donde se observa un mayor grado de digestión de la doble cadena por parte de la DNasa I y que podrían coincidir con elementos reguladores de la transcripción (Weintraub y Groudine 1976; Stalder *et al* 1980). Actualmente se sabe que estos sitios de hipersensibilidad son típicamente, al menos dos órdenes de magnitud más sensibles al corte por nucleasas que el resto de la cromatina y no deben confundirse con la sensibilidad generalizada al corte por nucleasas que presentan los genes que están siendo

transcritos y que puede extenderse por varias kilobases (Boyes y Felsenfeld 1996). Un sitio de hipersensibilidad típico contiene sitios de unión para una variedad de diferentes factores de transcripción y muy probablemente factores y co-factores que atraen actividades de remodelaje de la cromatina. La accesibilidad a nucleasas aumenta en estos sitios por la ausencia de nucleosomas, una organización nucleosomal más relajada, el ensamblaje parcial de éstos o su desplazamiento, y por la unión de un factor de transcripción que distorsiona a la cromatina favoreciendo su "apertura" en esa región o en regiones adyacentes (Boyes y Felsenfeld 1996). Se ha propuesto la posibilidad de que los factores de transcripción y los nucleosomas compitan por la unión en un lugar dado del genoma (como por ejemplo un promotor o un enhancer), e induzcan la formación y mantenimiento de un sitio de hipersensibilidad. El ejemplo más claro que apoya este modelo se basa en que mutaciones en el sitio de unión al factor GATA-1 en el ADN correlacionan con una disminución cuantitativa en la formación del sitio de hipersensibilidad (Boyes y Felsenfeld 1996). Por lo tanto se ha propuesto que debe existir un proceso dinámico de interacción proteína-ADN, con elementos que remodelan a la cromatina y la formación de un sitio de hipersensibilidad (Boyes y Felsenfeld 1996). Una constante para los sitios de hipersensibilidad a la DNasa I es que coinciden, invariablemente, con elementos estructurales o de regulación, como por ejemplo, regiones delimitadoras de un dominio o "insulators", enhancers, regiones de control del locus (LCR's como veremos más adelante) silenciadores y promotores.

Los sitios de hipersensibilidad pueden clasificarse en 3 diferentes tipos:

1) **Los sitios específicos.** Pueden aparecer antes de la transcripción y generalmente persisten incluso al término de la misma, son tejido específicos y a este tipo corresponden los promotores, enhancers y silenciadores (Blackwood y Kadonaga 1998; Kamakaka 1997). Los promotores son secuencias de ADN que se encuentran generalmente de - 40 a + 40 nucleótidos relativos al inicio de la transcripción, conocida también como región "core", a estos se unen la maquinaria basal de transcripción que

contiene a la RNA Polimerasa II, factores generales de transcripción y co-activadores. Los promotores controlan la direccionalidad y el inicio de la transcripción y responden a activadores o represores transcripcionales. Además del "core", existe una región proximal que va de - 50 a - 200 que algunos han llamado la región proximal del promotor. Esta región cuando es clonada en múltiplos cerca de un promotor puede conferirle una actividad de enhancer. Por su parte los enhancers ó aumentadores son secuencias en *cis* que se encuentran a una distancia relativa del inicio de la transcripción, tienen una organización modular, es decir, que son sitios de unión para más de un factor y son capaces de aumentar la transcripción de una manera independiente de su orientación, es decir, son funcionales tanto en el costado 3' como 5' de un gen (Blackwood y Kadonaga 1998). Los módulos de los enhancers desempeñan funciones específicas al activar la transcripción de un gen durante una etapa del desarrollo determinada, controlando los niveles de expresión del gen (Blackwood y Kadonaga 1998). Igual que los enhancers, los silenciadores tienen una organización modular, representada por los diferentes sitios de unión a las proteínas, pero a diferencia de éstos, regulan negativamente la transcripción (Kamakaka 1997).

2) **Las regiones de control de un dominio o "Locus Control Regions (LCR's)".** Otro tipo de sitio de hipersensibilidad son los comúnmente llamados LCR's, los cuales están conformados por varios sitios de hipersensibilidad, y contienen múltiples secuencias de unión para activadores transcripcionales. La diferencia entre los LCR's y los enhancers es que los primeros actúan en una etapa previa a la transcripción de los genes haciéndolos competentes para ser transcritos ya que presentan una actividad tipo enhancer y además, se ha propuesto que contribuyen a la apertura de la cromatina por diferentes mecanismos (Bulger y Groudine, 1999). Otra característica de los LCR's que los diferencia de los enhancers, es que estos están usualmente más alejados de los genes que controlan, en ocasiones hasta 16 kb (Blackwood y Kadonaga 1998; Tabla II). Más adelante se discutirá sobre el papel de los LCR's en el control de la transcripción dentro de un dominio.



3) **Los sitios de hipersensibilidad constitutivos.** Estos sitios no son tejido específicos, es decir que en un mismo locus pueden estar presentes en diferentes tipos celulares y su presencia es independiente de la expresión genética. Estos sitios se han encontrado en algunos casos enmarcando a un gen o grupo de genes y se ha propuesto que corresponden a los límites físicos y funcionales de los dominios transcripcionalmente activos (Burgess-Beusse *et al.* 2002; Labrador *et al.* 2002). Estos límites son regiones en el ADN que funcionan como secuencias neutras, es decir que no presentan actividad silenciadora, promotora o de potenciador y por lo tanto no afectan la transcripción positiva o negativamente. Estas regiones delimitadoras bloquean o aíslan la influencia de elementos de regulación de la transcripción que se encuentran fuera del locus al cual están flanqueando y se ha propuesto que contribuyen al mantenimiento de un locus activo. A estas secuencias se les ha llamado "insulators" o delimitadores. A continuación describiremos en mayor detalle a los LCR's.

#### **Regiones de control del locus ó LCR's**

Los LCR's son un conjunto de sitios de hipersensibilidad al corte por la DNasa I, que corresponden a elementos reguladores tejido específicos en el ADN y que fueron descritos primeramente en el locus  $\beta$ -globina de humano (Li *et al.* 2002; Fig 4). El estudio de la regulación de estos genes es importante ya que la falta de expresión de éstos causa una enfermedad conocida como talasemias que son una clase de anemias muy severas. Al estudiar el locus  $\beta$ -globina de individuos afectados por esta enfermedad, se identificaron una serie de deleciones río arriba del primer gen embrionario  $\epsilon$  del dominio, las cuales disminuían dramáticamente la expresión de todos los genes del locus, aunque éstos estuviesen intactos, es decir, sin mutaciones (Tuan *et al.* 1985). Este descubrimiento suponía por lo tanto, la presencia de elementos de regulación en la región 5' del locus. Al analizar la estructura de la cromatina de esta

los problemas de silenciamiento de transgenes y de los vectores diseñados para terapia génica (Chow *et al* 2002; Recillas-Targa *et al* 2004). Pronto los investigadores concluyeron que, en efecto, los LCR's son útiles para este fin, pero su rango de acción es restringido ya que su función es tejido específica. Se han descrito LCRs en el locus de  $\beta$ -globina de otros vertebrados con diferente número de sitios de hipersensibilidad (Martin *et al* 1996). A partir del descubrimiento del LCR  $\beta$ -globina se han descrito LCR's en otros loci como el de TCR $\alpha$  de células T, en el locus CD2 en humanos y en el locus de la lisozima de pollo entre otros (ver Tabla II).

LOCUS O GENE	ORGANISMO
Gene CD2 Gene de la adenosina desaminasa Genes de pigmentos visuales Gene S100 $\beta$ Gene de la keratina 18 Gene del MHC clase I HLA G Gene CD4 Locus $\alpha/\delta$ del receptor de células T Gene de la apolipoproteína B Locus E/C-I de la apolipoproteína Gene de la cadena pesada de la $\beta$ miosina Genes C4A y B del complemento Gene de la hormona del crecimiento Locus de la cadena pesada Ig Genes 1 y 2 del IgG $\alpha$ Gene HLA-B7 del MHC clase I Gene de la desmina	Humano
Gene CD34 Gene de la glicoporina Locus $\lambda$ 5-V <sub>preB1</sub> Cadena pesada de la Inmunoglobulina $\mu$ Genes I y II de la metalotioneína	Ratón

Gene Ea del MHC clase II  
Gene de la tirosinasa  
locus  $\beta$ -globina  $\alpha$ -fetoproteína

Genes Kallikrein  
Gene de la proteína acídica de Whey  
Gene de la aldolasa C  
Gene LAP (C/EBP  $\beta$ )

Rata

Gene de la lisozima  
Locus  $\beta$ -globina

Pollo

Gene de la  $\beta$ -lactoglobulina

Oveja

**Tabla II. "Locus Control Regions" en distintos loci genómicos.** LCR's conocidos hasta la fecha y su localización (Tomado de Valadez-Graham y Recillas-Targa, 2000).

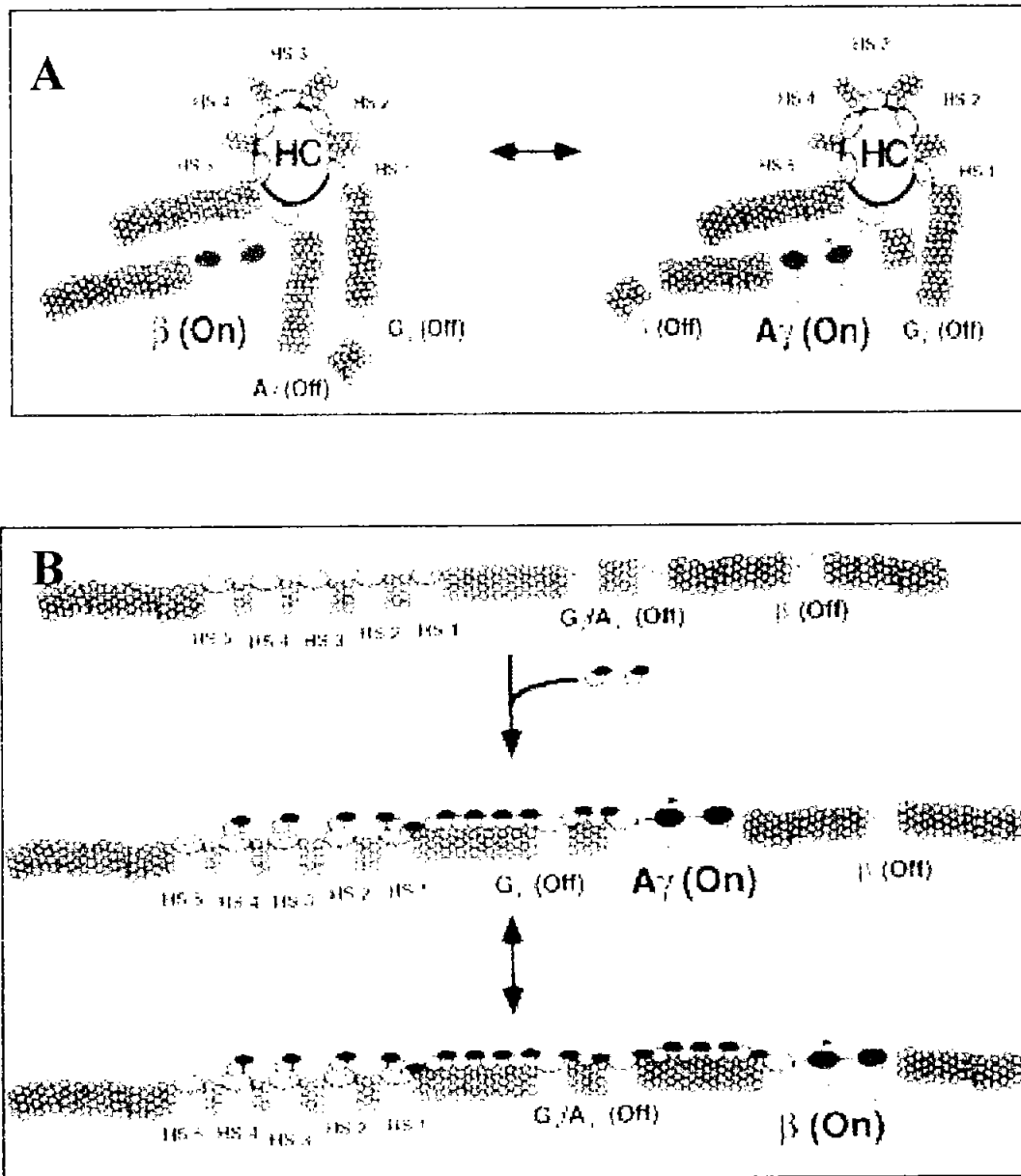
Los LCR's se definen funcionalmente por las siguientes características: 1) son capaces de conferir altos niveles de expresión a un transgen independientemente del sitio de integración y dependiente del número de copias integradas del transgen (Martin *et al* 1996; Kioussis y Festenstein 1997; Fraser y Grosveld 1998). Esto significa que en ausencia de un LCR, la expresión de un transgen integrado en un contexto cromatínico, estaría sujeta a los efectos intrínsecos de la estructura de la cromatina, causados por los distintos sitios de integración (efecto de posición). Por ejemplo, si el transgen se integra en o cerca de una región de heterocromatina (que tiene una estructura altamente compactada y por lo tanto represora) como podría ser un centrómero o un telómero, su expresión decaería a lo largo del tiempo o nunca sería expresado (Festenstein *et al* 1996); 2) regulan la expresión genética durante el desarrollo, y 3) contribuyen a la apertura de la cromatina, para permitir la transcripción de los genes del locus (Bungert *et al* 1999; Tanimoto *et al* 1999; Jackson *et al* 1996).

A pesar del gran número de investigaciones realizadas en laboratorios de todo el mundo, en la actualidad aún se sabe poco acerca de los mecanismos por los cuales los LCRs influyen en la estructura de la cromatina y la expresión de los genes contenidos en un dominio. Sin embargo, se han propuesto a la fecha, cuatro modelos principales para explicar la función de los LCRs:

- 1) **Formación de asas cromatínicas ó "looping"**. En este modelo, el LCR actuaría como un holocomplejo para estimular la transcripción de genes individuales dentro del locus y a lo largo del desarrollo. El LCR se acercaría al promotor del gen a activar formando un asa o "loop", reclutando toda la maquinaria de transcripción (Fig. 5A). Se ha propuesto un posible contacto entre el promotor y el LCR mediada por interacciones entre distintas proteínas reguladoras y que son tejido y tiempo específicas; (Bulger y Groudine 1999; Engel y Tanimoto 2000). Como se verá más adelante evidencias experimentales recientes favorecen este modelo.
- 2) **Acción a través de facilitadores ó modelo de "linking"**. Sugiere la cooperación de varios factores en *trans*, los cuales son expresados en un estadio de diferenciación específico y que serían reclutados por el LCR (Fig. 5B). Este modelo propone la formación de una cadena o relevo, mediada por factores facilitadores, que contribuiría a fomentar una estructura de la cromatina favorable acercando elementos de control para la transcripción de un solo gene durante una etapa determinada (Dorsett 1999; Bulger y Groudine 1999).
- 3) **Encarrilamiento o "tracking"**. Otro modelo que apoya la hipótesis de que el LCR contribuye a la apertura de la cromatina, es el llamado "tracking" el cual propone un reclutamiento del complejo de transcripción y de diversos factores en el LCR y su posterior encarrilamiento en la cromatina, desplazándose a lo largo de ésta con la supuesta ayuda de complejos de remodelaje hasta llegar a los promotores de los genes que serán expresados. Este modelo no implica que la polimerasa forzosamente vaya transcribiendo, más bien representa una forma de reclutamiento de la

maquinaria de remodelaje de la cromatina y como consecuencia el acceso de la maquinaria transcripcional (Tang *et al* 2002).

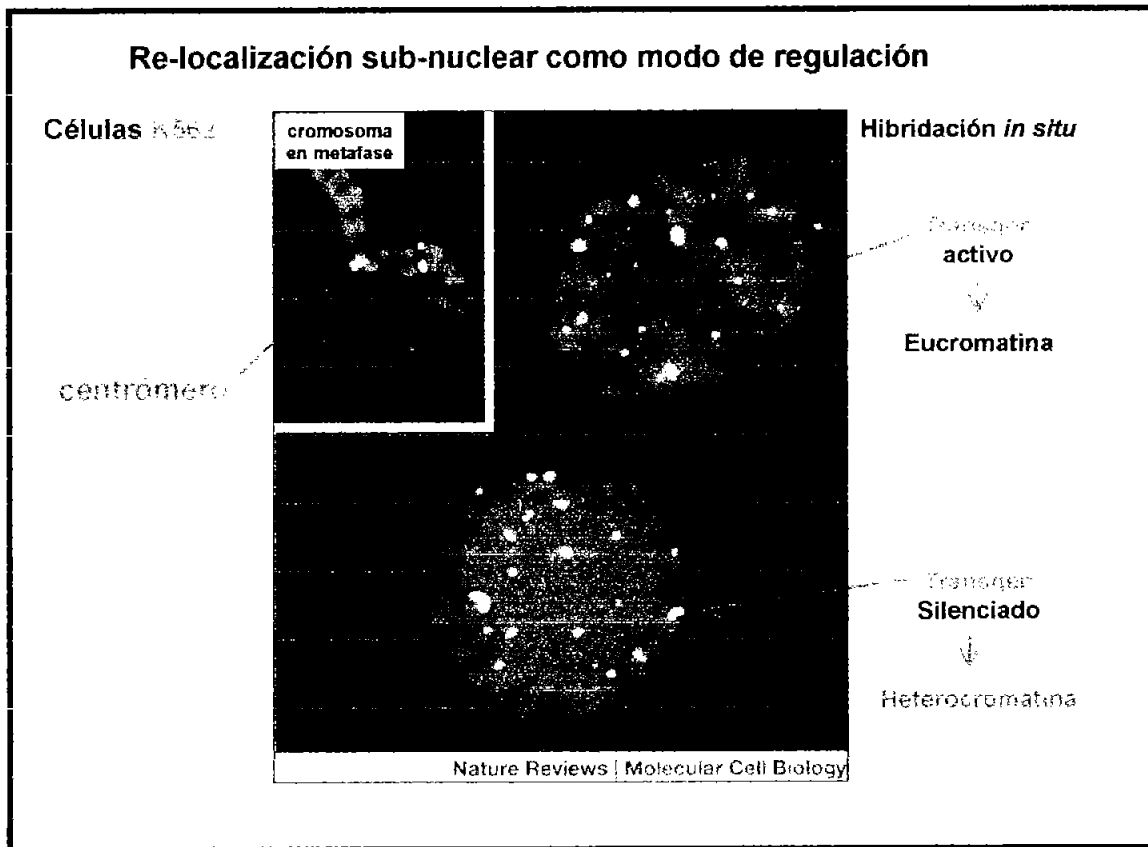
- 4) **Transcritos intergénicos:** Desde los años 60's Scherrer y colaboradores propusieron la presencia de transcritos gigantes policistrónicos abarcando regiones génicas e intergénicas cuya función sería la de contribuir a la formación de asas cromatínicas y resguardar la integridad de los RNA's pre-mensajeros y su procesamiento (Scherrer 2003). Recientemente se ha confirmado la existencia de transcritos intergénicos, es decir, en regiones no-codificantes en el dominio  $\beta$ -globina de humano. Se desconoce la función precisa de estos transcritos intergénicos pero se ha demostrado una correlación directa entre su presencia y la apertura de la cromatina en el contexto del dominio. Esta apertura de la cromatina mediada por estos transcritos intergénicos favorece la expresión regulada de los genes de manera individual y a lo largo del desarrollo (Bulger y Groudine 1999; Ashe *et al.* 1997; Gribnau *et al.* 2000 ).



**Fig. 5 Modelos propuestos para el funcionamiento del LCR.** A) "Looping": en este modelo, el LCR interactúa directamente con el promotor a través de contactos mediados por el holocomplejo e interacciones con factores que se unen de manera específica a los promotores individuales. B) "Linking": este modelo propone el uso de proteínas que actúen en *trans*, ó proteínas de relevo facilitadoras que interactúan directamente con otras proteínas del promotor, activándolo  
 Tomado de Bulger y Groudine 1999.

Dos grupos de investigación independientes descubrieron otra alternativa analizando la posibilidad de que los enhancer's y LCR's contribuyan a la regulación del dominio al relocalizarlo lejos de regiones de heterocromatina al interior del núcleo. En el primer trabajo de investigación se tomó a uno de los sitios de hipersensibilidad, específicamente el 5'HS2 del locus  $\beta$ -globina humano, el cual anteriormente había sido caracterizado como un enhancer y en el que se identificaron varios sitios de unión a factores transcripcionales como GATA-1, NF-E2, AP1 y EKLF, y se realizaron mutaciones en sitios específicos para la unión a estos factores, aboliendo su unión. La meta de este estudio era tratar de contestar dos preguntas: primero, si la pérdida de la unión de factores abolía la actividad del enhancer y segundo, si estas mutaciones afectaban la localización subnuclear de un transgen.

Se valieron de una técnica de recombinación homóloga para integrar a un transgen cuya expresión era dirigida por el 5'HS2 completo o con las diversas mutaciones en los sitios de unión a factores, actuando en una región cercana al centrómero. Visualizaron al transgen en todos los casos mediante la técnica de FISH (hibridación *in situ* fluorescente) y comprobaron su citolocalización; cuando el enhancer se encontraba intacto, de una manera estadísticamente significativa, demostraron que el transgen era relocalizado lejos del centrómero y se expresaba adecuadamente, lo que no pasaba en presencia del enhancer mutado el cuál se mantuvo asociado a regiones centroméricas. Concluyeron que se requiere un enhancer funcional tanto para suprimir el silenciamiento genético como para relocalizar a un gene lejos de la heterocromatina centromérica durante la interfase. Y que la supresión del silenciamiento requiere de la relocalización del transgen lejos de regiones de heterocromatina (Francastel *et al.* 1999; Fig. 6).



**Fig. 6 Relocalización sub-nuclear.** Se observa que cuando se encuentra activo el transgen co-localiza con regiones de euromatina, parte superior derecha (en verde). Cuando el transgen es silenciado, su localización dentro del núcleo cambia, co-localizando ahora con zonas de heterocromatina (en rojo). Tomado de Fancastel *et al* 2000.

Basados en estos resultados se propuso un modelo en el cual los factores que se unen al enhancer, posiblemente actúan reclutando y translocando al transgen a compartimentos nucleares o territorios donde existen factores y elementos basales de la transcripción en grandes concentraciones. A su vez no se descarta la posibilidad de que alguno de los factores del enhancer reclute a complejos de acetilación, metilación o de remodelaje de la cromatina completando el ciclo de control que lleve a la

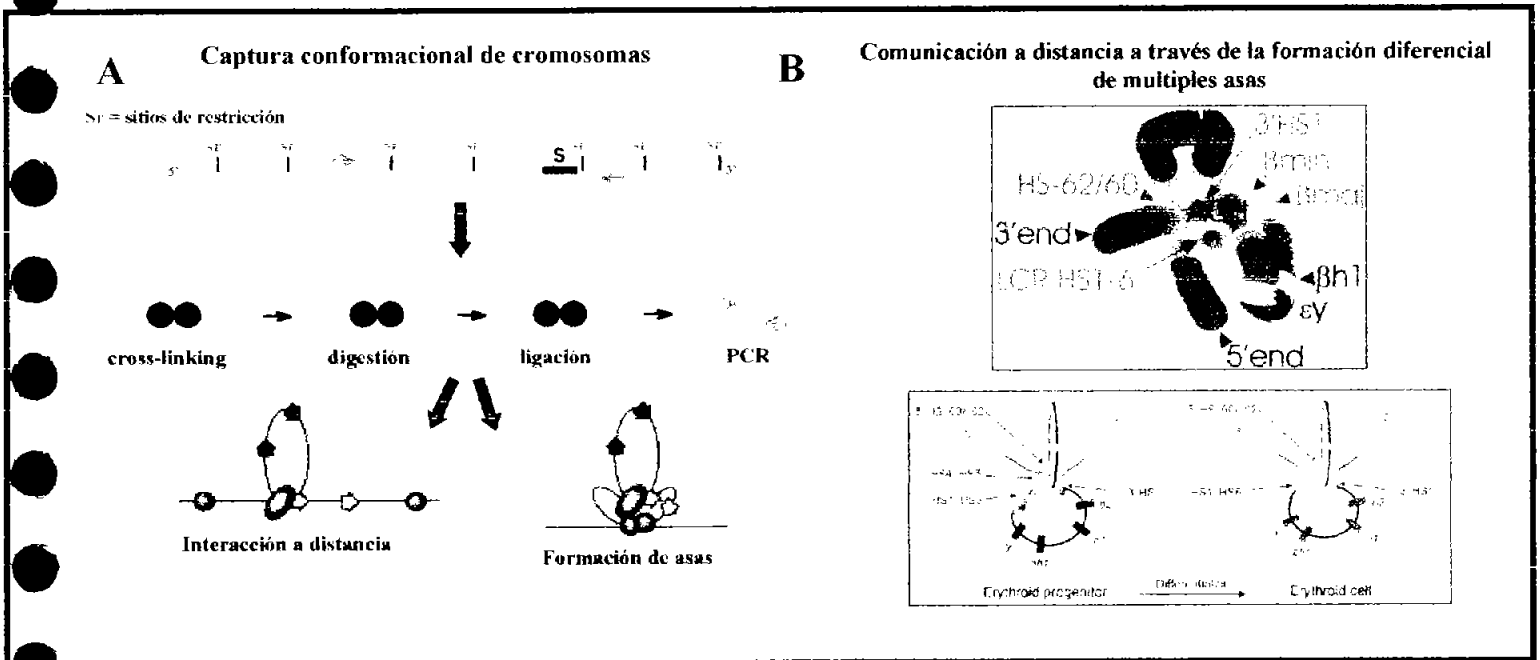


expresión regulada de un gen (Ragoczy *et al* 2003). Cabe señalar que se desconocen los mecanismos por los cuáles ocurre este tipo de relocalización al interior del núcleo.

En el segundo trabajo se analizó la posibilidad de que el LCR también estuviese involucrado en la re-localización sub-nuclear. El grupo de Groudine se valió de líneas celulares que tenían por un lado al LCR completo del locus  $\beta$ -globina de humano (con los 6 sitios de hipersensibilidad) y por otro la deleción presente en la talasemia hispánica  $\Delta$ HS1-HS5, y una línea celular donde ellos generaron una deleción más pequeña que abarcaba  $\Delta$ HS2-5. Por medio de FISH analizaron si la ausencia del LCR afectaba a la localización del locus y encontraron que el locus silvestre y la deleción  $\Delta$ HS2-5, se localizaban lejos del centrómero y tenían a las histonas H3 y H4 hiperacetiladas a lo largo del locus (Francastel *et al* 2000; Fig 6). Estos datos indicaban que en realidad el LCR no era necesario para la localización del locus lejos de la heterocromatina centromérica, ya que la línea celular con la deleción parcial del LCR ( $\Delta$ HS2-5), también se localizaba lejos del centrómero y además la sensibilidad al corte por la DNasa I se mantenía. Por otra parte, la hiperacetilación de las histonas H3 y H4 es característica de un locus  $\beta$ -globina que está abierto (Hebbes *et al* 1994; Litt *et al.* 2001). Anteriormente, se había predicho que ésta era una característica que dependía del LCR, pero con los datos anteriores se demostró que el LCR no es necesario para la acetilación de histonas a nivel del dominio. Por lo que infirieron que para la localización del locus lejos de la heterocromatina centromérica mediaba una estructura más abierta y una hiperacetilación generalizadas del locus y que estas actividades eran independientes del LCR. Por otro lado, la acetilación de H3 cerca del LCR en el locus silvestre, que es transcripcionalmente activo, es mucho mayor cuando el LCR se encontraba intacto. Estos datos sugirieron que el LCR no actúa directamente en mantener una estructura abierta de la cromatina, ni ayuda a alejar al locus de entornos heterocromáticos y por lo tanto represores, más sin embargo podría actuar a otro nivel para regular la expresión del locus.

Esto crea una controversia en cuanto a las funciones propuestas para el LCR, ya que en estudios realizados por el grupo Proudfoot (1997) y de manera independiente por el equipo de Fraser (2000) y colaboradores, demostraron la presencia de transcritos intergénicos los cuales son dependientes de la integridad del LCR. Estos dos grupos han propuesto que los transcritos intergénicos contribuyen a la apertura de la cromatina, ya que están presentes a lo largo del locus, son estadio específicos y sólo se encuentran en células donde el locus es expresado, es decir células eritroides (Schübeler *et al* 2000). Una de las explicaciones de los resultados que indican que el LCR no es necesario para la apertura del dominio, es que la regulación de la expresión del locus se da en varias etapas y posiblemente niveles, y que la combinación de los diferentes factores juega un papel importante para la regulación del locus.

Muy recientemente y gracias al desarrollo de dos nuevas estrategias experimentales: la captura de la conformación cromosomal ó 3C, y el RNA-TRAP (del inglés "Tagging and Recovery of Associated Proteins") han surgido nuevas evidencias que apoyan el modelo de "looping" como el modelo de acción de los LCR's (Dekker *et al.* 2002; Carter *et al* 2002). La técnica de 3C consiste en hacer digestiones de cromatina previamente tratada con formaldehído, el tratamiento con este químico hace que se formen puentes ("crosslinking" ) entre las proteínas y la cromatina en la célula, permitiendo que se conserven sus interacciones. Al digerir la cromatina, en particular en la base de las asas y posteriormente hacer una ligación intramolecular, las moléculas de ADN que están cerca o adyacentes (que de manera natural se encuentran alejadas) podrán ser ligadas y así es posible evidenciar mediante la técnica de reacción en cadena de la polimerasa (PCR) secuencias de ADN que estén cerca en ese momento (Fig. 7A).



**Fig. 7 Captura conformacional de cromosomas (3C).** **A)** Protocolo experimental. En la parte superior ejemplo del dominio  $\alpha$ -globina de pollo. S, silenciador; E, enhancer; sr, sitio de restricción. Las flechas c indican la localización (distante) de secuencias que corresponden a "primers" que serán utilizados para un amplificación por PCR una vez realizado el experimento de entrecruzamiento y ligación *in situ*. La predicción teórica sugiere la formación de una asa o de múltiples asas acercando elementos de regulación, en este caso al LCR, a los genes a ser activados (parte inferior). **B)** Para el dominio  $\beta$ -globina humano y mediante el uso de esta estrategia experimental, se ha demostrado la formación de múltiples asas formando el "centro de cromatina activa" o ACH y que su composición puede variar dependiendo del estadio de diferenciación.

De esta forma los laboratorios de Grosveld y de Laat pudieron confirmar que existen interacciones diferenciales entre los sitios de hipersensibilidad del locus  $\beta$ -globina de ratón, los promotores de los genes del dominio y el sitio de hipersensibilidad 3'HS1 del mismo locus, y que estas interacciones ocurren únicamente en células donde el locus  $\beta$ -globina de ratón se encuentra transcripcionalmente activo (Tolhuis *et al.* 2002; Palstra *et al.* 2003). Además se llegó a la conclusión de que el LCR actúa como un holocomplejo en células eritroides al activar los genes globina, esto quiere decir que la mayoría de los sitios de hipersensibilidad están involucrados en la regulación de estos genes cuando están activos, dejando fuera del holocomplejo a genes que en un momento de la diferenciación eritroide son inactivos, esto lo hace mediante la formación de múltiples asas generando lo que han llamado el "centro de cromatina activa" o ACH (del inglés "Active Chromatin Hubb"; Fig. 7B). Otro dato interesante es que al parecer no todos los sitios de hipersensibilidad están involucrados en la activación de la transcripción, tres de ellos 3'HS1 y 5'HS 60-62 juegan un papel estructural, un aspecto muy interesante es que todos estos sitios de hipersensibilidad que conforman al ACH unen al factor llamado CTCF. Este factor es esencial para la actividad de bloqueo de enhancer de los "insulators" y se hablará de él más adelante. El grupo de Victor Corces ha encontrado otra proteína llamada Su(Hw) (supresor de *Hairy Wing*) en *Drosophila*, la cual es indispensable para la actividad de "insulator" del elemento *gypsy* y la cuál tiene 12 dedos de zinc. Se ha demostrado que esta proteína tiene un papel estructural muy importante ya que es capaz de re-localizar zonas donde se encuentran genes que deben ser transcritos y de definir dominios (Byrd y Corces 2003).

En base a estos datos se ha propuesto que CTCF quizás pueda actuar como Su(Hw) teniendo además un papel estructural. Esta predicción no parece alejarse de la realidad dado que Felsenfeld en una comunicación personal nos ha comentado que CTCF interacciona con la proteína nucleofosmina la cuál atrae hacia la región perinucleolar las secuencias que unen a CTCF. Felsenfeld propone que la

función de CTCF es muy similar a la de gypsy en cuanto a su participación en la posible formación de asas o de una topología favorable para la formación regulada de un dominio.

Mediante la técnica de "RNA-TRAP" se pudo determinar que el enhancer HS2 del LCR  $\beta$ -globina de ratón, es capaz de interactuar con el gen globina HBB que se encuentra activo transcripcionalmente *in vivo*, este gen se localiza a 50 kb del 5'HS2. La técnica consiste en una modificación del RNA-FISH (del inglés "Fluorescence *in situ* hybridization"); en la cuál se realiza una hibridación *in situ*, pero el oligonucleótido está acoplado a digoxigenina, de manera que puede ser reconocido por un anticuerpo acoplado a peroxidasa, el cuál posteriormente es revelado mediante una reacción química al añadir biotina al tejido, posteriormente se analiza el resultado con microscopía de fluorescencia y con una inmunoprecipitación de cromatina, de esta manera se obtiene un valor estadístico. Esta técnica es utilizada actualmente para definir interacciones entre elementos de control y los genes, el mismo grupo está desarrollando la misma técnica pero utilizando sondas para ADN-FISH, esto permitiría diferenciar entre las interacciones que son acopladas a la transcripción y a las que no lo son.

En conclusión los LCR's están formados por sitios de hipersensibilidad que actúan sobre cada gen que necesite ser activado en un momento específico del desarrollo, y son los factores transcripcionales que se unen a cada sitio de hipersensibilidad los que contribuyen a la formación de compartimentos donde es reclutada activamente la polimerasa II (pol II) y que contribuyen a la transcripción, o bien que llevan a los sitios de hipersensibilidad a actuar como elementos estructurales importantes para la formación de un holocomplejo. Finalmente resulta indudable que los LCR's deben ser considerados no sólo como elementos de regulación sino también su función es esencial en la formación de un dominio transcripcionalmente activo.

A continuación analizaremos las características de los sitios de hipersensibilidad constitutivos, los que en la mayoría de los casos corresponden a secuencias delimitadoras de los dominios también conocidos como “insulators”.

### **Límites de los dominios**

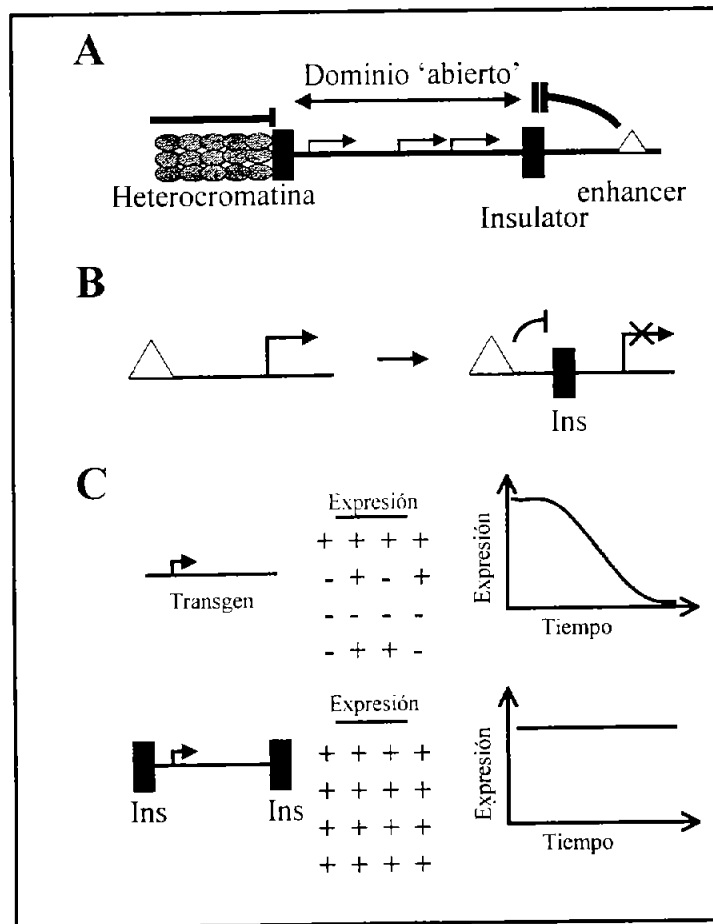
A diferencia de los LCR's los “insulators” o al menos algunos de ellos y en particular el cHS4 del dominio  $\beta$ -globina de pollo (Fig. 15), participan en el establecimiento de una región de acetilación de histonas, que contribuye a la apertura de la cromatina, a la escala de un dominio (Litt *et al.* 2001).

Los genes activos e inactivos en el genoma, se localizan en loci independientes que tienen diferentes programas de expresión durante el desarrollo celular. Si tomamos en cuenta que los sitios de hipersensibilidad descritos anteriormente correlacionan con secuencias reguladoras de la expresión genética y que estos elementos actúan sobre promotores que se encuentran localizados a grandes distancias, es necesario que existan elementos que aseguren que estas regiones de control actúen sólo sobre sus promotores y no sobre promotores de otros loci vecinos.

Cuando los genes de un dominio necesitan ser expresados, el dominio necesita adoptar una estructura de la cromatina más abierta, para esto es necesario que existan elementos que les indiquen el momento preciso para “abrirse” y que los ayuden a mantenerse “abiertos” para que los eventos finos que regulan la expresión de los genes puedan llevarse a cabo posteriormente.

Los “insulators” son secuencias que en la mayoría de los casos correlacionan con la presencia de sitios de hipersensibilidad constitutivos en el límite de algunos dominios y se definen como

secuencias neutras que protegen a un dominio de los efectos de la cromatina adyacente, previniendo la “promiscuidad” entre elementos de regulación de otros dominios y que contribuyen a formar y mantener el locus “abierto” cuando está siendo transcrito (West *et al.* 2002; Burgess-Beusse *et al.* 2002; Razin *et al.* 2003; Fig. 8A).



**Fig. 8. Actividades Funcionales de los “insulators”.** A) Los “insulators” son capaces de mantener un dominio abierto protegiéndolo de la invasión de la heterocromatina y de la promiscuidad con elementos de regulación de dominios adyacentes. B) Actividad de bloqueo de enhancer. un enhancer actúa sobre su promotor y lo activa, al colocar el “insulator” entre el enhancer y el promotor, se bloquea el efecto del enhancer y por lo tanto el promotor no es activado. C) Protección contra el efecto de posición. Al integrar un transgen al genoma, su expresión depende del sitio de integración y generalmente decae a lo largo del tiempo, al flanquear el transgen con los “insulators”, se suprime el efecto de posición, la expresión del transgen es homogénea y se mantiene a lo largo del tiempo.

Los "insulators" fueron descritos originalmente en los límites de los genes divergentes *heat-shock* (*Hsp70*), en el locus 87A7 de *Drosophila*. Kellum y Schedl analizaron dos secuencias que correlacionaban con sitios de hipersensibilidad constitutivos, estos elementos fueron nombrados *scs* y *scs'* (del inglés: "*specialized chromatin structures*") (Kellum y Schedl 1992). Para determinar las propiedades de "insulators" se hicieron construcciones que llevaban al gen *white* controlado por un promotor mínimo, y utilizando estas construcciones generaron moscas transgénicas con un patrón de color de ojos muy variable, esto como consecuencia del efecto de posición causado por el medio ambiente de la cromatina en los distintos sitios de integración. Al flanquear este gene con elementos *scs*, el color de los ojos de las moscas fué homogéneo; es decir, se suprimió el efecto de posición (Kellum y Schedl 1991). Así se demostró la primera característica funcional de los "insulators": que tiene que ver con la capacidad de estas secuencias para proteger a un transgen del entorno cromatínico en el cual se integre (Fig 8C). Por otra parte, recientemente se han identificado proteínas que se unen a *scs* y a *scs'*. A *scs* se une una proteína codificada por el gene *Zw5* y a *scs'* se unen dos proteínas: BEAF-32A y BEAF-32B (Hart *et al* 1997; Gaszner *et al* 1999). Se ha demostrado que estas proteínas son indispensables para la función de bloqueo de enhancer y apoyan el concepto de dominios ya que se han inmunolocalizado en los límites de los "puffs" de los cromosomas politénicos de las glándulas salivales de las larvas de *Drosophila* (Fig. 1).

Otro de los "insulators" mejor estudiados es *gypsy*, un retrotransposón de *Drosophila*. Esta secuencia es capaz de bloquear la acción de un enhancer sobre su promotor solamente cuando se localizaba entre el enhancer y el promotor (Cai y Levine 1997; Shen *et al* 1994; Fig. 8B). El retrotransposón *gypsy* es responsable de una gran variedad de mutaciones que afectan genes a lo largo del genoma de *Drosophila*. El mejor ejemplo está dado por el gen *yellow*, el cual es un gen no esencial y su producto está involucrado en la pigmentación de la cutícula de distintos órganos en la mosca de la



fruta (Golovnin *et al.* 2003). La expresión de este gen a lo largo del desarrollo es controlado por una serie de enhancers tejido específicos, los cuales se encargan de regular la expresión de *yellow* diferencialmente durante el desarrollo y pigmentar distintas estructuras de la mosca. Dichos enhancers controlan la expresión de *yellow* en la cutícula, alas y estructuras larvales pigmentadas, están localizados en la región 5' relativa al sitio de inicio de la transcripción del gen. La inserción de *gypsy* en la región 5' de *yellow* entre el enhancer que controla la expresión del gen en la cutícula y el que controla las estructuras larvales, bloquea la interacción del primero con el promotor del gen, pero no afecta al enhancer larval que se localiza más próximo al promotor, es decir, entre el transposón y el gen (Gdula *et al.* 1996). Un análisis más detallado de *gypsy* ha permitido identificar la región dentro del retrotransposón responsable de la actividad de bloqueo de enhancer. Esta región contiene una secuencia repetida de 8 nucleótidos que se encuentra flanqueada por regiones ricas en adeninas (A) y timinas (T). Cuando esta secuencia es insertada en cualquier lugar del extremo 5' del gen *yellow*, los enhancers que están río arriba de esta secuencia no tienen efecto alguno sobre el promotor, en cambio los que se encuentran río abajo del octámero repetido, siguen siendo funcionales. Además, al colocar estas secuencias flanqueando al transgen *white*, su expresión es más homogénea es decir, le confieren protección el efecto de posición (Gdula *et al.* 1996). Estudios genéticos demostraron que el fenotipo de las mutaciones causadas por *gypsy*, puede ser revertido al mutar al gen "*suppressor of hairy wing*" (*Su(Hw)*). La proteína del mismo nombre se caracterizó como un componente del "insulator" *gypsy*. La proteína *Su(Hw)* tiene 12 dedos de zinc, interacciona con las secuencias repetidas dentro del retrotransposón *gypsy* y cualquier mutación en el sitio de unión de esta proteína y que como consecuencia impida su unión al "insulator", hace que la actividad de bloqueo de enhancer se pierda. Posteriormente se identificó otro componente peptídico llamado "*modifier of mdg4*" (*mod(mdg4)*), esta proteína carece de un dominio de unión al ADN, es incapaz de unirse al "insulator" *gypsy in vitro*, y se ha encontrado que colocaliza con la proteína *Su(Hw)* y demostrado que interacciona con esta proteína

directamente. En ausencia de mod(mdg4) el "insulator" pierde su unidireccionalidad y adquiere propiedades de silenciador, inactivando a los enhancers del gen *yellow* tanto río abajo cómo río arriba del sitio donde se insertó (Shen *et al* 1994; Cai y Levine 1997). En resumen, *gypsy* tiene dos componentes principales identificados hasta ahora: Su(Hw) y mod(mdg4), sin los cuales sus funciones de "insulator" se ven alteradas (Mongelard *et al.* 2002). Los casos de *gypsy* y de *scs* y *scs'* no son los únicos en que se asocian proteínas a un "insulator", como veremos más adelante también en los "insulators" de vertebrados, se ha demostrado la asociación de la proteína llamada CTCF (Bell *et al* 1999).

ELEMENTO	LOCUS	ORGANISMO	BLOQUEO DE ENHANCER	PROTECCIÓN CONTRA EFECTO DE POSICIÓN	PROTECCIÓN CONTRA EL SILENCIAMIENTO	PROTEÍNA ASOCIADA	REFERENCIA
HMR-R	HMR	Levadura	ND	ND	+	SMC, SMC3	(Donze et al 1999)
HMR-L	HMR	Levadura	ND	ND	+	ND	(Donze et al 1999)
UAS	TEF2	Levadura	+	ND	+	Rap1	(De Sanctis et al 2002)
STARs	Telómeros	Levadura	ND	ND	+	Tbf1, Reb1p	(Fourel et al. 1999)
Fab-7	Abdominal-B	<i>Drosophila</i>	+	ND	+	ND	(Mihaly et al. 1997)
Fab-8	Abdominal-B	<i>Drosophila</i>	+	ND	+	ND	(Barges et al. 2000)
Mcp	Abdominal-B	<i>Drosophila</i>	ND	ND	+	ND	(Karch F 1994)
scs	Gen hsp 70	<i>Drosophila</i>	+	+	-	Zw5	(Kellum R 1992)
scs'	Gen hsp 70	<i>Drosophila</i>	+	+	ND	BEAF-32	(Dunaway et al 1997)
gypsy	Transposon	<i>Drosophila</i>	+	+	+	Su(Hw), Mod (mdg4)	(Gerasimova TI 1998)
eve	Gen even-skipped	<i>Drosophila</i>	+	ND	ND	GAGA	(Ohtsuki and Levine 1998)
5'HS4	$\beta$ -globina	Pollo	+	+	-	CTCF	(Bell AC 1999)
sns	Gen de histonas	Erizo de mar	+	ND	ND	GAGA	(Melfi et al. 2000)
HS2-6	TCR $\alpha/\delta$	Humano	+	ND	ND	ND	(Zhong and Krangel 1999)
BEAD-1	TCR $\alpha/\delta$	Humano	+	ND	-	CTCF	(Modin, Pedersen, and Duch 2000)
3'HS	$\beta$ -globina	Pollo	+	-	ND	CTCF	(Saitoh et al. 2000)
DMD	Igf2/H19	Humano	+	-	ND	CTCF	(Kanduri C 2000)
Ars	Aril-sulfatasa	Erizo de mar	+	+	ND	SpGCF1	(Akasaka K 1999)
AAVS1	AAVS1	Adenovirus	+	+	+	ND	(Ogata, Kozuka, and Kanda 2003)
SF1	Scr-ftz	<i>Drosophila</i>	+	+	ND	GAGA	(Belozarov et al 2003)
$\alpha$ EHS-1.2	$\alpha$ -globina	<i>Drosophila</i>	+	?	?	CTCF	(Valadez-Graham et al 2004)
$\alpha$ EHS-1.4	$\alpha$ -globina	<i>Drosophila</i>	+	?	?	CTCF	(Valadez-Graham et. Al 2004)

Tabla III. "Insulators" conocidos a la fecha. ND: no detectada, +: si tiene esa actividad, -: carece de esa actividad.

El primer "insulator" descubierto en vertebrados es el que delimita al dominio  $\beta$ -globina de pollo en su extremo 5' (5'cHS4) (Chung *et al.* 1993; Fig. 15). En este locus se identificó un sitio de hipersensibilidad constitutivo (5'cHS4). Río arriba de este sitio de hipersensibilidad en el locus de pollo, existe una zona de cromatina altamente compactada, resistente a la digestión por DNasa I (Prioleau *et al.* 1999; Ghirlando *et al.* 2004). Se pensó que el "insulator" 5'cHS4 pudiera definir el límite del dominio  $\beta$ -globina de pollo, ya que hacia el extremo 3' de este sitio de hipersensibilidad, se identificaron otros sitios de hipersensibilidad eritroide específicos que en conjunto forman al LCR, adicionalmente el 5'cHS4 sólo permite la apertura de la cromatina hacia 3' (que es donde se encuentran los genes del locus) y no hacia la región 5' (donde está la estructura de la cromatina altamente compactada) (Chung *et al.* 1993; Prioleau *et al.* 1999; Fig 15). Posteriormente se analizó este elemento y se descubrió que presenta las características funcionales de los "insulators", es decir protege a un transgen del efecto de posición y es capaz de bloquear la acción de un "enhancer" sobre su promotor (Chung *et al.* 1996; Pikaart *et al.* 1998). Investigaciones recientes han confirmado el concepto de dominio a partir del descubrimiento de un segundo "insulator" en el costado 3' del mismo dominio, lo cual demuestra que este dominio se encuentra enmarcado a ambos lados por "insulators" (Saitoh *et al.* 2000; Fig. 15). La actividad funcional del "insulator" 5' del gene  $\beta$ -globina pudo reducirse de un fragmento de ADN de 1.2 kb, que contenía el sitio de hipersensibilidad, a un fragmento de ADN de sólo 250 pb al que se le llamó "core" (Chung *et al.* 1999), dentro de las 250 pb existe un fragmento de 42 pb que es necesario y suficiente para la función de bloqueo de "enhancer" (Bell *et al.* 1999; Recillas-Targa *et al.* 1999). En 1999 Felsenfeld y colaboradores purificaron e identificaron una proteína que se

unía a este sitio y se descubrió que esta proteína era CTCF ("CCCTC-binding factor"). Para caracterizar la función de CTCF se hicieron mutaciones en el sitio de unión de esta proteína al "insulator" y se descubrió que una menor afinidad de unión de este factor disminuía la capacidad del "insulator" de bloquear la actividad de un enhancer sobre su promotor *in vivo* (Bell *et al.* 1999). La proteína CTCF es altamente conservada en vertebrados, es una proteína de 82 kDa compuesta por 11 dedos de zinc y está involucrada en la activación y el silenciamiento transcripcional cuando está presente en promotores (Klenova *et al.* 2002; Ohlsson *et al.* 2001). Se ha sugerido que la actividad de bloqueo de enhancer mediada por CTCF es un componente funcional de los límites de los dominios de los vertebrados; ya que podría estar controlando señales de elementos de regulación de dominios adyacentes y manteniendo abierto al dominio cuando los genes deben expresarse. La proteína CTCF es una proteína multifuncional y muy relevante para la actividad de los "insulators" de vertebrados, más adelante se describirá su estructura y diferentes funciones (ver capítulo sobre CTCF en esta tesis).

Una de las evidencias más contundentes de la relevancia de los "insulators" y la participación de CTCF en la regulación génica, es el reciente descubrimiento de una actividad de bloqueo de enhancer metilo sensible en el dominio improntado *Igf2/H19* en ratón y humano (Hark *et al.* 2000; Bell y Felsenfeld 2000). En este dominio de más de 100 kb, el gen *Igf2* se localiza a 80-90 kb del gen *H19* (Fig. 9). Ambos genes comparten dos enhancers localizados en el extremo 3' de los genes *Igf2* y *H19* respectivamente (Fig. 9). Evidencias experimentales demostraron la presencia de una región localizada río arriba del gen *H19* (es decir, entre los genes *Igf2* y *H19*) con la capacidad de ser metilada de manera diferencial dependiendo del tipo de alelo (Tilghman 1999). A esta región se le ha denominado región de metilación diferencial (RMD), la cual se encuentra metilada exclusivamente en el alelo paterno. Trabajos realizados por el grupo de Tilghman sugirieron originalmente que la expresión alelo específica de estos genes está dada por un mecanismo de competencia entre los enhancers distales y los promotores particulares a cada gen. Recientemente un cúmulo de estudios experimentales encabezados

por los grupos de Tilghman y Felsenfeld demostraron una actividad tipo “insulator” en la región RMD (Kanduri *et al.* 2000; Hark *et al.* 2000; Bell y Felsenfeld 2000; Recillas-Targa 2002). Esta actividad resulta ser regulada por la interacción del factor CTCF, cuya función es, en este caso, dependiente de la metilación. Para el alelo paterno, donde sólo se expresa *Igf2*, la región RMD se encuentra altamente metilada lo cual impide la unión de CTCF y por lo tanto no se presenta la actividad de bloqueo de enhancer, permitiendo la expresión de *Igf2*. Por el contrario, en el alelo materno, donde sólo se expresa *H19*, RMD no está metilado, lo cual permite la unión de CTCF. Esta unión contribuye a una actividad de bloqueo de enhancer la cual evita que los enhancers activen a *Igf2* y se concentren en la regulación y expresión de *H19*, únicamente en el alelo materno (Kanduri *et al.* 2000; Hark *et al.* 2000; Bell y Felsenfeld 2000). De esta forma se explica la especificidad en la expresión monoalélica de cada gen del locus *Igf2/H19* (Hark *et al.* 2000; Bell y Felsenfeld 2000). Datos experimentales recientes utilizando la novedosa estrategia de RNA de interferencia dirigida contra CTCF, demostraron que CTCF es esencial para la metilación diferencial de la región RMD y que su ausencia ocasiona la hipermetilación del ADN en ambos alelos (Fedoriw *et al.* 2004). Por lo tanto se propone que CTCF protege contra la metilación del ADN en el alelo materno (Fedoriw *et al.* 2004).

En resumen todas estas evidencias apoyan el concepto de dominios y la contribución de los “insulators” en la estructura y regulación de los mismos que en este caso es dependiente de la metilación diferencial del ADN en la región RMD.

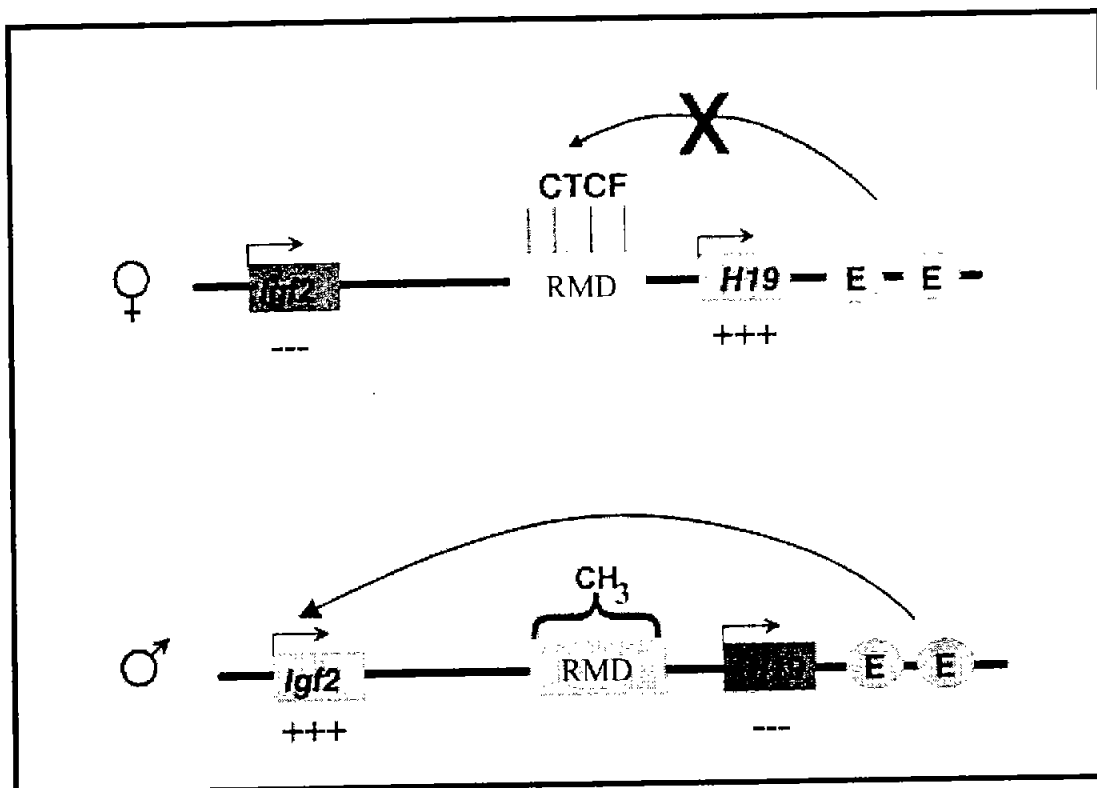


Fig. 9 El locus improntado *Igf2/H19*. En el alelo materno *Igf2* está silenciado, pero en el alelo paterno se expresa. La región RMD tiene 4 sitios de unión a la proteína CTCF, cuando está metilada (en el alelo paterno), se impide la unión del factor CTCF y se permite la activación del gen *Igf2* por los enhancers endodérmicos.

## Modelos de acción de los "insulators"

Hasta la fecha no ha sido posible determinar los mecanismos exactos de las funciones de los "insulators" y los modelos propuestos tienen una estrecha relación con la acción de los enhancers. Existen tres modelos propuestos sobre cómo el "insulator" lleva a cabo la función de bloqueo de enhancer: 1) en el primero llamado "looping" se propone que el "insulator", interrumpe los contactos protéicos entre el promotor y el enhancer (Fig. 10A), 2) en el segundo modelo llamado "tracking" se propone que al colocar al "insulator" entre el enhancer y el promotor, el "insulator" bloquea el paso del complejo de transcripción impidiendo de esta manera la activación del promotor (Fig. 10A); 3) otro modelo propuesto es el llamado de bloqueo estérico, en el cual, al colocar al "insulator" entre el enhancer y el promotor, los factores protéicos que se unen al "insulator", impedirían la unión de los factores del enhancer, pero hasta la fecha esto no ha sido posible comprobar de manera clara alguno de estos modelos (Fig. 10A). Anteriormente Geyer propuso otro modelo llamado "decoy" en el cual se propuso que el "insulator" titulaba los factores necesarios para la actividad de los enhancers reclutándolos o secuestrándolos y de esta manera se inactivaba la función del enhancer (Parnell y Geyer 2000; Fig. 10A). Sin embargo, actualmente se tienen varias evidencias experimentales que van en contra de este modelo (Cai y Levine 1995; Recillas-Targa, Bell y Felsenfeld 1999).

Experimentos realizados en transfecciones transitorias con plásmidos circulares y lineales mostraron que los "insulators" no alteran en ninguna forma a los enhancers y que su actividad de bloqueo es dependiente de su posición y requieren de proteínas específicas para realizarla (Recillas-Targa *et al* 1999). Aunque estos resultados se inclinan mayoritariamente a que el enhancer realiza su función por medio del "tracking", ya que el "insulator" bloquearía esta función de manera estérica, a la fecha no existe un estudio que favorezca a uno de los modelos en particular.

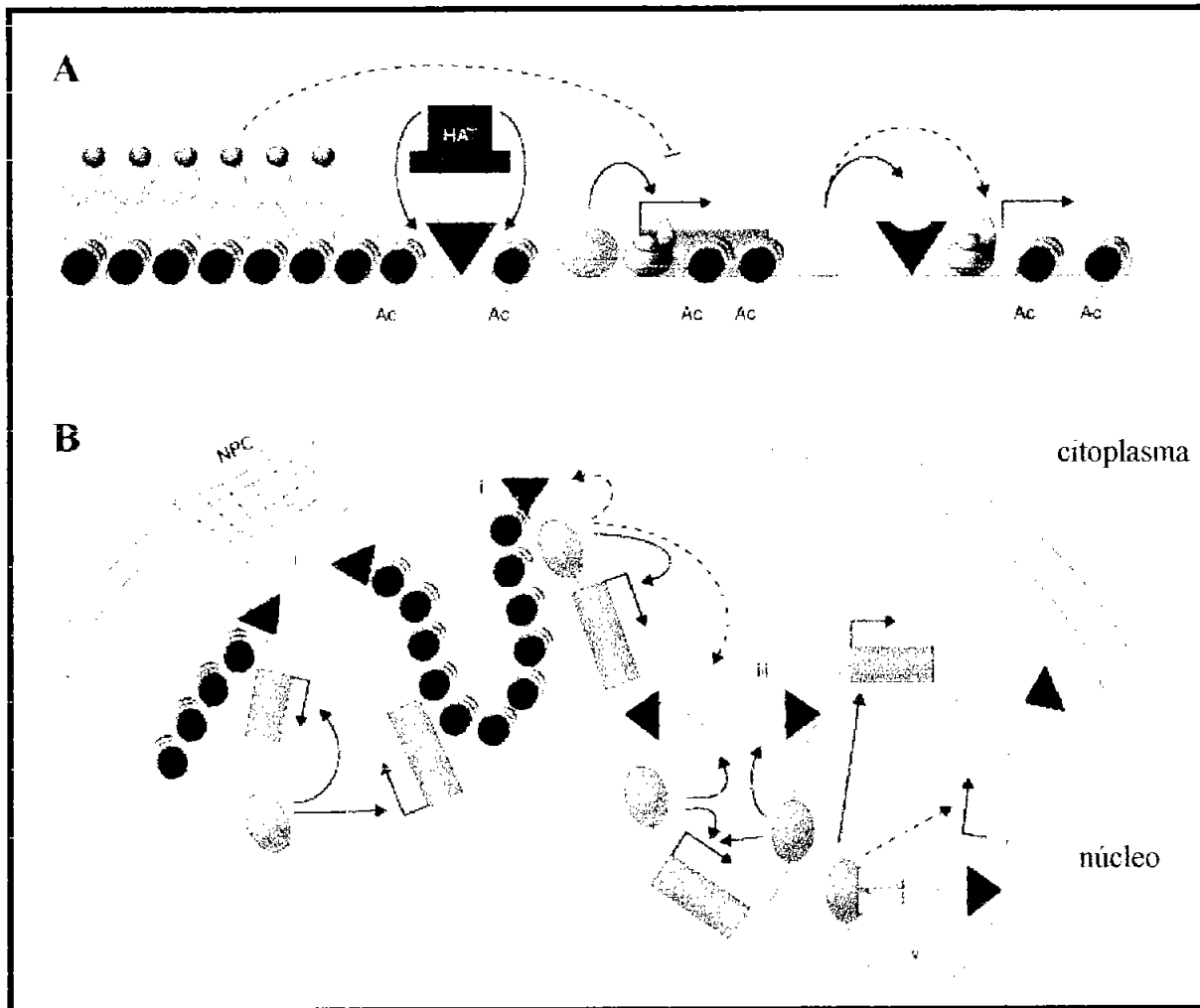
La segunda característica funcional de los "insulators" es la de proteger a un transgen del efecto de posición cuando se encuentran a ambos lados de éste. En 1998 Felsenfeld y colaboradores



demonstraron que el "insulator" 5'cHS4 del locus  $\beta$ -globina de pollo es capaz de proteger a un transgen en células en cultivo del efecto de posición y de la desacetilación y la metilación del locus que son sinónimo de represión (Pikaart *et al.* 1998). Basados en estos datos y en los de los "insulators" de *Drosophila*, se propusieron, entre otros, dos modelos sobre como el "insulator" podría funcionar para proteger a un dominio del efecto de posición. En el primero, sugieren que el "insulator" funciona como una barrera para el acceso de los complejos de metilación y desacetilación, quizás secuestrando desacetilasas de histonas e impidiendo que actúen sobre el dominio, en el segundo modelo, podrían reclutar activamente acetilasas de histonas (Fig. 10A) (Pikaart *et al.* 1998; Razin *et al.* 2003; Recillas-Targa 2002). Anteriormente mencionamos que al integrar un transgen en el genoma, su expresión está sujeta al entorno cromatínico en el cual se integra. Si el transgen se integra en una región de heterocromatina, la expresión del gen es silenciada, ya que la heterocromatina "invade" al transgen, creando una estructura de la cromatina represora la cual impide su transcripción (Fig 10B). Recientemente en levadura se implementó una técnica llamada "boundary-trap", mediante esta técnica se demostró que existe un contacto físico entre componentes del poro nuclear y el locus HML y que esto bloquea la represión, formando un dominio activo entre un región silenciadora; este estudio demuestra que la actividad de barrera puede estar ligada a contactos con la periferia nuclear (Ishii *et al.* 2002).

De este concepto se desprende otra actividad en la cuál los "insulators", que pueden actuar como una barrera física que impida la "invasión" de regiones de heterocromatina al transgen. Salta a la vista que dadas las propiedades de los "insulators" nos encontramos ante un novedoso tipo de elementos de regulación que pueden contribuir a evitar la variabilidad y extinción en la expresión de genes con fines terapéuticos (West *et al.* 2002).

En conclusión, los "insulators" son elementos que ayudan a organizar el genoma en dominios cromosomales al servir como límites entre genes o loci que se expresan diferencialmente durante el desarrollo, protegiéndolos de los efectos de la cromatina adyacente (ya sea heterocromatina o elementos de regulación en loci adyacentes), ayudando a mantener los dominios "abiertos" para que puedan ser transcritos y quizás tengan funciones todavía no descubiertas como por ejemplo regulando transcritos intergénicos.



**Fig. 10 Modelos propuestos para el funcionamiento de los "insulators".** A) Modelos para la actividad de bloqueo de enhancer. Un "insulator" (triángulos negros) podrían reclutar complejos protéicos (óvalos amarillos); estos complejos podrían reclutar actividades de HAT's con la capacidad de acetilar histonas para prevenir la propagación de cromatina represora (en gris). La acetilación de las histonas interfiere con el establecimiento de una marca que asocia complejos silenciadores. En la derecha el modelo de bloqueo de la actividad del enhancer y el modelo de "decoy" donde el "insulator" "atrapa" las señales del enhancer (óvalo verde). B) Relación entre los "insulators" y los dominios topológicos. Se muestran posibles interacciones de los "insulators" al formar las asas o "loops". Dominios independientes podrían ser definidos a través de las interacciones entre los "insulators" y componentes del poro nuclear (i); la lámina nuclear (ii) o mediante interacciones con complejos protéicos (verde y amarillo) (iii). En cada caso, la formación de dominios restringe interacciones regulatorias y la propagación del silenciamiento a un solo dominio. Algunos "insulators" podrían bloquear la actividad de los enhancers al interactuar con factores transcripcionales (iv). Tomado de (Kuhn y Geyer 2003).

## Las Secuencias de Unión a la Matriz Nuclear o MAR's

Otros elementos que delimitan los dominios, son las llamadas secuencias de unión a la matriz nuclear ("Matrix Attachment Regions o Scaffold Attachment Regions, MAR's o SAR's"), las cuales son elementos en *cis* que están involucrados en la organización de los dominios estructurales asociados con sitios de hipersensibilidad, que se encuentran enmarcando regiones que presentan una sensibilidad generalizada al corte por la DNasa I y que tienen la capacidad de unión a la matriz nuclear. A diferencia de los "insulators", no todas las MAR's confieren protección contra el efecto de posición, ni tienen una expresión dependiente del número de copias y la gran mayoría no tiene función de bloqueo de enhancer (Bode *et al* 2000; Allen *et al* 2000). Estas secuencias son ricas en A's y T's, no tienen una secuencia consenso y se identificaron como sitios de unión a la matriz nuclear, el ADN de estas regiones tiene un surco menor más angosto, es propensa a desenrollarse y se dobla con mayor facilidad. Estas secuencias generalmente son reconocidas por la histona H1 y por la topoisomerasa II (Inoue *et al.* 1999). Estas dos proteínas contribuyen a compactar la fibra de cromatina en "loops" o bien forman heterocromatina, por el contrario, otras proteínas como MATH-20 y HMG-IFI, se unen cooperativamente a estas regiones y pueden desplazar a las proteínas que inducen la compactación y de esta manera crear accesibilidad. Tres excepciones a esta regla son las encontradas en la MAR del locus de lisozima de pollo, llamadas elementos A, las cuales se encuentran en 5' y 3' delimitando el locus (von Kries *et al* 1990). Estas MAR's son capaces de proteger a un transgen contra el efecto de posición y tienen actividad de bloqueo de enhancer (Stief *et al.* 1989). Otro ejemplo clásico es el de las MAR's 5' y 3' del gene humano de la apolipoproteína B (apo B), con sólo la actividad de protección contra el efecto de posición, sin embargo no son capaces de bloquear a un enhancer (Kalos y Fournier 1995). Existen otros elementos MAR en el gen  $\mu$  de la inmunoglobulina, éstos fueron encontrados flanqueando el enhancer y el promotor del gen. Posteriormente se estudió el papel que jugaban estos

elementos en la expresión del gene  $\mu$  de la inmunoglobulina. Se generaron ratones transgénicos los cuales llevaban construcciones del gene  $\mu$  con o sin las secuencias MAR's, además, analizaron el efecto de la metilación del ADN previo a la transfección (en células en cultivo) en la actividad del gen  $\mu$  con o sin las MAR's. Encontraron que el enhancer  $\mu$  es capaz de activar a un promotor distal en transfecciones estables en células en cultivo, pero no en ratones transgénicos, sin embargo, al estar en combinación con las MAR's, el enhancer fué capaz de activar al gen también en los ratones transgénicos. Además el enhancer en combinación con una MAR indujeron accesibilidad en una región cromatínica. Las secuencias MAR's fueron capaces además, de localizar al transgen a la matriz nuclear y finalmente obtuvieron un dato relevante e inédito en el cual demostraron que las secuencias MAR's podían revertir la metilación del ADN, y gracias a esto se evitó la represión génica. Sin embargo, las MAR's por sí solas son incapaces de ejercer estas funciones por lo que se propuso que la acción de las MAR's era dependiente del enhancer (Namciu *et al* 1998; Nabirochkin *et al* 1998; Hart 1998)

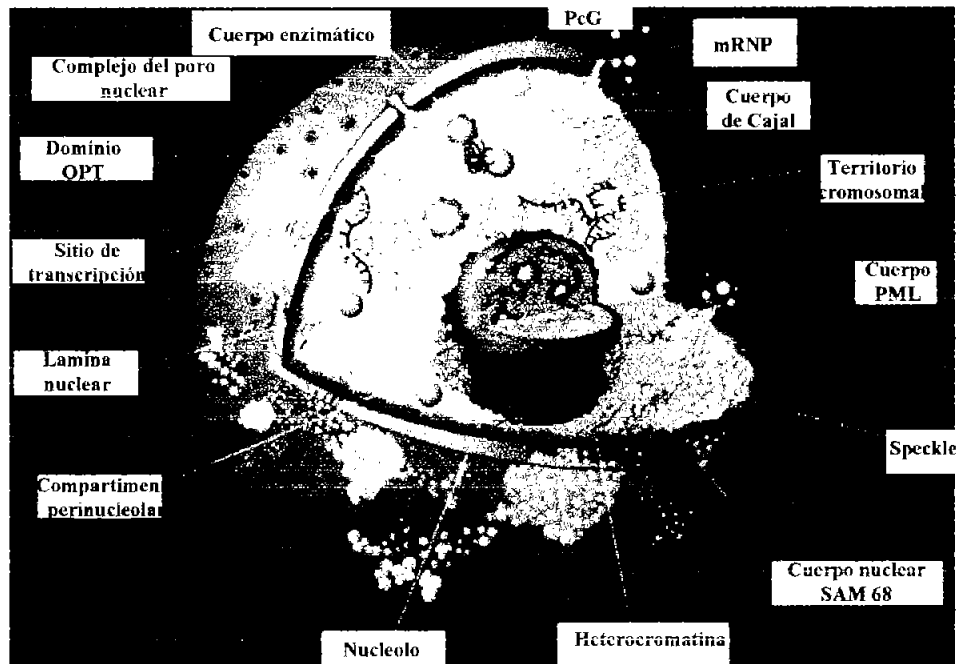
Además de las MAR's, se ha demostrado que en el genoma se presentan naturalmente secuencias repetidas ricas en A-T, las cuales presentan algunas propiedades que pudieran estar involucradas en la definición de dominios y a las cuales resulta difícil atribuirles un actividad funcional (Moreau *et al* 1982; Forrester *et al* 1999). Por otra parte, también existen secuencias repetidas como la familia CR1 que ha sido localizada en los límites de algunos dominios como en los extremos 5' y 3' del locus  $\beta$ -globina de pollo, y se ha propuesto que las repetidas de A's y T's junto con las CR1 y las secuencias Alu pudieran formar una estructura diferente de cromatina que contribuyera a delimitar el dominio (Kondo e Issa 2003; Haas 1997).

Al analizar las propiedades de las secuencias delimitadoras como "insulators" o las secuencias MAR's, se hace evidente la necesidad de seguir estudiando estas secuencias ya que forman parte fundamental de la organización de los genes en dominios transcripcionalmente activos sin descartar su

participación en la organización topológica de dominios génicos. A continuación hablaremos sobre las interacciones y la cooperación de estos límites para que los dominios puedan ser regulados.

### **Territorios Nucleares**

Hace algún tiempo se pensaba que el núcleo era una estructura donde la transcripción, la replicación y todas las actividades de procesamiento del ADN, se llevaban a cabo en cualquier lugar de éste. Actualmente se entiende que esto no puede ser cierto ya que se sabe que el núcleo tiene una estructura tridimensional y que esta estructura se adapta a distintas circunstancias, como por ejemplo en respuesta a diferentes tipos de estrés o durante las distintas etapas del ciclo celular para facilitar las funciones celulares, nucleares y de segregación o incluso se ve afectada en células tumorales (Fig. 11).



**Fig. 11 Territorios nucleares.** Se muestra la representación de un núcleo en interfase, y sus componentes. PcG, miembros de la familia Polycomb de proteínas; PML, cuerpo promielocítico; OPT dominio transcripcional donde son abundantes las proteínas Oct-1 y PTF; "speckles" sitios donde se encuentran los factores de "splicing" del ARN pre-mensajero; SAM 68 ,cuerpo nuclear que se cree está involucrado en el metabolismo del ARN; RNP, ribonucleoproteínas. (Tomado de Spector 2001).

En un núcleo en interfase, los cromosomas están organizados en zonas llamadas territorios cromosomales, los cuáles, basados en modelos propuestos por varios grupos, sugieren que se encuentran separados por canales llamados dominios o regiones intercromosomales (Lukasova *et al* 2002; Cremer y Cremer 2001). En esta etapa del ciclo celular, los cromosomas son altamente dinámicos y se ha visto que los genes activos transcripcionalmente se encuentran más cerca de los canales intercromosomales aledaños a la superficie de los territorios y con una mayor superficie de contacto con éstos (Lamond y Earnshaw 1998; Ragozy *et al* 2003). Se ha comprobado que los cromosomas se mueven dentro de su territorio y se ha propuesto que la explicación a estos cambios en

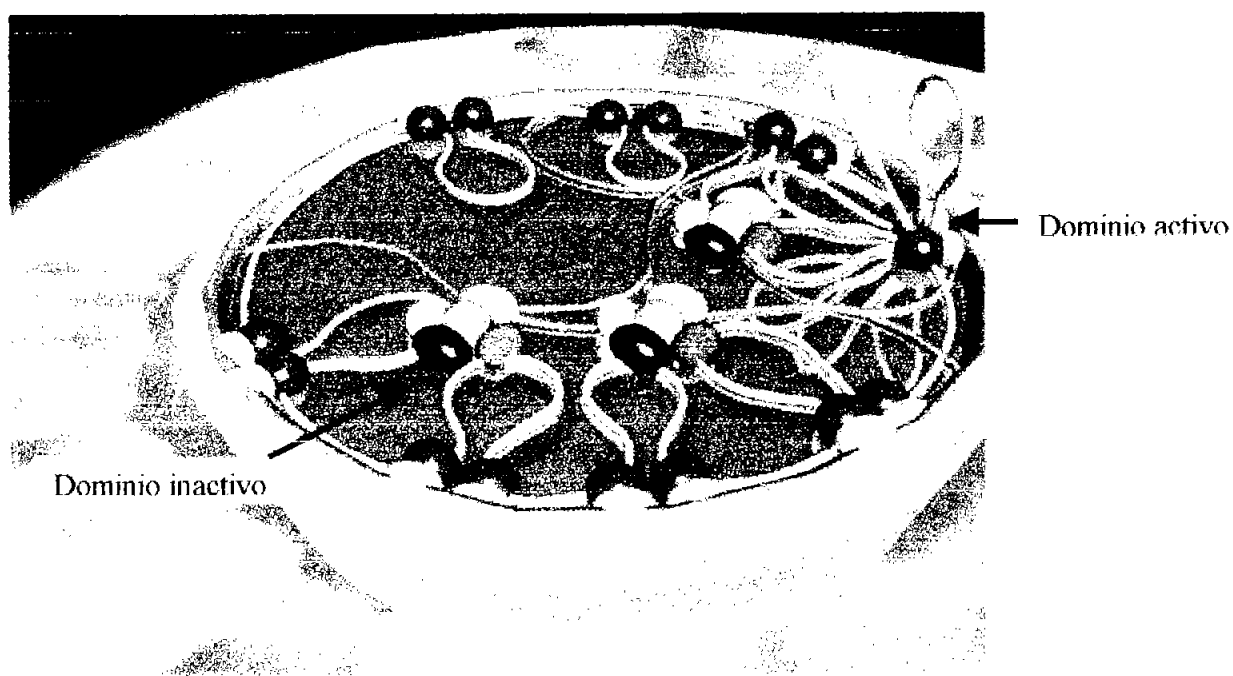
su localización sub-nuclear están determinados por la abundancia de factores involucrados en el procesamiento, la transcripción y la replicación en focos localizados aledaños a la periferia de estos territorios (Parada y Misteli 2002). Para explicar este concepto, existe un ejemplo clásico que demuestra el movimiento o reclutamiento de genes a una región en las células B de ratón que es rica en una proteína llamada Ikaros (Kirstetter *et al* 2002) Esta proteína es capaz de inactivar a los genes específicos del linaje linfoide a través de su relocalización a una región centromérica de heterocromatina. Mientras que los genes que necesitan ser transcritos permanecen en zonas favorables del nucleoplasma, Ikaros recluta a los genes que deben ser reprimidos a sitios de heterocromatina para inactivarlos (Lamond y Earnshaw 1998). Esta represión involucra el movimiento de loci específicos a un compartimiento nuclear donde son silenciados activamente (Cremer y Cremer 2001).

En levaduras se ha observado que los complejos que participan en el silenciamiento genético, particularmente la familia de las proteínas SIR (Fig. 12), se localizan en los telómeros cerca de la periferia nuclear donde se encuentran en compartimentos. Se han identificado por lo menos dos proteínas que ayudan a que esto ocurra; una de ellas es el complejo Ku, el cual al ser mutado en cualquiera de sus dos subunidades impide la posición perinuclear del telómero y la otra es Rap I, la cuál recluta a Sir4 al telómero la cuál a su vez va reclutando las demás proteínas del complejo Sir (Thrower y Bloom 2001; Mishra y Shore 1999; Luo *et al* 2002). Además se ha demostrado que la represión de genes reporteros baja a medida que se alejan de los telómeros (donde se encuentran los complejos represores (Marshall *et al* 1997; Wei *et al* 1998). Estas evidencias apoyan y asocian el concepto de territorios y dominios.





localizarse en regiones centrales del núcleo. Lo anterior correlaciona con la inactivación del locus estudiado (Fig. 13). Este mismo grupo ha demostrado una clara relación entre la presencia del “insulator” *gypsy* y una localización sub-nuclear cercana a la periferia del núcleo, siendo en esta zona donde el dominio enmarcado por el “insulator” *gypsy* se vuelve activo transcripcionalmente (Fig. 13). Las evidencias van más allá dado que demostraron que Su(Hw), el componente central de la acción del “insulator” *gypsy*, co-localiza con una alta frecuencia con la proteína laminina, componente estructural de la membrana nuclear (Gerasimova *et al* 1998).



**Fig. 13 Modelo de la organización del “insulator” *gypsy*.** El “insulator” *gypsy* mediante Su(Hw) (en azul), y mod (mdg4) (en verde), se une a se une a la lamina nuclear promoviendo la formación de los “loops” o asas cromáticas (en naranja), generalmente se encuentran en la periferia del núcleo. Proteínas represoras del grupo Polycomb encuentran los sitios de unión en los PRE (elementos de respuesta a polycomb) y no co-localizan con las proteínas componentes del “insulator” *gypsy*.

Recientemente el modelo de Corces ha sido apoyado debido a nuevos experimentos de su laboratorio, en los cuáles demostró que al colocar un elemento *gypsy* entre dos elementos *gypsy* origina la formación de un nuevo “loop” o asa cromatínica, formando de esta manera dos “loops” pequeños (Fig. 13). Esto además de apoyar su modelo, le da a las proteínas involucradas, particularmente a *Su(Hw)* una nueva relevancia al tener un papel en la formación y topología de los “loops” cromatínicos. Recientemente se ha descrito que CTCF es capaz de unirse a la matriz nuclear, esta característica le permitiría en cierta manera contribuir a la organización de la cromatina y quizás a la formación de nuevos “loops” como *gypsy* (Byrd y Corces 2003; Dunn *et al* 2003).

Las evidencias mencionadas demuestran que en el núcleo debe existir una organización tal que permita la correcta regulación de loci independientes, facilitando el remodelaje de la cromatina inicialmente a nivel de un dominio para que posteriormente se active o se reprima la transcripción en un momento determinado del desarrollo y que los límites de los dominios son elementos importantes para que esta organización se lleve a cabo de manera más eficiente. Además, los nuevos datos surgidos a partir de estudios citológicos apoyan la idea de la necesidad de tener elementos estructurales, ya sean asociados a membrana, poro nuclear o matriz nuclear, para el anclaje de secuencias que favorezca una topología óptima y así llevar a cabo la expresión regulada de los genes contenidos en un dominio.

### **El factor transcripcional CTCF**

La proteína CTCF se identificó primeramente como un factor capaz de unirse al promotor de *c-myc* de pollo (Klenova *et al* 1993), en este promotor su sitio de unión es rico en nucleótidos CG's. Posteriormente fue identificado en una región rica en AT's del silenciador del gen de la lisozima de pollo, como un factor que inducía la represión transcripcional en respuesta a V-erbA, el receptor

tiroideo o al receptor de ácido retinoico y al que bautizaron como Nep1 (Arnold *et al* 1996). Finalmente se demostró que Nep1 y CTCF eran la misma proteína aunque la secuencia del sitio de unión al ADN identificado en el promotor c-myc y el caracterizado en el silenciador del gen de la lisozima era diferente (Burcin *et al.* 1997).

### **Estructura peptídica y génica de CTCF**

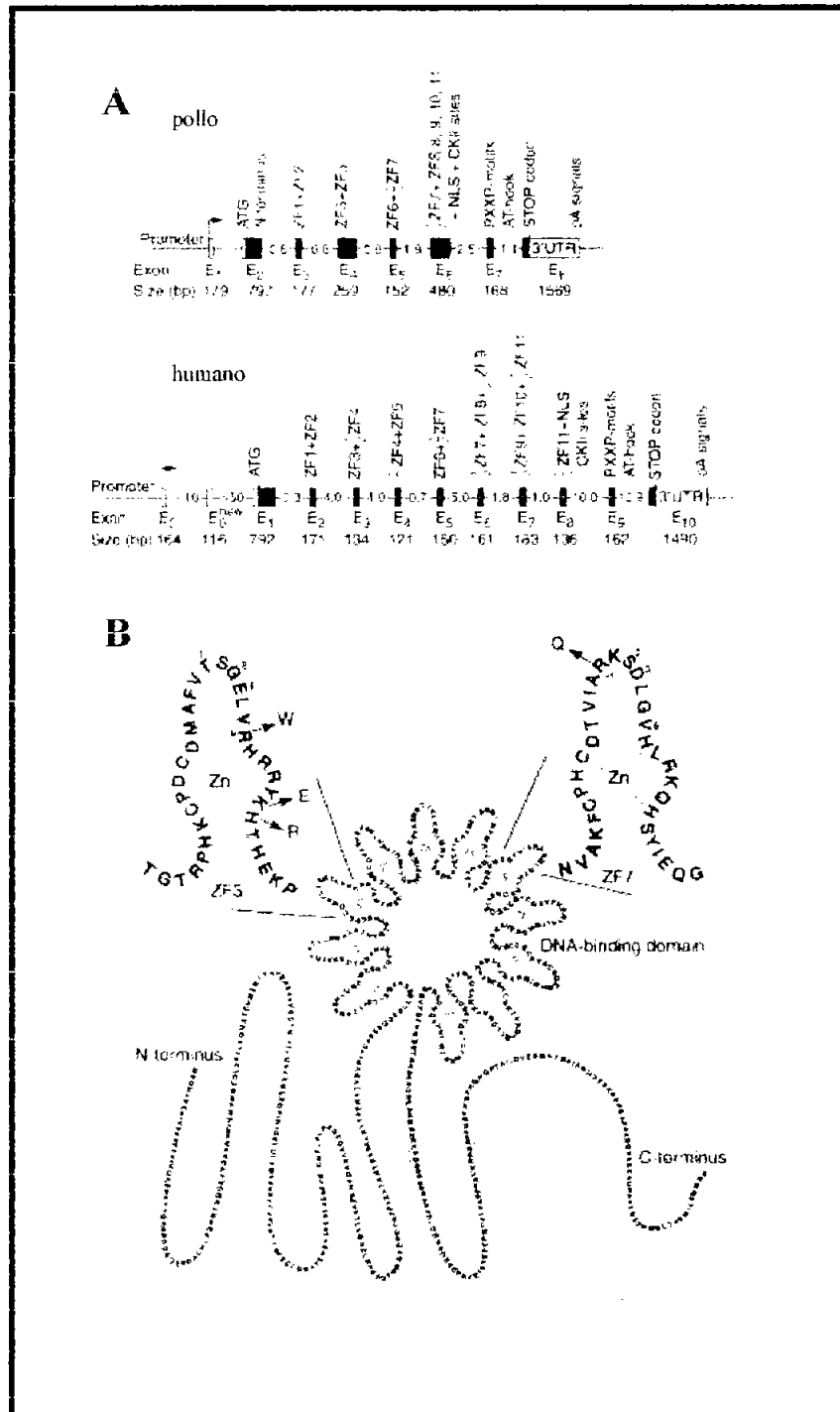
El factor CTCF es un factor altamente conservado y tiene una identidad del 100% entre distintos organismos en su dominio de unión al ADN, es decir en los dedos de zinc. CTCF tiene 11 dedos de zinc, 10 de la clase C<sub>2</sub>H<sub>2</sub> y uno de la clase C<sub>2</sub>HC (Klenova *et al* 1993; Ohlsson *et al* 2001). Al hacer una comparación entre el gen *CTCF* de pollo y el de mamíferos, en este último se identificaron 5 intrones adicionales y una gran cantidad de repetidas ALU y LINE al interior de los intrones, sin embargo, estas diferencias no cambian el marco de lectura; esto es importante, por que se ha visto que las secuencias ALU son características de los primates y en la mayoría de los casos ocasionan enfermedades debido a sus inserciones o cambian el patrón de metilación o de expresión de algunos genes, más no en este caso (Batzer 2002). Los exones que van del E2 al E9 de *CTCF* de mamíferos son relativamente pequeños, del exón E2 al E8 están codificados los 11 dedos de zinc y en el E9 una secuencia de unión al ADN rica en AT (KKRRGRP) (Ohlsson *et al* 2001; Fig. 14A). Los exones E1 al E10 contienen dominios represores; el onceavo dedo de zinc del tipo C<sub>2</sub>HC, es similar al usado por las proteínas “Friend of Gata-1” (FOG) las cuáles se unen al factor GATA-1 utilizando este tipo de dedo de zinc; esto es relevante ya que indica que CTCF probablemente pueda unirse a otros co-factores que ayuden a su función. Por su parte algunos dedos de zinc puedan estar mediando interacciones con otras proteínas, mientras que otros se encargan de la unión al ADN. La versatilidad de esta proteína está claramente dada por el uso diferencial de los dedos de zinc para unirse a distintas secuencias de unión en el ADN, y es esta unión diferencial la que provoca diversas respuestas biológicas (Filippova *et al.*

1996). Se ha propuesto que los dedos de zinc son capaces de unirse a varios factores como YB-1, YY1 y RNP-K (Chernukhin *et al.* 2000).

En el extremo *trans*-repressor localizado en la región terminal carboxilo de CTCF se encuentra un dominio conservado SKKEDSSDSE, donde la proteína puede fosforilarse en los 4 residuos de serinas. La cinasa responsable de la fosforilación es la CKII (Fig. 14B). Se sabe que al prevenir la fosforilación de las serinas se potencia la represión de los promotores blanco de CTCF. La fosforilación puede ser requerida para modular las actividades de CTCF mediadas a su vez por cambios conformacionales, homodimerizaciones o interacciones con otras proteínas.

Adyacentes a los sitios de fosforilación se encuentran 2 repetidas PXXP, características de un dominio SH3 el cuál media interacciones proteína-proteína. Datos no publicados del grupo de Lobanenkov indican que una de las proteínas que son capaces de unirse a este dominio es BIN1 (“Myc-binding protein 1”), la cuál se piensa que es un supresor tumoral. Lo que podría indicar que juega un papel importante junto con CTCF para controlar el ciclo celular (ver CTCF y represión).

Adicionalmente la proteína CTCF es capaz de dimerizarse e incluso trimerizarse, aunque la importancia biológica de esto todavía no ha sido estudiada. (Yusufzai *et al* 2004).



**Fig. 14. El gen CTCF y su estructura peptídica** **A)** Se muestra el locus del gen *CTCF* en pollo y en humano. Las cajas negras representan los exones, las flechas indican el sitio de inicio de la transcripción. **B)** Proteína CTCF, se muestran los 11 dedos de zinc, en rojo se muestra las regiones de fosforilación del carboxilo terminal, y en verde la secuencia rica en AT responsables de la unión al ADN.

## CTCF y represión

Otro reporte publicado y que tiene que ver con una proteína que se une a CTCF es el co-represor Sin3a, el cuál interactúa con los dedos de zinc de CTCF mediante 2 dominios: el dominio PAH3, el cuál previamente se había identificado como un sitio que interacciona con SMRT y con TFIIB, y el dominio carboxilo de 200 aminoácidos. Se ha propuesto que CTCF a través de Sin3a recluta una actividad de desacetilasa de histonas (Lutz *et al* 2000). Por otra parte en el silenciador de la lisozima de pollo, CTCF actúa de manera conjunta con el receptor de la hormona tiroidea (Lutz *et al* 2000). Por otra parte, el silenciamiento del gen *myc* mediado por CTCF llevó a pensar que esta proteína pudiera estar involucrada en la proliferación celular (Chernukhin *et al.* 2000).

La sobreexpresión de CTCF afecta negativamente la proliferación celular, CTCF lleva a la célula a un arresto en el ciclo celular pero sin inducir apoptosis. Sin embargo estos efectos observados no pueden explicarse únicamente por la represión de *myc*. Existe una gran variedad de genes regulados por CTCF, algunos de ellos involucrados en ciclo celular, como *E2F1*, *p53*, *p19/ARF*, *p16/INK4a* y *p27*. La expresión ectópica de CTCF induce un arresto en la proliferación celular sin provocar apoptosis en varias líneas celulares, ellos fueron capaces de mantener líneas celulares vivas sin proliferar por varios días (Rasko *et al.* 2001).

Estos hallazgos indican que CTCF puede regular la progresión del ciclo celular en varios niveles, lo que apoya las evidencias generadas por el grupo de Lobanenkov que indican que CTCF puede actuar como un supresor tumoral (Filippova *et al.* 2002; Takai *et al.* 2001; Ulaner *et al* 2003). Otro dato interesante surge al mutar los dedos de zinc que se unen a las regiones ricas en GC o que son importantes para la unión de CTCF a los promotores de distintos genes, no tiene efecto alguno en la unión de este factor a los sitios en donde CTCF funciona como un “insulator”, esto concuerda con el

hecho de que la unión de diferentes dedos de zinc dicta las diferentes respuestas biológicas para un gen o dominio determinado y que estas respuestas dependen del contexto genómico (Filippova *et al.* 2002).

En conclusión y debido a que CTCF está involucrado en diversos procesos, se le ha nombrado como un factor multifuncional, ya que puede formar parte integral de los “insulators”, y además conducir a las células a distintas respuestas biológicas al actuar como activador o represor.

### **CTCF y BORIS**

Hasta hace poco se pensaba que CTCF era el único factor con las características descritas, pero Loukinov y colaboradores descubrieron un segundo factor que tenía un 100% de identidad en los dedos de zinc, el cuál puede reconocer a las mismas secuencias de unión de CTCF, pero que difiere en sus extremos terminales amino y carboxilo (Loukinov *et al.* 2002). Este factor fue identificado y clonado de células provenientes de ratón y de humano, pero aparentemente está ausente en el pollo. Este gen fue nombrado *BORIS* (“Brother Of Regulator of Imprinted Sites”), presenta un patrón de expresión altamente específico, concentrándose en una subpoblación de espermatoцитos involucrados en la reprogramación de las marcas de metilación durante el desarrollo del espermatozoide. Un dato interesante, es que CTCF no se expresa en los espermatoцитos y que en ese momento la metilación de las citosinas es nula, lo que podría indicar que a diferencia de CTCF, BORIS no juega un papel en el mantenimiento de los genes improntados (Fedoriw *et al.* 2004).

La activación de BORIS se da en la última mitosis de la línea germinal cuando CTCF se activa BORIS se inactiva, y se ha propuesto que esto ocurre en asociación con el borrado y re-establecimiento de las marcas de metilación epigenéticas. BORIS a diferencia de CTCF se encuentra mayoritariamente en citoplasma, y se ha propuesto que su importación al núcleo tiene un papel regulatorio que no es compartido por CTCF, aunque todavía se desconoce cuál es ese papel (Loukinov *et al.* 2002). Se ha propuesto que en células transformadas se presenta una expresión aberrante de BORIS que compite con



CTCF por la regulación de sus genes blanco, o bien que interfiere con la propia regulación del promotor del gen *CTCF*. Un ejemplo de esto se ha visto en células normales de glándula mamaria, en las cuáles no se detectan niveles del ARN mensajero de BORIS, mientras que en células de tumores de glándula mamaria, se ha encontrado expresión de ARN mensajero de BORIS en un 50% de los tumores analizados (Klenova *et al* 2002). La expresión ectópica de BORIS en células normales induce proliferación y lleva a la transformación celular. El hecho de que BORIS se haya encontrado aberrantemente expresado en diversos tipos de cáncer y que además esté amplificado en una región asociada a cáncer (20q13) indican que quizás BORIS esté redirigiendo una reprogramación epigenética en sitios de unión a CTCF, y que quizás coopere con co-factores distintos a los de CTCF para interferir con las funciones normales de CTCF en el control del ciclo celular (Klenova *et al* 2002).

En conclusión es importante el recordar que CTCF es un factor multifuncional y que aún falta determinar sus mecanismos de acción incluyendo la actividad de bloqueo de enhancer. Estudios donde se encuentren a sus co-factores en distintos contextos, serán importantes para el entendimiento de las funciones de esta proteína.

El hecho de que la cromatina sea una estructura altamente represora, da relevancia a la existencia de los elementos que son capaces de remodelarla para poder regular la transcripción génica. Entre los elementos que regulan la transcripción revisamos a los LCR's, enhancers y los "insulators". Estos últimos tienen especial y novedosa importancia ya que son los encargados de delimitar a los dominios, de cooperar con distintos elementos para mantener a los dominios accesibles y protegerlos de los efectos de la cromatina adyacente. La organización del genoma en dominios independientes y la cooperación de los elementos inherentes a ellos para su regulación específica ha llevado a proponer a los dominios como uno de los niveles iniciales de regulación de la expresión génica. A continuación se describirán dos de los dominios mejor estudiados tanto a nivel de la estructura de la cromatina cómo de la regulación de la expresión de sus genes a lo largo del desarrollo.

## Los dominios $\alpha$ y $\beta$ -globina de pollo

Como ya hemos descrito, el genoma eucariote se encuentra organizado en dominios cromatínicos. Los dominios transcripcionalmente activos se caracterizan por una sensibilidad generalizada a la digestión por la nucleasa DNasa I, lo que indica que tienen una estructura más “abierta” de la cromatina y por lo tanto más accesible a los factores de transcripción. Estos dominios activos tienen límites bien definidos, lo que sugiere que los dominios representan unidades estructurales de cromatina, y su organización no afecta regiones aledañas. Los límites representan las fronteras entre diferentes estructuras de la cromatina (Recillas-Targa y Razin 2001). El ejemplo más claro de la organización del genoma en dominios, surge al estudiar el dominio  $\beta$ -globina de pollo (Burgess-Beusse *et al* 2002; Razin *et al* 2003; Fig. 15).

Este dominio en particular se encuentra activo cuando sus genes están siendo expresados en células eritroides, y por lo tanto la estructura de la cromatina está “abierta”, lo que correlaciona con una acetilación generalizada de los extremos amino-terminales de las histonas (Hebbes *et al* 1994; Litt *et al.* 2001); mientras que en tipos celulares no-eritroides, la estructura de la cromatina del dominio se encuentra “cerrada” y por lo tanto inaccesible a los factores transcripcionales. Así pues, existe una conexión entre la sensibilidad generalizada al corte por la DNasa I, la acetilación de las histonas y la activación del dominio (Martín *et al* 1996; West *et al* 2002).

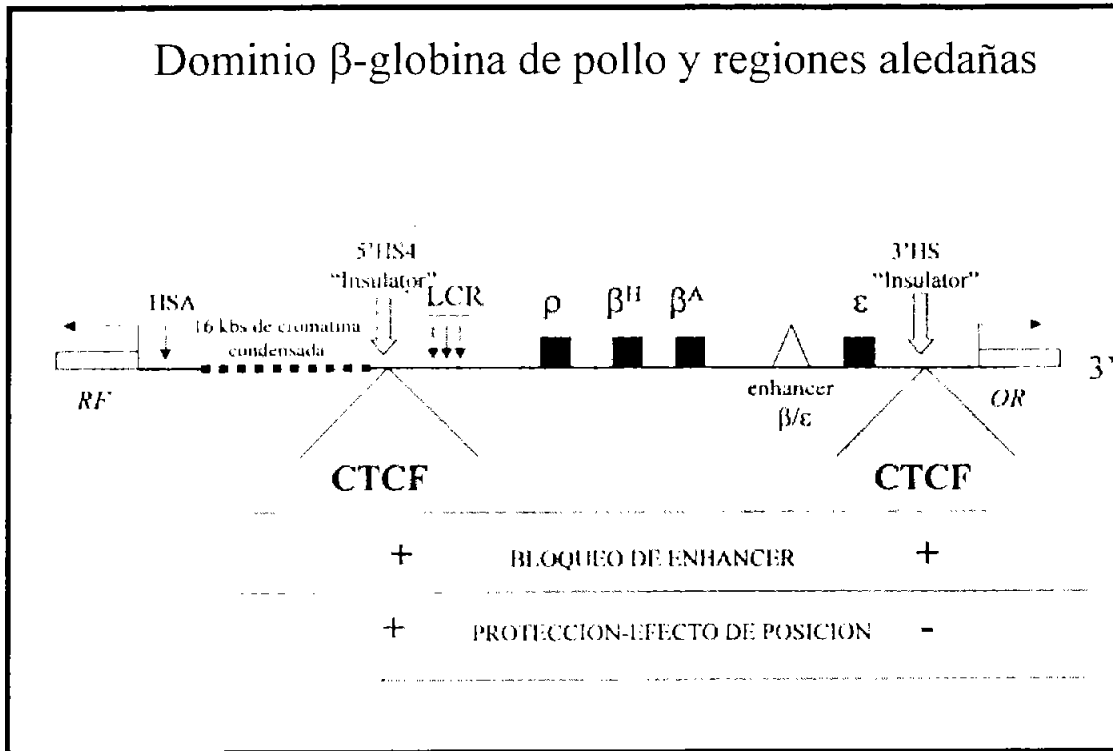
El dominio  $\beta$ -globina de pollo es considerado como un dominio de los llamados “fuertes”; estos dominios tienen límites bien definidos y la presencia de elementos que definen el límite del dominio como los insulators o elementos barrera es un factor decisivo para la formación y el mantenimiento de estos dominios. Otros ejemplos de dominios “fuertes” son el locus 87A7 delimitado por los insulators

scs y scs' y locus de mating type HML-HMR en *S. cerevisiae* (Razin *et al* 2003; Kellum 1992; Donze *et al.* 1999).

El dominio  $\beta$ -globina de pollo se extiende por 30 kb y contiene un gen embrionario ( $\epsilon$ ), uno fetal ( $\rho$ ) y 2 genes adultos ( $\beta^A$  y  $\beta^H$ ). Entre el gen adulto  $\beta^A$  y el gen embrionario se ha identificado a un enhancer eritroide específico ( $\beta^{A/\epsilon}$ ), el cuál actúa sobre ambos genes (Recillas-Targa y Razin 2001). En el extremo 5' no-codificante del dominio, se identificaron cuatro sitios de hipersensibilidad al corte por la DNasa I. Estos sitios forman parte del LCR del dominio el cuál funciona activando a los genes globina dependiendo del estadio de diferenciación celular y etapa del desarrollo (Li *et al.* 2002). El sitio de hipersensibilidad 5' cHS4 constituye el límite 5' del dominio (Burgess-Beusse *et al.* 2002; Reitman 1990). Río arriba del "insulator" cHS4 hay una región de cromatina altamente condensada, esta región de 16 kb es rica en secuencias repetidas conocidas como CR1 y el ADN se encuentra altamente metilado, río arriba de las 16 kb de cromatina condensada se encuentra un gen que codifica para el receptor del folato (Prioleau *et al.* 1999; Ghirlando *et al* 2004). En el extremo 3' del dominio se ha encontrado otro sitio de hipersensibilidad llamado 3' HS, y éste sitio define el límite del dominio en su extremo 3'. Río abajo del límite, a 2.3 kb, se encuentra el gen para el receptor olfatorio COR3' $\beta$ , el cuál aparentemente posee un patrón de expresión cerebro específico (Saitoh *et al.* 2000). Así pues, tenemos que el dominio  $\beta$ -globina de pollo tiene en sus extremos a dos sitios de hipersensibilidad constitutivos, el 5'cHS4 y el 3'HS, los cuáles han sido caracterizados como "insulators" corroborando la organización del genoma eucariota en dominios (Burgess-Beusse *et al.* 2002; Fig 15).

El sitio de hipersensibilidad 5' cHS4 ha sido identificado y caracterizado como un elemento tipo "insulator" que tiene las dos propiedades funcionales de estos elementos, la de bloquear la comunicación entre un enhancer y un promotor y la capacidad de proteger a un transgen de los efectos de posición creados por el entorno cromatínico del sitio de integración del transgen (Burgess-Beusse *et*

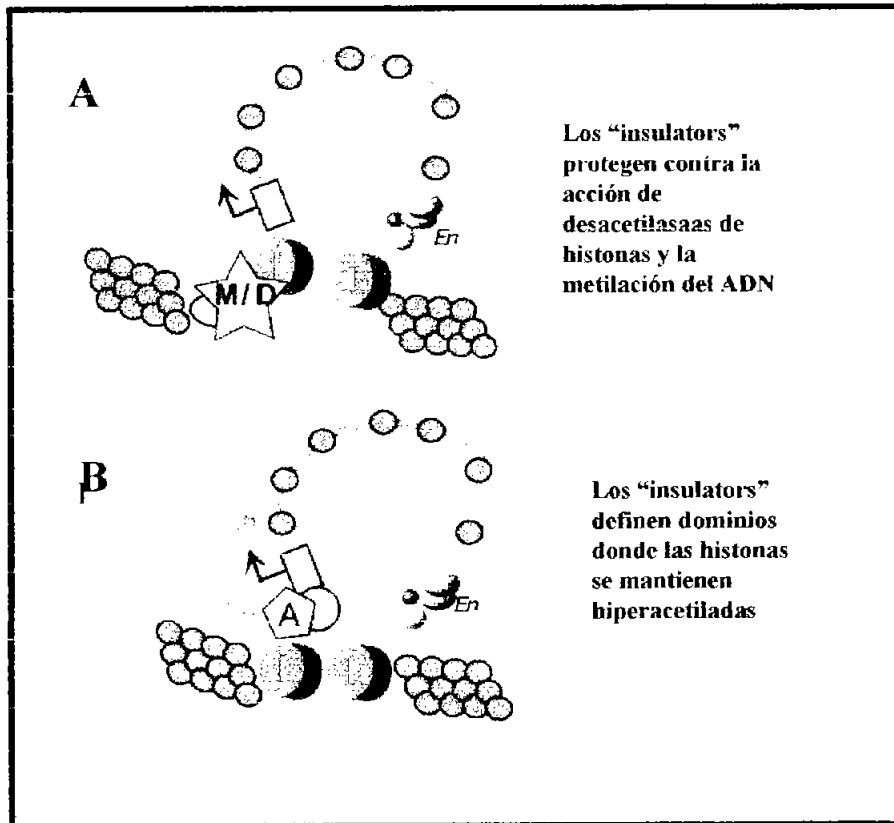
*al.* 2002; Chung *et al* 1997; Pikaart, Pikaart *et al* 1998). El “insulator” 3’HS tiene sólo la propiedad de bloqueo de enhancer, ya que es incapaz de proteger a un transgen de los efectos de posición (Saitoh *et al.* 2000; Recillas-Targa y Razin 2001).



**Fig. 15 Organización genómica del dominio  $\beta$ -globina de pollo y dominios aledaños.** Se señalan la posición de los “insulators” 5’ cHS4 y 3’ HS, a los cuales se une la proteína CTCF y las diferentes actividades que presentan. HS, sitio de hipersensibilidad; *RF*, gene del receptor del folato; *OR*, gene del receptor olfatorio.

De manera interesante, la propiedad de bloqueo de enhancer de ambos “insulators” es dependiente de la unión del factor CTCF (Bell *et al* 1999). La presencia de estos elementos a ambos lados del dominio apoyan la hipótesis de que están involucrados en la formación y el establecimiento de los dominios cromatínicos adicionalmente se ha planteado la posibilidad de su interacción física formando así la base de una asa cromatínica que contenga al dominio  $\beta$ -globina de pollo.

Recientemente los modelos propuestos acerca de la manera como los límites de los dominios regulan la apertura y cerrado de los mismos, relacionan a los “insulators” y a las secuencias MAR’s con complejos de remodelaje, metilación y acetilación/desacetilación de histonas. Felsenfeld propuso en 1998 dos modelos que explican cómo los límites de un dominio influyen en su formación y mantenimiento (Fig. 16). El primer modelo considera a los “insulators” como responsables de evitar o bloquear el acceso al dominio, de enzimas capaces de generar una cromatina compacta, como por ejemplo desacetilasas de histonas o metil-transferasas de ADN (Fig. 16A) En el segundo modelo, los “insulators” actuarían reclutando acetil-transferasas de histonas y/o complejos de remodelaje ATP-dependientes para favorecer una estructura de la cromatina “abierta” a nivel de un dominio (Pikaart *et al* 1998; Fig. 16B). Recientemente se han encontrado evidencias en el “insulator” 5’HS4  $\beta$ -globina de pollo que indican la presencia de una actividad de acetilasa de histonas (Mutskov *et al* 2003). Además Felsenfeld propuso que los límites eran capaces de bloquear la acción de desacetilasas y metilasas, para impedir que éstas actuaran sobre el dominio y mantenerlo abierto. Por otra parte, al reunir todos los estudios acerca de los límites de los dominios (en particular de los “insulators”), existen evidencias que sugieren que las funciones de estos límites dependen del contexto cromatínico en el cual se encuentren, es decir, la actividad de bloqueo de enhancer está destinada a evitar la promiscuidad entre los elementos de regulación a distancia de dominios adyacente como sucede en el caso del “insulator” 3’ del locus  $\beta$ -globina de pollo. Este “insulator” sólo presenta actividad de bloqueo de enhancer, por el contrario el “insulator” 5’cHS4 del mismo locus presenta ambas actividades (bloqueo de “enhancer” y protección contra el efecto de posición), se ha propuesto que esto se debe a que río arriba del “insulator” se encuentra una región de heterocromatina y por lo tanto debe de evitar que esta estructura represora invada el dominio a través de un efecto de heterocromatinización (Recillas-Targa y Razin 2001; Burgess-Beusse *et al* 2002; Fig 8).



**Fig. 16 Actividades propuestas de los insulators.** I, “insulators”; M/D, metil-transferasas, desacetilasas; A, acetilasas; En, enhancer

Los dominios son entidades dinámicas, y el ambiente genómico tiene una gran influencia en su organización y función. Sobre el dominio actúan diversos factores como su localización dentro del núcleo, el cuál se sabe que está compartimentalizado y que esta compartimentalización tiene efectos directos sobre la estructura de la cromatina y la regulación transcripcional, de esta manera encontraremos dominios que quizás no tengan limites bien definidos, o bien que no necesiten tener limites o “insulators” (Razin *et al* 2003; Isogai y Tjian 2003).

## El dominio $\alpha$ -globina de pollo

Un ejemplo de un dominio que aparentemente no tiene límites bien definidos dado que se encuentra en una región genómica constitutivamente abierta, es el dominio  $\alpha$ -globina de pollo, este dominio contiene un gen embrionario ( $\pi$ ) y dos genes adultos ( $\alpha^A$  y  $\alpha^D$ ) (Fig 17).

La región 5' no codificante, entre 2.5 y 5.5 kb río arriba del gen embrionario  $\pi$ , se ha encontrado un origen de replicación, el cuál correlaciona con la presencia de un enhancer débil, activo en células no-eritroides (Razin *et al* 1991). De 3 a 4.5 kb río arriba del gen  $\pi$  se ha identificado una isla CpG, la cuál tiene un efecto negativo sobre la transcripción de un gen reportero en transfecciones transitorias (Yudinkova y Razin 2002). Se ha pensado que la región rica en CG's, la cual tiene un efecto negativo sobre la transcripción podría estar modulando la actividad transcripcional de los genes cuando no necesiten ser expresados y que a su vez sería necesaria para mantener una estructura abierta de la cromatina, ya que las islas CpG's que se encuentran en promotores de genes generalmente están protegidas de la metilación (Razin *et al* 2000; Hendrich y Tweedie 2003).

A lo largo de todo el dominio existen múltiples sitios de hipersensibilidad (Fig 17). De 11 a 15 kb río arriba a partir del gen  $\pi$ , se encontraron 3 sitios de hipersensibilidad, uno constitutivo, enmarcado por dos sitios eritroide-específicos, uno de estos sitios co-localiza con un elemento tipo MAR (Razin *et al* 1999). En el extremo 3' del dominio, 2 kb río abajo del gen adulto  $\alpha^A$ , han sido identificadas secuencias repetidas tipo CR1, adicionalmente este dominio tiene secuencias ricas en A/T, localizadas en ambos extremos del dominio y algunas co-localizan con secuencias tipo MAR's (Farache *et al* 1990). Experimentos realizados por Razin y colaboradores, han concluido que estas MAR's organizan al dominio en un "loop" de aproximadamente 20 kb (Razin *et al* 1991). Por otra parte en el extremo 3' se identificó por homología de secuencias entre pato y pollo, una región

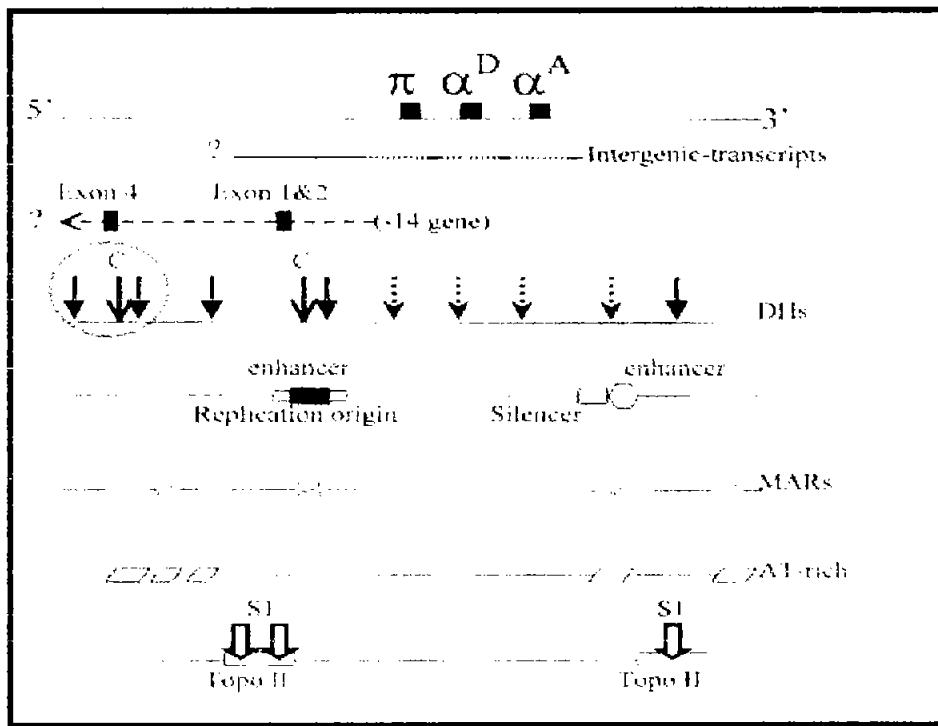
altamente conservada, la cuál corresponde a un silenciador y a un enhancer, estos elementos tienen sitios de unión a factores eritroide específicos como GATA-1 y a factores de transcripción generales como EKLF, NF-E2 y Sp1 (Recillas-Targa 1993).

Este dominio a diferencia del dominio  $\beta$ -globina de pollo conserva una estructura de la cromatina “abierta” en células no-eritroides donde los genes no son activos, esto puede deberse a que en el extremo 5' del dominio existe la presencia de secuencias que poseen más del 70% de homología con los exones 1, 2 y 4 de un gen constitutivo llamado -14, este gen fue encontrado originalmente en el locus  $\alpha$ -globina humano. El exón 4 correlaciona con los 3 sitios de hipersensibilidad del extremo 5' del dominio. El gen -14 se transcribe en células eritroides y en otros tipos celulares en dirección opuesta a los genes globina (Sjakste *et al* 2000; Vyas *et al* 1995). El producto del gen es ubicuo pero hasta el momento se desconoce su función. Se ha propuesto que estos transcritos intergénicos puedan contribuir a mantener la estructura de la cromatina abierta, ya que se sabe que durante la transcripción, la RNA polimerasa II es capaz de reclutar complejos con actividades de HAT's, y por lo tanto estarían contribuyendo a la acetilación de las histonas y a la formación de eucromatina (Gribnau *et al* 2000).

La estructura del dominio  $\alpha$ -globina de pollo es interesante ya que representa un dominio que no parece tener límites bien definidos, con una aparente organización constitutivamente “abierta” de la cromatina y pertenece al grupo de los dominios llamados “débiles”. Se piensa que estos dominios se establecen gracias a los elementos de regulación via la unión altamente específica de proteínas regulatorias a sus secuencias blanco. En este modelo de dominios “débiles”, la presencia de “insulators” o barreras se piensa que podrían regular la actividad de elementos como enhancers o silenciadores que afecten la expresión de los genes del dominio, o bien podrían tener una función que no haya sido descrita previamente, como por ejemplo la de regular a los transcritos intergénicos; esto es



particularmente relevante para este trabajo, ya que nosotros nos enfocamos a describir el extremo 5' del dominio  $\alpha$ -globina de pollo (Recillas-Targa *et al* 2000; Razin *et al* 2003).



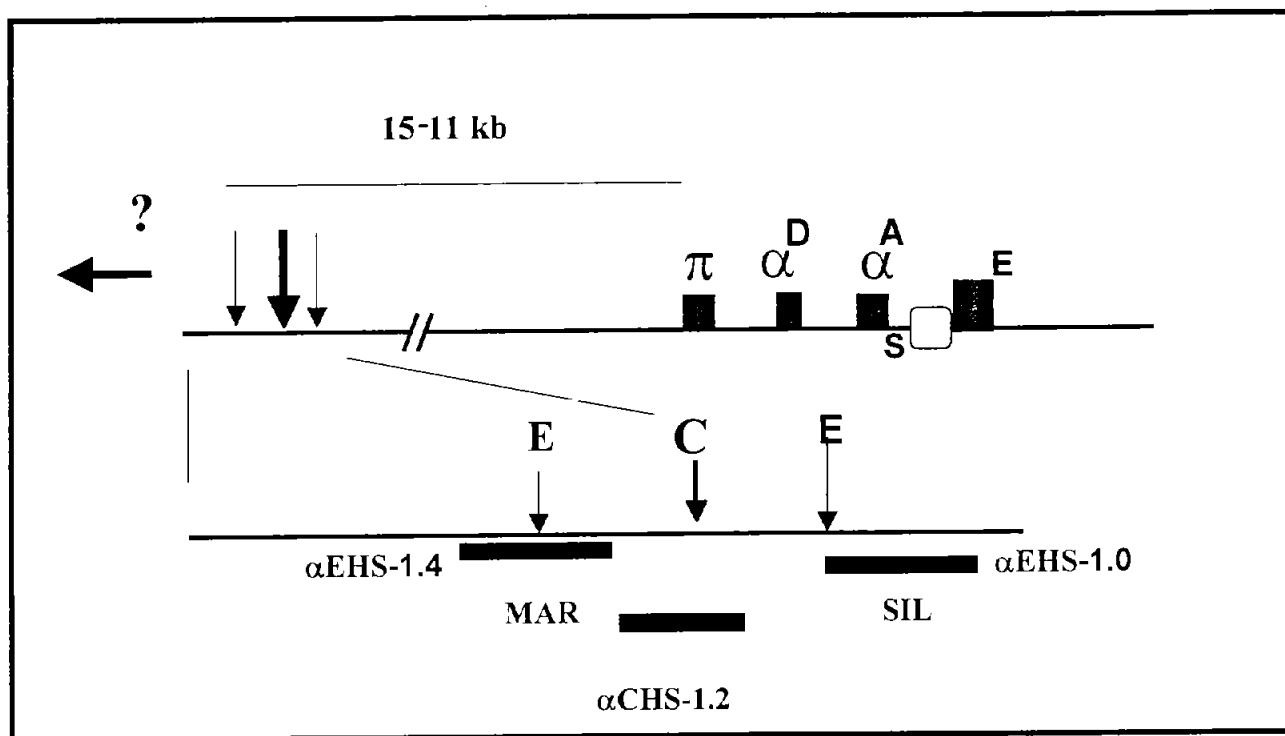
**Fig. 17. Dominio  $\alpha$ -globina de pollo.** Se muestran los 3 genes del dominio, los transcritos intergénicos aunque no está definido su sitio de inicio dentro del dominio. También se muestra la posición de los exones del gen -14, sobrelapando al exón 4 del gen -14, se encuentran los 3 sitios de hipersensibilidad que serán estudiados en el contexto del presente trabajo de tesis, y los demás sitios de hipersensibilidad que se encuentran a lo largo de todo el dominio. Se muestran los elementos de regulación encontrados, el enhancer débil hacia 5' y el silenciador y enhancer localizados hacia el extremo 3'. Además se ejemplifican las MAR's las cuáles correlacionan con las regiones ricas en A/T, y los sitios de corte a la enzima topoisomerasa II. Se resalta en un círculo rojo la zona 5' no-codificante que es estudio de esta tesis (Tomado de Recillas-Targa 2000; Recillas Targa y Razin 2001).

## Planteamiento del Problema

El dominio  $\alpha$ -globina de pollo está constituido por un gen embrionario  $\pi$  y dos genes adultos  $\alpha^D$  y  $\alpha^A$  (Fig. 17). Razin y colaboradores identificaron tres sitios de hipersensibilidad a la DNasa I en la región 5', localizados entre 11 y 15 kb río arriba del grupo de genes  $\alpha$ -globina (Fig. 18; Razin *et al* 1999). Sus observaciones muestran que dos sitios de hipersensibilidad a la DNasa I específicos del estado eritroide, enmarcan un sitio constitutivo (Razin *et al* 1999). Una zona posee la capacidad de unirse a la matriz nuclear *in vitro* (MARs) y una de ellas además tiene una actividad silenciadora (Fig. 18; Razin *et al.* 1999). La característica de unión a la matriz nuclear es importante dentro del contexto de nuestras investigaciones, como una alternativa de estudio, ya que como se ha dicho, se han identificado actividades tipo "insulator" que corresponden a secuencias de unión a la matriz nuclear como el caso de las secuencias MAR de lisozima de pollo o de la apolipoproteína B en humano (Stief *et al.* 1989; Namsiu *et al.* 1998). En estos dos casos estas secuencias confieren además protección a un transgen contra el efecto de posición. Las secuencias conteniendo estos sitios (un fragmento de ADN de 6.6 kb) fueron secuenciadas y clonadas. En el trabajo de Razin y colaboradores se analizaron las actividades funcionales de estos fragmentos y encontraron que no tenían actividad promotora ni de enhancer, aunque uno de los sitios eritroide específicos presentaba una actividad silenciadora ( $\alpha$ EHS-1.0; Fig. 18); demostraron que la secuencia que se encuentra en el sitio constitutivo (fragmento Bgl II-Sau3A de 900 pb; Razin *et al.* 1999), tiene una actividad propuesta y llamada de "terminador" (Razin *et al.* 1999). Esto lo comprobaron haciendo transfecciones transitorias, en donde la secuencia de 900 pb fue colocada en un plásmido entre el promotor  $\alpha^D$  y el enhancer de SV40. Con ésta construcción se observó una disminución en la actividad del gen reportero, atribuida a su actividad de "terminador" estos experimentos fueron hechos utilizando plásmidos circulares permitiendo que el enhancer actúe por el extremo 5' sobre el promotor. Los datos pueden ser interpretados como el bloqueo de la interacción entre el enhancer y el promotor (Recillas-Targa *et al.* 1999). Esta interpretación alternativa de los datos nos permite pensar que el fragmento de 900 pb pudiera tener una actividad tipo "insulator".

De manera preliminar realizamos un análisis de las secuencias de los 3 fragmentos que incluyen a los sitios de hipersensibilidad y encontramos que en el fragmento que incluye al sitio de hipersensibilidad constitutivo (llamado  $\alpha$ CHS-1.2; Fig 18), y el fragmento que corresponde al sitio de hipersensibilidad eritroide que tiene una actividad de unión a la matriz nuclear ( $\alpha$ EHS-1.4) se hallaron

posibles sitios de unión al factor CTCF (Bell *et al.* 1999; Razin *et al.* 1999) el cuál como se ha mencionado anteriormente es indispensable para la actividad de bloqueo de enhancer por parte de los “insulators”. Como se comentó anteriormente en la introducción se sabe que las MAR’s son capaces de tener actividad de bloqueo de enhancer también dependiente de CTCF y que generalmente se encuentran delimitando dominios (Bell *et al.* 1999). Por lo que suponemos que CTCF podría unirse a esas secuencias y participar en una actividad de bloqueo de enhancer característica de los insulators. Estos antecedentes nos llevaron a proponer la siguiente hipótesis.



**Fig. 18** Esquema del dominio α-globina de polio y la región 5' no codificante. En el costado 3' se encuentran dos elementos de regulación tipo silenciador (S) y enhancer (E). Entre 11-15 kb's arriba del primer gen se encuentra un grupo de sitios de hipersensibilidad a la DNasa I. Dos son eritroides (E) y uno constitutivo (C) (Razin *et al.* 1999). A partir de este estudio se les ha nombrado αEHS-1.4; αCHS-1.2 y αEHS-1.0. Una zona de unión a la matriz nuclear (MAR) y una zona con actividad silenciadora (Sil) en transfecciones transitorias ha sido localizada en el fragmento de 6.6 kb. Lo más relevante para el presente proyecto es que la secuencia que proponemos con un posible “insulator” se localiza exactamente en la zona del sitio constitutivo αCHS-1.2(Razin *et al.* 1999).

## HIPÓTESIS

Los “insulators” tienen la propiedad funcional de bloquear la interacción entre un enhancer y un promotor, propiedad dependiente del factor CTCF, únicamente cuando se encuentran entre el enhancer y el promotor. Generalmente correlacionan con sitios de hipersensibilidad constitutivos al corte por la DNasa I. Si los sitios de hipersensibilidad al corte por la DNasa I encontrados en el extremo 5' del dominio  $\alpha$ -globina de pollo son capaces de reducir la actividad de un gen reportero cuando están colocados entre el enhancer y el promotor, y basados en su localización a nivel de la organización genómica del dominio  $\alpha$ -globina de pollo, entonces proponemos que el elemento  $\alpha$ CHS-1.2 puede formar parte del límite del dominio y posee propiedades de “insulator”.

## OBJETIVOS

- 1.- Determinar si el sitio constitutivo de hipersensibilidad a la DNasa I  $\alpha$ CHS-1.2 y los sitios eritroides de hipersensibilidad  $\alpha$ EHS-1.4 y  $\alpha$ EHS-1.0, los cuáles se localizan de 11 a 15 kbs río arriba del primer gen del dominio  $\alpha$ -globina de pollo; tienen la propiedad de bloqueo de enhancer característica de los “insulators”.
- 2.- Determinar si el factor CTCF, responsable de la actividad de bloqueo en otros “insulators” de vertebrados, es capaz de unirse a estas regiones y contribuir a dicha propiedad.
- 3.- Determinar si el factor CTCF participa *in vivo* a lo largo de la diferenciación eritroide.
- 4.- Establecer si esta secuencia tiene las propiedades para proteger a un transgen de los efectos de posición causados por los distintos sitios de integración en el genoma.

## MATERIALES Y METODOS

### *Plásmidos y construcciones:*

El plásmido p $\alpha^0$  PUC contiene un fragmento de 6.6 kb que comprende a los tres sitios de hipersensibilidad eritroides (Razin *et al* 1999), este plásmido fue usado como templado para obtener los fragmentos correspondientes al  $\alpha$ EHS-1.4,  $\alpha$ CHS-1.2,  $\alpha$ EHS-1.0. Estos fragmentos fueron obtenidos por la reacción en cadena de la polimerasa (PCR) los oligonucleótidos utilizados para amplificar cada fragmento fueron:

Para el  $\alpha$ EHS-1.4: a) ccggggaagcttgcgactgttgcacctctaa, b) ccggggaagcttgcgactttcatgaagt; para el  $\alpha$ CHS-1.2: a)ccggggaagcttgcgacgtccctcctttcaaatetaac, b) ccggggaagcttgcgacctatttcagatgtgatca; para el  $\alpha$ EHS-1.0: a)ccggggaagcttgcgacgtgctcagcata, b) ccggggaagcttgcgacaattgccttctga; adicionalmente se amplificó un fragmento incluido dentro del  $\alpha$ CHS-1.2, el  $\alpha$ CHS-0.6 los oligonucleótidos utilizados fueron: a) ccggggaagcttgcgacgggttcacagagaagacttg, b) ccggggaagcttgcgacagccacaaaatagga.

Las condiciones de la PCR fueron: 50 segundos a 94°C, 1 min. a 55°C, 1 min. a 72°C, por 30 ciclos. Todos los primers contienen un sitio de restricción para las enzimas Sall y HindIII; estos sitios fueron útiles para clonar los fragmentos en el vector pAcatE; el sitio Sall se utilizó para clonar a los fragmentos entre el enhancer  $\beta/\epsilon$  y el promotor  $\beta^A$ , mientras que el sitio HindIII para clonarlos “afuera” (Recillas-Targa *et al.* 1999). El vector pALUCE se aisló a partir de la digestión del vector pGL3basic (Promega) con BglII y BamHI, de esta manera se aisló el gen de la luciferasa completo más la señal de poly (A). Este fragmento de 2214 bp se subclonó en el vector pAcatE el cuál fue previamente digerido con BamHI para eliminar el gen reportero CAT.

### *Cultivo Celular*

Todos los medios y sueros fueron obtenidos de Invitrogen. Las células DT40 son células linfoides pre-B transformadas por el virus de la leucosis aviar (ALV), estas fueron cultivadas en 75% DMEM suplementado con 50 mM  $\beta$ -mercaptoetanol (Invitrogen), 5% suero de pollo, 10% suero fetal bovino, 10% Tryptose phosphate broth y penicilina-estreptomina.

LA TESIS NO SALE  
DE LA BIBLIOTECA

Las células 6C2 son pre-eritroblastos transformados de pollo fueron cultivadas en medio  $\alpha$ -MEM suplementado con 2% suero de pollo, 10% suero fetal bovino, 1mM buffer HEPES (Sigma) y 50 mM  $\beta$ -mercaptoetanol. Las células HeLa (células de epiteliales de cáncer cervical) y las células COS-1 (células de riñón de mono verde africano transformadas por SV40) fueron cultivadas en DMEM suplementado con 10% suero fetal bovino más antibióticos.

Las células eritroides de pollo de 10 días de gestación (RBC), después de la transfección se mantuvieron a 37°C en medio L-15 suplementado con suero fetal bovino al 5%, y 2% suero de pollo más antibióticos sin CO<sub>2</sub>. Todas las células excepto las RBC fueron cultivadas en una atmósfera de 5% CO<sub>2</sub> a 37°C.

#### *Ensayo de Transfección Transitoria*

Las células de embriones de pollo de 10 días de gestación fueron obtenidas de las venas apicales de los embriones. Estas células se lavaron 2 veces con PBS 1X y fueron contadas. Se utilizaron  $1 \times 10^8$  células por transfección con 6  $\mu$ g de cada plásmido linealizado con XmnI y 6  $\mu$ g del plásmido pTK  $\beta$ -gal para normalizar la transfección. Las células se resuspendieron en 500  $\mu$ l de PBS 1 X y se electroporaron utilizando celdas de electroporación de Bio-Rad (ver apéndice) a 500  $\mu$ F y 450 V en un electroporador (Gene-Pulser II de Bio-Rad). Después de la electroporación las células se incubaron en hielo por 5 min. Y posteriormente se les añadió 1 ml de medio L-15. 24 horas después de la transfección se lisaron las células con 300  $\mu$ l del buffer de lisis del kit  $\beta$ -galactosidase enzyme assay kit (Promega), los extractos se incubaron a 65°C para inactivar las desacetilasas y se determinó la actividad de  $\beta$ -galactosidasa, las actividades de CAT se determinaron agregando un mismo volumen del sustrato CAT ( para 10 ml: 3 ml Tris pH 8.0 1M, 140  $\mu$ l de cloramfenicol 250 mM, 250  $\mu$ l Acetil CoA 1mM, 80  $\mu$ l Acetil CoA [<sup>3</sup>H] 250 mCi) y se transfirieron a un vial para el contador de centelleo con 5 ml del líquido de centelleo Bray (POPOP 0.0125%, PPO 0.055% en Tolueno). Se incubaron los viales por 24 horas a 60°C y se contó cada vial en el contador Beckman (modelo LS 6000 SC) por dos minutos.

Al utilizar el vector pALUCE, que lleva como gen reportero a la luciferasa, se normalizó la actividad de la transfección, co-transfectando 200  $\mu$ g del plásmido pRLCMV. La actividad de luciferasa se determinó a las 24 o 48 horas utilizando el kit Dual luciferase (Promega), las unidades relativas de luciferasa se midieron en el luminómetro TD 20/20 (Turner Designs).

### *Extractos Nucleares de células RBC de 10 días*

Se obtiene la sangre de las venas apicales de los embriones de pollo de 10 días y se colecta en PBS + 2 mM EDTA frío, se centrifugan 10 min a 5000 rpm en una centrifuga Labofuge 400 R (Heraeus) y se lavan 2 veces con cinco volúmenes de PBS. Posteriormente se resuspenden en 20 volúmenes de PBS 0.67X +2 mM EDTA, se incuban en hielo por 10 minutos. Se cuentan las células, se centrifugan y se resuspenden en 5 volúmenes de buffer A (10 mM HEPES pH 7.9, 10 mM NaCl, 1.5 mM MgCl<sub>2</sub>, añadir al final 1mM DTT, 0.7 µg/ml pepstatina, 0.5 µg/ml leupeptina, 0.5 mM PMSF). Transcurrido el tiempo de incubación, las células se homogenizan mediante 12-15 golpes con el pistón B ("tight") en un homogenizador Dounce. corroborar la lisis celular al microscopio. Centrifugar 20 min. a 15 000 rpm en rotor SS34 y resuspender el pellet en 3 ml de buffer C (20 mM HEPES pH7.9, 0.42 M NaCl, 1.5 mM MgCl<sub>2</sub>, 0.2 mM EDTA, 25% glicerol e inhibidores y DTT como en buffer A) por cada 5 x 10<sup>9</sup> células, lentamente con una pipeta, transferir al homogenizador y dar 12 golpes, se agitan lentamente por 30 minutos en el hielo en el mismo homogenizador (sin acortar el tiempo en este paso). Centrifugar 30 min. en SS-34 a 15 000 rpm. Tomar el sobrenadante y dializar contra 6 litros (o contra más de 50 volúmenes) de buffer D (20 mM HEPES pH 7.9, 0.1 M NaCl, 1.5 mM MgCl<sub>2</sub>, 0.2 mM EDTA, 20% glicerol, 0.5 mM DTT e inhibidores como en buffer A). Las proteínas fueron cuantificadas por Bradford (Bio-Rad) y congeladas en nitrógeno líquido.

### *Ensayo de Movilidad Electroforética*

Los oligonucleótidos se marcaron radioactivamente con [ $\gamma$  <sup>32</sup>P] dATP. 0.5 pmol (20,000 cpm de cada oligonucleótido) posteriormente se alinearon y fueron incubados con 20 µg de diferentes extractos nucleares en buffer GS 2X (40 mM HEPES pH7.9, 100 mM KCl, 10 mM MgCl<sub>2</sub>, 10% glycerol, más 1 µg of poly dI-dC como competidor inespecífico y 1 mM DTT) en un volumen final de 20 µl , por 15 min en hielo y 15 min a temperatura ambiente. Las competencias se llevaron a cabo con 20 o 40 pmol de oligonucleótido frío, la incubación con estos oligonucleótidos se hizo por dos etapas: 15 min en hielo y 15 min a temperatura ambiente, antes de agregarle el oligonucleótido marcado radiactivamente. Las muestras fueron analizadas en un gel de poliacrilamida al 5 o 6%, se secaron y se expusieron a una película de rayos X (KODAK Biomax) la cuál fué revelada 24 horas después.

Oligonucleotidos utilizados para los ensayos de retardo: Se muestra para cada caso solo una cadena, para el experimento la cadena reversa se alineó con su contraparte incubándola por 10 minutos a 90°C

e incubando posteriormente a temperatura ambiente por 10 minutos en buffer de alineamiento 2X (20 mM Tris-HCl pH8.0, 100 mM NaCl, 1 mM EDTA). Los oligonucleótidos marcados y alineados se purificaron por electroforesis en gel de poliacrilamida (PAGE), todos los oligonucleótidos se muestran en orientación 5'-3'.

**M9)** TTT CCA TAT AGA ACA ATG TAA ATT CTT TCC TAG AGT TAA GCA GGG CTT TTT  
CAT AAG CTG; **C10)** CAT GGC CAG CAG GCC CTC CTT GGT GCC CAC CTT GCT GCA GTG  
GGA CTG CGC TGC TCC GGA; **C11)** CCT CAG TGA GCT GCC AGG GGA AGC TCA TAG TTG  
CAG GAT GGG AGC AGT GTG AGG TTT GAT; **C12)** TGT GTG CCT CCA TCA GAC CAC CCC  
CTC CAC CAT GCA GGT ACC CCA GAT AAC ACA GTA TCA; **C13)** TGG GTG TGT TCC ACA  
CCC ACC TTA TCT CTC GGC CAC AAG CAG GCC CTC AGT GCC CAC AGA; **C14)** GGC TGC  
TCT GCT CCA AAC AGG ATG CCC AGC ACA GCT GAG GGC TAG CAG AGA GGA GGG  
GAG AGT

#### *Sobre-expresión de CTCF de pollo*

El plásmido pSG5-CTCF fue donado amablemente por la Dra. Elena Klenova de la Universidad de Essex, Reino Unido; este plásmido se transfectó en células COS-1 por electroporación a 100 V y 950  $\mu$ F; 24 horas después las células se lisaron con 2 ciclos de congelación y descongelación en buffer RIPA (150 mM NaCl, 1% Nonidet P-40, 0.5% deoxycholate, 0.1% SDS y 50 mM Tris-HCl pH 8.0 con 10  $\mu$ g/ml inhibidores de proteasas, leupeptina, PMSF y pepstatina A). Posteriormente se centrifugan y los lisados se almacenaron a  $-70^{\circ}\text{C}$ . Este mismo protocolo fue utilizado para la preparación de los extractos totales de células HeLa.

#### *Transcripción/ traducción in vitro.*

El plásmido pSG5-CTCF fue utilizado para transcribir/traducir la proteína CTCF de pollo, esto se llevó a cabo con el kit TnT a partir del promotor T7 (Promega).



### *Generación del anticuerpo policlonal anti-CTCF de pollo*

La region terminal amino no conservada entre pollo, ratón y humano de CTCF de pollo que comprende los aminoácidos 86 al 233, fue clonada en el vector pet-28a (+) el cuál se obtuvo de Novagen. La expression del péptido se indujo con IPTG y la purificación se llevo a cabo a través de una columna cargada con níquel (Ni (+)). Una vez que se purificó la proteína, los anticuerpos fueron generados inmunizando conejos machos de la cepa Nueva Zelanda. El título del anticuerpo fué determinado por la técnica de ELISA.

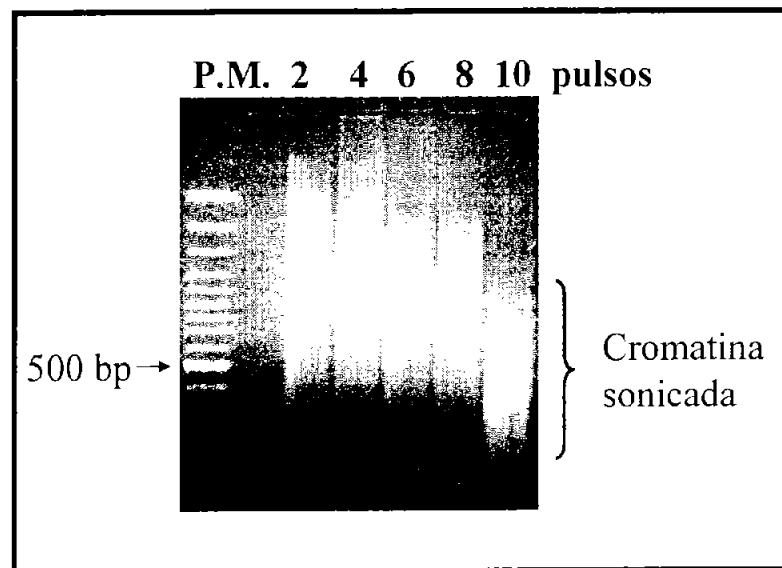
### *Western Blot*

Las células se cosecharon y resuspendieron en buffer de lisis (20 mM HEPES pH 7.9, 400 mM KCl, 20% glicerol, 2 mM DTT más inhibidores de proteasas); se congelaron y descongelaron dos veces y las proteínas fueron cuantificadas por el ensayo de Bradford (Bio-Rad). Después de separar las proteínas por electroforesis en gel de poliacrilamida desnaturalizante, las proteínas se transfirieron a membranas PVDF (Millipore). Las membranas se bloquearon con leche descremada al 5% toda la noche, se incubaron con el anticuerpo primario correspondiente entre 2 horas y toda la noche a 4°C, se lavaron extensivamente y se incubaron con el anticuerpo secundario correspondiente por 1-2 horas, se hicieron lavados y se revelaron con el kit de Super Signal (PIERCE), posteriormente fueron expuestos a una película de rayos X por 5-10 min y se revelaron.

### *Ensayo de Inmunoprecipitación de la cromatina (ChIP)*

Este protocolo fué implementado de acuerdo a Farnham y colaboradores con modificaciones menores. Células 6C2, DT40 o RBC, fueron tratadas con formaldehído (Sigma) por 10 minutos, se añadió una concentración final de 125 mM de glicina sólida para detener la reacción de entrecruzamiento ó “crosslinking” y se incubaron las células por 5 min a temperatura ambiente. Se lavaron las células con PBS 1X (8 g NaCl, 0.2 g KCl, 1.44 g Na<sub>2</sub>HPO<sub>4</sub>, y 0.24 g KH<sub>2</sub>PO<sub>4</sub> para 1 litro) y se resuspendieron en 10 ml de buffer de lisis (5 mM PIPES pH 8.0, 8 mM KCl, 0.5% NP40 más inhibidores de proteasas), se incubaron por 10 minutos en hielo y en los casos que fue necesario, se homogenizaron las células con un homogenizador tipo dounce con el pistón B hasta obtener los núcleos. Los núcleos se recuperaron por centrifugación y fueron resuspendidos en 2 ml del buffer de lisis de núcleos (50 mM Tris-HCl pH8.0, 10 mM EDTA, 1% SDS más inhibidores de proteasas), después de una incubación de

10 min en hielo, se sonicó la cromatina en un Sonicador Branson modelo 250 con una punta de 3 mm por 30 segundos con salida 5, o hasta que se obtuvo cromatina en un rango entre 500 y 800 bp (Fig. A). Los lisados se centrifugaron 2 veces a máxima velocidad a 4°C, y se recuperó el sobrenadante. En este punto se toma una alícuota de la cromatina y se marca como INPUT. La cromatina se dividió en alícuotas que representan  $2 \times 10^7$  células y se llevó a un volumen de 1.5 ml con buffer IP (0.01% SDS, 1.1% Triton X 100, 1.2 mM EDTA, 17.7mM Tris HCl pH 8.1, 167 mM NaCl), y se incubó overnight a 4°C con un anticuerpo irrelevante (anti-X, goat anti-mouse IgG Santa Cruz Biotechnology) como control negativo, y con 20 µl de suero pre-inmune o 5, 10, 15 and 20 µl de anti-pollo CTCF-IFC de un título 1:10 000, o 10 µg de anti-CTCF N17 (Santa Cruz Biotechnology).



**Fig. A Estandarización de los pulsos de sonicación necesarios para obtener cromatina entre 500 y 800 bp.** Gel de agarosa que muestra la cromatina de RBC con distintos pulsos de sonicación. Con 10 pulsos se obtuvo cromatina entre 500 y 800 bp.

Al día siguiente se añadieron 20 µl de proteína A/G plus agarosa (Santa Cruz) y se incubó 5 horas a temperatura ambiente o bien toda la noche a 4°C. Se centrifugó la muestra y el sobrenadante se

recuperó y se etiquetó como la fracción no unida (UB). La fracción inmunoprecipitada se lavó una vez con 1.5 ml de buffer RIPA (50 mM Tris pH 8.0, 150 mM NaCl, 0.1% SDS, 0.5% deoxycholate, 1% NP-40, 1 mM EDTA mM), una vez con buffer de Alta Sal (50 mM Tris pH 8.0, 500 mM NaCl, 0.1% SDS, 0.5 % Deoxicolato. 1% NP-40, 1 mM EDTA más inhibidores de proteasas); una vez con buffer de LiCl (50 mM Tris pH 8.0, 1 mM EDTA, 250 mM LiCl, 1% NP-40, 0.5% deoxicolato más inhibidores de proteasas) y dos veces con buffer TE (10mM Tris- HCL, 1mM EDTA pH 8.0). Se añadieron 5 µg de RNasa A por muestra y se incubaron 30 min a 37°C, las muestras se incubaron toda la noche a 67°C para revertir el entrecruzamiento; al día siguiente se añadieron 60 µg de proteinase K (Sigma), y se incubaron 2 horas a 45°C. Se extrajo el ADN 2 veces con fenol:cloroformo y una vez con cloroformo, las muestras se precipitaron añadiendo un volumen 1/10 de NaCl 5M, 20 µg de glicógeno y 3 volúmenes de etanol 100%, se incubaron toda la noche a -20 C. Al día siguiente las muestras se centrifugaron se lavaron una vez con etanol 70% y se resuspendieron en 40 µl de agua desionizada y se hicieron las respectivas PCRs radioactivas.

#### *PCR's radioactivas*

Se diseñaron dos pares de oligonucleótidos, el primero amplifica una región de 120 bp del  $\alpha$ EHS-1.4 que contiene al sitio M9 las secuencias son: 0022, 5'-cagtctgtcaggaagaaaga; 0023, 5'-tgtggcctggttttagctg; el segundo par de oligonucleotidos amplifican una secuencia de 301 bp que contiene un fragmento del enhancer  $\beta$ -globina de pollo (110-558 bp después de la señal de poly-A del gen  $\beta^A$ ) los cuáles fueron usados como control negativo: 0026, 5'-ctgggtgggggcaggt; 0027, 5'-ctgctttgctgccctgtg. Las PCRs se hicieron a una temperatura de alineamiento de 56.3°C por 20 ciclos, y se corrieron en un gel de poliacrilamida al 6% a 150 V en 0.5X TBE. Los geles se secaron y se expusieron a una película de rayos X (KODAK BIO-MAX) por 24 a 72 horas y revelados.

## **RESULTADOS**

“Insulators” dependientes de CTCF en el costado 5' del dominio  $\alpha$ -globina de pollo

recuperó y se etiquetó como la fracción no unida (UB). La fracción inmunoprecipitada se lavó una vez con 1.5 ml de buffer RIPA (50 mM Tris pH 8.0, 150 mM NaCl, 0.1% SDS, 0.5% deoxycholate, 1% NP-40, 1 mM EDTA mM), una vez con buffer de Alta Sal (50 mM Tris pH 8.0, 500 mM NaCl, 0.1% SDS, 0.5 % Deoxicolato, 1% NP-40, 1 mM EDTA más inhibidores de proteasas); una vez con buffer de LiCl (50 mM Tris pH 8.0, 1 mM EDTA, 250 mM LiCl, 1% NP-40, 0.5% deoxicolato más inhibidores de proteasas) y dos veces con buffer TE (10mM Tris- HCL, 1mM EDTA pH 8.0). Se añadieron 5 µg de RNasa A por muestra y se incubaron 30 min a 37°C, las muestras se incubaron toda la noche a 67°C para revertir el entrecruzamiento; al día siguiente se añadieron 60 µg de proteinase K (Sigma), y se incubaron 2 horas a 45°C. Se extrajo el ADN 2 veces con fenol:cloroformo y una vez con cloroformo, las muestras se precipitaron añadiendo un volumen 1/10 de NaCl 5M, 20 µg de glicógeno y 3 volúmenes de etanol 100%, se incubaron toda la noche a -20 C. Al día siguiente las muestras se centrifugaron se lavaron una vez con etanol 70% y se resuspendieron en 40 µl de agua desionizada y se hicieron las respectivas PCRs radioactivas.

#### *PCR's radioactivas*

Se diseñaron dos pares de oligonucleótidos, el primero amplifica una región de 120 bp del  $\alpha$ EHS-1.4 que contiene al sitio M9 las secuencias son: 0022, 5'-cagtctgcaggaagaaaga; 0023, 5'-tgtggcctggttagctg; el segundo par de oligonucleotidos amplifican una secuencia de 301 bp que contiene un fragmento del enhancer  $\beta$ -globina de pollo (110-558 bp después de la señal de poly-A del gen  $\beta^A$ ) los cuáles fueron usados como control negativo: 0026, 5'-ctgggtggggcaggt; 0027, 5'-ctgctttgctgcctgtg. Las PCRs se hicieron a una temperatura de alineamiento de 56.3°C por 20 ciclos, y se corrieron en un gel de poliacrilamida al 6% a 150 V en 0.5X TBE. Los geles se secaron y se expusieron a una película de rayos X (KODAK BIO-MAX) por 24 a 72 horas y revelados.

## **RESULTADOS**

"Insulators" dependientes de CTCF en el costado 5' del dominio  $\alpha$ -globina de pollo

En el laboratorio el interés principal se centra en comprender como la arquitectura de la cromatina es capaz de modular la regulación de la expresión genética a lo largo del desarrollo. La necesidad del genoma de organizarse en cromatina, trae como consecuencia una estructura represora para la actividad transcripcional, la cuál debe remodelarse para permitir la entrada de la maquinaria basal de transcripción y de diferentes factores transcripción y lograr reconocer y unirse a sus secuencias blanco.

Se ha propuesto que los genes están organizados en dominios cromosómicos, los cuáles tienen límites bien definidos. Existen evidencias de que estos límites o “insulators” juegan papeles preponderantes en la regulación de los genes dentro de un dominio, y cooperan con los elementos de regulación del dominio para activar o reprimir a los genes de una manera regulada y localizada (West *et al.* 2002).

En particular para el caso del dominio  $\alpha$ -globina de pollo aún se desconoce la presencia y necesidad de tener límites bien definidos. Este dominio está localizado en una región genómica cuya estructura de la cromatina se encuentra permanentemente abierta, y se desconoce su entorno cromatínico tanto en el extremo 5' como en el 3'.

Con el fin de conocer más acerca de la regulación de los genes globina y de la importancia de la organización del genoma eucariota en dominios, hemos decidido estudiar la posibilidad de que este dominio en su costado 5' tenga elementos tipo “insulator”.

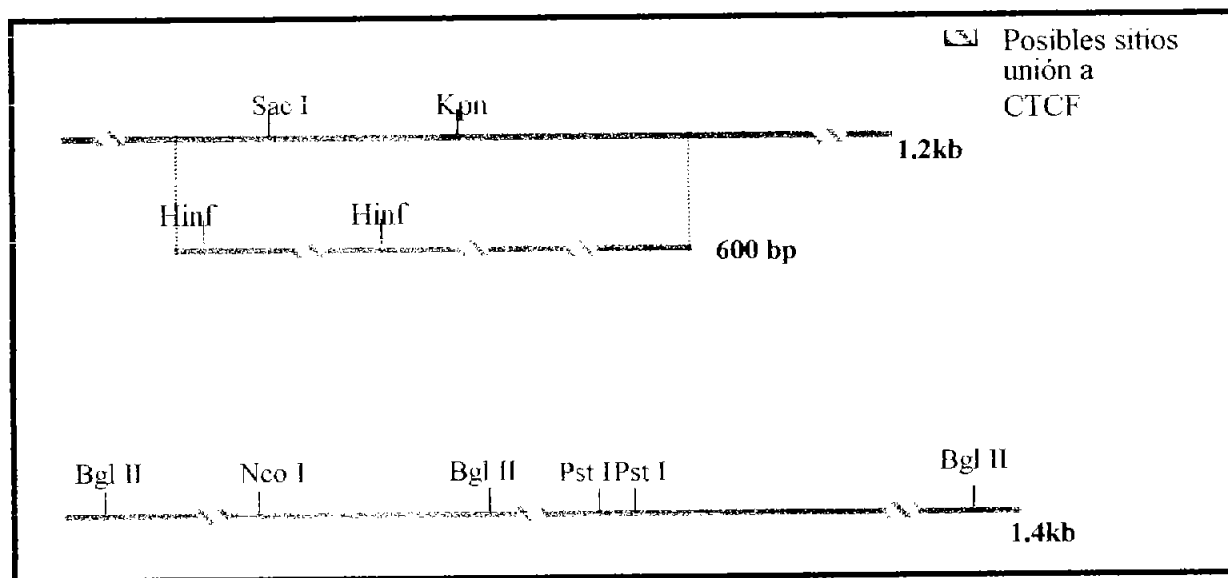
El Dr. Sergey Razin, nos proporcionó los plásmidos que contienen la región de 6.6 kb que contiene los 3 sitios de hipersensibilidad localizados entre 11 y 15 kb a partir del primer gen del dominio, los cuáles utilizamos para obtener las diferentes regiones  $\alpha$ EHS-1.4,  $\alpha$ EHS-1.0 y  $\alpha$ CHS-1.2, y poder hacer los experimentos subsecuentes y a quién agradezco esta donación, también quiero agradecer a el L. en IBB

Héctor Rincón Arano por la ayuda prestada durante la elaboración del anticuerpo anti-CTCF de pollo y quien hizo el Western Blot presentado en el artículo.

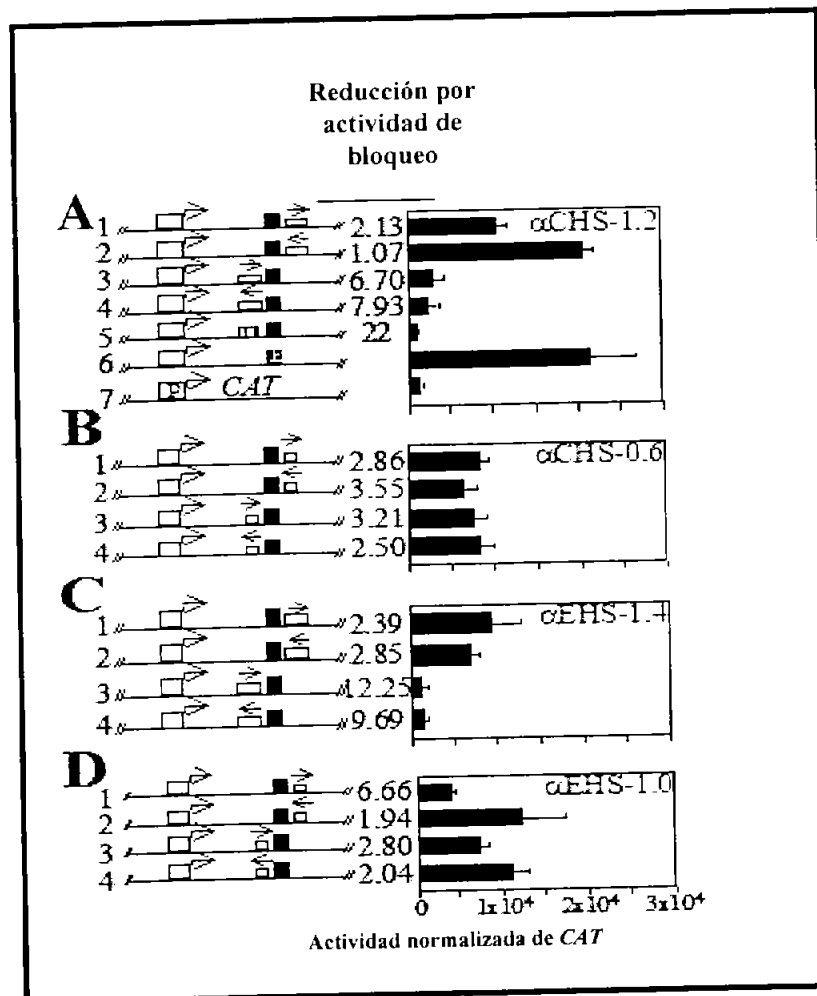
### **I. Análisis de la actividad de bloqueo de enhancer de los fragmentos que corresponden a los tres sitios de hipersensibilidad.**

Para analizar la actividad de bloqueo de enhancer de los sitios de hipersensibilidad, primeramente se realizó una reacción en cadena de la polimerasa para obtener los fragmentos correspondientes a los sitios de hipersensibilidad. Se amplificaron un fragmento de 1.4 kb correspondiente al sitio de hipersensibilidad eritroide 5' (nombrado  $\alpha$ EHS-1.4), un fragmento de 1.2 kb y uno de 600 bp correspondientes al sitio de hipersensibilidad constitutivo (nombrados  $\alpha$ CHS-1.2 y  $\alpha$ CHS-0.6), esto se hizo ya que derivado del análisis de secuencias nucleotídicas se encontró que el fragmento de 600 bp contenía la mayor cantidad de posibles sitios de unión al factor CTCF y otros factores de transcripción como GATA-1 (Fig. 19) y finalmente un fragmento de ADN de 1.0 kb correspondiente al sitio de hipersensibilidad eritroide 3' (nombrado  $\alpha$ EHS-1.0) fue también amplificado; cabe resaltar que en este fragmento no se encontraron posibles sitios de unión para el factor CTCF. Estos fragmentos de ADN fueron clonados en el plásmido pAcatE, el cuál lleva al promotor  $\beta^A$  y al enhancer  $\beta/\epsilon$  del dominio  $\beta$ -globina de pollo dirigiendo la expresión del gen reportero *CAT*, los fragmentos fueron seleccionados en ambas orientaciones en el sitio Sall, es decir "entre" el enhancer y el promotor, o bien "fuera" en el sitio HindIII, la razón de clonarlas en ambas orientaciones es saber si la propiedad de bloqueo de enhancer pudiese tener alguna direccionalidad preferente. La actividad de bloqueo de enhancer sólo deberá observarse cuando los fragmentos estén entre el enhancer y el promotor, el clonar las secuencias en el sitio HindIII ayuda a diferenciar claramente a los bloqueadores de enhancers de los silenciadores, ya que de tratarse de un silenciador, disminuiría la actividad del gen reportero también estando clonado

en este sitio ya que los silenciadores funcionan independientemente de su posición (Recillas-Targa *et al.* 1999; Fig. 20). Para evitar la comunicación entre el enhancer y el promotor a través del extremo 5' del promotor los plásmidos fueron linearizados en el sitio XmnI, interrumpiendo el gen de resistencia a la ampicilina y transfectados de manera transitoria.



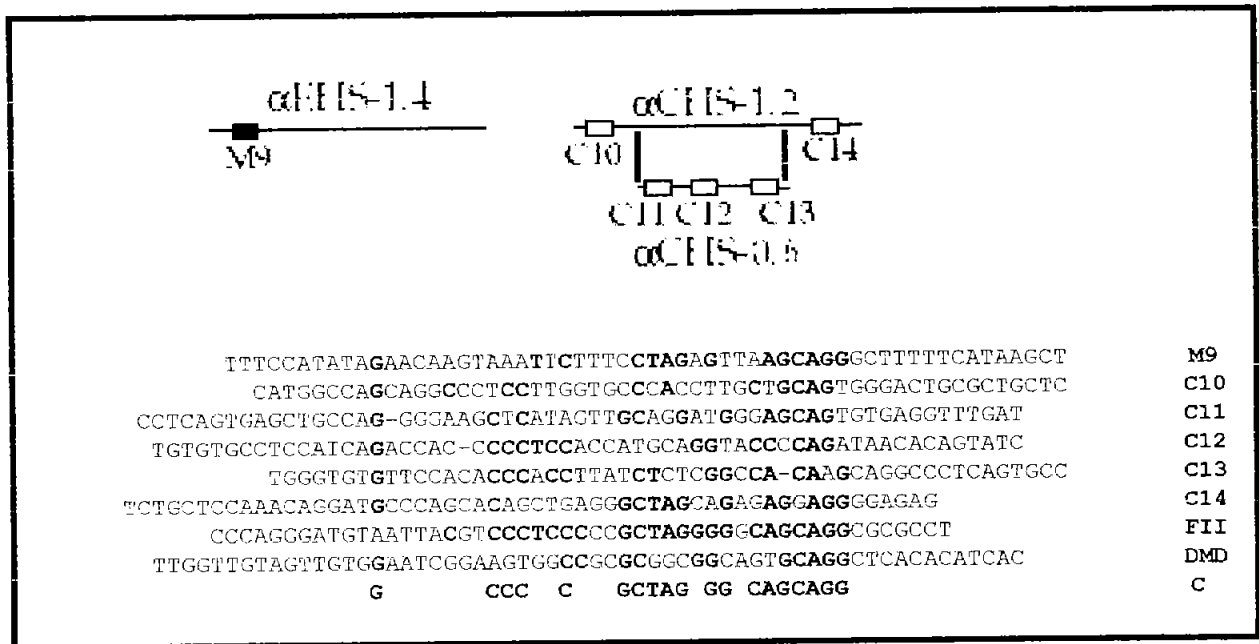
**Fig. 19. Posibles sitios de unión a la proteína CTCF definidos por un análisis de secuencia.** Se esquematiza el fragmento de 1.2 kb correspondiente al sitio constitutivo  $\alpha$ CHS-1.2, el fragmento de 600 bp correspondiente al  $\alpha$ CHS-0.6 y al fragmento de 1.4 kb correspondiente al fragmento  $\alpha$ EHS1.-4, se muestran algunos sitios de restricción. Los cuadros de rayas azules representan los posibles sitios de unión de la proteína CTCF definidos teóricamente por un análisis de las secuencias nucleotídicas de la región.



**Fig. 21 Ensayo de Bloqueo de Enhancer.** Resultados de las transfecciones transitorias realizadas con los plásmidos que contienen a los diferentes fragmentos que incluyen a los sitios de hipersensibilidad. A) Actividad de bloqueo del elemento αCHS1.2; B) Actividad de cloqueo de enhancer del elemento αCHS0.6; C) Actividad de bloqueo de enhancer del elemento αEHS1.4; D) Actividad de bloqueo de enhancer del elemento αEHS1.0. Las barras numeradas como 6 y 7 representan la actividad al añadir el elemento enhancer a la construcción y actividad basal del gen reportero (construcción con el promotor sólo), respectivamente. La barra 5 representa la actividad de bloqueo de enhancer del "insulator" β-globina de pollo. Las barras 1 y 2 representan al elemento clonado "fuera" en distintas orientaciones y las barras 3 y 4 representan al elemento clonado "entre" el enhancer y el promotor, respectivamente. La reducción por la actividad del bloqueo se presenta en la parte izquierda de cada barra, el error estandar se muestra en cada barra.



sitios de unión a este factor (Filippova *et al.* 1996; Bell *et al.* 1999; Lutz *et al.* 2000; Saitoh *et al.* 2000). Dada la relevancia del factor CTCF en la actividad de los “insulators” de vertebrados, decidimos hacer una búsqueda y análisis de las secuencias de los tres sitios de hipersensibilidad  $\alpha$ CHS-1.2,  $\alpha$ EHS-1.4 y  $\alpha$ EHS-1.0, con la finalidad de determinar la posible presencia de CTCF en la zona. Para hacer los alineamientos tratando de encontrar posibles secuencias de unión a este factor, decidimos usar como molde el sitio de unión al ADN del factor CTCF identificado en el “insulator” 5'cHS4 del dominio  $\beta$ -globina de pollo conocido como FII (Bell *et al.* 1999). En este “insulator” el sitio de unión a CTCF tiene cuatro bases (CTAG) las cuáles se ha demostrado mediante ensayos funcionales y de retardo que son esenciales para la unión del factor; adicionalmente utilizamos el sitio de unión de CTCF del locus improntado de ratón *Igf2/H19* (DMD; Hark *et al.* 2000; Bell y Felsenfeld 2000). Tomando en cuenta todo esto realizamos los alineamientos con distintos programas computacionales (DNAMAN, LALIGN). En la figura 22 se muestran los alineamientos obtenidos. Es importante señalar que los alineamientos tienen un porcentaje de identidad mayor del 60% y que en el fragmento  $\alpha$ EHS-1.0 no se encontraron alineamientos con una identidad mayor al 20%, lo que correlaciona con que ese fragmento no presenta una actividad de bloqueo de enhancer.



**Fig. 22 Búsqueda de secuencias de unión al factor CTCF.** M9 secuencia nucleotídica del sitio encontrado en el fragmento  $\alpha$ EHS-1.4, sitios C10-14 encontrados en el fragmento  $\alpha$ CHS-1.2, Sitio FII de unión de CTCF en el insulador del dominio  $\beta$ -globina de pollo, Sitio DMD de unión de CTCF en el locus improntado Igf2/H19 de ratón, y c posible consenso que se deriva de este alineamiento.

Las secuencias presentadas en la figura 22 son por lo tanto posibles sitios de unión a CTCF y no aseguran que la proteína realmente se una a estos sitios por lo que esta unión se comprobó mediante un ensayo de retardo en gel utilizando como sonda a estas secuencias. Aquí es importante señalar que se ha observado que el factor CTCF necesita al menos 60 pb para poder unirse a su secuencia en el ADN (Bell *et al.* 1999), por lo que los oligonucleótidos diseñados para el ensayo de retardo fueron de al menos 60 pares de bases (ver material y métodos). Inicialmente el ensayo de retardo fue realizado con extractos nucleares de células eritroides de embriones de pollo de 10 días ó RBC's (Fig. 23).

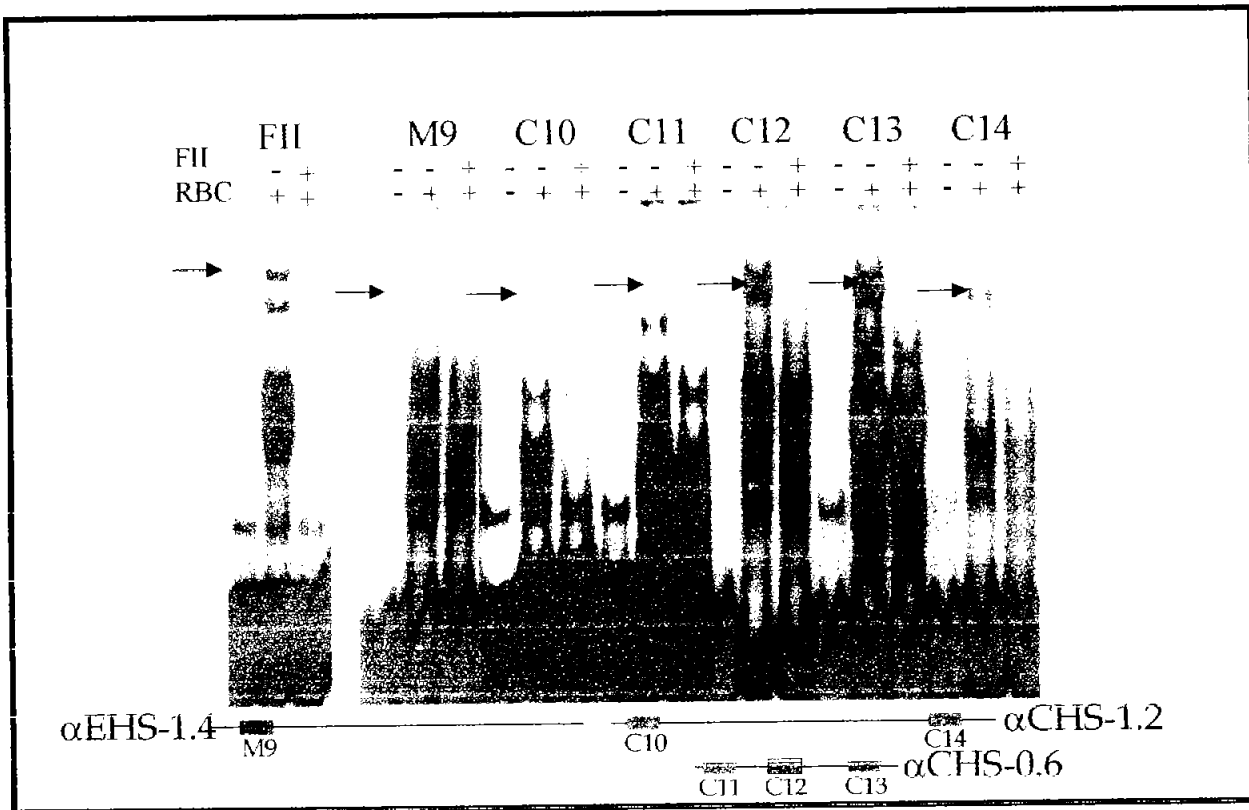
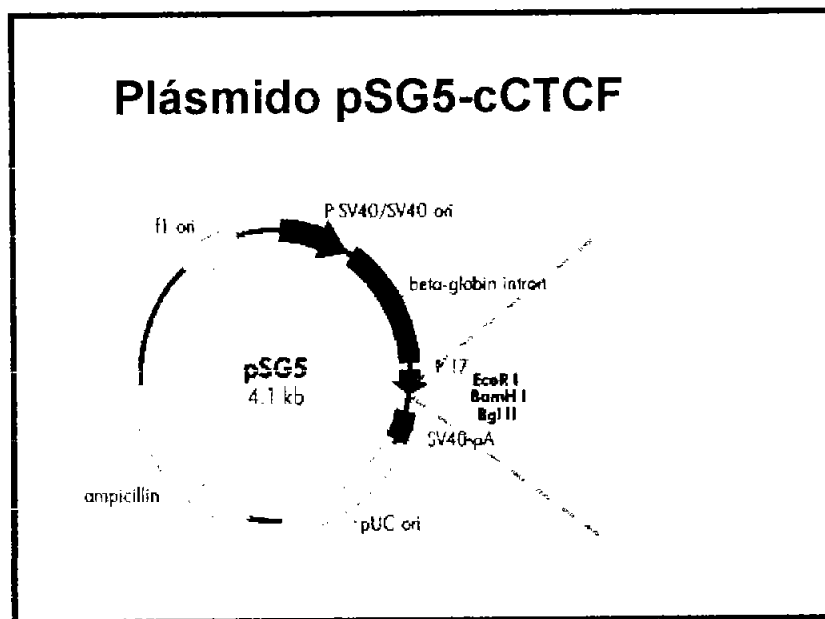


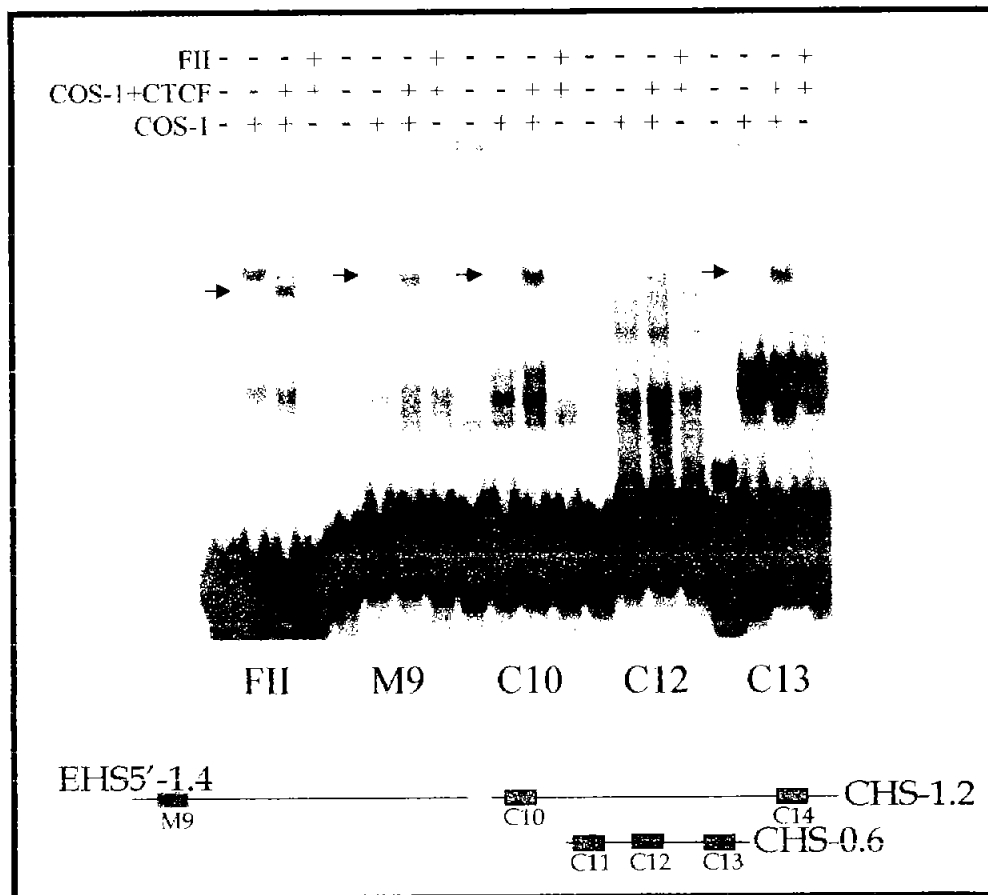
Fig. 23 Ensayo de retardo con extractos nucleares de células eritroides de embriones de pollo de 10 días (RBC). (FII) secuencia de unión al factor CTCF del “insulator” 5’cHS4  $\beta$ -globina de pollo. (-) sin extractos ó sin competidor, (+) con extractos o con 20 mM de competidor, las flechas indican los complejos retardados que son específicamente competidos con la sonda fría FII.

Este ensayo de retardo nos indicó que las secuencias identificadas eran capaces de unir a proteínas nucleares que eran competidas con un oligonucleótido frío (no marcado) correspondiente al FII del “insulator”  $\beta$ -globina de pollo el cuál reconoce con una alta afinidad a CTCF (Bell et al 1999).

Para corroborar que la unión fuera específica del factor CTCF, decidimos sobre-expresar el factor CTCF de pollo en un contexto celular no-eritroide. La Dra. Elena Klenova de la Universidad de Essex, nos proporcionó un vector que expresa el ADNc de CTCF de pollo (Fig. 24), el cual se transfectó en células COS-1, derivadas de riñón de mono verde, para sobre-expresar la proteína y obtener extractos enriquecidos con la proteína CTCF de pollo, para posteriormente preparar extractos nucleares de las células transfectadas, y se realizaron los ensayos de retardo en gel (Fig. 25; Filippova *et al.* 1996).



**Fig. 24 Plásmido pSG5-CTCF que contiene al ADNc de CTCF de pollo.** El plásmido lleva un promotor eucariote SV40 y el promotor T7, el cual fue útil al realizar la transcripción/traducción *in vitro*.



**Fig. 25** Ensayo de retardo en gel utilizando extractos de células COS-1 donde se sobre-expresó al ADNc de CTCF de pollo. Las flechas indican los complejos formados con las distintas sondas y que son competidos específicamente con el oligonucleótido FII frío.

Con este ensayo de retardo en donde se sobre-expresó la proteína CTCF, se confirmó la unión específica de CTCF *in vitro* a esas secuencias. A partir de este punto nos interesamos en estudiar más a fondo al sitio M9, este sitio está incluido dentro del fragmento  $\alpha$ EHS-1.4, el cual corresponde a la región con propiedades de unión a la matriz nuclear. Decidimos concentrarnos en este sitio por las siguientes 3 razones: 1) Es el consenso más cercano a FII, 2) tienen la mayor actividad de bloqueo de enhancer; y 3) tiene la propiedad de unión a la matriz nuclear. Se ha observado que las MAR's son

capaces de definir los límites de los dominios y algunas de ellas tienen propiedades de “insulators”. Esto es importante en el contexto de este proyecto ya que al parecer el dominio  $\alpha$ -globina de pollo no tienen límites bien definidos, y se encuentra en un contexto de la cromatina permanentemente abierto, por lo que definir el límite del dominio resulta importante.

Nos interesó saber si CTCF era capaz de unirse a la secuencia M9 por sí mismo o si necesitaba de proteínas accesorias, para lo cual se realizó una transcripción/traducción *in vitro* de la proteína utilizando al vector pSG5-CTCF donde el ADNc puede ser transcrito a partir del promotor T7 (Fig. 24). Este ensayo comprobó que al sitio M9 contenido dentro del fragmento  $\alpha$ EHS-1.4 se puede unir la proteína CTCF sin requerir de ninguna proteína adicional (Fig. 26). Además se demostró en un ensayo de retardo utilizando extractos de células COS-1 que sobre-expresan a CTCF que al mutagenizar el sitio CTAG el cual era indispensable para la unión de CTCF al “insulator” del dominio  $\beta$ -globina de pollo, se pierde la unión de la proteína (Fig. 27).

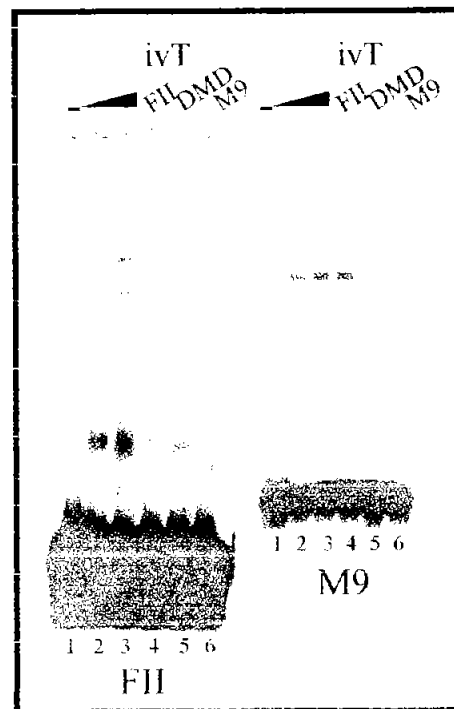


Fig 27. Ensayo de retardo con la secuencia M9 mutada y extractos de células COS-1. Del lado izquierdo se observa la secuencia del sitio M9 y el sitio en la secuencia donde se diseñó la mutación. Del lado derecho se observa el gel de retardo; Carril 1 sonda M9 marcada, Carril 2 sonda M9 con extractos de células COS-1, carril 3 sonda M9 con extractos de células COS-1 en donde se sobre-expresó la proteína CTCF de pollo (nótese el aumento del complejo), carril 4 competencia con FII frío, carriles 5-8 se observa la pérdida de la formación del complejo con la sonda mutada  $\Delta$ M9.

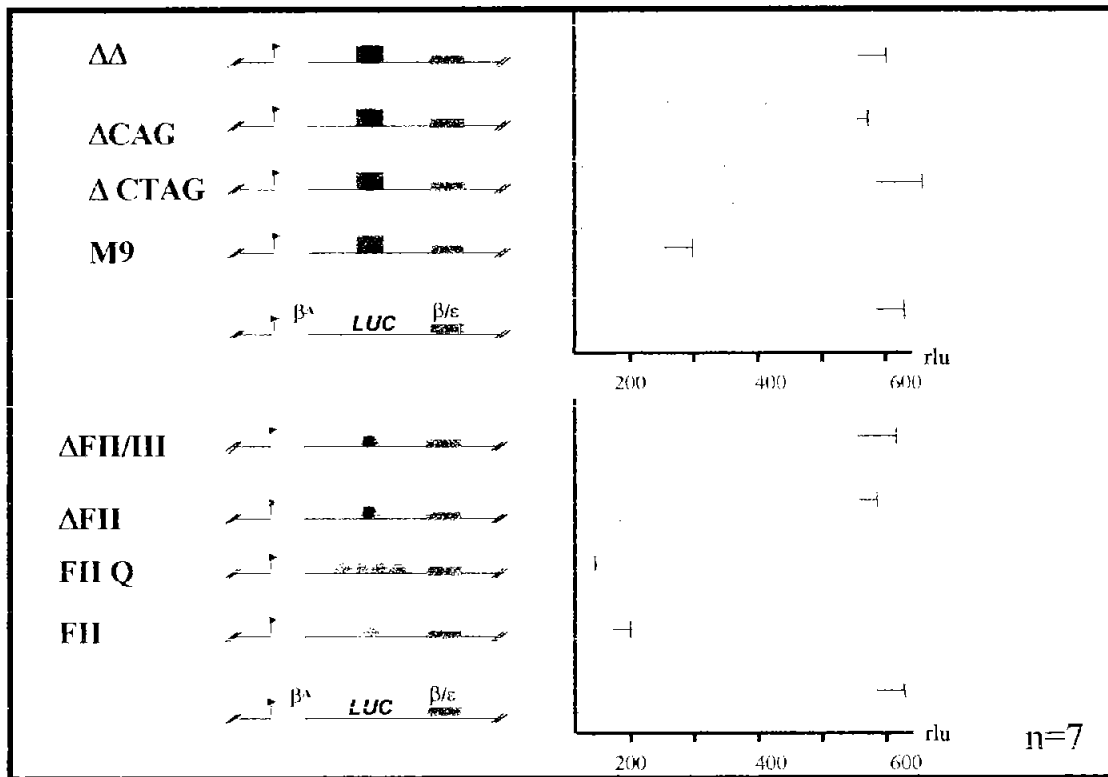
Una observación interesante se encuentra en el carril 3 de la figura 27, donde además del complejo de retardo señalado se ven dos complejos con menor movilidad. Este patrón fue observado en otros experimentos y nos hace pensar en la multimerización de CTCF o de su interacción con otros factores. Consideramos este punto muy interesante y en un futuro deseamos demostrar experimentalmente si estas posibles interacciones tienen consecuencias funcionales.

Derivado de los datos obtenidos decidimos iniciar la búsqueda del elemento mínimo que contuviese la actividad de bloqueo de enhancer en los elementos  $\alpha$ EHS-1.4 y  $\alpha$ CHS-1.2. Para confirmar funcionalmente la mutación hicimos el ensayo de bloqueo de enhancer utilizando a la secuencia M9 silvestre y las mutantes  $\Delta$ tag y  $\Delta$ cag ésta última se diseñó por que se observó que también era una secuencia que estaba conservada entre el sitio M9 y el FII, y la doble mutante  $\Delta$ cag  $\Delta$ tag; únicamente los 60 pares de bases correspondientes a los oligonucleótidos diseñados para los ensayos de retardo fueron clonados en el vector pALucE (ver material y métodos). La decisión de cambiar de gen reportero fué debido a que este método es más sensible y así podríamos observar mejor las diferencias en las actividades; una ventaja adicional fué que al tener a la luciferasa como gen reportero podríamos usar el sistema de "Dual luciferase" para normalizar las transfecciones (ver material y métodos). El ensayo de bloqueo fue realizado también con diferentes construcciones que incluyen al sitio FII del "insulator"  $\beta$ -globina de pollo. Se utilizó la secuencia FII silvestre (50 bp); FII/FIIQ (con cuatro copias del "insulator");  $\Delta$ FII la cuál tiene mutado las bases CTAG, esenciales

para la unión del factor CTCF; FII/FIII secuencia que contiene el sitio de unión a CTCF y finalmente  $\Delta$ FII/FIII secuencia que tiene la mutación que impide la unión del factor pero que tiene íntegra la secuencia del FIII. El usar estas construcciones nos permitió comparar la actividad de bloqueo de enhancer del sitio M9 con un "insulator" ya conocido y además como control para validar nuestro ensayo.

En la figura 28 se observa que el fragmento FII tiene una actividad de bloqueo de enhancer la cuál se pierde al utilizar las construcciones con las mutaciones en FII, lo que valida nuestro ensayo; la secuencia del sitio M9 tiene una actividad de bloqueo de enhancer por sí sola, aunque esta actividad es menor a la del "insulator"  $\beta$ -globina de pollo (FII). Además pudimos confirmar que la actividad de bloqueo de enhancer del fragmento M9 es dependiente del factor CTCF, ya que la mutante  $\Delta$ ctag pierde la actividad de bloqueo, así mismo, al mutar las bases cag ( construcción  $\Delta$ cag), se pierde la actividad de bloqueo de enhancer de manera similar a la construcción  $\Delta$ ctag, lo que indica que estamos mutando el mismo sitio de unión al factor, este dato fue corroborado al analizar la actividad de bloqueo de enhancer de la doble mutante (Fig. 28 construcción  $\Delta\Delta$ ), ya que observamos que el efecto de pérdida de bloqueo de enhancer no es aditivo. Estos resultados confirmaron que este sitio requiere únicamente de CTCF para la actividad de bloqueo de enhancer. Adicionalmente con este ensayo se demostró también que el sitio M9, conserva la actividad de bloqueo de enhancer aunque al parecer es menor que la del fragmento completo ( $\alpha$ EHS-1.4) y que constituye el elemento mínimo donde está contenida la actividad (Fig. 28).

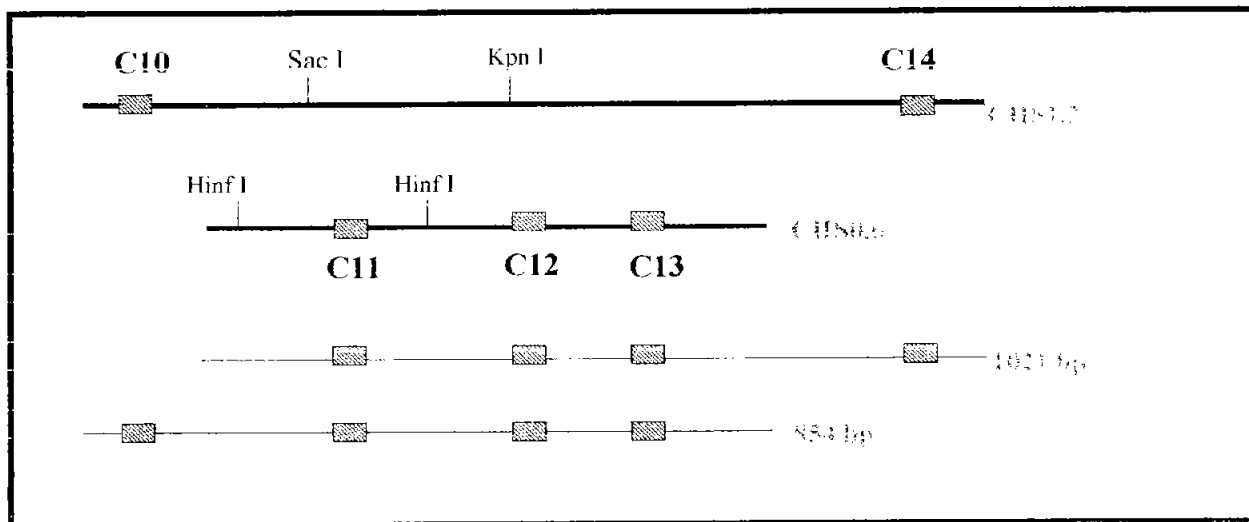




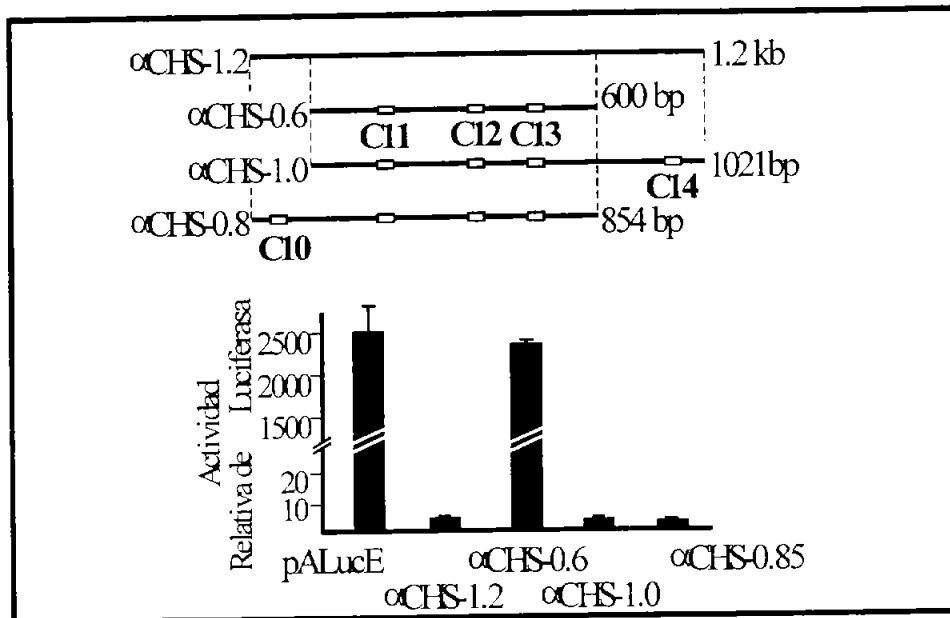
**Fig. 28 Ensayo de bloqueo de enhancer.** Transfección transitoria en cultivos primarios de eritrocitos de embriones de pollo de 10 días de gestación. En la gráfica superior se observa que el fragmento M9 conserva la capacidad de bloqueo de enhancer, mientras que las mutantes la pierden. En la gráfica inferior se muestran las actividades de bloqueo de enhancer de las diferentes construcciones control que incluyen al "insulator"  $\beta$ -globina de pollo.

Como ya se ha mencionado en este punto empezamos a buscar el elemento mínimo que contuviera la actividad de bloqueo de enhancer en el fragmento  $\alpha$ EHS-1.4, y en el fragmento  $\alpha$ CHS-1.2. Con el resultado de la figura 24 se confirmó que el fragmento mínimo que contiene la actividad de bloqueo de enhancer incluye los 60 pares de bases del sitio M9, aunque previos a este experimento se realizaron otras construcciones que dividían al fragmento  $\alpha$ EHS-1.4 y se ensayaron en transfecciones transitorias y se presentan en el capítulo de perspectivas.

En cuanto al fragmento de 1.2 kb, se analizaron dos sub-fragmentos que se muestran en la figura 29, éstos sub-fragmentos fueron clonados en el vector pALucE y se realizó el ensayo de bloqueo de enhancer (Fig. 30).



**Fig. 29** Búsqueda del elemento mínimo dentro del “insulator”  $\alpha$ CHS-1.2. Sub-división del fragmento  $\alpha$ CHS-1.2, los cuadros azules muestran los sitios de unión al factor CTCF identificados por homología de secuencias con el “insulator”  $\beta$ -globina de pollo y el sub-fragmento  $\alpha$ CHS-0.6 y los sub-fragmentos de 1021 bp (contiene los sitios C10 al C13 de unión al factor CTCF) y de 854 bp (contiene los sitios C11 al C14 de unión a CTCF).



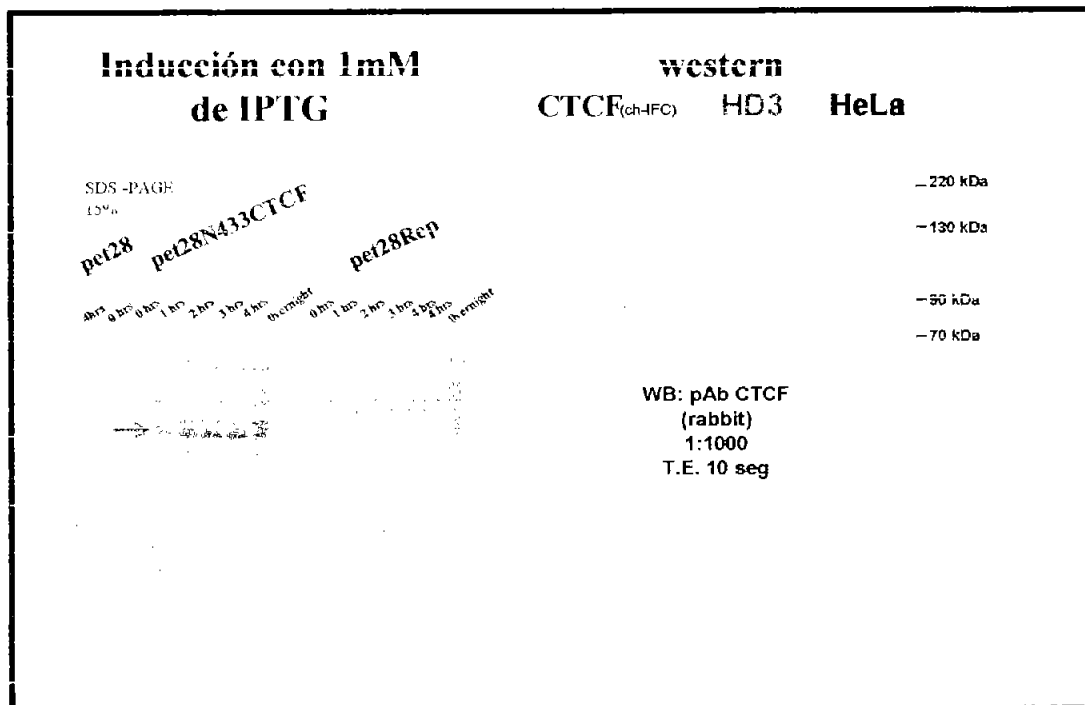
**Fig. 30** Ensayo de bloqueo de enhancer del fragmento  $\alpha$ CHS-1.2 y sus sub-fragmentos. Transfección transitoria en RBC donde se observa que la actividad de bloqueo de enhancer se pierde al ensayar el fragmento de 600 bp ( $\alpha$ CHS-0.6), lo que corrobora el ensayo de bloqueo de enhancer en donde se utilizó como gen reportero a CAT. Nótese que los demás sub-fragmentos conservan la actividad significativa de bloqueo de enhancer.

Con este ensayo corroboramos que al deletar las secuencias que rodean al fragmento de 600 bp se pierde la actividad de bloqueo de enhancer, y que cualquiera de los dos sub-fragmentos que incluye las secuencias C10 ó C14 conservan la actividad de bloqueo de enhancer. Estos datos sugieren que se requiere del fragmento completo para conservar la actividad de bloqueo de enhancer.



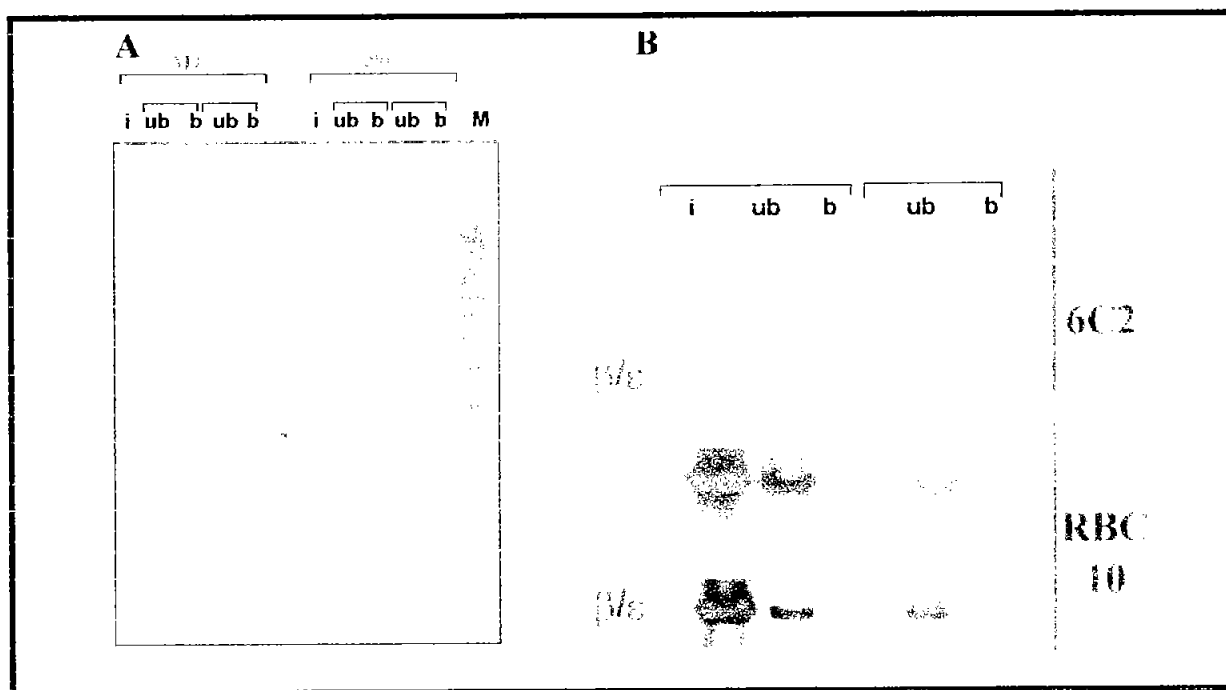
**Fig. 31 Generación del anticuerpo anti-CTCF de pollo.** Se muestra la región terminal amino de CTCF de pollo no conservada entre pollo, ratón y humano. Esta región se clonó en el plásmido pET-28 y se obtuvo el vector pet28(b)N433 CTCF.

Este plásmido fue utilizado para sobre-producir la proteína en bacteria ya que tiene un promotor inducible por IPTG, el péptido además tiene una secuencia de histidinas en el extremo terminal amino el cuál sirvió para purificar la proteína mediante cromatografía a través de una columna de níquel (Fig. 31 y 32). Posteriormente se inmunizaron conejos para obtener el anticuerpo (ver material y métodos) y en colaboración con el Lic. en IIB Héctor Rincón Arano se realizó el Western –blot para comprobar que el anticuerpo reconociera específicamente a la proteína CTCF de pollo, y que no tuviera reacción cruzada con la proteína de humano, además de estimar el título del anticuerpo (Fig. 32).



**Fig. 32 Caracterización por Western-blot del anticuerpo anti-CTCF específico de pollo.** En la parte izquierda se muestra el gel donde se indujo la expresión del péptido (flecha roja), en la parte superior derecha se muestra el Western Blot, se utilizó el péptido CTCF purificado como control positivo, una línea celular eritroide de pollo (HD3) y una línea celular de humano HeLa, nótese que el anticuerpo únicamente reconoce la proteína CTCF de pollo (aprox. 130 kDa)

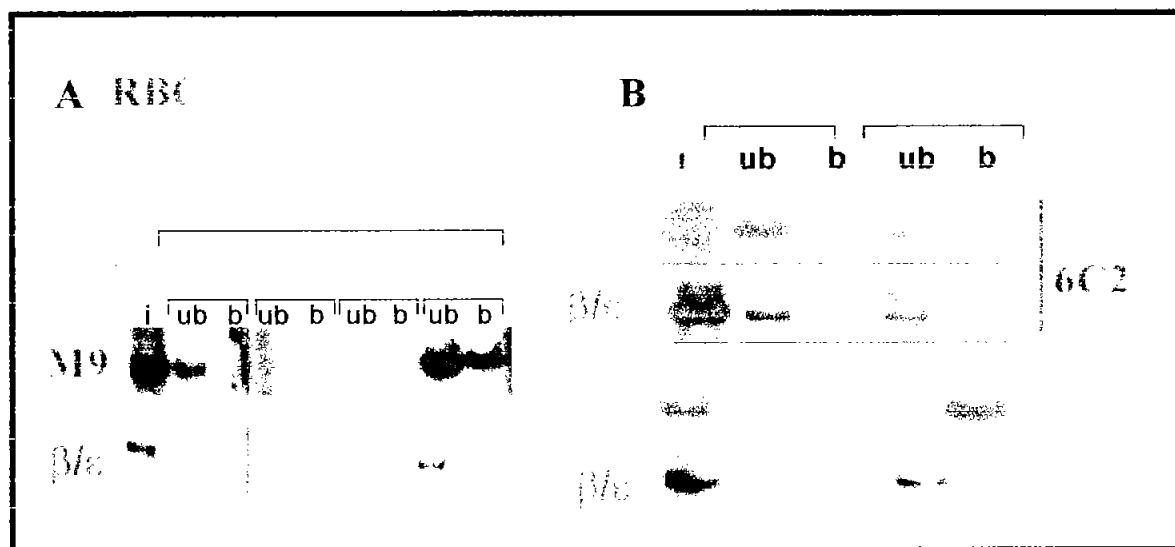
Con este anticuerpo hicimos los ensayos de inmunoprecipitación de la cromatina. Este ensayo permite determinar si una proteína está unida *in vivo* a un sitio en particular (ver material y métodos). Se diseñaron oligonucleótidos que amplifican una zona de 120 bp que incluye al sitio M9 y como control negativo se amplificó una región del enhancer del dominio  $\beta$ -globina de pollo de 301 bp, en cuya región existen varios factores transcripcionales pero no está la proteína CTCF. Además se añadió un control de inmunoprecipitación, es decir un anticuerpo irrelevante en este caso anti-IgG ( $\alpha$ -X). Inicialmente estos ensayos fueron realizados con un anticuerpo comercial CTCFN17 (ver material y métodos; Fig. 33).



**Fig. 33 Ensayo de Inmunoprecipitación de la cromatina (ChIP).** A) (i) "input"; (ub) fracción no-unida; (b) fracción unida.  $\beta/s$  enhancer  $\beta$ -globina de pollo; (anti-X) anticuerpo irrelevante; ChIP en 6C2 pre-eritroblastos de pollo; RBC 10 células eritroides de embriones de pollo de 10 días de gestación. Gel de acrilamida al 5% de la reacción de PCR se muestra un gel típico en donde se ve únicamente la banda específica amplificada en la fracción unida al inmunoprecipitar con el anticuerpo comercial contra CTCF que tiene reacción cruzada con la proteína anti-CTCF de pollo. B) Ensayo de ChIP utilizando al anticuerpo comercial anti-CTCF (sc).

Los resultados obtenidos mediante el ensayo de ChIP nos indicaron que la proteína CTCF se encontraba asociada *in vivo* a la secuencia M9 en diferentes estadios de diferenciación eritroides de pollo.

Una vez que obtuvimos el anticuerpo anti-CTCF de pollo estos ensayos se repitieron en distintos estadios de la diferenciación eritroide y en una línea celular linfóide DT40 (Fig. 34).



**Fig. 34** Ensayo de inmunoprecipitación de la cromatina utilizando el anticuerpo específico contra CTCF de pollo. DT40 línea celular linfoide (misma nomenclatura que en la figura 33). **A)** Inmunoprecipitación realizada con distintas concentraciones del anticuerpo anti-CTCF de pollo, se observa que a mayor concentración de anticuerpo el enriquecimiento de la banda es mayor, sin embargo al inmunoprecipitar con suero de conejo pre-inmune no se ve un enriquecimiento en la fracción unida (pre-im). **B)** ChIP's en la línea celular 6C2 y en la línea linfoide DT40, nótese la presencia de CTCF en la línea celular no-eritroide.

Con estos datos concluimos que CTCF está presente en el sitio M9 en todos los estadios de diferenciación eritroide y además en una línea celular linfoide, lo que parece indicar que CTCF juega *in vivo* un papel importante para la función del nuevo "insulator"  $\alpha$ EHS-1.4 en la organización del dominio. Esto es interesante ya que recordemos que el sitio M9 está dentro del sitio de



hipersensibilidad eritroide  $\alpha$ EHS-1.4, lo que indica que CTCF podría estar jugando un papel estructural en la región de unión a la matriz nuclear adicional a su actividad de bloqueo de enhancer (ver discusión).

## DISCUSIÓN

En este estudio se identificaron dos nuevos "insulators" en el costado 5' del dominio  $\alpha$ -globina de pollo, cuya actividad de bloqueo de enhancer depende del factor multifuncional CTCF. Esta parece ser una característica única de este dominio, como se ha comentado anteriormente, el dominio  $\alpha$ -globina de pollo está localizado en una conformación cromatínica permanentemente "abierta" (de la cromatina), por lo que su regulación debe ser diferente a la del dominio  $\beta$ -globina de pollo (Razin *et al.* 2003). El haber encontrado estas dos actividades en un dominio constitutivamente abierto resulta relevante, ya que a diferencia del dominio  $\beta$ -globina de pollo en el que el "insulator" 5'HS4 define una transición entre una región de eucromatina y una de heterocromatina (16 kb; Burgess-Beusse *et al.* 2002). Desconocemos si en el dominio  $\alpha$ -globina de pollo los "insulators" identificados podrían estar definiendo una zona de transición en la estructura de la cromatina del dominio.

El caso del "insulator"  $\beta$ -globina de pollo, 5'HS4, es el de un "insulator" clásico el cual aparentemente necesita tener las dos propiedades funcionales, la actividad de bloqueo de enhancer y la de protección contra el efecto de posición. La actividad de protección contra el efecto de posición del 5'HS4, es independiente del factor CTCF, esto indica que la actividad está representada por otro factor u otros factores, y se ha propuesto que alguno de los factores eritroides o no-eritroides que se unen dentro del "core" de 250 bp del "insulator"  $\beta$ -globina de pollo sean los responsables de esta actividad (Recillas-Targa *et al.* 2002).

Cabe señalar que no todos los "insulators" tienen ambas propiedades funcionales, por ejemplo el "insulator" 3'HS del dominio  $\beta$ -globina de pollo únicamente tiene la actividad de bloqueo de enhancer y carece de la propiedad de protección contra el efecto de posición. Se ha propuesto que esto ocurre por que el "insulator" 3'HS está en un entorno cromatínico diferente y que probablemente únicamente requiera la actividad de bloqueo de enhancer en particular por que en el dominio adyacente se localiza un gen cuyo patrón de expresión es específico de cerebro y no se presenta una región de heterocromatina similar a la identificada en el costado 5' del dominio (Fig. 15; Saitoh *et al.* 2000;

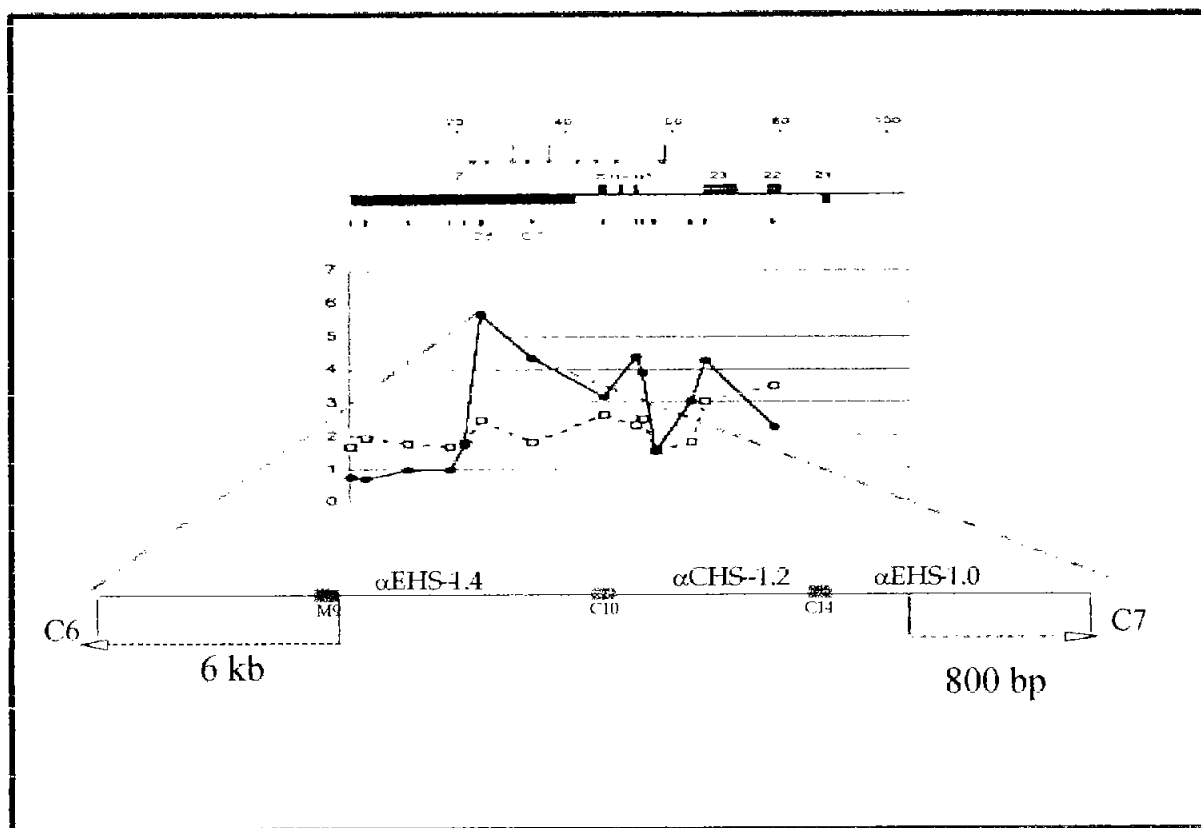
Burgess-Beusse *et al.* 2002). Lo que indican estos datos es que no todos los “insulators” funcionan de la misma manera y que probablemente el entorno cromatínico y la organización genómica de la región tengan una influencia directa sobre sus funciones.

En el contexto de la presente investigación, un dato relevante fue encontrar la presencia *in vivo* de la proteína CTCF en el “insulator” contenido en el sitio de hipersensibilidad eritroide específico  $\alpha$ EHS-1.4 en células eritroides y en la línea celular linfoide DT40. Existen reportes de otros “insulators” que se encuentran en sitios de hipersensibilidad tejido específicos (Farrell *et al.* 2002). El hecho de que la proteína CTCF esté además presente en una línea celular linfoide indica que probablemente no sea esta la proteína responsable de la formación del sitio de hipersensibilidad y que estén involucradas otras proteínas eritroide-específicas en esta función. Una interpretación alternativa contempla la posibilidad de que la unión de CTCF al “insulator”  $\alpha$ EHS-1.4 en células DT40, signifique que no nos encontramos en presencia de un sitio de hipersensibilidad eritroide específico, cómo fue originalmente definido, sino por el contrario que se trate de un sitio de hipersensibilidad constitutivo. Lo anterior apoyaría aún más la presencia de elementos tipo “insulator” en la zona. Para el caso de este “insulator” se ha encontrado por análisis de secuencia la presencia de otros factores, en particular un sitio consenso para la unión del factor transcripcional Oct-1 (datos no presentados). La posible participación del Oct-1 parece relevante dado que se sabe que este factor ayuda a la actividad de protección contra el efecto de posición de los LCR’s y recientemente ha surgido como un factor importante en el diseño de vectores para el tratamiento de las talasemias (Bharadwaj *et al.* 2003). Así pues, es posible que este u otros factores sean los responsables y/o contribuyan a la formación de los sitios de hipersensibilidad.

El sitio de unión a CTCF en el  $\alpha$ EHS-1.4 llamado M9 está dentro de una zona con la capacidad de unirse a la matriz nuclear (Razin *et al.* 1999). Las MAR’s son secuencias ricas en A’s y T’s, pueden estar delimitando dominios y se ha propuesto que contribuyen a la formación de las asas o “loops” cromatínicos. Se han encontrado otras MAR’s con actividades de “insulators” un ejemplo de esto es la MAR 3’ de la apolipoproteína de pollo, la cuál está delimitando al dominio y además tiene la capacidad de bloqueo de enhancer dependiente del factor CTCF (Antes *et al.* 2001). Recientemente se ha descrito que la proteína CTCF es capaz de unirse a la matriz nuclear (Dunn *et al.* 2003), el hecho de que hayamos encontrado *in vivo* a esta proteína en este sitio y que además correlacione con la actividad de MAR de la zona, apoya el argumento de que quizás CTCF en este dominio pueda tener un papel

adicional a su actividad de bloqueo de enhancer como por ejemplo un papel estructural que contribuya a la organización del dominio.

Hasta ahora y con la evidencia presentada no podemos decir con certeza que estos "insulators" representan el límite 5' del dominio. Anguita y colaboradores han hecho recientemente un análisis comparativo de los niveles de acetilación de la lisina 5 de la histona H4 y de la lisina 14 de la histona H3 del costado 5' del dominio  $\alpha$ -globina de pollo, humano, ratón y pez globo (Anguita *et al.* 2001; Fig. 35 ver "primers" C6 y C7) Ellos definieron una zona de alta acetilación de las histonas, que va desde 6 kb río arriba del  $\alpha$ EHS-1.4, hasta 0.8 kb antes del término de la secuencia  $\alpha$ EHS-1.0. Interesantemente ellos encuentran una transición en los niveles de acetilación de la lisina-5 de la histona H4, y de la lisina 14 de la histona H3, en la zona del "primer" C6 (Fig. 35).



**Fig. 35 Esquema que sitúa los niveles de acetilación de las histonas en una región genómica de 100 kb que incluye al dominio  $\alpha$ -globina de pollo.** La región de 6.6 kb objeto de nuestro estudio se encuentra entre los oligonucleótidos marcados como C6 y C7, que incluye una zona de alta acetilación de histonas. En la gráfica la línea continua representa la acetilación de la lisina 5 de la histona H4, y la línea punteada la acetilación de la lisina 16 de la histona H4 (tomado de Anguita *et al.* 2001).

Al tomar en cuenta estos datos, podemos concluir que los tres sitios de hipersensibilidad están localizados en una zona de alta acetilación de histonas y cerca del pico de transición de la acetilación. Estos “insulators” representan buenos candidatos para determinar o formar parte del límite 5’ del dominio, sin embargo no se puede descartar la posible presencia de otros “insulators” río arriba del  $\alpha$ EHS-1.4, co-localizando con el punto más alto de acetilación (C6).

Otra característica particular del dominio  $\alpha$ -globina de pollo es la presencia de transcritos intergénicos, los cuáles se transcriben en sentido opuesto a los genes globina y los cuáles se ha propuesto que ayudan a mantener abierto el dominio; por otra parte, el exón 4 del gen constitutivo -14 se encuentra sobrelapando al sitio  $\alpha$ EHS-1.4 (Vyas *et al* 1995), proponemos que quizá estos “insulators” estén regulando las señales que dan lugar a los transcritos intergénicos o bien que estén regulando la expresión del gen constitutivo (Gribnau *et al* 2000; Sjakste *et al* 2000). Siendo que este es un dominio constitutivamente abierto, puede que se necesiten varios bloqueadores de enhancer que ejerzan sus funciones de manera divergente para controlar elementos de regulación eritroides y no-eritroides.

Con respecto al “insulator” encontrado en el sitio constitutivo,  $\alpha$ CHS-1.2, se comprobó la unión de la proteína CTCF *in vitro* y se determinó la actividad de bloqueo de enhancer de este elemento, interesantemente, al probar un sub-fragmento contenido dentro del  $\alpha$ CHS-1.2, el  $\alpha$ CHS-0.6, la actividad de bloqueo de enhancer se pierde, esto indica que el contexto es importante. también en esta zona se encontró por análisis de secuencias la presencia de posibles sitios de unión a factores eritroide-específicos y ubicuos como Sp1, lo que podría sugerir la presencia de un bloqueador de enhancer que pudiera estar regulado por los factores eritroide específicos, y consecuentemente por el estadio de diferenciación celular, esto ha sido sugerido anteriormente (Farrell *et al.* 2002; Lutz *et al.* 2003); y representaría un avance para el entendimiento del mecanismo de acción de estos elementos.

En conclusión, hemos descrito que los sitios de hipersensibilidad  $\alpha$ EHS-1.4 y  $\alpha$ CHS-1.2 tienen una actividad de bloqueo de enhancer, y que esta actividad depende del factor CTCF. La relevancia de este trabajo reside en el hecho de haber encontrado dos “insulators” independientes en el extremo 5’ de

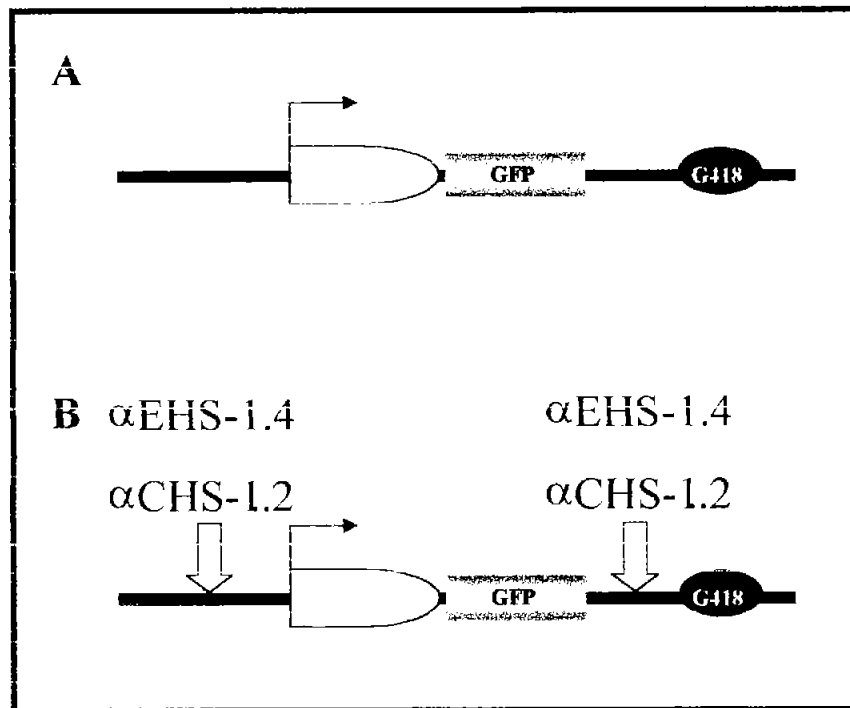
un dominio cromatínico “abierto” constitutivamente, este es el primer reporte en donde se ha descrito la presencia de dos actividades de bloqueo de enhancer adyacentes; el haber comprobado la presencia de la proteína CTCF *in vivo* en el sitio M9 tanto en células eritroides como no eritroides, y de que su unión no dependa de la formación de un sitio de hipersensibilidad, llevan a proponer un modelo en el que probablemente los factores de regulación eritroide específicos puedan estar regulando a estos “insulators”.

## PERSPECTIVAS

El descubrimiento de estos dos nuevos “insulators” abre un campo para el estudio de la regulación del dominio  $\alpha$ -globina de pollo, sobre la contribución de estos “insulators” a la organización y transcripción de los genes de este dominio, a continuación se presentan algunos experimentos y propuestas que pueden contribuir al desarrollo de futuras investigaciones en el laboratorio.

**Determinar si los “insulators”  $\alpha$ EHS-1.4 y  $\alpha$ CHS-1.2 tienen la capacidad de proteger a un transgen contra los efectos de posición:**

En un primer intento por determinar si estos “insulators” tienen la propiedad de protección contra el efecto de posición, se han clonado los sitios  $\alpha$ EHS-1.4 y  $\alpha$ CHS-1.2 en el vector pG $\alpha$ D que lleva como gen reportero a la proteína verde fluorescente (GFP) y el promotor  $\alpha^D$  (Fig. 36).

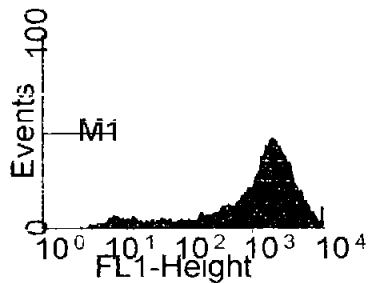
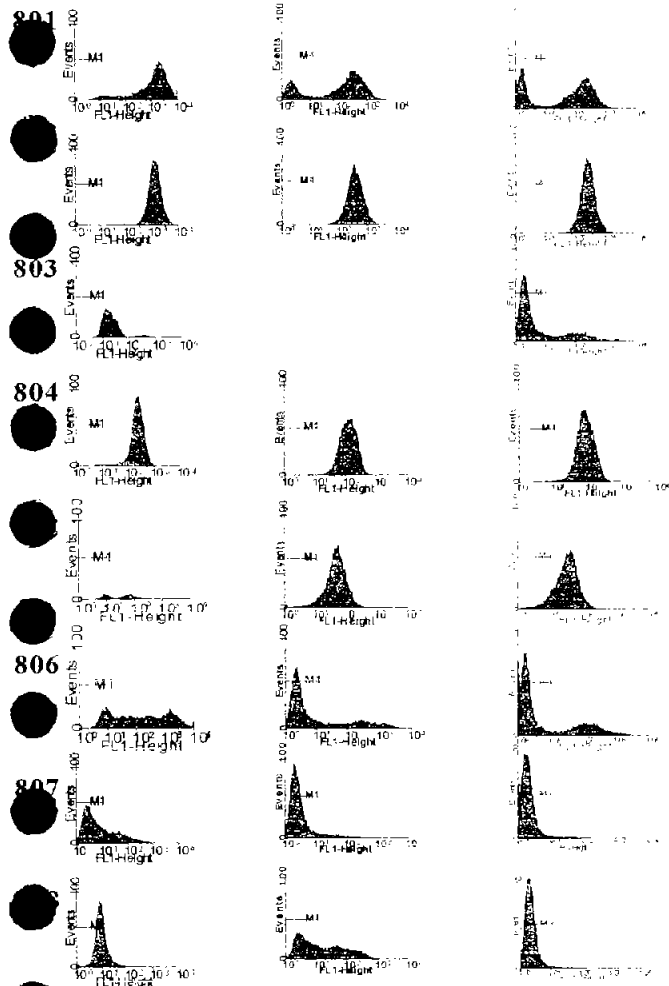
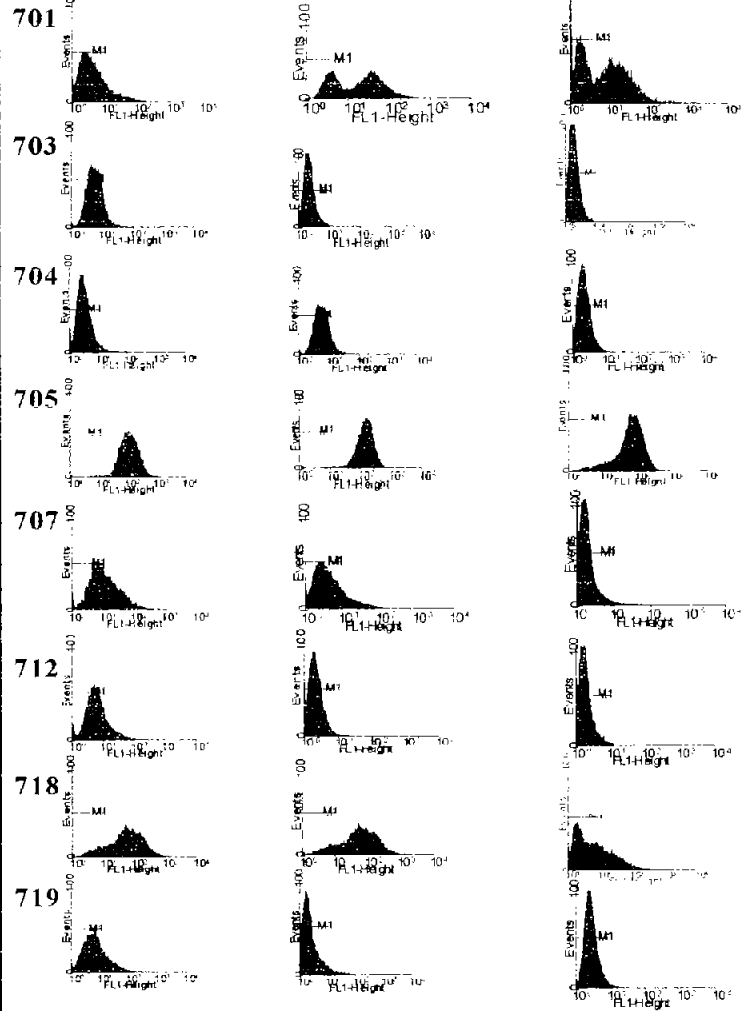


**Fig. 36 Vectores utilizados para evaluar la propiedad de protección contra el efecto de posición.** A) Vector pGαD construido por Héctor Rincón Arano; B) Las flechas muestran la posición en la que fueron clonados los insulators.

Estos vectores se transformaron en eritroblastos de pollo (HD3), en donde el promotor  $\alpha^D$  es activo. Posteriormente se seleccionaron clones estables en las cuáles el transgen se está expresando, estas clones se mantienen en cultivo durante 100 días y se analiza la actividad del gen reportero (fluorescencia) mediante citometría de flujo siguiendo la emisión de fluorescencia generada por la proteína verde fluorescente (GFP). Anteriormente se ha descrito que el silenciamiento de genes reporteros está dado por la desacetilación de histonas y por la metilación del ADN (Pikaart *et al.* 1998; Recillas-Targa *et al.* 2002).

Si existe una actividad de protección contra el efecto de posición, la actividad del transgen se mantendrá a lo largo del tiempo, o por más tiempo al compararla con la actividad del plásmido control (pGαD) cuyo transgen no se encuentra protegido por las secuencias “insulator”.

Datos preliminares muestran que el "insulator"  $\alpha$ CHS-1.2 es incapaz de proteger a un transgen de los efectos de posición (Fig. 37). De hecho se observa que en estas clonas la expresión del transgen decae más rápido que en las células control, estos datos podrían indicar que hay un elemento silenciador en la secuencia  $\alpha$ CHS-1.2 que está actuando para silenciar más rápidamente la expresión del gen reportero en estas clonas. En estas construcciones solamente se tiene una copia del "insulator" a cada lado del promotor y del transgen, datos en la literatura muestran que dos copias en tandem a cada lado del transgen confieren una mayor protección contra el efecto de posición (Pikaart *et al.* 1998). Se propone que se clonen dos copias de estos elementos a cada lado del transgen y se evalúe la protección contra el efecto de posición con estas nuevas construcciones.

**A****B****pGα<sup>D2</sup>**clona **Día cero****día 20****día 40****C****pGα<sup>D</sup>CHS1.2**clona **Día cero****día 20****día 40**



**Fig. 37 Expresión de la GFP evaluada por citometría de flujo o FACS.** A) Gráfica representativa: En el eje de las ordenadas (X) se muestra la intensidad de la fluorescencia, que es una medida directa de la expresión del transgen GFP en escala logarítmica. En el eje de las abscisas (Y) se muestra el número de células (se contaron 10000 células por clona); la zona M1 delimitada por una línea horizontal, representa la población de células que carecen de fluorescencia y en las que por lo tanto, el transgen está silenciado; B) Clonas representativas de las clonas seleccionadas a las que se le integró el plásmido control (sin “insulators”); C) En el panel derecho se observan clonas representativas donde el transgen se encuentra enmarcado por una copia del “insulator”  $\alpha$ CHS-1.2.

**Determinar si estos “insulators” son capaces de interactuar físicamente con otras regiones del dominio  $\alpha$ -globina de pollo formando asas cromatínicas.**

Se sabe que los dominios cromatínicos no son entidades estáticas. Recientemente han surgido evidencias de la formación de “loops” o asas entre “insulators” (Blanton *et al.* 2003) y entre LCR's y promotores de los genes globina (Tolhuis *et al.* 2002). La demostración de estas interacciones se han hecho utilizando la técnica de captura de la conformación cromosomal descrita anteriormente (ver capítulo introductorio sobre LCR's en esta tesis). Los resultados obtenidos mediante esta técnica, apoyan el modelo de “looping” para la actividad del LCR, y apoyan el modelo de la formación de un “loop” entre “insulators” del extremo 5' y 3' (mediado por la interacción entre dos o más proteínas) del dominio para delimitarlo.

En experimentos realizados en el contexto de esta tesis hemos encontrado una asociación de los factores YY1 y GATA-1 a la zona de 120 bp que contiene a la secuencia M9 del “insulator”  $\alpha$ EHS-1.4, es decir al sitio de unión para el factor CTCF (Fig. 38).

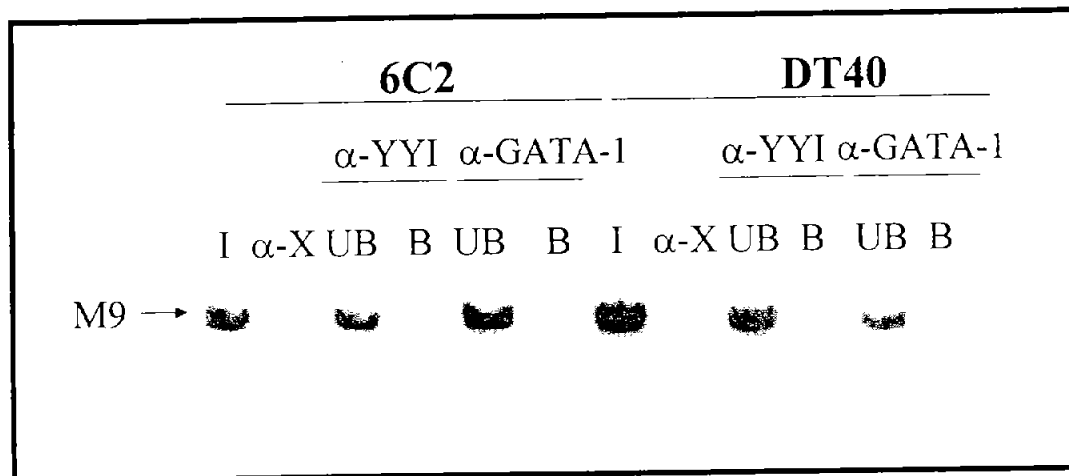


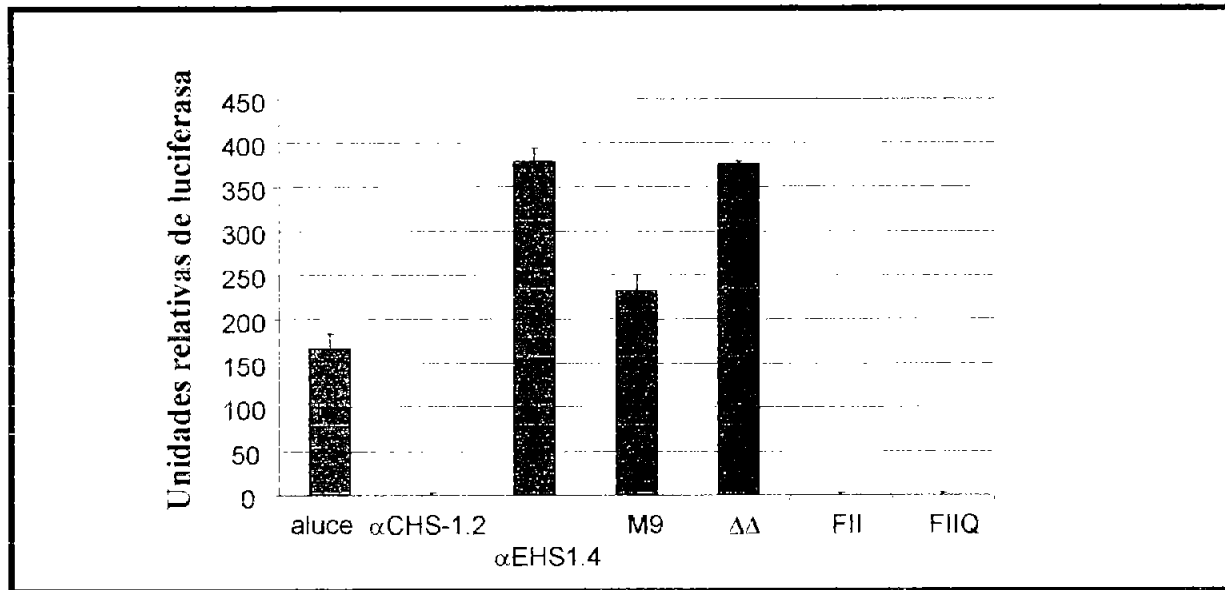
Fig. 38 Ensayo de Inmunoprecipitación de la cromatina de los factores YY1 y GATA-1 en el fragmento de 120 bp que contiene el sitio de unión al factor transcripcional CTCF (secuencia M9). I: "Input"; B: fracción unida; UB:fracción no unida;  $\alpha$ -X: anticuerpo irrelevante; 6C2: pre-eritroblastos de pollo; DT40: línea celular linfoide.

Interesantemente observamos un enriquecimiento de las bandas únicamente en las fracciones unidas de las células 6C2 y no así en células no-eritroides (línea linfoide DT40). Al hacer un análisis de secuencia se encontró que en los 120 bp no existen sitios de unión evidentes para estos dos factores. (GATA-1 y YY1) Esto puede sugerir varias cosas: Reportes no publicados del grupo de Lobanenkov dicen que la proteína CTCF es capaz de interactuar con el factor YY1, a su vez este factor es capaz de interactuar con el factor eritroide GATA-1 (ver anexo 1). Pudiera ser que la proteína CTCF esté interactuando con estos factores y a través de ellos recluta actividades de remodelaje de la cromatina determinadas para actuar en un momento determinado de la diferenciación eritroide.

En este sentido consideramos que se tendrían que hacer experimentos que probaran la interacción entre estas dos proteínas (YY1 y CTCF) haciendo por ejemplo fusiones GST con los distintos dominios de las proteínas, y comprobar su interacción *in vitro* mediante co-

inmunoprecipitaciones. Para lo cual se tendrían que generar anticuerpos específicos contra los factores YY1 y GATA-1 de pollo (ver anexo 1). Otra posibilidad es que estos elementos estén mediando la formación de un asa cromatínica entre el “insulator”  $\alpha$ EHS-1.4 y el enhancer localizado en el costado 3' del dominio. Lo anterior está apoyado por los resultados experimentales de nuestro laboratorio donde hemos demostrado que a este enhancer se unen los factores YY1, GATA-1 y Sp1, y que estos factores están solamente presentes en líneas celulares eritroides (Rincón-Arano *et al. en preparación*). Un dato adicional tiene que ver con reportes en la literatura indican que la proteína Oct-1 es capaz de interaccionar con la proteína Sp1 (Gunther *et al* 2000). Datos preliminares sugieren fuertemente la presencia de un sitio de unión al factor Oct-1 muy cercano al sitio de unión a CTCF. Sería interesante evaluar mediante la técnica de conformación cromosomal, si se está llevando a cabo un contacto entre estas dos regiones, cuáles son las proteínas mediadoras de estas interacciones, si este “loop” o asa se lleva a cabo únicamente en las líneas eritroides y determinar en que estadios de diferenciación está ocurriendo.

Otro dato preliminar es que el “insulator”  $\alpha$ EHS-1.4 pierde la propiedad de bloqueo de enhancer en células 6C2 (pre-eritroblastos; Fig 39).; mientras que el “insulator”  $\alpha$ CHS-1.2 continúa teniendo una actividad de bloqueo de enhancer en esta línea celular (Fig 39). Este dato apoya la hipótesis de que el “insulator”  $\alpha$ EHS-1.4 es un “insulator condicional” y que pudiese ser inactivado por la unión de otros factores eritroide-específicos (cómo GATA-1 y YY1). Nosotros demostramos en esta tesis que el factor CTCF es capaz de interaccionar *in vivo* con su sitio de unión en el “insulator”  $\alpha$ EHS-1.4 en las células 6C2 (Fig. 39), la presencia *in vivo* de CTCF en ese sitio y la pérdida de la capacidad de bloqueo de enhancer en estas células apoya la hipótesis de que CTCF tiene una actividad alterna a la de bloqueo de enhancer, y podría estar actuando como un factor estructural al contribuir a la unión con la matriz nuclear.



**Fig. 39 Ensayo de Bloqueo de enhancer en células 6C2.** La gráfica muestra la actividad del gen reportero luciferasa con las distintas construcciones que llevan a los “insulators” identificados en este trabajo  $\alpha$ CHS-1.2,  $\alpha$ EHS-1.4, el elemento mínimo M9, la doble mutante  $\Delta\Delta$ , o bien el “insulator” del dominio  $\beta$ -globina de pollo en una copia FII o en cuatro copias FIIQ. Obsérvese que la actividad de bloqueo de enhancer del “insulator”  $\alpha$ EHS1.4, se pierde en este estadio de diferenciación eritroide.

El abordar estas preguntas permitirá proponer un modelo de acción de los “insulators” y proporcionará información sobre la contribución de los factores eritroides para regular los genes del dominio y las actividades de los “insulators”.

**Determinar si los “insulators” descritos determinan el límite 5’ del dominio  $\alpha$ -globina de pollo:**

De acuerdo a los resultados de Anguita y colaboradores, río arriba de los “insulators” descritos en este trabajo, se encuentran otros sitios de hipersensibilidad al corte por la DNasa I, los cuáles podrían contener otros elementos “insulators”, para probar esta hipótesis se tendría que proceder a la

clonación de las secuencias río arriba del  $\alpha$ EHS-1.4, analizar la estructura de la cromatina de esa zona. determinar si esos sitios de hipersensibilidad tienen propiedades de “insulators” y se tendrían adicionalmente que analizar las modificaciones de las histonas, como la metilación en los residuos específicos como la metilación de la lisina 9 de la histona H3 y particularmente de marcas de heterocromatina como la desacetilación de las histonas para corroborar de esta forma la presencia en esta zona de una transición entre una organización de la cromatina “abierta” (en el dominio  $\alpha$ -globina) y una cerrada (dominio adyacente).

### **Análisis detallado del fragmento de 120 bp que contiene al sitio M9**

Se ha determinado por análisis de secuencias que existen otros factores como Oct-1 que interaccionan en la región M9, para confirmar estos datos se propone primero hacer un ensayo de digestión a la DNasa I (footprint *in vitro*), sobre el fragmento de ADN de 120 bp, utilizando extractos de células eritroides, 6C2, HD3 y RBC correspondientes a distintos estadios de diferenciación eritroide y determinar qué secuencias se protegen de la digestión por proteínas nucleares de extractos representativos de los diversos estadios. Además se propone que en el sitio M9 se determinen las bases nucleotídicas que son contactadas por la proteína CTCF llevando a cabo un ensayo de interferencia por metilación; este ensayo a su vez podría hacerse utilizando extractos nucleares de los estadios de diferenciación eritroide, y con extractos de una línea celular linfóide (DT40) si los contactos de CTCF cambian durante la diferenciación, esto apoyaría la hipótesis de que esta proteína está interaccionando con diferentes factores durante la diferenciación, que es una proteína que no se encuentra estática sino realizando diversas funciones dentro de la célula y en particular dentro del dominio  $\alpha$ -globina de pollo.

Adicionalmente se propone estudiar si estos “insulators” son capaces de actuar en conjunto, para lo cuál actualmente se están realizando construcciones que contengan al  $\alpha$ EHS-1.4 y al  $\alpha$ CHS-1.2. estas construcciones serán ensayadas transitoriamente, y se comprobará si tienen la capacidad de ser mejores bloqueadores de enhancer que individualmente.

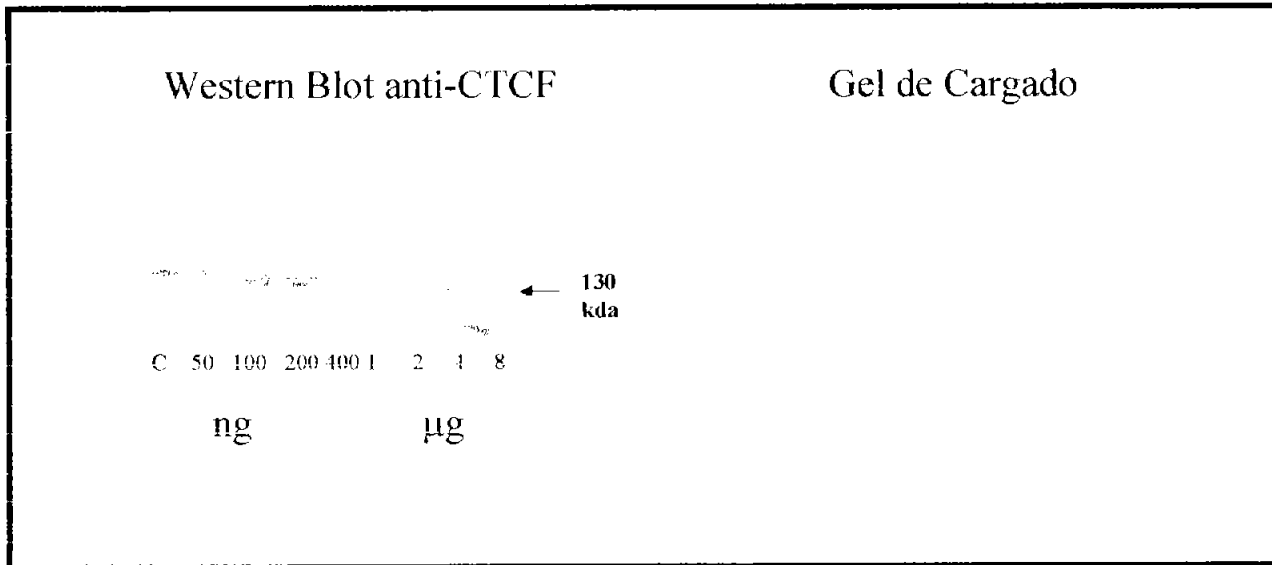
### **Estudio de la importancia del factor CTCF en la regulación de los genes del dominio $\alpha$ -globina de pollo y del gen constitutivo.**

Recientemente con el avance en los conocimientos sobre el RNA de interferencia (RNAi), han surgido novedosas técnicas que permiten disminuir los niveles de una proteína para estudiar sus efectos *in vivo*. En colaboración con el L. en IBB Fabián Flores Jasso del grupo del Dr. Luis Vaca se ha

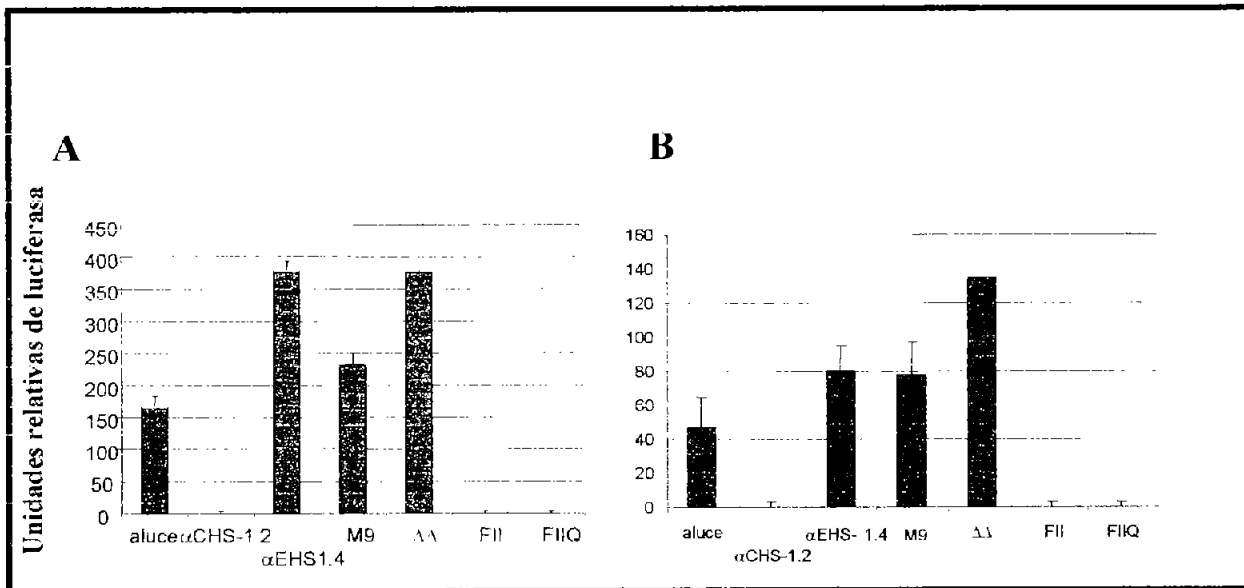
desarrollado un método para disminuir los niveles de la proteína CTCF de pollo mediante RNA de interferencia.

Utilizando el vector BB4CMV2, se clonó el extremo terminal amino de la proteína CTCF de pollo entre dos promotores T7, gracias a los cuáles se puede llevar a cabo una transcripción *in vitro* del extremo amino-terminal de CTCF. El RNA obtenido se alineó para obtener una doble cadena, fue cuantificado y posteriormente fue transfectado en células 6C2 en diferentes concentraciones. Las células se recuperaron a las 24 y 48 horas y se realizó un Western Blot con anticuerpos dirigidos contra la proteína CTCF. A las 24 horas no se observaron cambios en la concentración de la proteína (datos no mostrados), sin embargo a las 48 horas las células muestran una disminución en la concentración de la proteína CTCF (Fig. 40). Estos datos deben ser corroborados mediante RT-PCR del transcrito de CTCF. Sin embargo muestran que los niveles de la proteína pueden bajar después de la transfección del RNA de doble cadena.

Con este antecedente realizamos un ensayo de bloqueo de enhancer para corroborar funcionalmente que en los "insulators" identificados la actividad de bloqueo de enhancer dependiera de los niveles de la proteína CTCF presentes en las células (Fig. 41). En este ensayo se transfectaron las células 6C2 con el RNA de doble cadena correspondiente al extremo terminal amino de la proteína CTCF de pollo (Fig. 41), 48 horas después se realizó el ensayo de bloqueo de enhancer. Los resultados muestran que en las células que fueron transfectadas previamente con el RNA de doble cadena la actividad del gen reportero al transfectar con el vector control pALucE decae considerablemente comparar unidades relativas de luciferasa en ambas gráficas. Estos datos muestran que el RNA de doble cadena está afectando los niveles de expresión de otras proteínas además de la proteína CTCF. Esto puede llegar a ocurrir cuando se utiliza un RNA de doble cadena tan grande como el nuestro (aprox. 500 bp).



**Fig. 40** Ensayo de RNAi contra CTCF. En el extremo izquierdo se muestra el Western Blot anti-CTCF en células 6C2 realizado a las 48 horas post-transfección con el iRNA contra CTCF. C, control, ng, nanogramos; μg, microgramos del RNA de doble cadena transfectado a las células 6C2. Del lado derecho se muestra la membrana transferida para corroborar que el cargado de las proteínas fuese parejo.



**Fig. 41** Ensayo de bloqueo de enhancer **A)** Actividad de bloqueo de enhancer de los “insulators” αEHS-1.4 y αEHS-1.2, y FII en células 6C2 ; **B)** Actividad de bloqueo de enhancer en células 6C2 transfectadas 48 horas antes con RNA de doble

cadena correspondiente al extremo terminal amino de la proteína CTCF de pollo, nótese que la actividad del gen luciferasa disminuye considerablemente en todos los casos (ver escala).

Para resolver este problema se necesitaría diseñar un RNA de doble cadena pequeño (21 bp) y transfectar éste a las células. Si este RNA pequeño funciona entonces sería interesante analizar lo siguiente:

1. Para corroborar si la proteína CTCF está involucrada de alguna forma en la regulación de la expresión de los genes del dominio  $\alpha$ -globina de pollo, se analizarían los niveles de expresión de los genes a lo largo de la diferenciación mediante RT-PCR, utilizando células en donde se hayan abatido los niveles de CTCF mediante el método anteriormente descrito. El hacer este experimento tendría como finalidad saber si la ausencia de esta proteína afecta la expresión de los genes del dominio. Como control podría utilizarse a la línea celular linfoide DT40, en la que no debiera de existir expresión de los genes.
2. Podría realizarse un ensayo de inmunoprecipitación de la cromatina en el sitio M9 para determinar si el abatimiento de los niveles de la proteína es suficiente para la pérdida de la unión de CTCF a sus sitios en los "insulators".
3. Así mismo podría tratar de definirse si esta proteína afecta de alguna manera a los transcritos intergénicos presentes a lo largo del extremo 5' del dominio, para lo cuál se utilizarían células en las que se abate la expresión de la proteína CTCF y se buscaría la presencia de estos transcritos.
4. Se ha propuesto que un posible mecanismo de acción de los "insulators" podría ser atrayendo actividades remodeladoras de la estructura de la cromatina, por lo que el analizar los niveles de acetilación de las histonas en la región de estudio en las células donde se abata la expresión de CTCF e incluso corroborar estos niveles con los datos obtenidos por el grupo de Higgs, sería de extrema importancia ya que ayudaría a determinar si el reclutamiento de estas actividades está regulado también por CTCF.

Sería interesante el llevar a cabo estos experimentos, ya que podría estudiarse *in vivo* el papel que juega la proteína CTCF e incluso estos experimentos podrían determinar si esta proteína está involucrada en la regulación de los genes del dominio  $\alpha$ -globina de pollo, y si la ausencia de esta proteína afecta la estructura del dominio, analizando los posibles cambios en la expresión de los genes, la presencia de los transcritos intergénicos característicos de este dominio y los patrones de modificación en las histonas.



## BIBLIOGRAFIA

- Ahmad K y Henikoff S. (2002). Histone H3 variants specify modes of chromatin assembly. *Proc. Natl. Acad. Sci. USA* **99** ( suppl. 4):16477-84.
- Akasaka K, Nishimura A, Takata K, Mitsunaga K, Mibuka F, Ueda H, Hirose S, Tsutsui K y Shimada H. (1999). Upstream element of the sea urchin arylsulfatase gene serves as an insulator. *Cell Mol Biol (Noisy-le-grand)* **45**:555-65.
- Allen GC, Spiker S y Thompson WF. (2000). Use of matrix attachment regions (MARs) to minimize transgene silencing. *Plant Mol Biol.* **43**:361-76.
- Anguita E, Colin A, Wood W, Turner B, y Higgs D. (2001). Identification of a conserved erythroid specific domain of histone acetylation across the  $\alpha$ -globin gene cluster. *Proc. Natl. Acad. Sci.* **98**:12114-19.
- Arnold R, Burcin M, Kaiser B, Muller M y Renkawitz R. (1996). DNA bending by the silencer protein NeP1 is modulated by TR and RXR. *Nucl. Acids. Res.* **24**:2640-7.
- Ausio J y Abbott D (2002) The many tales of a tail: carboxyl-terminal tail heterogeneity specializes histone H2A variants for defined chromatin function. *Biochem.* **41**:4945-9.
- Barges S, Mihaly J, Galloni M, Hagstrom K, Muller M, Shanower G, Schedl P, Gyurkovics H y Karch F. (2000). The Fab-8 boundary defines the distal limit of the bithorax complex iab-7 domain and insulates iab-7 from initiation elements and a PRE in the adjacent iab-8 domain. *Development* **127**:779-90.
- Bassing CH, Suh H, Ferguson DO, Chua KF, Manis J, Eckersdorff M, Gleason M, Bronson R, Lee C y Alt FW. (2003). Histone H2AX: a dosage-dependent suppressor of oncogenic translocations and tumors. *Cell* **114**:359-70.
- Batzer MA y Deininger PL. (2002). Alu repeats and human genomic diversity. *Nat Rev Genet* **3**:370-9.

- Bell A y Felsenfeld G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405**:482-5.
- Bell A, West A y Felsenfeld G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* **98**:387-96.
- Belozzerov V, Majumder P, Shen P y Cai H. (2003). A novel boundary element may facilitate independent gene regulation in the Antennapedia complex of Drosophila. *EMBO J.* **22**:3113-21.
- Blackwood EM y Kadonaga JT. (1998). Going the distance: a current view of enhancer action. *Science* **281**:61-3.
- Bode J, Benham C, Knopp A y Mielke C. (2000). Transcriptional augmentation: modulation of gene expression by scaffold/matrix-attached regions (S/MAR elements). *Crit Rev Eukaryot Gene Expr.* **10**:73-90.
- Boyes, J y G Felsenfeld. (1996). Tissue-specific factors additively increase the probability of the all-or-none formation of a hypersensitive site. *EMBO J.* **15**:2496-2507.
- Bulger M, y Groudine M. (1999). Looping versus linking: toward a model for long-distance gene activation. *Genes Dev.* **13**:2465-77.
- Burcin M, Arnold R, Lutz M, Kaiser B, Runge D, Lottspeich F, Filippova GN, Lobanenko VV y Renkawitz R. (1997). Negative protein 1, which is required for function of the chicken lysozyme gene silencer in conjunction with hormone receptors, is identical to the multivalent zinc finger repressor CTCF. *Mol. Cell. Biol.* **17**:1281-88.
- Burgess-Beusse B, Farrell C, Gaszner M, Litt M, Mutskov V, Recillas-Targa F, Simpson M, West A y Felsenfeld G. (2002). The insulation of genes from external enhancers and silencing chromatin. *Proc. Natl. Acad. Sci. USA* **99**:16433-37.
- Byrd K, y Corces V. (2003). Visualization of chromatin domains created by the gypsy insulator of Drosophila. *J. Cell Biol.* **162**:565-74.
- Cai H y Levine M. (1995). Modulation of enhancer-promoter interactions by insulators in the Drosophila embryo. *Nature* **376**:533-6.
- Cai H y Levine M. (1997). The gypsy insulator can function as a promoter-specific silencer in the Drosophila embryo. *EMBO J.* **16**:1732-41.
- Carter D, Chakalova L, Osborne CS, Dai YF y Fraser P. (2002). Long-range chromatin regulatory interactions in vivo. *Nature Genetics* **32**:623-6.
- Chernukhin I, Shamsuddin S, Robinson A, Carne A, Paul A, El-Kady A, Lobanenko V y Klenova E. (2000). Physical and Functional Interaction between Two Pluripotent Proteins, the Y-box DNA/RNA-binding Factor, YB-1, and the Multivalent Zinc Finger Factor, CTCF. *J. Biol. Chem.* **275**:29915-21.
- Chow CM, Athanassiadou A, Raguz S, Psiouri L, Harland L, Malik M, Aitken MA, Grosveld F y Antoniou M. (2002). LCR-mediated, long-term tissue-specific gene expression within replicating episomal plasmid and cosmid vectors. *Gene therapy* **9**:327-36.
- Chung J, Bell A, y Felsenfeld G. (1997). Characterization of the chicken  $\beta$ -globin insulator. *Proc. Natl. Acad. Sci. USA* **94**:575-80.
- Cremer T y Cremer C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet* **2**:292-301.
- De Sanctis V, La Terra S, Bianchi A, Shore D, Burderi L, Di Mauro E y Negri R. (2002). In Vivo Topography of Rap1p-DNA Complex at Saccharomyces cerevisiae TEF2 UASRPG During Transcriptional Regulation. *J. Mol. Biol.* **318**:333-49.
- Dekker J, Rippe K, Dekker M y Kleckner N. (2002). Capturing Chromosome Conformation. *Science* **295** 1306-11.

- Spector D. (2003). The dynamics of chromosome organization and gene regulation. *Ann. Rev. of Biochem.* **72**:573-608.
- Donze D, Adams C, Rine J y Kamakaka R. (1999). The boundaries of the silenced HMR domain in *Saccharomyces cerevisiae*. *Genes Dev.* **13**:698-708.
- Dorsett, D. (1999). Distant liaisons: long-range enhancer-promoter interactions in *Drosophila*. *Curr Opin Genet Dev.* **9**:505-14.
- Dunaway M, Hwang JY, Xiong M, y Yuen HL. (1997). The activity of the scs and scs' insulator elements is not dependent on chromosomal context. *Mol. Cell. Biol.* **17**:182-189.
- Dunn K, Zhao H y Davie J. (2003). The insulator binding protein CTCF associates with the nuclear matrix. *Experimental Cell Research* **288**:218-23.
- Engel JD yTanimoto K. (2000). Looping, linking, and chromatin activity: new insights into  $\beta$ -globin locus regulation. *Cell* **100**:499-502.
- Ericsson C, Grossbach U, Bjorkroth B y Daneholt B. (1990). Presence of histone H1 on an active Balbiani ring gene. *Cell* **60**:73-83.
- Farache G, Razin SV, Targa FR y Scherrer K. (1990). Organization of the 3'-boundary of the chicken  $\alpha$  globin gene domain and characterization of a CR 1-specific protein binding site. *Nucl. Acids. Res.* **18**:401-9.
- Fedoriw A, Stein P, Svoboda P, Schultz R y Bartolomei M. (2004). Transgenic RNAi Reveals Essential Function for CTCF in H19 Gene Imprinting. *Science* **303**:238-40.
- Felsenfeld G y Groudine M. (2003). Controlling the double helix. *Nature* **421**:448-53.
- Felsenfeld G, Boyes J, Chung J, Clark D y Studitsky V. (1996). Chromatin structure and gene expression. *Proc. Natl. Acad. Sci.* **93**:9384-88.
- Festenstein R, Tolaini M, Corbella P, Mamalaki C, Parrington J, Fox M, Miliou A, Jones M y Kioussis D. (1996). Locus control region function and heterochromatin-induced position effect variegation. *Science* **271**:1123-5.
- Filippova G, Chen-Feng Qi, Jonathan E. Ulmer, James M. Moore, Michael D. Ward, Ying J. Hu, Dmitri I. Loukinov, Elena M. Pugacheva, Elena M. Klenova, Paul E. Grundy, Andrew P. Feinberg, Anne-Marie Cleton-Jansen, Elna W. Moerland, Cees J. Cornelisse, Hiroyoshi Suzuki, Akira Komiya, Annika Lindblom, Francoise Dorion-Bonnet, Paul E. Neiman, Herbert C. Morse, III, Steven J. Collins y Victor V. Lobanenkoy. (2002). Tumor-associated Zinc Finger Mutations in the CTCF Transcription Factor Selectively Alter Its DNA-binding Specificity. *Cancer Res* **62**:48-52.
- Filippova, GN, Fagerlie S, Klenova E, Myers C, Dehner Y, Goodwin G, Neiman P, Collins SJ y Lobanenkoy V. (1996). An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Mol. Cell. Biol.* **16**:2802-13.
- Fischle W, Wang Y y Allis C. (2003). Binary switches and modification cassettes in histone biology and beyond. *Nature* **425**:475-9.
- Fischle W, Wang Y y Allis C. (2003). Histone and chromatin cross-talk. *Current Opinion in Cell Biology* **15**:172-83.
- Fourel G, Revardel E, Koering C y Gilson E. (1999). Cohabitation of insulators and silencing elements in yeast subtelomeric regions. *EMBO J.* **18**:2522-37.
- Francastel C, Schubeler D, Martin DI yGroudine M. (2000). Nuclear compartmentalization and gene activity. *Nat. Rev. Mol Cell Biol.* **1**:137-43.
- Gaszner M, Vazquez J y Schedl P. (1999). The Zw5 protein, a component of the scs chromatin domain boundary, is able to block enhancer-promoter interaction. *Genes Dev.* **13**:2098-2107.

- Gdula D, Gerasimova T y Corces V. (1996). Genetic and molecular analysis of the gypsy chromatin insulator of *Drosophila*. *Proc. Natl. Acad. Sci. USA* **93**:9378-83.
- Gerasimova TI y Corces VG. (1998). Polycomb and trithorax group proteins mediate the function of a chromatin insulator. *Cell* **92**:511-21.
- Ghirlando R, Litt MD, Prioleau MN, Recillas-Targa F y Felsenfeld G. (2004). Physical properties of a genomic condensed chromatin fragment. *J.Mol. Biol* **336**:597-605.
- Golovnin A, Birukova I, Romanova O, Silicheva M, Parshikov A, Savitskaya E, Pirrotta V, Georgiev P. (2003). An endogenous Su(Hw) insulator separates the yellow gene from the Achaete-scute gene complex in *Drosophila*. *Development* **130**:3249-3258.
- Gribnau J, Diderich K, Pruzina S, Calzolari R, Fraser P. (2000). Intergenic transcription and developmental remodeling of chromatin subdomains in the human  $\beta$ -globin locus. *Mol Cell* **5**:377-86.
- Grosveld F, van Assendelft GB, Greaves DR y Kollias G. (1987). Position-independent, high-level expression of the human  $\beta$ -globin gene in transgenic mice. *Cell* **51**:975-85.
- Gunther M, Laithier M y Brison O (2000) A set of proteins interacting with transcription factor Sp1 identified in a two-hybrid screening. *Mol Cell Biochem* **210**:131-42.
- Haas NB, Grabowski JM, Sivitz AB y Burch JB. (1997). Chicken repeat 1 (CR1) elements, which define an ancient family of vertebrate non-LTR retrotransposons, contain two closely spaced open reading frames. *Gene* **197**:305-9.
- Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM y Tilghman SM. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* **405**:486-9.
- Hart CM y Laemmli UK. (1998). Facilitation of chromatin dynamics by SARs. *Curr Opin Genet Dev* **8**:519-25.
- Hart, CM, K Zhao y UK Laemmli. (1997). The scs' boundary element: characterization of boundary element-associated factors. *Mol. Cell. Biol.* **17**:999-1009.
- Hassan AH, Prochasson P, Neely KE, Galasinski SC, Chandy M, Carrozza MJ y Workman JL. (2002). Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. *Cell* **111**:369-79.
- Hebbes TR, Clayton AL, Thorne AW, Crane-Robinson C. (1994). Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken  $\beta$ -globin chromosomal domain. *EMBO J.* **13**:1823-30.
- Hendrich B y Tweedie S. (2003). The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends in Genetics* **19**:269-77.
- Ishii K, Arib G, Lin C, Van Houwe G y Laemmli UK. (2002). Chromatin boundaries in budding yeast: the nuclear pore connection. *Cell* **109**:551-62.
- Isogai Y y Tjian R. (2003). Targeting genes and transcription factors to segregated nuclear compartments. *Current Opinion in Cell Biology* **15**:296-303.
- Jenuwein T, y Allis C. (2001). Translating the Histone Code. *Science* **293**:1074-80.
- Kamakaka R. (1997). Silencers and locus control regions: opposite sides of the same coin. *Trends in Biochemical Sciences* **22**:124-28.
- Kamakaka RT y Thomas JO. (1990). Chromatin structure of transcriptionally competent and repressed genes. *EMBO J.* **9**:3997-4006.
- Kanduri C, Pant V, Loukinov D, Pugacheva E, Qi CF, Wolffe A, Ohlsson R y Lobanenkov VV. (2000). Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. *Curr Biol.* **10**:853-6.

- Karch F, Galloni M, Sipos L, Gausz J, Gyurkovics H y Schedl P. (1994). Mcp and Fab-7: molecular analysis of putative boundaries of cis-regulatory domains in the bithorax complex of *Drosophila melanogaster*. *Nucl. Acids. Res.* **22**:3138-46.
- Kellum R y Schedl P. (1991). A position-effect assay for boundaries of higher order chromosomal domains. *Cell* **64**:941-50.
- Kellum R. (1992). A group of scs elements function as domain boundaries in an enhancer-blocking assay. *Mol. Cell Biol.* **12**:2424-31.
- Kirstetter P, Dierich T, Kastner P y Chan S. (2002). Ikaros is critical for B cell differentiation and function. *Eur J Immunol* **32**:720-30.
- Klenova E, III Morse H, Ohlsson R y Lobanenkova V. (2002). The novel BORIS + CTCF gene family is uniquely involved in the epigenetics of normal biology and cancer. *Seminars in Cancer Biology* **12**:399-414.
- Klenova EM, Nicolas RH, Paterson HF, Carne AF, Heath CM, Goodwin GH, Neiman PE y Lobanenkova VV. (1993). CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-Zn-finger protein differentially expressed in multiple forms. *Mol Cell Biol.* **13**:7612-24.
- Kondo Y y Issa J. (2003). Enrichment for Histone H3 Lysine 9 Methylation at Alu Repeats in Human Cells. *J. Biol. Chem.* **278**:27658-662.
- Kuhn E, y Geyer P. (2003). Genomic insulators: connecting properties to mechanism. *Current Opinion in Cell Biology* **15**:259-65.
- Labrador M y Corces VG. (2002). Setting the boundaries of chromatin domains and nuclear organization. *Cell* **111**:151-4.
- Lachner M, O'Carroll D, Rea S, Mechtler K y Jenuwein T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**:116-20.
- Lachner M y Jenuwein T. (2002). The many faces of histone lysine methylation. *Current Opinion in Cell Biology* **14**:286-98.
- Lamond A y Earnshaw W. 1998. Structure and Function in the Nucleus. *Science* **280**:547-53.
- Li Q, Peterson K, Fang X y Stamatoyannopoulos G. (2002). Locus control regions. *Blood* **100**:3077-86.
- Litt M, Simpson M, Gaszner M, Allis C y Felsenfeld G. (2001). Correlation Between Histone Lysine Methylation and Developmental Changes at the Chicken  $\beta$ -Globin Locus. *Science* **293**:2453-55.
- Loukinov D, Elena Pugacheva, Sergei Vatolin, Svetlana D. Pack, Hanlim Moon, Igor Chernukhin, Poonam Mannan, Erik Larsson, Chandrasekhar Kanduri, Alexander A. Vostrov, Hengmi Cui, Emily L. Niemitz, John E. J. Rasko, France M. Docquier, Malathi Kistler, Joseph J. Breen, Zhengping Zhuang, Wolfgang W. Quitschke, Rainer Renkawitz, Elena M. Klenova, Andrew P. Feinberg, Rolf Ohlsson, Herbert C. Morse, III y Victor V. Lobanenkova. (2002). BORIS, a novel male germ-line-specific protein associated with epigenetic reprogramming events, shares the same 11-zinc-finger domain with CTCF, the insulator protein involved in reading imprinting marks in the soma. *Proc. Natl. Acad. Sci.* **99**:6806-11.
- Luger K, Mader A, Richmond R, Sargent F y Richmond T. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**:251-60.
- Lukasova E, Kozubek S, Kozubek M, Falk M y Amrichova J. (2002). The 3D structure of human chromosomes in cell nuclei. *Chromosome Res.* **10**:535-48.
- Luo, Kunheng, Miguel A. Vega-Palas y Michael Grunstein. (2002). Rap1-Sir4 binding independent of other Sir, yKu, or histone interactions initiates the assembly of telomeric heterochromatin in yeast. *Genes Dev.* **16**:1528-39.

- Lutz M, Baniahmad A y Renkawitz R. (2000). Modulation of thyroid hormone receptor silencing function by co-repressors and a synergizing transcription factor. *Biochemical Soc Trans* **28**:386-9.
- Lutz M, Burke LJ, Barreto G, Goeman F, Greb H, Arnold R, Schultheiss H, Brehm A, Kouzarides T, Lobanenko V y Renkawitz. (2000). Transcriptional repression by the insulator protein CTCF involves histone deacetylases. *Nucl. Acids. Res.* **28**:1707-13.
- Martin DI, Fiering S y Groudine M. (1996). Regulation of  $\beta$ -globin gene expression: straightening out the locus. *Curr Opin Genet Dev.* **6**:488-95.
- McKittrick E, Gafken P, Ahmad K y Henikoff S. (2004). Histone H3.3 is enriched in covalent modifications associated with active chromatin. *Proc. Natl. Acad. Sci. USA* **19**:
- Melfi R, Palla F, Di Simone P, Alessandro C, Cali L, Anello L y Spinelli G. (2000). Functional Characterization of the Enhancer Blocking Element of the Sea Urchin Early Histone Gene Cluster Reveals Insulator Properties and Three Essential cis-acting Sequences *Journal of Molecular Biology* **304**:753-763.
- Mihaly, J, Hogga I, Gausz J, Gyurkovics H y Karch F. (1997). In situ dissection of the Fab-7 region of the bithorax complex into a chromatin domain boundary and a Polycomb-response element. *Development* **124**:1809-20.
- Mishra K y Shore D. (1999). Yeast Ku protein plays a direct role in telomeric silencing and counteracts inhibition by rif proteins. *Curr Biol.* **9**:1123-6.
- Mizuguchi G, Shen X, Landry J, Wu W, Sen S y Wu C. (2004). ATP-Driven Exchange of Histone H2AZ Variant Catalyzed by SWI1 Chromatin Remodeling Complex. *Science* **303**:343-48.
- Modin C, Pedersen F y Duch M. (2000). Lack of Shielding of Primer Binding Site Silencer-Mediated Repression of an Internal Promoter in a Retrovirus Vector by the Putative Insulators scs, BEAD-1, and HS4. *J. Virol.* **74**:11697-707.
- Mongelard F, Labrador M, Baxter E, Gerasimova T y Corces V. (2002). Trans-splicing as a Novel Mechanism to Explain Interallelic Complementation in *Drosophila*. *Genetics* **160**:1481-87.
- Nabirochkin S, Ossokina M y Heidmann T. (1998). A Nuclear Matrix/Scaffold Attachment Region Co-localizes with the Gypsy Retrotransposon Insulator Sequence. *J. Biol. Chem.* **273**:2473-79.
- Namciu S, Blochlinger K y Fournier R. (1998). Human Matrix Attachment Regions Insulate Transgene Expression from Chromosomal Position Effects in *Drosophila melanogaster*. *Mol. Cell. Biol.* **18**:2382-91.
- Ogata T, Kozuka T y Kanda T. (2003). Identification of an Insulator in AAVS1, a Preferred Region for Integration of Adeno-Associated Virus DNA. *J. Virol.* **77**:9000-7.
- Ohlsson T, Renkawitz R y Lobanenko V. (2001). CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends in Genetics* **17**:520-27.
- Ohtsuki S y Levine M. (1998). GAGA mediates the enhancer blocking activity of the eve promoter in the *Drosophila* embryo. *Genes Dev.* **12**:3325-30.
- Parada L y Misteli T. (2002). Chromosome positioning in the interphase nucleus. *Trends in Cell Biology* **12**:425-32.
- Park J, Cosgrove M, Youngman E, Wolberger C y Boeke JD. (2002). A core nucleosome surface crucial for transcriptional silencing. *Nature Genetics* **32**:273-9.
- Parnell T, y Geyer P. (2000). Differences in insulator properties revealed by enhancer blocking assays on episomes. *EMBO J.* **19**:5864-74.
- Pikaart M, Recillas-Targa F y Felsenfeld G. (1998). Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators. *Genes Dev.* **12**:2852-62.

- Prioleau M, Nony P, Simpson M y Felsenfeld G. (1999). An insulator element and condensed chromatin region separate the chicken  $\beta$ -globin locus from an independently regulated erythroid-specific folate receptor gene. *EMBO J.* **18**:4035-48.
- Ragoczy T, Telling A, Sawado T, Groudine M y Kosak ST. (2003). A genetic analysis of chromosome territory looping: diverse roles for distal regulatory elements. *Chromosome Res.* **11**:513-25.
- Rasko J, Klenova E, Leon J, Filippova G, Loukinov D, Vatolin S, Robinson A, Hu Y, Ulmer J, Ward M, Pugacheva E, Neiman P, Morse H, Collins S y Lobanenkov V. (2001). Cell Growth Inhibition by the Multifunctional Multivalent Zinc-Finger Factor CTCF. *Cancer Res* **61**:6002-7.
- Razin S, Ioudinkova E y Scherrer K. (2000). Extensive methylation of a part of the CpG island located 3.0-4.5 kbp upstream to the chicken  $\alpha$ -globin gene cluster may contribute to silencing the globin genes in non-erythroid cells. *Journal of Molecular Biology* **299**:845-52.
- Razin SV, Farrell CM y Recillas-Targa F. (2003). Genomic domains and regulatory elements operating at the domain level. *Int Rev Cytol* **226**:63-125.
- Razin, SV, P Petrov y R Hancock. (1991). Precise Localization of the  $\alpha$ -Globin Gene Cluster Within One of the 20- to 300-Kilobase DNA Fragments Released by Cleavage of Chicken Chromosomal DNA at Topoisomerase II Sites in vivo: Evidence that the Fragments are DNA Loops or Domains. *Proc. Natl. Acad. Sci. USA* **88**:8515-19.
- Razin SV, Shen K, Ioudinkova E y Scherrer K. (1999). Functional analysis of DNA sequences located within a cluster of DNase I hypersensitive sites colocalizing with a MAR element at the upstream border of the chicken  $\alpha$ -globin gene domain. *J Cell Biochem.* **74**:38-49.
- Razin SV, Vassetzky YS Jr, Kvartskhava AI, Grinenko NF y Georgiev GP. (1991). Transcriptional enhancer in the vicinity of a replication origin within the 5' region of the chicken  $\alpha$ -globin gene domain. *J Mol Biol* **217**:595-8.
- Recillas-Targa F y Razin SV. (2001). Chromatin domains and regulation of gene expression: familiar and enigmatic clusters of chicken globin genes. *Crit Rev Eukaryot Gene Expr.* **11**:227-42.
- Recillas-Targa F, Bell A y Felsenfeld G. (1999). Positional enhancer-blocking activity of the chicken  $\beta$ -globin insulator in transiently transfected cells. *PNAS* **96**:14354-59.
- Redon C, Pilch D, Rogakou E, Sedelnikova O, Newrock K y Bonner W. (2002). Histone H2A variants H2AX and H2AZ. *Current Opinion in Genetics & Development* **12**:162-69.
- Reina-San-Martin B, Difilippantonio S, Hanitsch L, Masilamani R, Nussenzweig A y Nussenzweig M. (2003). H2AX Is Required for Recombination Between Immunoglobulin Switch Regions but Not for Intra-Switch Region Recombination or Somatic Hypermutation. *J. Exp. Med.* **197**:1767-78.
- Reitman M y Felsenfeld G. (1990). Developmental regulation of topoisomerase II sites and DNase I-hypersensitive sites in the chicken  $\beta$ -globin locus. *Mol Cell Biol.* **10**:2774-86.
- Saitoh N, Bell A, Recillas-Targa F, West A, Simpson M, Pikaart M y Felsenfeld G. (2000). Structural and functional conservation at the boundaries of the chicken  $\beta$ -globin domain. *EMBO J.* **19**:2315-22.
- Santos-Rosa H, Schneider R, Bannister AJ, Sherriff J, Bernstein BE, Emre NC, Schreiber SL, Mellor J y Kouzarides T. (2002). Active genes are tri-methylated at K4 of histone H3. *Nature* **419**:407-11.
- Scherrer K (2003). Historical review: The discovery of 'giant' RNA and RNA processing: 40 years of enigma. *Trends in Biochemical Sciences* **28**:566-71.
- Shen B, Kim J y Dorsett D. (1994). The enhancer-blocking suppressor of Hairy-wing zinc finger protein of *Drosophila melanogaster* alters DNA structure. *Mol Cell Biol.* **14**:5645-52.

- Sjakste N, Iarovaia OV, Razin SV, Linares-Cruz G, Sjakste T, Le Gac V, Zhao Z, Scherrer K. (2000). A novel gene is transcribed in the chicken  $\alpha$ -globin gene domain in the direction opposite to the globin genes. *Mol Gen Genet.* **262**:1012-21.
- Smith M (2002) Histone variants and nucleosome deposition pathways. *Mol Cell.* **9**:1158-60.
- Spector D. (2001). Nuclear domains. *J Cell Sci* **114**:2891-93.
- Stalder J, Larsen A, Engel J, Dolan M, Groudine M y Weintraub H. (1980). Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNaseI. *Cell* **20**:451-60.
- Strahl BD y Allis CD. (2000). The language of covalent histone modifications. *Nature* **403**:41-5.
- Struhl K. (1998). Histone acetylation and transcriptional regulatory mechanisms. *Genes Dev.* **12**:599-606.
- Takai D, Gonzales F, Tsai Y, Thayer M y Jones P. (2001). Large scale mapping of methylcytosines in CTCF-binding sites in the human H19 promoter and aberrant hypomethylation in human bladder cancer. *Hum. Mol. Genet.* **10**:2619-26.
- Tang Y, Liu D, y Liang C. (2002) Further understanding of the  $\beta$ -globin locus regulation at the molecular level: looping or linking models? *Genes Cells* **7**:889-900.
- Targa FR, de Moura Gallo CV, Huesca M, Scherrer K y Marcaud L. (1993). Silencer and enhancer elements located at the 3'-side of the chicken and duck  $\alpha$ -globin-encoding gene domains. *Gene* **129**:229-37.
- Thomas JO (1984) The higher order structure of chromatin and histone H1. *J Cell Sci Suppl.* **1**:1-20.
- Thrower D y Bloom K (2001). Dicentric Chromosome Stretching during Anaphase Reveals Roles of Sir2/Ku in Chromatin Compaction in Budding Yeast. *Mol. Biol. Cell* **12**:2800-12.
- Tilghman, SM. (1999). The sins of the fathers and mothers: genomic imprinting in mammalian development. *Cell* **96**:185-93.
- Tuan, D., A. Abeliovich, M. Lee Oldham y D. Lee. (1987). Identification of regulatory elements of human  $\beta$ -like globin genes. *Progress in Clinical and Biological Research* **251**:211-20.
- Tuan, D, Solomon W, Li Q y London I. (1985). The " $\beta$ -like-globin" gene domain in human erythroid cells. *Proceedings of the National Academy of Sciences of the United States of America* **82**:6384-8.
- Ulaner GA, Vu TH, Li T, Hu JF, Yao XM, Yang Y, Gorlick R, Meyers P, Healey J, Ladanyi M y Hoffman AR. (2003). Loss of imprinting of IGF2 and H19 in osteosarcoma is accompanied by reciprocal methylation changes of a CTCF-binding site. *Hum. Mol. Genet.* **12**:535-49.
- von Kries JP, Phi-Van L, Diekmann S y Stratling WH. (1990). A non-curved chicken lysozyme 5' matrix attachment site is 3' followed by a strongly curved DNA sequence. *Nucl. Acids. Res.* **18**:3881-5.
- Vyas P, Vickers MA, Picketts DJ y Higgs DR. (1995) Conservation of position and sequence of a novel, widely expressed gene containing the major human alpha-globin regulatory element. *Genomics.* **29**: 679-89.
- Wanner, G y H. Formanek. (2000). A New Chromosome Model. *Journal of Structural Biology* **132**:147-61.
- Wei Y, Mizzen C, Cook R, Gorovsky M y Allis CD (1998) Phosphorylation of histone H3 at serine 10 is correlated with chromosome condensation during mitosis and meiosis in Tetrahymena. *Proc. Natl. Acad. Sci.* **95**:7480-94.
- Weintraub H y Groudine M. (1976). Chromosomal subunits in active genes have an altered conformation. *Science* **193**:848-56.
- West A, Gaszner M y Felsenfeld G. (2002). Insulators: many functions, many mechanisms. *Genes Dev.* **16**:271-88.



Widom J (1998). Chromatin structure: linking structure to function with histone H1. *Curr. Biol.* 8:R788-91.

Yudinkova ES y Razin SV. (2002). A block of CpG dinucleotides that are differentially methylated in erythroid and nonerythroid cells is located upstream of the cluster of chicken  $\alpha$ -globin genes. *Dokl Biochem Biophys* 384:179-80.

Yusufzai TM, Tagami H, Nakatani Y y Felsenfeld G.(2004) CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. *Mol. Cell* 13:291-8.

Zhong X, y Krangel M. (1999). Enhancer-Blocking Activity Within the DNase I Hypersensitive Site 2 to 6 Region Between the TCR  $\alpha$  and Dad1 Genes. *J Immunol* 163:295-300.

# CTCF-dependent enhancer blockers at the upstream region of the chicken $\alpha$ -globin gene domain

Viviana Valadez-Graham, Sergey V. Razin<sup>1</sup> and Félix Recillas-Targa\*Instituto de Fisiología Celular, Departamento de Genética Molecular, Universidad Nacional Autónoma de México, Apartado Postal 70-242, México D.F. 04510, México and <sup>1</sup>Laboratory of Structural and Functional Organization of Chromosomes, Institute of Gene Biology of the Russian Academy of Sciences, Vavilov Str. 34/5, 119334 Moscow, Russia

Received January 6, 2004; Revised and Accepted February 3, 2004

## ABSTRACT

The eukaryotic genome is partitioned into chromatin domains containing coding and intergenic regions. Insulators have been suggested to play a role in establishing and maintaining chromatin domains. Here we describe the identification and characterization of two separable enhancer blocking elements located in the 5' flanking region of the chicken  $\alpha$ -globin domain, 11–16 kb upstream of the embryonic  $\alpha$ -type  $\pi$  gene in a DNA fragment harboring a MAR (matrix attachment region) element and three DNase I hypersensitive sites (HSs). The most upstream enhancer blocking element co-localizes with the MAR element and an erythroid-specific HS. The second enhancer blocking element roughly co-localizes with a constitutive HS. The third erythroid-specific HS present within the DNA fragment studied harbors a silencing, but not an enhancer blocking, activity. The 11 zinc-finger CCCTC-binding factor (CTCF), which plays an essential role in enhancer blocking activity in many previously characterized vertebrate insulators, is found to bind the two  $\alpha$ -globin enhancer blocking elements. Detailed analysis has demonstrated that mutation of the CTCF binding site within the most upstream enhancer blocking element abolishes the enhancer blocking activity. The results are discussed with respect to special features of the tissue-specific  $\alpha$ -globin gene domain located in a permanently open chromatin area.

## INTRODUCTION

The eukaryotic genome is partitioned into active and inactive chromatin domains that are likely to constitute targets for special domain-level regulatory mechanisms (1,2). It is becoming increasingly evident that maintaining chromatin domain configuration with the possible regulatory consequences should be understood in the context of nuclear compartmentalization (1–5). Chromatin insulators are con-

sidered to be among the key players in this regulation (2,6,7). Insulators are functionally defined by two properties: (i) the ability to interfere with enhancer–promoter communication and (ii) the capacity to protect a transgene against position effects caused by local chromatin structure at some of the randomly chosen integration sites.

In vertebrates, the enhancer blocking activity of insulators is largely dependent on CCCTC-binding factor (CTCF), an 11 zinc-finger nuclear factor (8,9). One of the clearest examples of CTCF-dependent enhancer blocking activity is in the imprinted *Igf2/H19* loci, where the conditional (regulated by CpG methylation) binding of CTCF to the imprinting-choice region determines the imprinting status of paternal and maternal alleles via allele-specific enhancer blocking activity and the chicken  $\beta$ -globin *cHS4* insulator (2,7,10,11).

In contrast to the domains of  $\beta$ -globin genes, the domains of  $\alpha$ -globin genes (in chicken and other vertebrates studied) remain in an open chromatin configuration in both erythroid and non-erythroid cells (1,7,12,13). This might be explained by localization of  $\alpha$ -globin gene domains in gene-dense genomic regions containing many housekeeping genes, including the so-called ‘–14’ gene overlapping the upstream area of the  $\alpha$ -globin domain (14–16).

Two clusters of DNase I hypersensitive sites (HSs) (each of them including a constitutive HS and several erythroid-specific ones) were found previously in the upstream area of the chicken  $\alpha$ -globin domain (15). Several lines of indirect evidence suggested that the first of these groups, located 11–15 kb upstream of the embryonic  $\alpha$ -type globin gene, could represent the upstream boundary of the chicken  $\alpha$ -globin domain (15). Previous observations using transient transfections demonstrated that the DNA fragments harboring these HSs possess neither promoter nor enhancer activity (15). At this point it is important to recall that the majority of the insulator elements are neutral in terms of their transcriptional influence (6). Moreover, placing some of these fragments between the SV40 promoter and a reporter gene caused strong repression of reporter gene activity, which could suggest an enhancer blocking activity, characteristic of one of the previously defined insulator properties (6,15). In addition, the most upstream erythroid-specific HS was found to possess properties of a matrix attachment region (MAR), these elements are frequently present at the borders of different

\*To whom correspondence should be addressed. Tel: +52 55 5622 5674; Fax: +52 55 5622 5630; Email: frecilla@ifc.unam.mx

tissue-specific gene domains, some of which possess insulator features, such as in the chicken lysozyme domain (15,17–19). In the present study we tested the DNA fragments harboring the above three HSs for the presence of an enhancer blocking activity. This activity was found in DNA fragments harboring the two most upstream of the three HSs studied. Furthermore, the corresponding DNA fragments were found to bind CTCF *in vitro*, and at least one of these sites interacts with CTCF *in vivo* in both erythroid and non-erythroid cells. Mutant analysis carried out on the corresponding enhancer blocking element has demonstrated that binding of CTCF is essential for the enhancer blocking activity. These findings strongly suggest that the most upstream group of HSs present in the 5' flank of the chicken  $\alpha$ -globin domain could separate distinct chromatin subdomains in a constitutively open genomic region.

## MATERIALS AND METHODS

### Plasmids and constructs

The recombinant clone containing the 6.6 kb BamHI fragment harboring the three most upstream HSs of the chicken  $\alpha$ -globin domain has been described previously (15). Four sub-fragments of the insertion were PCR-amplified: a 1.4 kb DNA fragment harboring the 5' erythroid-specific HS ( $\alpha$ EHS-1.4), 1.2 and 0.6 kb DNA fragments harboring the constitutive HS with flanking sequences of different lengths ( $\alpha$ CHS-1.2 and  $\alpha$ CHS-0.6) and a 1 kb DNA fragment harboring the 3' erythroid-specific HS ( $\alpha$ EHS-1.0). Each pair of primers contained SalI and HindIII sites necessary for cloning the PCR-amplified DNA fragments in the pAcate vector (20,21). Oligonucleotides used for amplification of the regions  $\alpha$ EHS-1.4,  $\alpha$ CHS-1.2,  $\alpha$ CHS-0.6 and  $\alpha$ EHS-1.0 were as follows. For  $\alpha$ EHS-1.4: a, 5'-CCGGGGAAGCTTGTGCGACTGTTTGCACCTCTAA-3'; b, 5'-CCGGGGAAGCTTGTGCGACTTTCATGAAGTT-3';  $\alpha$ CHS-1.2: a, 5'-CCGGGGAAGCTTGTGCGACTGCGACGTCCCTCCTTTCAAATCTAAC-3'; b, 5'-CCGGGGAAGCTTGTGCGACCTATTTTCAGATGTGATCA-3';  $\alpha$ CHS-0.6: a, 5'-CCGGGGAAGCTTGTGCGACGGGTTACAGAGAAGACTTG-3'; b, 5'-CCGGGGAAGCTTGTGCGACGCCACAAAATAGGA-3'; and  $\alpha$ EHS-1.0: a, 5'-CCGGGGAAGCTTGTGCGACGTGCTCAGCATA-3'; b, 5'-CCGGGGAAGCTTGTGCGACAATTGCCCTTCTGA-3'. The new pALucE luciferase reporter vector was constructed by digesting the pGL3-basic vector (Promega) with the BglII and BamHI restriction enzymes to obtain the luciferase gene with the poly(A) signal. This 2214 bp DNA fragment was subcloned in the pAcate vector, which was digested with BamHI to eliminate the chloramphenicol acetyltransferase (CAT) gene. The pALucE vector retains the SalI site located between the gene and the  $\beta$ /e enhancer. All tested fragments were cloned in this site.

### Cell culture

Chicken cell lines DT40 [pre-B cells transformed by avian leukosis virus (ALV)], 6C2 (transformed pre-erythroblasts) and AEV-transformed erythroblasts [line LSCCHD3 (HD3)] were grown as described previously (22,23). HeLa and COS-1 cells (African green monkey kidney cells transformed by SV40) were grown in DMEM supplemented with 10% fetal

bovine serum. Ten-day chicken embryonic red blood cells (RBCs) were cultivated as described previously (20).

### Transient enhancer blocking assay

RBCs were obtained directly from the apical veins of chicken embryos.  $1 \times 10^8$  cells were used per transfection with 6  $\mu$ g of each linearized plasmid and 6  $\mu$ g of pTk $\beta$ -gal plasmid (Stratagene). The latter served as an internal marker to normalize the efficiency of transfection. The RBCs were resuspended in 500  $\mu$ l of PBS buffer and electroporated at 500  $\mu$ F with 450 V. CAT activities were measured as described previously (20,21). In experiments with constructs bearing firefly luciferase and *Renilla* luciferase gene reporters,  $1 \times 10^8$  cells were transfected with 1  $\mu$ g of each plasmid and with 200 ng of the *Renilla* plasmid (pRL-CMV). Cells were resuspended in 500  $\mu$ l of PBS buffer and electroporated at 500  $\mu$ F and 450 V, luciferase activities were determined 48 h later using the Dual luciferase kit (Promega). Relative luciferase units were measured in a TD 20/20 luminometer (Turner Designs).

### Chicken CTCF over-expression

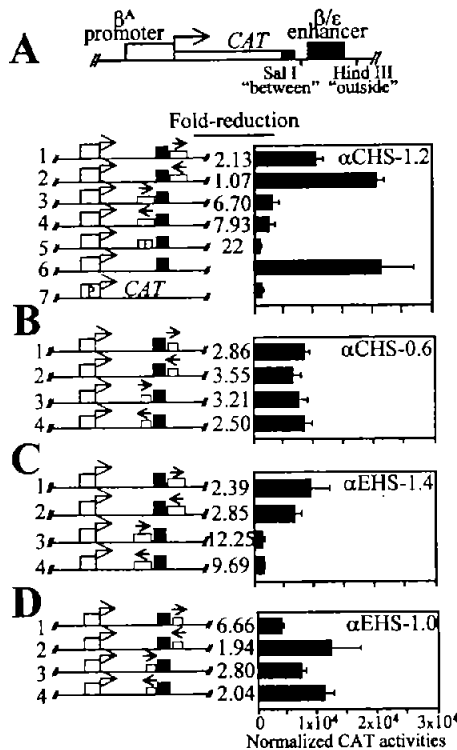
Plasmid pSG5-CTCF (kindly provided by Dr Elena Klenova) was transfected into COS-1 cells by electroporation at 100 V and 950  $\mu$ F. 24 h later, cells were lysed by freezing and thawing for 30 min at 4°C in RIPA buffer (150 mM NaCl, 1% Nonidet P-40, 0.5% deoxycholate, 0.1% SDS and 50 mM Tris-HCl pH 8.0) with 10  $\mu$ g/ml of protease inhibitors (leupeptin, PMSF and pepstatin A from Sigma). After centrifugation, lysates were stored at -70°C. This same protocol was used for preparation of HeLa total cell extracts.

### *In vitro* transcription/translation

The full-length chicken CTCF cDNA was *in vitro* transcribed and translated from the pSG5-cCTCF vector, using the TnT reticulocyte lysate-coupled *in vitro* transcription-translation system (Promega).

### Electrophoretic mobility shift assay

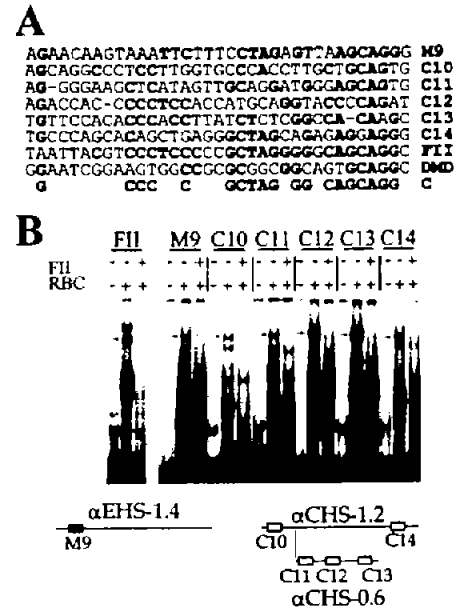
The EMSA assay was performed as previously described (8). Competitions were carried out with 100 pmol of cold oligonucleotide. The following oligonucleotides (all purified by PAGE) were used in different mobility shift experiments (in all cases only one of the complementary chains is shown): M9, 5'-TTTCCATATAGAACAATGTAAATTCCTTCCTAGAGTTAAGCAGGGCTTTTTCATAAGCTG-3'; C10, 5'-CATGGCCAGCAGGCCCTCCTTGGTGCCACCTTGCTGCAGTGGGACTGCGCTGTCCCGGA-3'; C11, 5'-CCTCAGTGAGCTGCCAGGGGAAGCTCATAGTTGCAGGATGGGAGCAGTGTGAGGTTTGTAT-3'; C12, 5'-TGTGTGCCCTCCATCAGACCACCCCTCCACCATGCAGGTACCCCAGATAACACAGTATCA-3'; C13, 5'-TGGGTGTGTCCACACCCACCTTATCTCTCGGCCACAAGCAGGCCCTCAGTGCCACAGA-3'; C14, 5'-CCAAACAGGATGCCAGCACAGCTGAGGGCTAGCAGAGAGGAGGGGAGAGT-3'; FII, 5'-CCCAGGGATGTAATTACGTCCCTCCCGCTAGGGGGCAGCAGGCGCGCT-3' (8); and DMD, 5'-TTGGTTGTAGTTGTGGAATCGGAAGTGGCCGCGGGCGGCAGTGCAGGCTCACACATCACAGCCCAGC-3' (10). The following oligonucleotides have SalI and



**Figure 2.** Transient enhancer blocking activity of the HSs located at the 5' upstream region of the chicken  $\alpha$ -globin locus. Transient transfections of linearized plasmids in chicken 10-day-old embryonic erythrocytes. The scheme of the basic construct with the  $\beta^A$  promoter and  $\beta^E$  enhancer is shown at the top of the figure. (A) The constitutive HS ( $\alpha$ CHS-1.2) tested 'between' and 'outside' in both orientations. Positions of the tested fragments relative to the enhancer are shown schematically at the left of the histogram. Construct 5 contains four copies of the FII/FIII sites of the  $\beta$ -globin insulator (striped box). Constructs 6 and 7 represent pAcatE and enhancer-less pAcat used to normalize the results. (B–D) Results obtained correspondingly with the  $\alpha$ CHS-0.6,  $\alpha$ EHS-1.4 and  $\alpha$ EHS-1.0 fragments. All designations are the same as in (A). Data shown in histograms represent the average of seven independent experiments and error bars show the SEM in all cases.

Since several putative protein-binding sites were identified around the middle of the 1.2 kb sequence in the vicinity of the constitutive HS, we decided to test a shorter DNA fragment representing the central part of the  $\alpha$ CHS-1.2 DNA fragment (see Fig. 1). Unexpectedly, this shorter 600 bp DNA fragment ( $\alpha$ CHS-0.6) showed a moderate (3-fold) reduction in CAT reporter activity in both orientations and locations (Fig. 2B). Taken together, these data show that the  $\alpha$ CHS-1.2 element possesses an enhancer blocking activity that is lost when the fragment is delimited to a 600 bp central part (see below). Instead, the  $\alpha$ CHS-0.6 fragment shows a moderate silencing activity.

In the next set of experiments the fragments harboring the two erythroid-specific HSs ( $\alpha$ EHS-1.4 and  $\alpha$ EHS-1.0; see Fig. 1) were studied. The  $\alpha$ EHS-1.4 DNA fragment was found to possess the enhancer blocking activity, in addition to a modest silencing activity. Indeed, a strong reduction in the reporter CAT gene activity was observed when the  $\alpha$ EHS-1.4 DNA fragment was placed between the  $\beta^A$  promoter and the  $\beta^E$  enhancer (Fig. 2C, constructs 3 and 4). The enhancer

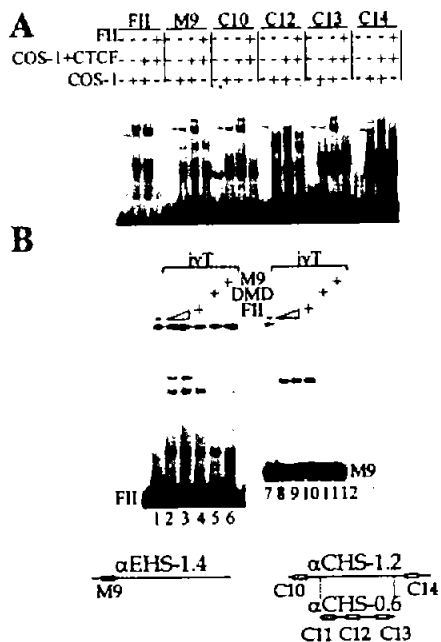


**Figure 3.** Sequence identification of conserved CTCF sites and binding activities. (A) Sequences of putative DNA binding sites for CTCF clustered in the  $\alpha$ EHS-1.4 and  $\alpha$ CHS-1.2 elements. Conserved nucleotides are shown in bold. The DMD sequence corresponds to a CTCF binding site of the differentially methylated domain from the mouse *Igf2/H19* imprinted locus (10). (B) Electrophoretic mobility shift assay using nuclear extracts from chicken RBCs. Each retarded complex was competed with 100-fold molar excess of FII unlabeled DNA. The location of each binding site is shown below.

blocking effect in this case was even more prominent than that observed for the  $\alpha$ CHS-1.2 DNA fragment. In contrast, the  $\alpha$ EHS-1.0 fragment harboring the 3' erythroid-specific HS behaves like a classical silencer, since the activity of the CAT gene was similarly reduced whether this fragment was placed between the enhancer and promoter or in the 'outside' position (Fig. 2D). Thus, the most upstream HSs,  $\alpha$ EHS-1.4 and  $\alpha$ CHS-1.2, are enhancer blocker elements, and the  $\alpha$ EHS-1.4 shows a stronger ability to interfere with enhancer-promoter action.

### CTCF interacts with the $\alpha$ EHS-1.4 and the $\alpha$ CHS-1.2 enhancer blocking elements

Since enhancer blocking activity is a property of insulators and it has been previously demonstrated that CTCF interacts with many known vertebrate insulators, we decided to analyze the DNA sequence of the region under study for the presence of potential CTCF-binding motifs (6,8,9,25). Identification of such motifs represents a difficult task, since the differential use of the 11 zinc-fingers allows CTCF to bind divergent sequences having no close similarity (8,9,25). We fixed a CTAG sequence as the starting motif of similarity with upstream and downstream CG-rich sequences and compared our sequences with the FII CTCF-binding site present in the chicken  $\beta$ -globin insulator (8,25). Computer-assisted analysis allowed us to identify six potential binding sites for CTCF (named M9 and C10–14; Fig. 3A). The M9 site was found in the  $\alpha$ EHS-1.4 element while the  $\alpha$ CHS-1.2 DNA fragment harbored several putative CTCF-binding motifs (C10–C14).



**Figure 4.** DNA-protein interactions of overexpressed and *in vitro* transcription/translated chicken CTCF. (A) The gel shift experiments show binding of nuclear extracts from cells transfected with the empty (COS-1) and the pSG5-CTCF vectors (COS-1+CTCF) to the FII, M9 and C10–C14 labeled oligonucleotides. The retarded complexes were competed with 100-fold molar excess of the chicken  $\beta$ -globin cHS4 FII insulator site. (B) Gel-shift experiments using CTCF obtained by *in vitro* transcription/translation (ivT). Labeled FII and M9 sites were used as probes and 25-fold molar excesses of the FII, M9 and mouse DMD sites were used as cold competitors as indicated above the slots. Open triangles over lanes 2, 3 and 8, 9 represent increasing amounts of *in vitro* transcription/translated CTCF.

Most importantly, no apparent CTCF-binding motifs were found in the  $\alpha$ EHS-1.0 fragment that did not show enhancer blocking activity (Fig. 2D).

Based on the above analysis, we synthesized six pairs of 60 bp long oligonucleotides harboring the putative CTCF binding sites (Fig. 3A). These fragments were tested in gel-shift assays with nuclear extracts prepared from chicken RBCs (Fig. 3B). Using the FII site from the chicken  $\beta$ -globin insulator as a positive control and a specific CTCF competitor, we found that all the predicted sites are able to produce a retarded complex to different extents, similar to the one seen with FII. These retarded complexes were efficiently competed with 100-fold molar excess of the FII site (Fig. 3B). Similarly, the CTCF-binding motif from the DMD region of the mouse *Igf2/H19* locus competed the retarded bands formed by the M9 motif in a specific fashion, while the oligonucleotide bearing an Sp1 binding site behaved similarly to a non-specific competitor (data not shown).

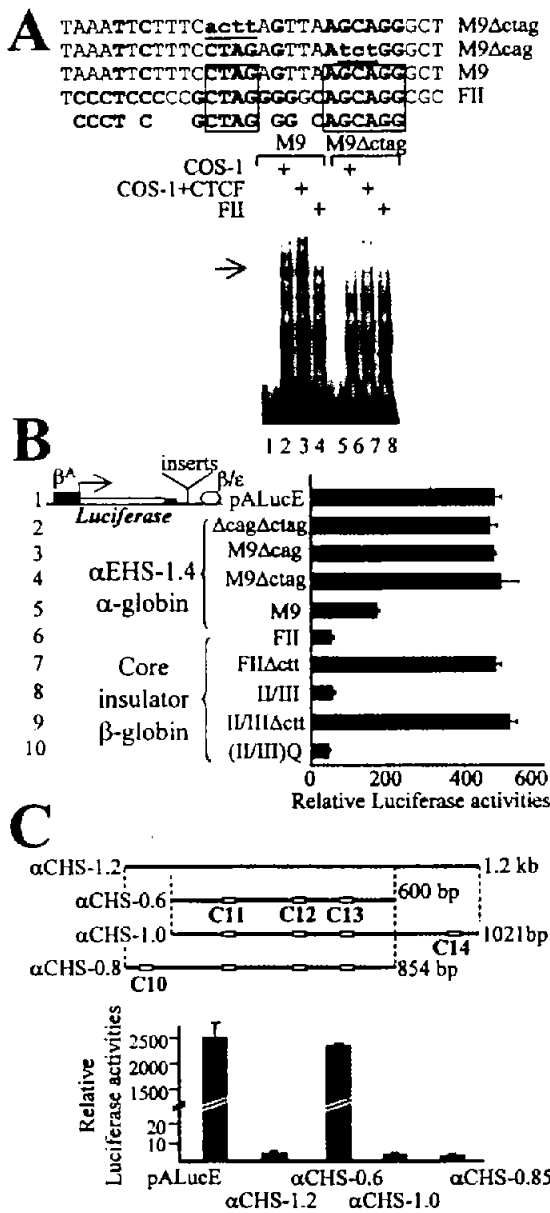
To confirm our conclusions, the extracts with enriched CTCF concentrations [prepared from COS-1 cells transiently transfected with a chicken CTCF expressing plasmid (pSG5-CTCF expression vector)] were made. Two parallel sets of band-shift experiments were carried out using extracts from COS-1 cells transfected with empty pSG5 and pSG5-CTCF plasmids. The results of gel-shift experiments confirmed the binding of CTCF to the M9 and C10–C14 sequences (Fig. 4A).

Indeed, with all binding sequences tested the intensity of one of the retarded bands (indicated by arrows in Fig. 4A) increased significantly when the CTCF-rich extracts were used, and consistently those bands were all effectively competed by the cold FII site. Based on our observation demonstrating stronger enhancer blocking activity of the  $\alpha$ EHS-1.4 element (Fig. 2C), we decided to focus our attention on the characterization of the M9 binding site. Thus, to further confirm that CTCF binds to the M9 motif, we carried out a gel-shift assay using the FII and M9 sequences as probes and *in vitro* transcribed/translated full-length CTCF (Fig. 4B). A specific retarded complex was observed in each case and the retarded band was competed, although with different efficiencies, by the FII, DMD and M9 sequences. To estimate differences in CTCF-binding affinities between the FII and M9 sites, competitions were done at low molar excess of cold competitors (25-fold). These experiments demonstrated that, at low concentrations, FII was not able to efficiently compete CTCF bound to the M9 sequence. In contrast, M9 clearly competed CTCF bound to FII (Fig. 4B, compare lane 4 and 10). This observation requires further investigation but may reflect differences in binding affinities or in the mode of binding that can correlate with differences in enhancer blocking strength of different insulators, as recently observed for the human and mouse  $\beta$ -globin insulators (25). In those domains the enhancer blocking activity is, at least partially, dependent on CTCF affinity for these DNA binding sequences (25). Thus we confirmed the *in vitro* interaction of CTCF with the M9 site of the  $\alpha$ EHS-1.4 enhancer blocking element.

#### Enhancer blocking activity depends on CTCF binding sites and additional sequences

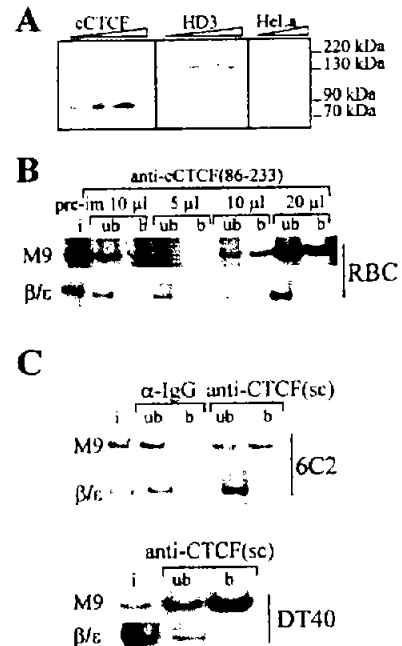
To further study a possible role of CTCF in enhancer blocking activity, we tested the binding capacity of the M9 site alone and three mutant versions of this element lacking important conservative elements (Fig. 5A and B). The specificity of CTCF binding to M9 was confirmed by loss of CTCF interaction to the mutant M9 $\Delta$ cag, M9 $\Delta$ tag and the double mutant M9 $\Delta$ cag $\Delta$ tag) by gel-shift assay (Fig. 5A and data not shown). In order to functionally characterize the M9 CTCF binding site we used luciferase gene as a reporter, allowing better quantitation of the results. The  $\beta$ -globin insulator FII site known to bind CTCF, the FII/FIII sites in single and four copies and the mutant versions of those sites (Fig. 5B, constructs 6–10) were used as controls. The results presented in Figure 5B demonstrate that the M9 site alone (construct 5) possesses a prominent enhancer blocking effect, although less strong than the enhancer blocking effect of the FII site. Most importantly, mutations abolishing binding of CTCF to the M9 site (Fig. 5A) resulted in a complete loss of the enhancer blocking activity (constructs 2–4). Thus, the enhancer blocking activity of the  $\alpha$ EHS-1.4 element is dependent at least in part on CTCF.

In the next set of experiments the enhancer blocking activity of the CTCF binding sites present in the 1.2 kb DNA fragment was studied (Fig. 5C). As expected (see Fig. 2B) the central 600 bp fragment did not possess any positional enhancer blocking activity (Fig. 5C, construct  $\alpha$ CHS-0.6). In contrast, when either 5' or 3' flanking sequences were added to this fragment, we recovered a significant enhancer blocking



**Figure 5.** Detailed analysis of the enhancer blocking activities of the M9 site in the  $\alpha$ EHS-1.4 and  $\alpha$ CHS-1.2 enhancer-blocking elements using linearized pALucE luciferase reporter plasmids transiently transfected in chicken RBCs. (A) CTCF binding to the M9 and mutant M9 $\Delta$ ctag sites. Each experiment was carried out with extracts from normal COS-1 cells and CTCF-overexpressing COS-1 cells. Lanes 4 and 8 show competition with 100-fold molar excess of the FII motif. (B) Enhancer blocking activity of the M9 site of the  $\alpha$ EHS-1.4 element. Results are presented as relative luciferase activities and error bars show the SEM. Data from constructs 1–5 are the average of seven independent experiments and constructs 6–10 of four. (C) Sub-division of the  $\alpha$ CHS-1.2 fragment harboring the second enhancer blocking element. In the scheme above the histogram, putative CTCF-binding elements are shown with open boxes indicated as C10–C14. Data in the histogram represent an average from six independent experiments.

activity equivalent to the activity of the entire  $\alpha$ CHS-1.2 element (Fig. 5C, constructs  $\alpha$ CHS-1.0,  $\alpha$ CHS-0.85 and  $\alpha$ CHS-1.2). These results suggest that the C10 and C14 CTCF



**Figure 6.** *In vivo* interaction of CTCF with the M9 binding site in erythroid and non-erythroid cells. To further confirm the interaction of CTCF with the M9 site, a ChIP experiment was carried out. (A) Characterization of the chicken cCTCF(86–233) antibody by western blot. Bacterially produced peptide [cCTCF(86–233)] and full-length peptide from total HD3 and HeLa cell extracts. (B) ChIP assay using chicken RBCs. PCR amplification of the test fragments (M9 and  $\beta^A/\epsilon$  enhancer) using as template DNA input fraction (i) and DNA recovered from immunoprecipitated fractions bound (b) versus unbound (ub) by the chicken CTCF-specific antibody. In control experiments a preimmune serum was used (pre-im, 10  $\mu$ l). Note the increase of PCR-amplified M9 test fragment in bound fractions obtained with increased amount of CTCF-specific antibody. (C) ChIP assay in the same conditions except that the ChIP was done with a commercial CTCF antibody [anti-CTCF(sc)]. CTCF binds to the M9 site in chicken pre-erythroblast 6C2 cells and in non-erythroid DT40 cells. An irrelevant antibody was also used for ChIP showing no amplification product ( $\alpha$ -IgG). These data are representative of three independent experiments.

binding sites, possibly in combination with other sites, are relevant for the overall  $\alpha$ EHS-1.2 enhancer blocking activity.

### CTCF binds *in vivo* to the M9 site

To find out if CTCF is bound *in vivo* to the M9 sequence, a ChIP assay was employed. To carry out these experiments, we first produced chicken-specific polyclonal antibodies against a non-conserved region of chicken CTCF, adjacent to the zinc-fingers [cCTCF(86–233); Fig. 6A]. Antibody specificity was confirmed by western blot (Fig. 6A). These antibodies were used to precipitate the chromatin fragments bearing bound CTCF. The concentration of the M9 motif in the precipitated fraction was compared with the concentration of a chicken reference sequence, the chicken  $\beta^A/\epsilon$  enhancer element which does not interact with CTCF *in vivo* (21,26). The relative representations of both fragments in the immunoprecipitated DNA fraction were assayed using PCR with two sets of primers permitting the amplification of a 120 bp DNA fragment bearing the M9 motif and a 301 bp DNA fragment bearing the  $\beta^A/\epsilon$  enhancer (Fig. 6B and C). The ChIP assay

was first carried out using chicken RBCs. The fragment bearing the M9 motif was clearly enriched (as compared to the fragment bearing the  $\beta^A/\epsilon$  enhancer) in the chromatin fraction precipitated by our anti-cCTCF(86–233) antibody (Fig. 6B). Importantly, when 10  $\mu$ l of pre-immune fraction was used, no enrichment was observed. The same results were obtained when commercial anti-CTCF antibodies were used (Fig. 6C). ChIP was repeated with the chicken pre-erythroblast 6C2 transformed cell line and with non-erythroid DT40 cells (pre-B cell line). These experiments demonstrated that CTCF is bound *in vivo* to the M9 motif in all the cell types tested (Fig. 6B and C).

## DISCUSSION

This study describes the identification and characterization of two new enhancer blocking elements located in the upstream area of the chicken  $\alpha$ -globin domain with a possible function to separate control elements within the same open chromatin domain or sub-domains. It should be underlined that the present knowledge about vertebrate insulators is based on the results of detailed analysis of only a few examples (7,10,11). Characterization of any new insulators may significantly extend our knowledge about these important genomic elements. This seems to be the case as far as the results of the present study are concerned. The chicken  $\alpha$ -globin domain HSs are composed of two separable enhancer blocking activities. This is a distinguishing feature of the chicken  $\alpha$ -globin domain upstream insulator, as compared to the insulators present at upstream and downstream flanks of the vertebrate  $\beta$ -globin domains (7,8). The second unexpected, and thus distinguishing, feature of one of the two enhancer blocking elements described in the present work is that it co-localizes with an initially defined erythroid-specific HS. There are previously described insulators which co-localize with tissue-specific HSs such as the mouse 3'HS1 (25). We thus decided to characterize more precisely the enhancer blocking element localized in the erythroid-specific HS. This element (referred to throughout this paper as  $\alpha$ EHS-1.4, including the M9 motif) turned out to be a typical vertebrate enhancer blocker whose activity depends on the CTCF recognition sequence, which was found associated *in vivo* with the M9 motif in both immature erythroid and non-erythroid cells (DT40 cell line). It is also worth mentioning that this element co-localizes with the previously defined MAR sequence (15). Co-localization with MAR elements has been noted for some vertebrate insulators studied (2,17,19). Moreover it has recently been described that CTCF is able to bind to the nuclear matrix; the fact that CTCF is found *in vivo* in the DT40 cell line could mean that it plays a relevant structural role in the MAR function in addition to its enhancer blocking activity (27).

To understand a possible significance of the location of an enhancer blocker element in an erythroid-specific HS, one should consider special features of  $\alpha$ -globin domains as compared to  $\beta$ -globin domains (2,12,28). In contrast to the latter, the  $\alpha$ -globin domains are located in permanently open chromatin areas which contain several non-globin genes, including a housekeeping gene located very close to the  $\alpha$ -globin gene cluster and overlapping the  $\alpha$ -globin gene domain (14). To this end the original prediction is that the  $\alpha$ -globin

domain constitutes a typical example of weak genomic domains that do not have defined structural boundaries (2). Nevertheless, the transcriptional status of the domain as a whole is regulated by special, although as yet poorly understood mechanisms. Operation of these domain-level regulatory systems became especially clear when it was demonstrated that the domain as a whole (with flanking sequences) is characterized by increased levels of site-specific histone acetylation in erythroid cells but not in cells of other lineages (13,29,30). The LCR-like regulatory element of mammalian  $\alpha$ -globin domains was found 40 kb upstream of the first  $\alpha$ -globin gene in one of the so-called '-14' gene introns (14,31). Analysis of the distribution of conserved regions in the upstream areas of vertebrate  $\alpha$ -globin domains has permitted the prediction of the positions of similar regulatory elements in the chicken genome [about 20 kb upstream to  $\pi$  gene (32)]. The intriguing fact is that this regulatory region is located upstream of the enhancer blockers characterized in the present study. If indeed this regulatory region is involved in control of  $\alpha$ -globin gene expression, the enhancer blockers should be somehow inactivated in cells expressing  $\alpha$ -globin genes. Conditional insulators have been described in the literature (10,11,33,34). In the majority of the known cases, their activity is regulated by DNA methylation, which interfere with the CTCF binding. It is, however, easy to imagine a situation when CTCF binding would be prevented by binding of other sequence-specific factors (in our case erythroid-specific transcription factors or ubiquitous factors) close to the CTCF recognition site. To this end it might be of importance that the genomic region studied in the present work contains multiple recognition motifs for different erythroid-specific factors (15). However, our results indicate that CTCF is capable of interacting with the M9 recognition site by itself according to the *in vitro* transcription/translation results, and also in the presence of other proteins according to the gel-shift experiments performed with erythroid nuclear extracts.

Although the second enhancer blocking element ( $\alpha$ CHS-1.2) identified in our study co-localizes roughly with the constitutive HS, the short fragment ( $\alpha$ CHS-0.6) containing this constitutive HS does not possess enhancer blocking activity by itself, and binding of CTCF to the recognition motifs located at the flanks of the  $\alpha$ CHS-1.2 DNA fragment might be affected by binding of erythroid-specific transcription factors to the recognition sequences present within the two erythroid-specific HSs. If indeed erythroid-specific transcription factors can compete with the CTCF binding, a simple increase in the concentration of these factors occurring in the course of differentiation of erythroblast precursors would inactivate the enhancer blocker.

An alternative model has to do with the location of the newly defined enhancer blockers, and the bi-directional presence of intergenic transcripts over the studied area (Fig. 1) (14,16). One possible speculation in terms of the location and presence of two enhancer blockers in the 5' upstream area of the chicken  $\alpha$ -globin domain is related to those intergenic transcripts. Each opposing transcript should have regulatory elements that drive their initiation and elongation. Based on such supposition, the newly defined  $\alpha$ EHS-1.4 and  $\alpha$ CHS-1.2 elements could block non-specific regulatory signals affecting those transcripts. Such an idea

seems attractive since one insulator could act on the 5' strand and the other on the opposite strand based on the transcript directions and we have demonstrated that the enhancer blocking activity of these elements is independent of their orientation (see Fig. 2) (14,16). There are experiments currently in progress in our laboratory to test this hypothesis. In addition, we should recall that there is recent evidence supporting the presence of intergenic transcripts in chromatin domains and sub-domains that contribute to their formation and developmental regulation (35,36).

Interestingly, detailed comparative analysis of histone acetylation levels on the 5' side of the human, mouse, pufferfish and chicken  $\alpha$ -globin gene clusters have revealed histone acetylation transitions (29). For the chicken  $\alpha$ -globin domain this group has found the highest levels of H4 Lys-5 and H3 Lys-14 hyperacetylation overlapping the  $\alpha$ EHS-1.4 and  $\alpha$ CHS-1.2 enhancer blocking elements. They have defined an area with a high level of histone acetylation, which extends from 6 kb upstream of the  $\alpha$ EHS-1.4 [primer C6 (29)] through 200 bp downstream of the  $\alpha$ CHS-1.2 element [primer C7 (29)]. This area overlaps the three HSS under study but the transition on histone hyperacetylation is located 6 kb upstream of the  $\alpha$ EHS-1.4 (primer C6). Taking all this data into account we can conclude that the newly defined enhancer blockers are located within a zone of high levels of histone acetylation and 6 kb downstream of a sharp histone acetylation transition which could correspond to the boundary of the domain [see figures 6 and 7 from Anguita *et al.* (29)].

At this point we can only speculate how these enhancer blocking elements work in the chicken  $\alpha$ -globin domain. Nevertheless, the proposed models may constitute a starting point to design further experiments aimed at better understanding the regulation of  $\alpha$ -globin gene expression.

## ACKNOWLEDGEMENTS

We are indebted to the support of Alicia González Manjarrez, Jaime Mas-Oliva and members of their laboratories, Gary Felsenfeld for the chicken CHS4  $\beta$ -globin insulators plasmids, and Amy Weinmann and Peggy Farnham for ChIP advice. We are also grateful to Catherine M. Farrell, Paul Delgado-Olguín and Héctor Rincón-Arango for critical reading of the manuscript, useful suggestions and constant support. We thank Georgina Guerrero Avendaño for excellent technical assistance. This work was supported by the Dirección General de Asuntos del Personal Académico-UNAM (IN203200), the Consejo Nacional de Ciencia y Tecnología, CONACyT (33863-N), the Third World Academy of Sciences (TWAS, Grant 01-055 RG/BIO/LA) and Fundación Miguel Alemán. V.V.-G. was supported by a CONACyT Ph.D. fellowship. The work of S.V.R. was supported by a Grant on Physico-Chemical Biology given by the Presidium of the Russian Academy of Sciences.

## REFERENCES

- Recillas-Targa, F. and Razin, S.V. (2001) Chromatin domains and regulation of gene expression: familiar and enigmatic clusters of chicken globin genes. *Crit. Rev. Eukaryot. Gene Exp.*, **11**, 227–242.
- Razin, S.V., Farrell, C.M. and Recillas-Targa, F. (2003) Genomic domains and regulatory elements operating at the domain level. *Int. Rev. Cytol.*, **226**, 63–125.
- Gerasimova, T.I., Byrd, K. and Corces, V.G. (2000) A chromatin insulator determines the nuclear localization of DNA. *Mol. Cell*, **6**, 1025–1035.
- Cremer, T. and Cremer, C. (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.*, **2**, 292–301.
- Ishii, K., Arib, G., Lin, C., Van Houwe, G. and Laemmli, U.K. (2002) Chromatin boundaries in budding yeast: the nuclear pore connection. *Cell*, **109**, 552–562.
- West, A.G., Gaszner, M. and Felsenfeld, G. (2002) Insulators: many functions, many mechanisms. *Genes Dev.*, **16**, 271–288.
- Burgess-Beusse, B., Farrell, C., Gaszner, M., Litt, M., Mutskov, V., Recillas-Targa, F., Simpson, M., West, A. and Felsenfeld, G. (2002) The insulation of genes from external enhancers and silencing chromatin. *Proc. Natl Acad. Sci. USA*, **99** Suppl 4, 16433–16437.
- Bell, A.C., West, A.G. and Felsenfeld, G. (1999) The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell*, **98**, 387–396.
- Ohlsson, R., Renkawitz, R. and Lobanov, V. (2001) CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends Genet.*, **17**, 520–527.
- Hark, A.T., Schoenherr, C.J., Katz, D.J., Ingram, R.S., Levorso, J.M. and Tilghman, S.M. (2000) CTCF mediates methylation-sensitive enhancer-blocking activity at the *H19/Igf2* locus. *Nature*, **405**, 486–489.
- Bell, A.C. and Felsenfeld, G. (2000) Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature*, **405**, 482–485.
- Recillas-Targa, F. (2000) The chicken  $\alpha$ - and  $\beta$ -globin gene domains and their chromatin organization. *Cell. Mol. Biol. Lett.*, **5**, 451–467.
- Litt, M.D., Simpson, M., Recillas-Targa, F., Prioleau, M.-N. and Felsenfeld, G. (2001) Transitions in histone acetylation reveal boundaries of three separately regulated neighboring loci. *EMBO J.*, **20**, 2224–2235.
- Vyas, P., Vickers, M.A., Picketts, D.J. and Higgs, D.R. (1995) Conservation of position and sequence of a novel, widely expressed gene containing the major human  $\alpha$ -globin regulatory element. *Genomics*, **29**, 679–689.
- Razin, S.V., Shen, K., Ioudinkova, E. and Scherrer, K. (1999) Functional analysis of DNA sequences located within a cluster of DNase I hypersensitive sites colocalizing with a MAR element at the upstream border of the chicken  $\alpha$ -globin gene domain. *J. Cell. Biochem.*, **74**, 38–49.
- Sjakste, N., Iarovaia, O.V., Razin, S.V., Linares-Cruz, G., Sjakste, T., Le Gac, V., Zhao, Z. and Scherrer, K. (2000) A novel gene is transcribed in the chicken  $\alpha$ -globin gene domain in the direction opposite to the globin genes. *Mol. Gen. Genet.*, **262**, 1012–1021.
- Stief, A., Winter, D.M., Strätling, W.H. and Sippel, A.R. (1989) A nuclear DNA attachment element mediates elevated and position-independent gene activity. *Nature*, **341**, 343–345.
- Recillas-Targa, F., Razin, S.V., De Moura Gallo, C.V. and Scherrer, K. (1994) Excision close to matrix attachment regions of the entire chicken  $\alpha$ -globin gene domain by nuclease S1 and characterization of the framing structures. *Proc. Natl Acad. Sci. USA*, **91**, 4422–4426.
- Antes, T.J., Namciu, S.J., Fournier, R.E.K. and Levy-Wilson, B. (2001) The 5' boundary of the human apolipoprotein B chromatin domain in intestinal cells. *Biochemistry*, **40**, 6731–6742.
- Recillas-Targa, F., Bell, A.C. and Felsenfeld, G. (1999) Positional enhancer-blocking activity of the chicken  $\beta$ -globin insulator in transiently transfected cells. *Proc. Natl Acad. Sci. USA*, **96**, 14354–14359.
- Reitman, M. and Felsenfeld, G. (1988) Mutational analysis of the chicken  $\beta$ -globin enhancer reveals two positive-acting domains. *Proc. Natl Acad. Sci. USA*, **85**, 6267–6271.
- Prioleau, M.-N., Nony, P., Simpson, M. and Felsenfeld, G. (1999) An insulator element and condensed chromatin region separate the chicken  $\beta$ -globin locus from an independently regulated erythroid-specific folate receptor gene. *EMBO J.*, **18**, 4035–4048.
- Beug, H., Palmieri, S., Freudenstein, C., Zentgraf, H. and Graf, T. (1982) Hormone-dependent terminal differentiation in vitro of chicken erythroleukemia cells transformed by ts mutants of avian erythroblastosis virus. *Cell*, **28**, 907–919.
- Weinmann, A.S., Yan, P.S., Oberley, M.J., Huang, T.H. and Farnham, P.J. (2002) Isolating human transcription factor targets by coupling chromatin



- immunoprecipitation and CpG island microarray analysis. *Genes Dev.*, **16**, 235–244.
25. Farrell, C.M., West, A.G. and Felsenfeld, G. (2002) Conserved CTCF insulator elements flank the mouse and human  $\beta$ -globin loci. *Mol. Cell Biol.*, **22**, 3820–3831.
  26. Recillas-Targa, F., Pikaart, M.J., Burgess-Beusse, B., Bell, A.C., Litt, M.D., West, A.G., Gaszner, M. and Felsenfeld, G. (2002) Position-effect protection and enhancer blocking by the chicken  $\beta$ -globin insulator are separable activities. *Proc. Natl Acad. Sci. USA*, **99**, 6883–6888.
  27. Dunn, K.L., Zhao, H. and Davie, J.R. (2003) The insulator binding protein CTCF associates with the nuclear matrix. *Exp. Cell Res.*, **288**, 218–223.
  28. Knezetic, J.A. and Felsenfeld, G. (1989) Identification and characterization of a chicken  $\alpha$ -globin enhancer. *Mol. Cell Biol.*, **9**, 893–901.
  29. Anguita, E., Johnson, C.A., Wood, W.G., Turner, B.M. and Higgs, D.R. (2001) Identification of a conserved erythroid specific domain of histone acetylation across the  $\alpha$ -globin gene cluster. *Proc. Natl Acad. Sci. USA*, **98**, 12114–12119.
  30. Noma, K.-I., Allis, D.C. and Grewal, S.I.S. (2001) Transition in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. *Science*, **293**, 1150–1155.
  31. Higgs, D.R., Wood, W.G., Jarman, A.P., Sharpe, J., Lida, J., Pretorius, I.-M. and Ayyub, H. (1990) A major positive regulatory region located far upstream of the human  $\alpha$ -globin gene locus. *Genes Dev.*, **4**, 1588–1601.
  32. Flint, J., Tufarelli, C., Peden, J., Clark, K., Daniels, R.J., Hardison, R., Miller, W., Philipsen, S., Tan-Un, K.C., McMorrow, T. *et al.* (2001) Comparative genome analysis delimits a chromosomal domain and identifies key regulatory elements in the alpha globin cluster. *Hum. Mol. Genet.*, **15**, 371–382.
  33. Chao, W., Huynh, K.D., Spencer, R.J., Davidow, L.S. and Lee, J.T. (2002) CTCF, a candidate *trans*-acting factor for X-inactivation choice. *Science*, **295**, 345–347.
  34. Lutz, M., Burke, L.J., LeFevre, P., Myers, F.A., Thorne, A.W., Crane-Robinson, C., Bonifer, C., Filippova, G.N., Lobanekov, V. and Renkawitz, R. (2003) Thyroid hormone-regulated enhancer blocking cooperation of CTCF and thyroid hormone receptor. *EMBO J.*, **22**, 1579–1587.
  35. Ashe, H.L., Monks, J., Wijgerde, M., Fraser, P. and Proudfoot, N.J. (1997) Intergenic transcription and transinduction of the human  $\beta$ -globin locus. *Genes Dev.*, **11**, 2494–2509.
  36. Gribnau, J., Diderich, K., Pruzina, S., Calozolari, R. and Fraser, P. (2000) Intergenic transcription and developmental remodeling of chromatin subdomains in the human  $\beta$ -globin locus. *Mol. Cell*, **5**, 377–386.

# CARACTERÍSTICAS Y FORMACIÓN DE UN DOMINIO TRANSCRIPCIONALMENTE ACTIVO EN EUCARIOTES\*

Viviana Valadez Graham<sup>1</sup> y Félix Recillas Targa<sup>1</sup>

## RESUMEN

La necesidad funcional de compactar al genoma de una célula eucariota en cromatina para contenerlo dentro del núcleo restringe el acceso al ADN, de proteínas que regulan la actividad genética. Tanto los nucleosomas individuales como la organización del genoma en estructuras más complejas (como los cromosomas), limitan el acceso a factores de transcripción y en consecuencia reprimen la actividad genética. Es necesario por lo tanto que exista un remodelaje regulado de la cromatina que permita la accesibilidad de las proteínas a sus secuencias blanco en el ADN y así se lleve a cabo la expresión eficiente y ordenada de un gen. Para favorecer la expresión génica se ha propuesto que los genes se encuentran organizados en dominios dentro del genoma, los cuales son unidades transcripcionales cuya expresión es regulada de manera independiente por elementos inherentes a ellos, y cuya formación proponemos que representa uno de

los primeros niveles de regulación en células eucariotas a lo largo del desarrollo.

**PALABRAS CLAVE:** Cromatina, nucleosoma, dominios, sitios de hipersensibilidad a la DNasa I, LCR.

## ABSTRACT

In higher eukaryotes the genome is packaged into a compact and repressive chromatin structure. To allow small portion of the genome to be expressed in a given cell-type, tissue or moment, it is necessary a series of ordered chromatin remodeling events. On recent years it is more generally accepted that eukaryotic genome is organized into structural and functional discrete domains. In the present work we propose that chromatin domain formation represents one of the first steps in genome remodeling influencing the control of gene activation.

**KEY WORDS:** Chromatin, nucleosome, domains, DNase I hypersensitive sites, LCR.

## INTRODUCCIÓN

El genoma de una célula eucariota se encuentra asociado a proteínas llamadas histonas, las cuales contribuyen a empaquetar el ADN. Esta asociación del ADN con proteínas denominada cromatina, permite que la información genética pueda ser contenida dentro del núcleo. Hasta hace algunos años se estudiaba a la estructura de la cromatina y la activación transcripcional como dos procesos biológicos independientes, de hecho se pensaba

que la cromatina no jugaba un papel esencial dentro de la regulación transcripcional y que su papel era meramente estructural. Actualmente se entiende que la activación transcripcional de los genes de células eucariotas, se encuentra íntimamente asociada a cambios en la arquitectura de la cromatina. El empaquetamiento del ADN constituye *per se* una estructura represora que impide la transcripción genética, pero que a su vez es capaz de remodelarse para permitir la expresión

regulada de los genes. Este remodelaje en la cromatina favorece o impide de manera regulada el ensamblaje del complejo de iniciación de la transcripción asociado a la ARN polimerasa II en las regiones promotoras de genes, repercutiendo en la expresión de los mismos. En un nivel más elevado de complejidad, la transcripción de los genes eucariotes es modulada además por elementos distantes, es decir, que se encuentran alejados de las regiones promotoras de los mis-

\*Recibido: 6 de febrero de 2001. Aceptado: 17 de julio de 2001.

<sup>1</sup>Departamento de Genética Molecular, Instituto de Fisiología Celular, UNAM. Apartado Postal 70-242, C. P. 04510. México D. F. Fax 5622-5630. Correo E: frecilla@ifisiol.unam.mx

mos. Estos elementos en posición *cis* son sitios de unión a factores que en cooperación con diversos complejos son capaces de regular la transcripción genética positiva o negativamente y posiblemente contribuyen también a la remodelación de la cromatina.

En el contexto de la cromatina, estos sitios de unión a factores se encuentran en regiones que son hipersensibles al corte por nucleasas, como por ejemplo la DNasa I. La hipersensibilidad al corte por estas enzimas correlaciona con una estructura más abierta o accesible de la cromatina, lo que permite la unión de diferentes factores de transcripción a sus secuencias blanco en el ADN. En ocasiones, cuando los genes son activamente transcritos, esta accesibilidad al corte por endonucleasas puede extenderse por varias kilobases pudiendo incluir a uno o varios genes. A las regiones de la cromatina que se encuentran abiertas en un tiempo y espacio determinado del desarrollo celular se les ha denominado dominios transcripcionalmente activos.

En años recientes se ha propuesto que estos dominios tienen límites bien definidos y la transcripción del gen o genes contenidos en ellos es regulada diferencialmente a lo largo del desarrollo por elementos situados dentro de estos límites. Existen además, evidencias citológicas que apoyan el concepto de dominios, siendo la más evidente lo que ocurre durante la expresión genética en los cromosomas politénicos de las glándulas salivales de larvas de *Drosophila*. Al observar al microscopio los cromosomas politénicos, estos presentan un patrón de bandas e interbandas. En las bandas se encuentran los genes, los cuales al ser expresados generan una estructura conocida como "puff", la cual se ob-

serva como un ensanchamiento de la banda representativo de un relajamiento en la estructura de la cromatina. Cabe señalar que estas estructuras o "puffs" están acotados, es decir, que no invaden bandas adyacentes y que recientemente se ha propuesto, contribuyen a la formación y mantenimiento de un dominio transcripcionalmente activo. Así pues, se ha comprobado que en los "puffs" existe un dominio que está siendo regulado y transcrito por elementos inherentes a él y que no invaden otros dominios gracias a la presencia de sus límites.

Además de las evidencias bioquímicas y citológicas mencionadas, existe actualmente un cúmulo de datos experimentales recientes que apoyan el concepto de una organización del genoma eucariote en dominios transcripcionalmente activos. En esta revisión se analizarán los elementos y mecanismos que regulan la apertura y el cerrado de regiones específicas en el genoma. De igual manera se pondrá de manifiesto que el remodelaje de la cromatina es esencial para la formación de un dominio transcripcionalmente activo. Se discutirá primeramente aspectos relacionados sobre las proteínas involucradas en la compactación de la información genética, en segundo lugar se describirán los elementos de regulación de la transcripción, posteriormente se hará un análisis detallado sobre los límites de los dominios y por último se discutirá la manera cómo cooperan de manera coordinada todos estos elementos para formar, mantener y regular la expresión de los genes contenidos en un dominio. A su vez, se propone que la formación de un dominio representa probablemente uno de los niveles iniciales dentro de la regulación de la expresión genética.

## ESTRUCTURA DE LA CROMATINA

Para comprender cómo se forma un dominio transcripcionalmente activo, es necesario entender el contexto en el cual el ADN está estructurado en forma de cromatina y su relación con el núcleo celular. Si se extiende de manera lineal la molécula de ADN contenida en el interior del núcleo de una célula eucariota, su longitud varía entre 2 y 4 metros; tomando en cuenta que el diámetro del núcleo oscila entre 10 y 20 micras, se hace evidente que debe existir un sistema de empaquetamiento que compacte el genoma celular de manera que pueda ser contenido dentro del núcleo. Para empaquetar el ADN, la célula se vale de unas proteínas llamadas histonas las cuales forman un octámero cilíndrico sobre el cual el ADN es enrollado para formar lo que se conoce como nucleosoma. El nucleosoma tiene un diámetro de 6.5 nanómetros (nm), el cual está compuesto de un tetrámero central formado por dos copias de las histonas H3 y H4, flanqueado por dos heterodímeros de las histonas H2A y H2B. Las histonas insertan residuos de argininas en el surco menor del ADN que está en contacto con el octámero. Esta unión se estabiliza gracias a los extremos amino-terminales de las histonas, los cuales quedan expuestos por fuera del cilindro de histonas y son ricos en residuos de lisina. La abundancia en residuos de lisina en los extremos amino-terminales genera una carga neta positiva, la cual interacciona con la carga negativa del ADN estabilizando la unión del ADN al octámero de histonas, de esta manera el ADN queda enrollado sobre el octámero (1).

La estructura nucleosomal es considerada como el primer nivel estructural de empaquetamiento, donde 146 pares de bases (pb) de

ADN se enrollan en 1.65 vueltas alrededor del octámero de histonas. Los cilindros de histonas van posicionándose uno tras otro regularmente a lo largo del ADN, de manera independiente de secuencias nucleotídicas específicas formando una cadena de polinucleosomas conocida como collar de perlas o fibra de 10 nm.

En el siguiente nivel de estructuración de la cromatina, la histona H1, juega un papel tanto estructural como de regulación. Primero se toma en consideración la contribución de la histona H1 a la estructuración de la cromatina, para posteriormente presentar las evidencias de la participación de esta histona en la regulación transcripcional. La histona H1 contribuye a la formación de lo que se conoce como cromatosoma, el cual contiene al octámero de histonas, a la histona H1 y abarca 168 pb de ADN (20 pb más por la interacción con la histona H1). La histona H1 se localiza en la región internucleosomal e interactúa con el ADN que entra y sale del nucleosoma (ADN de unión), tiene un papel estructural fundamental, ya que contribuye a que la fibra de 10 nm escale al siguiente nivel de compactación del ADN en cromatina: la fibra de 30 nm o solenoide. La histona H1 estabiliza la fibra de 30 nm al fomentar contactos entre nucleosomas adyacentes (2). Se han propuesto varios modelos para explicar la formación de la fibra de 30 nm y hasta la fecha el más aceptado es el que sugiere que la histona H1 coopera para formar una estructura de hélice con 6 nucleosomas por vuelta. Cabe señalar que durante la mayor parte del ciclo celular, la cromatina permanece empaquetada en fibras de 30 nm de diámetro, aunque existen niveles más complejos de compactación siendo el más alto

el de los cromosomas, pero a la fecha se desconocen los mecanismos por los cuales se alcanzan los niveles más elevados de estructuración y compactación de la cromatina.

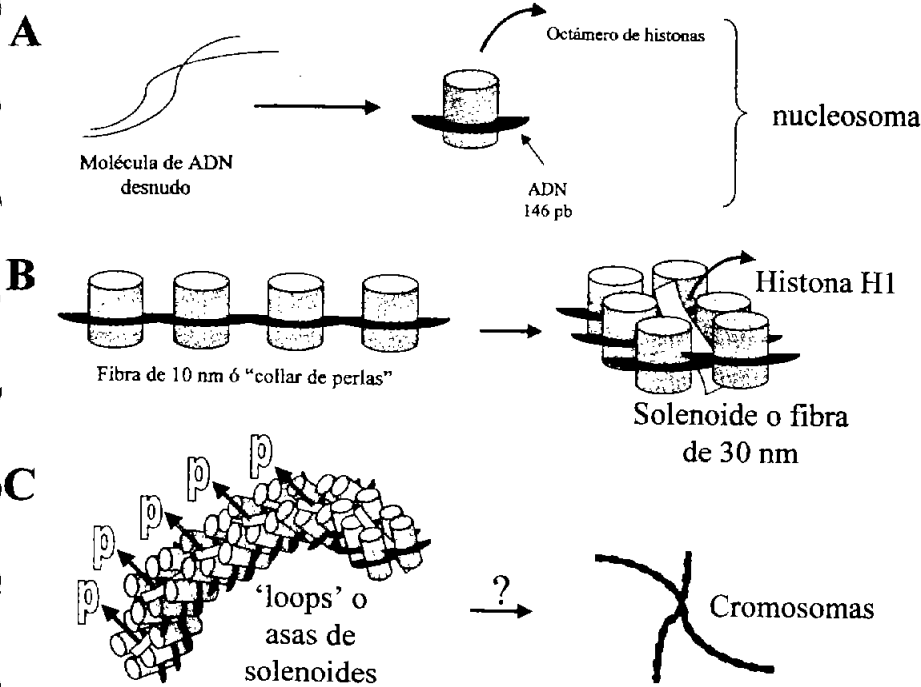
Cabe resaltar que la fibra de 30 nm es una estructura altamente compacta y consecuentemente represora. Anteriormente se pensaba que la fibra de 10 nm era el molde natural de la transcripción, ya que ésta constituía una estructura lo suficientemente abierta y accesible de la cromatina sobre la cual el complejo transcripcional podría realizar su función. Sin embargo, las siguientes evidencias sugieren que el molde real donde ocurre el remodelamiento de la cromatina para una subsecuente activación de la transcripción es la fibra de 30 nm:

—Las condiciones en las que se hicieron los estudios *in vitro* utilizando la fibra de 10 nm como molde de la transcripción, no son las mismas que existen en una célula *in vivo*. De hecho en condiciones fisiológicas la fibra de cromatina tiende a compactarse al siguiente nivel, en el cual se forman los llamados "loops" o asas de solenoides, los cuales se supone que se forman cuando las proteínas que se unen al ADN anclan la fibra de 30 nm a estructuras nucleares juntándolas, formando de esta manera el cuello de las asas. Cada asa de solenoides contiene entre 20,000 a 80,000 pb y alcanza una compactación de hasta 400 nm. Diversos cálculos teóricos han estimado que cada cromosoma humano contiene aproximadamente 2600 asas.

—Si se realiza una estimación teórica del tamaño que abarca el ADN descompactado es prácticamente imposible considerar a la fibra de 10 nm como el molde real de la transcripción debido a que en un momento dado de diferenciación de la

célula, los genes activos generarían una descompactación de la cromatina que impediría su inclusión dentro del núcleo. Además, en estudios hechos por microscopía electrónica, se ha observado que una descondensación por debajo del nivel de la fibra de 30 nm, sólo ocurre entre polimerasas que se encuentran en la etapa de extensión y cuando hay una gran densidad de éstas. Cuando las polimerasas terminan de extender y se encuentran lo suficientemente alejadas de la región que estaba siendo transcrita, la estructura de la cromatina regresa rápidamente a la fibra de 30 nm (3). Esta última observación explicaría la presencia de la histona H1 en dominios que están siendo transcritos ya que su presencia ayudaría a un pronto empaquetamiento de los genes recién transcritos al nivel de la fibra de 30 nm (Fig. 1).

Una muestra adicional de la relevancia de la histona H1 en estos niveles de estructuración es el efecto que tiene la fosforilación de H1 en 5 residuos de serina específicos, esta fosforilación ayuda a formar los cromosomas durante la mitosis. En este nivel, la compactación llega hasta 1400 nm (Fig. 1). En cromatina *de novo*, la histona H1 se encuentra moderadamente fosforilada, a medida que la cromatina madura, la histona H1 se vuelve altamente fosforilada. La fosforilación de H1 es mayor en células que se dividen rápidamente y menor en células que no están proliferando. Los niveles de fosforilación de la histona H1 son menores en G1 y mayores en la fase S del ciclo celular y en la mitosis; en esta última, la fosforilación aumenta específicamente durante la metafase, que es cuando los cromosomas se encuentran condensados. Asimismo, la histona H1 está presente en una menor densidad y con un



**Figura 1.** Niveles de empaquetamiento del ADN. A) El ADN desnudo se enrolla en 1.6 vueltas alrededor del octámero de histonas para formar un nucleosoma. B) En el siguiente nivel, los nucleosomas se van ordenando periódicamente a lo largo de la fibra de ADN formando lo que se conoce como collar de perlas o fibra de 10 nm; al participar la histona H1, los contactos entre los nucleosomas se hacen estables y se forma el solenoide o fibra de 30 nm. C) El siguiente nivel corresponde a las asas de solenoides, las cuales se forman cuando la histona H1 se fosforila, en este nivel la compactación llega hasta 400 nm, las P representan la fosforilación de la histona H1, después de este nivel todavía no se sabe como se llega a la compactación de los siguientes niveles hasta llegar a los cromosomas.

grado menor de fosforilación en dominios transcripcionalmente activos. Estos resultados han llevado a pensar que existe una relación directa entre la fosforilación de la histona H1 y la compactación de los cromosomas. Sin embargo, en otros sistemas donde la división es amitótica se ha encontrado que por el contrario, la histona H1 se desfosforila y esto permite la condensación de los cromosomas. La fosforilación de la histona H1 debilita la interacción de los extremos básicos de la proteína al ADN, pero por otra parte esta fosforilación acentúa la unión de la histona H1 a la cromatina. Dos de los ejemplos más representativos de que la defosforilación de la histona H1 contribuye

a la condensación de los cromosomas son los siguientes:

—En *Tetrahymena* la histona H1 se encuentra hiperfosforilada en células en interfase, y completamente defosforilada cuando los cromosomas se encuentran en el estado altamente condensado.

—En las fases terminales del desarrollo de eritrocitos de pollo, el núcleo está condensado en heterocromatina, lo que ocurre debido a la presencia de la histona H5 (homólogo de H1 en pollo). La histona H5 recién sintetizada está altamente fosforilada, pero cuando la cromatina del eritrocito se condensa, la histona H5 se encuentra defosforilada. Estos datos permiten pensar que todavía existe una paradoja

acerca de las funciones de la histona H1 y sus modificaciones postraduccionales como la fosforilación y su contribución a la estructuración de un dominio.

Además de su importante papel estructural, se ha propuesto que la histona H1 puede actuar en la regulación transcripcional. La regulación mediada por la histona H1, se lleva a cabo de diferentes maneras:

1) En sistemas experimentales *in vitro* se ha demostrado que la histona H1 se encuentra en intercambio activo, es decir, que no permanece siempre unida al ADN. Además, la fosforilación en sitios específicos de esta histona correlaciona con un nivel de empaquetamiento más elevado que el del solenoide durante la mitosis. Se sugiere que la histona H1 podría actuar modulando la exposición de un sitio en el ADN a proteínas reguladoras. Al estar ausente la histona H1, no existen contactos estables entre los nucleosomas adyacentes, por lo tanto se impide el siguiente nivel de empaquetamiento de la cromatina manteniéndose una estructura más laxa, lo que permite una mayor accesibilidad al ADN.

2) Recientemente, se han hecho estudios *in vitro* en los cuales se ha demostrado una interferencia estérica entre los extremos amino-terminales de la histona H1 y otras histonas como la H3 impidiendo la acetilación de sus extremos amino-terminales y por lo tanto que la cromatina se descompacte interfiriendo con la unión de factores reguladores de la estructura de la cromatina y de la transcripción genética.

Este conjunto de evidencias apoyan la idea de que el molde real de la transcripción es la fibra de 30 nm o solenoide y que la histona H1 juega un papel importante en la regulación de la transcripción de un dominio (3).

Hasta ahora se ha revisado la manera como las histonas empaquetan y organizan al ADN para que pueda ser contenido dentro de una célula en las distintas etapas del ciclo celular y es evidente que la estructuración del ADN en cromatina genera un obstáculo, por lo que ésta tiene que ser remodelada para permitir la regulación de la expresión genética. A continuación se analizará la presencia de regiones en la cromatina, las cuales son sitios accesibles para la unión de elementos que participan en la transcripción y en su regulación.

### SITIOS DE HIPERSENSIBILIDAD

En 1976 Weintraub y Groudine se preguntaron si la estructura de la cromatina en genes activos sería la misma que en genes inactivos. Para contestar a su pregunta, aislaron núcleos de células eritroides de pollo que expresan los genes de globina y núcleos de células que no expresan estos genes y los trataron con concentraciones crecientes de la nucleasa pancreática DNasa I, la cual reconoce la doble hebra del ADN y corta inespecíficamente en zonas libres de nucleosomas. Posteriormente, mediante una técnica de hibridación indirecta la cual actualmente ha sido reemplazada por el "Southern blot", determinaron la magnitud de la digestión. Observaron que en el ADN genómico de las células que expresaban los genes de globina, la cromatina del *locus* tenía un mayor grado de digestión, mientras que el *locus* de globina de las células que no expresaban los genes era resistente a la digestión. Esto hizo suponer que la estructura de la cromatina se alteraba cuando un gen es transcripcionalmente activo. Actualmente se sabe que la sensibilidad al corte por la DNasa I correlaciona con un relajamiento en la estructura

de la cromatina del *locus* de genes activos, aunque el *locus* continúe ensamblado en nucleosomas. Por lo general en estas regiones los nucleosomas se encuentran en menor densidad. En 1979 Wu y Elgin realizaron digestiones con DNasa I en cromatina de núcleos de células de *Drosophila* y encontraron en el "Southern blot" bandas más pequeñas que los fragmentos originales (4). A estas bandas se les denominó sitios de hipersensibilidad al corte por la DNasa I. Se propuso que estos sitios correspondían a regiones con una estructura alterada de la cromatina, donde se observa un mayor grado de digestión de la doble cadena por parte de la DNasa I y que podrían coincidir con elementos reguladores de la transcripción. Actualmente se sabe que estos sitios de hipersensibilidad son típicamente, al menos dos órdenes de magnitud más sensibles al corte por nucleasas que el resto de la cromatina y no deben confundirse con la sensibilidad generalizada al corte por nucleasas que presentan los genes que están siendo transcritos y que puede extenderse por varias kilobases (kb). Además, en estos sitios los nucleosomas están ausentes o parcialmente ensamblados. Un sitio de hipersensibilidad típico contiene sitios de unión para una variedad de diferentes factores de transcripción. El aumento de la sensibilidad a nucleasas aumenta en estos sitios por la ausencia de nucleosomas, el ensamblaje parcial de éstos o su desplazamiento, y por la unión de un factor de transcripción que distorsiona a la cromatina favoreciendo su apertura en esa región o en regiones adyacentes. Se ha propuesto la posibilidad de que los factores de transcripción y los nucleosomas compitan por la unión en un lugar dado del genoma (como por

ejemplo un promotor o un "enhancer"), e induzcan la formación y mantenimiento de un sitio de hipersensibilidad. El ejemplo más claro que apoya este modelo demuestra que mutaciones en el sitio de unión al factor GATA-1 en el ADN correlacionan con una disminución cuantitativa en la formación del sitio de hipersensibilidad. Por lo tanto se ha propuesto que debe existir un proceso dinámico de interacción proteína-ADN, con elementos que remodelan a la cromatina y la formación de un sitio de hipersensibilidad (5). Una constante para los sitios de hipersensibilidad a la DNasa I es que coinciden, invariablemente, con elementos estructurales o de regulación, como por ejemplo, regiones delimitadoras de un dominio, "enhancers", silenciadores y promotores.

En general se pueden dividir los sitios de hipersensibilidad en 3 diferentes tipos:

1) **Los sitios específicos.** Pueden aparecer antes de la transcripción y generalmente persisten incluso al término de la misma, son tejido específicos y a este tipo corresponden los promotores, "enhancers" y silenciadores. Los promotores son secuencias de ADN que se encuentran generalmente de - 40 a + 40 nucleótidos relativos al inicio de la transcripción, conocida también como región "core", a estos se unen la maquinaria basal de transcripción que contiene a la ARN Polimerasa II, factores generales de transcripción y coactivadores. Los promotores controlan la direccionalidad de la transcripción y responden a activadores o represores transcripcionales que actúan sobre ellos. Además del "core", existe una región proximal que va de - 50 a - 200 que algunos han llamado el promotor regulador. Esta región, cuando es clonada en múltímeros cerca de

un promotor, puede conferirle una actividad de "enhancer". Por su parte los "enhancers" o aumentadores son secuencias en *cis* que se encuentran a una distancia relativa del inicio de la transcripción, tienen una organización modular, es decir, que son sitios de unión para más de un factor y son capaces de aumentar la transcripción de una manera independiente de su orientación. Los módulos de los "enhancers" desempeñan funciones específicas al activar la transcripción de un gen durante una etapa del desarrollo determinada, controlando los niveles de expresión del gen (6). Igual que los "enhancers", los silenciadores tienen una organización modular, pero a diferencia de éstos, regulan negativamente la transcripción (7).

2) **Los sitios constitutivos.** La presencia de estos sitios precede a la activación transcripcional y no son tejido específicos, es decir que en un mismo *locus* pueden estar presentes en diferentes tipos celulares y su presencia es independiente de la expresión genética. Estos sitios se han encontrado en algunos casos enmarcando a un gen o grupo de genes y se ha propuesto que corresponden a los límites físicos y funcionales de los dominios transcripcionalmente activos. Estos límites son regiones en el ADN que funcionan como secuencias neutras, es decir que no presentan actividad silenciadora, promotora o de aumentador y por lo tanto no afectan la transcripción positiva o negativamente. Estas regiones delimitadoras bloquean o aíslan la influencia de elementos de regulación de la transcripción que se encuentran fuera del *locus* al cual están flanqueando y se ha propuesto que contribuyen al mantenimiento de un *locus* activo. A estas secuencias se les ha llamado "insulators" o delimitadores.

3) **Las regiones de control de un dominio o "Locus Control Regions (LCR)".** Otro tipo de sitio de hipersensibilidad son los comúnmente llamados LCRs, los cuales están conformados por varios sitios de hipersensibilidad, y contienen múltiples sitios de unión para activadores transcripcionales. La diferencia entre los LCRs y los "enhancers" es que los primeros actúan en una etapa previa a la transcripción de los genes haciéndolos competentes para ser transcritos ya que presentan una actividad tipo "enhancer" y además, se ha propuesto que contribuyen a la apertura de la cromatina por diferentes mecanismos. Otra característica de los LCRs que los diferencia de los "enhancers", es que estos están usualmente más alejados de los genes que controlan, en ocasiones hasta 16 kb. Más adelante se discutirá sobre el papel de los LCRs en el control de la transcripción dentro de un dominio.

Los LCRs son un conjunto de sitios de hipersensibilidad al corte por la DNasa I, que corresponden a elementos reguladores tejido específicos en el ADN y que fueron descritos primeramente en el *locus*  $\beta$ -globina de humano. El estudio de la regulación de estos genes es importante ya que la falta de expresión de éstos causa una enfermedad conocida como talasemias que son una clase de anemias muy severas. Diversos grupos encontraron, al estudiar el *locus*  $\beta$ -globina de individuos afectados por esta enfermedad, una serie de supresiones río arriba del primer gen embrionario  $\epsilon$  del dominio, las cuales disminuían dramáticamente la expresión de todos los genes del *locus*, aunque éstos estuviesen intactos. Este descubrimiento suponía la presencia de elementos de regulación en la región 5' del *locus*. Al analizar la estructura de la

cromatina de esta región en células normales, encontraron una serie de sitios de hipersensibilidad, a los cuales subsecuentemente se les llamó "Locus Control Region" o LCR. Ratones transgénicos que expresaban niveles bajos de los genes globina, a los que se añadió el transgen de la región 5' no-codificante que incluía al LCR, normalizaron los niveles de expresión de los genes y mantuvieron su especificidad a lo largo del desarrollo observándose además que la expresión de los genes era independiente del sitio de integración (8). Esta observación motivó estudios subsecuentes basados en la idea de que los LCRs pudieran contrarrestar los problemas de silenciamiento de los vectores diseñados para terapia génica. Pronto los investigadores concluyeron que, en efecto, los LCRs son útiles para este fin, pero su intervalo de acción es restringido ya que su función es tejido específica. Se han descrito LCRs en el *locus* de  $\beta$ -globina de otros vertebrados con diferente número de sitios de hipersensibilidad (Tabla I). A partir del descubrimiento del LCR  $\beta$ -globina se han descrito LCRs en otros *loci* como el de TCR $\alpha$  de células T, en el *locus* CD2 en humanos y en el *locus* de la lisozima de pollo, entre otros (Tabla I).

Los LCRs se definen funcionalmente por las siguientes características: 1) son capaces de conferir altos niveles de expresión a un transgen independientemente del sitio de integración (efecto de posición) y dependiente del número de copias integradas del transgen. Esto significa que en ausencia de un LCR, la expresión de un transgen integrado en un contexto cromatínico, estaría sujeta a los efectos intrínsecos de la estructura de la cromatina, causados por los distintos sitios de integración (8). Por ejemplo, si el transgen se

integra en o cerca de una región de heterocromatina (que tiene una estructura altamente compactada y por lo tanto represora) como podría ser un centrómero o un telómero, su expresión decaería a lo largo del tiempo o nunca sería expresado, 2) regulan la expresión genética durante el desarrollo, y 3) contribuyen a la apertura de la cromatina, para permitir la transcripción de los genes del locus (9).

A pesar del gran número de investigaciones realizadas en laboratorios de todo el mundo, en la actualidad aun se sabe poco acerca de los mecanismos por los cuales los LCRs influyen en la estructura de la cromatina y la expresión de los genes contenidos en un dominio. Sin embargo, se han propuesto a la fecha, cuatro modelos principales para explicar la función de los LCRs:

1) **“Looping”**. En este modelo, el LCR actuaría como un holocomplejo para estimular la transcripción de genes individuales dentro del locus y a lo largo del desarrollo. El LCR se acercaría al promotor del gen a activar formando un asa o “loop”, reclutando toda la maquinaria de transcripción. Se ha propuesto un posible contacto entre el promotor y el LCR mediada por interacciones entre distintas proteínas reguladoras y que son tejido y tiempo específicas (Fig. 2A) (10).

2) **“Linking”**. Sugiere la cooperación de varios factores en *trans*, los cuales son expresados en un estadio de diferenciación específico y que serían reclutados por el LCR. Este modelo propone la formación de una cadena o relevo, mediada por factores que contribuiría a fomentar una estructura de la cromatina favorable para la transcripción de un solo gene durante una etapa determinada (Fig. 2B).

3) **“Tracking”**. Otro modelo que apoya la hipótesis de que el LCR contribuye a la apertura de la cromatina, es el llamado “tracking”, el cual propone un reclutamiento del complejo de transcripción y de diversos factores en el LCR y su posterior encarrilamiento en la cromatina, desplazándose a lo largo de ésta con la ayuda de complejos de remodelaje hasta llegar a los promotores de los genes que serán expresados. Este modelo no implica que la polimerasa forzosamente vaya transcribiendo, más bien representa una forma de reclutamiento de la maquinaria de remodelaje de la cromatina y como consecuencia el acceso de la maquinaria transcripcional (Fig. 2C).

4) **Transcritos intergénicos**. Se ha identificado la presencia de transcritos intergénicos, es decir, en regiones no-codificantes en el do-

TABLA I

LOCUS O GENE	ORGANISMO
Gene CD2	Humano
Gene de la adenosina desaminasa	
Genes de pigmentos visuales	
Gene S100 $\beta$	
Gene de la queratina 18	
Gene del MHC clase I HLA G	
Gene CD4	
Locus $\alpha/\delta$ del receptor de células T	
Gene de la apolipoproteína B	
Locus E/C-I de la apolipoproteína	
Gene de la cadena pesada de la $\beta$ miosina	
Genes C4A y B del complemento	
Gene de la hormona del crecimiento	
Locus de la cadena pesada Ig	
Genes 1 y 2 del Ig C $\alpha$	
Gene HLA-B7 del MHC clase I	
Gene de la desmina	
Gene CD34	Ratón
Gene de la glicoporina	
Locus $\lambda 5-V_{preB1}$	
Cadena pesada de la Inmunoglobulina $\mu$	
Genes I y II de la metalotioneína	
Gene Ea del MHC clase II	
Gene de la tirosinasa	Rata
Locus $\beta$ -globina $\alpha$ -fetoproteína	
Genes de la calicreína	
Gene de la proteína ácida de Whey	
Gene de la aldolasa C	Pollo
Gene LAP (C/EBP $\beta$ )	
Gene de la lisozima	Oveja
Locus $\beta$ -globina	
Gene de la $\beta$ -lactoglobulina	

Ig = Inmunoglobulina, MHC = Complejo principal de Histocompatibilidad, LAP = leucina amino peptidasa.



minio  $\beta$ -globina de humano. Se desconoce la función precisa de estos transcritos intergénicos pero se ha demostrado una correlación directa entre su presencia y la apertura de la cromatina en el contexto del dominio. Esta apertura de la cromatina

mediada por estos transcritos intergénicos favorece la expresión regulada de los genes de manera individual (Fig. 2D) (11,12).

Aunque hasta la fecha no se ha podido demostrar del todo que el LCR contribuye a la apertura de la

cromatina, está claro que este elemento es absolutamente necesario para que la transcripción de los genes de los dominios donde se ha identificado, se lleve a cabo adecuadamente.

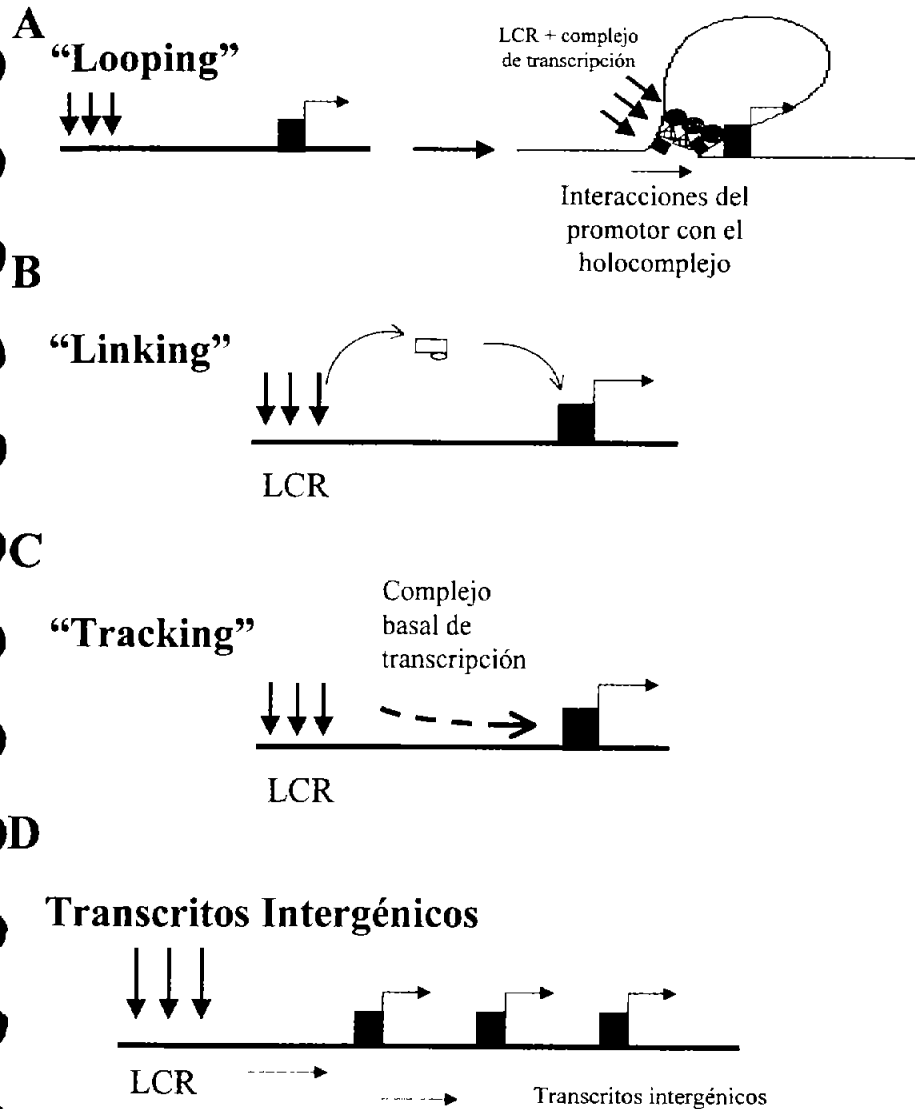
A continuación se analizarán las características de los sitios de hipersensibilidad constitutivos, los que corresponden a secuencias delimitadoras de los dominios.

### LÍMITES DE LOS DOMINIOS

Los genes activos e inactivos en el genoma, existen en *locus* independientes que tienen diferentes programas de expresión durante el desarrollo celular. Si tomamos en cuenta que los sitios de hipersensibilidad descritos anteriormente correlacionan con secuencias reguladoras de la expresión genética y que estos elementos actúan sobre promotores que se encuentran localizados a grandes distancias, es necesario que existan elementos que aseguren que estas regiones de control actúen sólo sobre sus promotores y no sobre promotores de otros *loci* vecinos.

Los "insulators" son secuencias que correlacionan con la presencia de sitios de hipersensibilidad constitutivos en el límite de los dominios y se definen como secuencias neutras que protegen a un dominio de los efectos de la cromatina adyacente, previniendo la promiscuidad entre elementos de regulación de otros dominios y que contribuyen a formar y mantener el *locus* abierto cuando está siendo transcrito (Fig. 3).

Los "insulators" fueron descritos originalmente en los límites de los genes divergentes *heat-shock* (*Hsp70*), en el *locus* 87A7 de *Drosophila*. Kellum y Schedl analizaron dos secuencias que correlacionaban con sitios de hipersensibilidad constitutivos, estos elementos



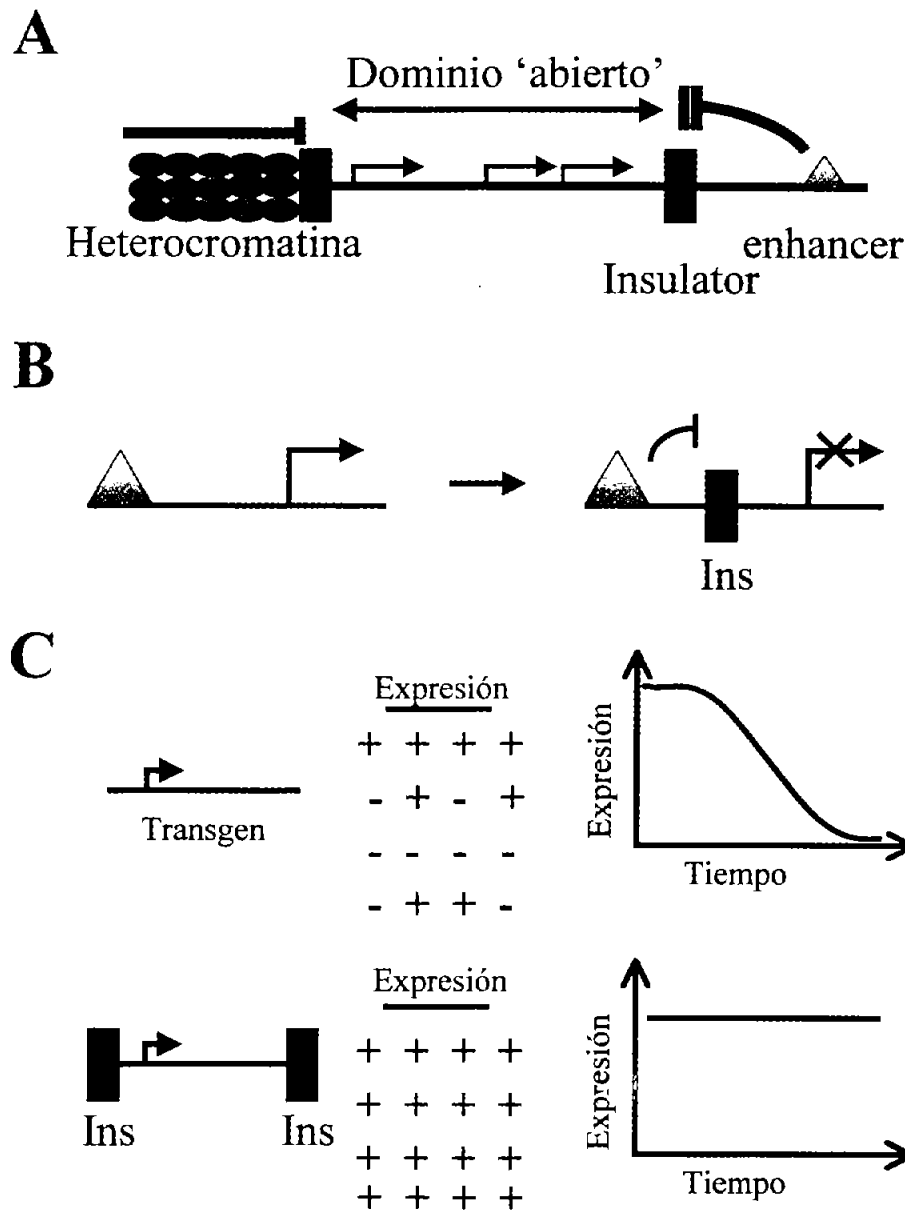
**Figura 2.** Modelos propuestos para el funcionamiento del LCR. A) "Looping": en este modelo, el LCR interactúa directamente con el promotor a través de contactos mediados por el holocomplejo e interacciones con factores que se unen de manera específica a los promotores individuales. B) "Linking": este modelo propone el uso de proteínas que actúen en trans, o proteínas de relevo que interactúan directamente con otras proteínas del promotor, activándolo. C) "Tracking": se propone que el LCR sirve como una entrada para que el holocomplejo se encarrile en la cromatina y se desplace hasta el promotor que va a ser activado. D) Transcritos Intergénicos: se ha propuesto que el LCR contribuye a la apertura de la cromatina, ya que se han encontrado transcritos intergénicos a lo largo de los loci regulados por LCRs (flechas en punteado) y estos transcritos son estadio y tejido específicos.

fueron nombrados *scs* y *scs'* (del inglés: "*specialized chromatin structures*"), se hicieron construcciones que llevaban al gen *white* controlado por un promotor mínimo, y utilizando estas construcciones genera-

ron moscas transgénicas con un patrón de color de ojos muy variable, esto como consecuencia del efecto de posición causado por el medio ambiente de la cromatina en los distintos sitios de integración. Al flan-

quear este gene con elementos *scs*, el color de los ojos de las moscas fue homogéneo; es decir, se suprimió el efecto de posición (13). Así se demostró la primera característica funcional de los "insulators": son capaces de proteger a un transgen del entorno cromatínico en el cual se integre (Tabla II).

Otra característica funcional de los "insulators" fue demostrada cuando se analizó a *gypsy*, un retrotransposón de *Drosophila*. Esta secuencia era capaz de bloquear la acción de un "enhancer" sobre su promotor solamente cuando se localizaba entre el "enhancer" y el promotor. Esta propiedad también se demostró para *scs* y *scs'*. En particular, el retrotransposón *gypsy* es responsable de una gran variedad de mutaciones que afectan genes a lo largo del genoma de *Drosophila*. Uno de estos genes es *yellow*, el cual es un gen no esencial y su producto está involucrado en la pigmentación de la cutícula de distintos órganos en la mosca de la fruta. La expresión de este gen a lo largo del desarrollo es controlado por una serie de "enhancers" tejido específicos, los cuales se encargan de regular la expresión de *yellow* diferencialmente durante el desarrollo y pigmentar distintas estructuras de la mosca. Los "enhancers" que controlan la expresión de *yellow* en la cutícula, alas y estructuras larvales pigmentadas, están localizados en la región 5' relativa del gen. La inserción de *gypsy* en la región 5' de *yellow* entre el "enhancer" que controla la expresión del gen en la cutícula y el que controla las estructuras larvales, bloquea la interacción del primero con el promotor del gen, pero no afecta al "enhancer" larval que se localiza más próximo al promotor, es decir, entre el transposón y el gen (14). Un análisis más deta-



**Figura 3.** Actividades Funcionales de los "insulators". A) Los "insulators" son capaces de mantener un dominio abierto protegiéndolo de la invasión de la heterocromatina y de la promiscuidad con elementos de regulación de dominios adyacentes. B) Actividad de bloqueo de "enhancer", un "enhancer" actúa sobre su promotor y lo activa, al colocar el "insulator" entre el "enhancer" y el promotor, se bloquea el efecto del "enhancer" y por lo tanto el promotor no es activado. C) Protección contra el efecto de posición, al integrar un transgen al genoma, su expresión depende del sitio de integración y generalmente decae a lo largo del tiempo, al flanquear el transgen con los "insulators", se suprime el efecto de posición, la expresión del transgen es homogénea y se mantiene a lo largo del tiempo.

TABLA II

ELEMENTO	LOCUS	ORGANISMO	BLOQUEO DE ENHANCER	PROTECCIÓN CONTRAEFFECTO DE POSICIÓN	PROTEÍNA ASOCIADA
HMR-R	HMR	Levadura	ND	ND	SMC, SMC3
HMR-L	HMR	Levadura	ND	ND	ND
UASrpg	Gene TEF2	Levadura	ND	ND	Rap1
STARs	Telómeros	Levadura	ND	ND	Tbf1p, Reb1p
Fab-7	Abdominal-B	<i>Drosophila</i>	+	ND	ND
Fab-8	Abdominal-B	<i>Drosophila</i>	+	ND	ND
Mcp	Abdominal-B	<i>Drosophila</i>	ND	ND	ND
scs	Gene hsp 70	<i>Drosophila</i>	+	+	Zw5
scs'	Gene hsp 70	<i>Drosophila</i>	+	+	BEAF-32
gypsy	Transposon	<i>Drosophila</i>	+	+	Su(Hw), Mod(mdg4)
promotor eve	Gene even-skipped	<i>Drosophila</i>	+	ND	GAGA
HS4 $\beta$ -globina	Locus $\beta$ -globina	Pollo	+	+	CTCF
sns	Gene temprano de histonas	Erizo de mar	+	ND	ND
HS2-6	TCR $\alpha/\delta$ -locus Dad-1	Humano	+	ND	ND
BEAD-1	Locus TCR $\alpha/\delta$	Humano	+	ND	CTCF
3'HS	Locus $\beta$ -globina	Pollo	+	-	CTCF
DMD	Locus H19/Igf2	Humano	+	-	ND
Ars	Gene de la arilsulfatasa	Erizo de mar	+	+	SpGCF1

ND: no detectada, +: sí tiene esa actividad, -: carece de esa actividad.

llado de *gypsy* ha permitido identificar la región dentro del retrotransposón que es responsable del bloqueo de "enhancer". Esta región contiene una secuencia repetida de 8 nucleótidos que se encuentra flanqueada por regiones ricas en adeninas (A) y timinas (T). Cuando esta secuencia es insertada en cualquier lugar del extremo 5' de *yellow*, los "enhancers" que están río arriba de esta secuencia no tienen efecto alguno sobre el promotor, en cambio los que se encuentran río debajo del octámero repetido, siguen siendo funcionales. Además, al colocar estas secuencias flanqueando al transgen *white*, se observó que lo protegían contra el efecto de posición. El fenotipo de las mutaciones causadas por *gypsy*, puede ser revertido al mutar al gen "*suppressor of hairy wing*" (*Su(Hw)*). La proteína del

mismo nombre se caracterizó como un componente del "insulator" *gypsy*. *Su(Hw)* tiene 12 dedos de zinc, interacciona con las secuencias repetidas dentro del retrotransposón *gypsy* y cualquier mutación en el sitio de unión de esta proteína y que como consecuencia impida su unión al "insulator", hace que el efecto de bloqueo de "enhancer" se pierda. Al realizar mutaciones en sitios diferentes a los que interaccionan con *gypsy* en *Su(Hw)*, se identificó otro componente "*modifier of mdg4*" (*Mod(mdg4)*), esta proteína carece de un dominio de unión al ADN, es incapaz de unirse al "insulator" *gypsy in vitro*, y se ha encontrado que colocaliza con la proteína *Su(Hw)* y demostrado que interacciona con esta proteína directamente. En ausencia de *Mod(mdg4)*, el "insulator" pierde su unidirec-

cionalidad y adquiere propiedades de silenciador, inactiva "enhancers" río abajo y río arriba del sitio donde se insertó. En resumen, *gypsy* tiene dos componentes principales identificados hasta ahora: *Su(Hw)* y *Mod(mdg4)*, sin los cuales sus funciones de "insulator" se ven alteradas. Por otra parte, recientemente se han identificado proteínas que se unen a *scs* y a *scs'*. A *scs* se une una proteína codificada por el gene *Zw5* y a *scs'* se unen dos proteínas: BEAF 32A y BEAF 32B. Se ha demostrado que estas proteínas son indispensables para la función de bloqueo de "enhancer" y apoyan el concepto de dominios, ya que se han inmunolocalizado en los límites de los "puffs" de los cromosomas politénicos de las glándulas salivales de las larvas de *Drosophila*. Los casos de *gypsy* y de *scs*, *scs'* no son

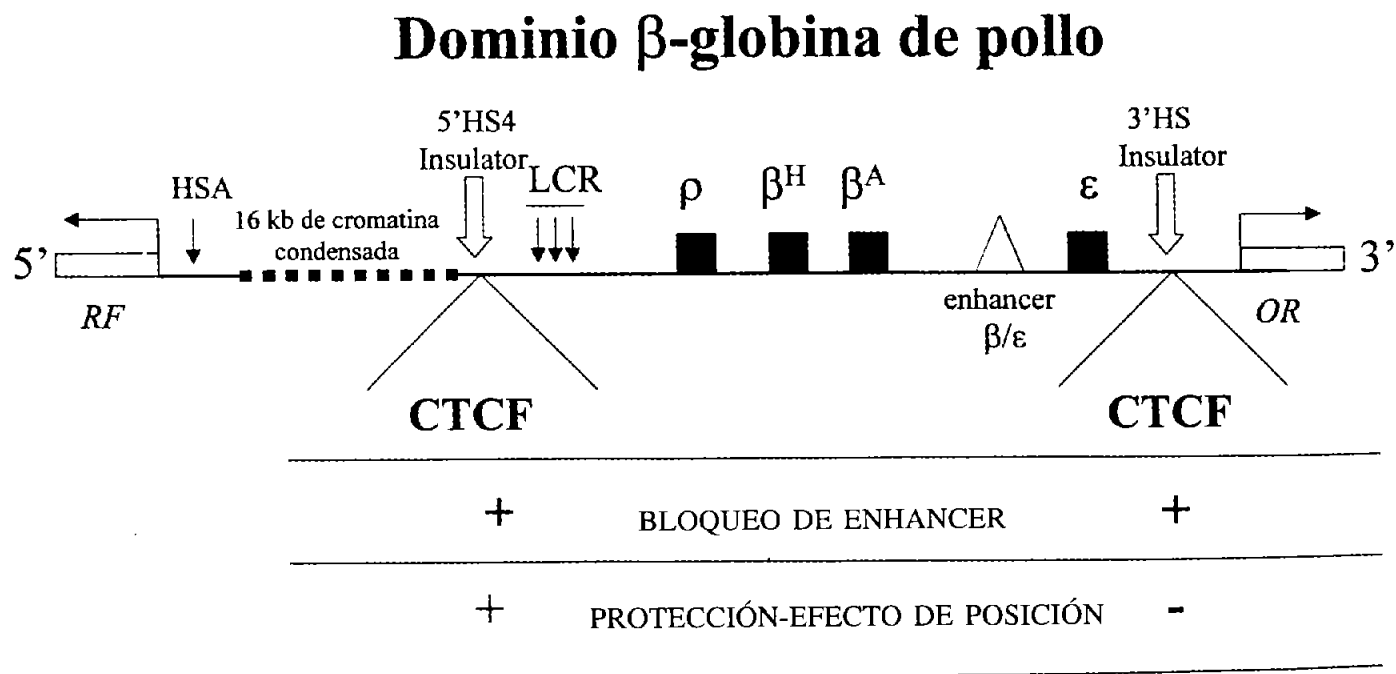
los únicos en que se asocian proteínas a un "insulator", como se verá también en los "insulators" de vertebrados, se ha demostrado la asociación de una proteína llamada CTCF (Tabla II).

El primer "insulator" descubierto en vertebrados es el que delimita al dominio  $\beta$ -globina de pollo en su extremo 5' (15). En este locus se identificó un sitio de hipersensibilidad constitutivo (5'HS4), que correspondía al sitio de hipersensibilidad 5'HS5 del mismo locus en humano y ratón. Además, río arriba de este sitio de hipersensibilidad en el locus de pollo, existe una estructura de cromatina altamente compactada, resistente a la digestión por DNasa I. Otra de las razones por las que se pensó que el 5'HS4 pudiera definir el límite del dominio  $\beta$ -globina, es que hacia el extremo 3' de este sitio de hipersensibilidad, se identificaron otros sitios de hipersensibilidad que corresponden al

LCR y este elemento sólo actuaba permitiendo la apertura de la cromatina hacia 3' (que es donde se encuentran los genes del locus) y no hacia la región 5' (donde está la estructura de la cromatina altamente compactada). Posteriormente se analizó este elemento y se descubrió que presenta las características funcionales de los "insulators", es decir protege a un transgen del efecto de posición y es capaz de bloquear la acción de un "enhancer" sobre su promotor. Investigaciones recientes han confirmado el concepto de dominio a partir del descubrimiento de un segundo "insulator" en el costado 3' del mismo dominio, lo cual demuestra que este dominio se encuentra enmarcado a ambos lados por "insulators" (Fig. 4). La actividad funcional del "insulator" 5' del gene  $\beta$ -globina pudo reducirse del fragmento de 1.2 kb que contenía el sitio de hipersensibilidad a un fragmento de tan sólo 250 pb al que se

le llamó "core", dentro de las 250 pb existe un fragmento de 42 pb que es necesario y suficiente para la función de bloqueo de "enhancer". En 1999 Felsenfeld y colaboradores identificaron y purificaron una proteína que se unía a este sitio y se comprobó que esta proteína era CTCF (CCCTC-binding factor). Para caracterizar la función de CTCF se hicieron mutaciones en el sitio de unión de esta proteína al "insulator" y se descubrió que una menor afinidad de unión de este factor disminuía la capacidad del "insulator" de bloquear la actividad de un "enhancer" sobre su promotor *in vivo*.

La proteína CTCF es altamente conservada en vertebrados, es una proteína de 82 kDa compuesta por 11 dedos de zinc y está involucrada en la activación y el silenciamiento transcripcional dependiendo del contexto en que se encuentre. Se han identificado sitios de unión para



**Figura 4.** Esquema del dominio  $\beta$ -globina de pollo. Se señalan la posición de los "insulators" 5' y 3', a los cuales se une la proteína CTCF y las diferentes actividades que presentan. HSA, sitio de hipersensibilidad; RF, gene del receptor del folato; OR, gene del receptor olfatorio.

CTCF en otros "insulators" localizados entre genes que se regulan de manera independiente en vertebrados. Por esta razón se ha sugerido que la actividad de bloqueo de "enhancer" mediada por CTCF es un componente funcional de los límites de los dominios de los vertebrados (16).

Una de las evidencias más contundentes de la relevancia de los "insulators" y la participación de CTCF en la regulación génica, es el reciente descubrimiento de una actividad de bloqueo de "enhancer" regulado por "imprinting" en el dominio *Igf2/H19* en ratón y humano. En este dominio de más de 100 kb, el gen *Igf2* se localiza a 80-90 kb del gen *H19*. Ambos genes comparten dos "enhancers" localizados en el extremo 3' de los genes *Igf2* y *H19* respectivamente. Evidencias experimentales demostraron la presencia de una región localizada río arriba del gen *H19* (es decir, entre los genes *Igf2* y *H19*) con la capacidad de ser metilada de manera diferencial dependiendo del tipo de alelo. A esta región se le ha denominado región de metilación diferencial (RMD), la cual se encuentra metilada exclusivamente en el alelo paterno. Trabajos realizados por el grupo de Tilghman (1991) sugirieron originalmente que la expresión alelo específico de estos genes está dada por un mecanismo de competencia entre los "enhancers" distales y los promotores particulares a cada gen. De esta forma se explicaba la especificidad en la expresión monoalélica de cada gen del locus *Igf2/H19*.

Recientemente un cúmulo de estudios experimentales encabezados por los grupos de Tilghman y Felsenfeld demostraron una actividad tipo "insulator" en la región RMD. Esta actividad resulta ser regulada

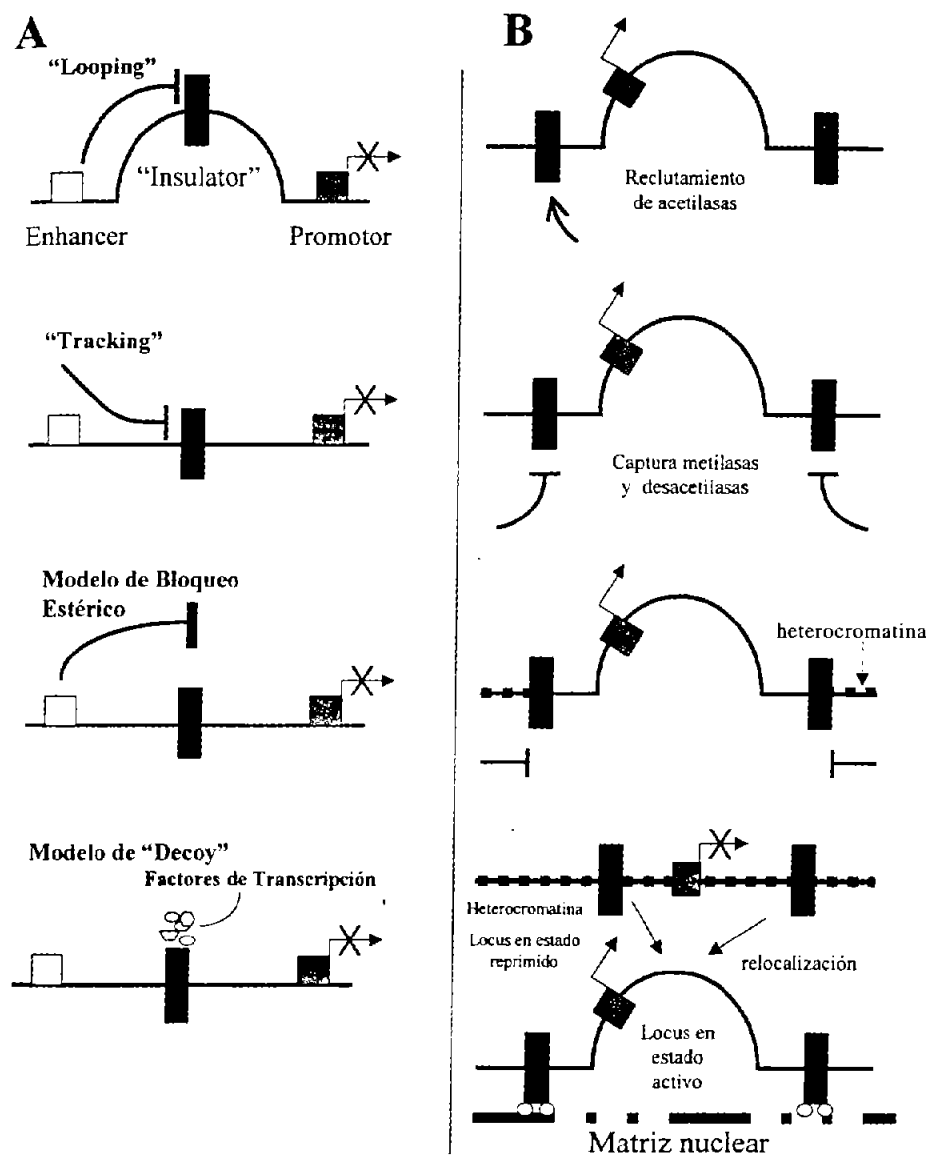
por la interacción del factor CTCF, cuya función es, en este caso, dependiente de la metilación. Para el alelo paterno, donde sólo se expresa *Igf2*, la región RMD se encuentra altamente metilada, lo cual impide la unión de CTCF y por lo tanto no se presenta la actividad de bloqueo de "enhancer", permitiendo la expresión de *Igf2*. Por el contrario, en el alelo materno, donde sólo se expresa *H19*, RMD no está metilado, lo cual permite la unión de CTCF. Esta unión contribuye a una actividad de bloqueo de "enhancer" la cual evita que los "enhancers" activen a *Igf2* y se concentren en la regulación y expresión de *H19*, únicamente en el alelo materno (17). Estas evidencias apoyan el concepto de dominios y la contribución de los "insulators" en la estructura y regulación de los mismos.

Hasta la fecha no ha sido posible determinar los mecanismos exactos de las funciones de los "insulators" y los modelos propuestos tienen una estrecha relación con el funcionamiento propio de los "enhancers". Existen tres modelos propuestos sobre cómo el "insulator" lleva a cabo la función de bloqueo de "enhancer": 1) en el primero, llamado "looping", se propone que el "insulator" interrumpe los contactos proteicos entre el promotor y el "enhancer" (Fig. 5A), 2) en el segundo modelo, llamado "tracking", se propone que al colocar al "insulator" entre el "enhancer" y el promotor, el "insulator" bloquea el paso del complejo de transcripción impidiendo de esta manera la activación del promotor; 3) otro modelo propuesto es el llamado de bloqueo estérico, en el cual, al colocar al "insulator" entre el "enhancer" y el promotor, los factores proteicos que se unen al "insulator" impedirían la unión de los factores del "en-

hancer", pero hasta la fecha esto no ha sido totalmente comprobado. Anteriormente se había propuesto otro modelo llamado "Decoy" en el cual se propuso que el "insulator" titulaba los factores necesarios para la actividad de los "enhancers" reclutándolos y de esta manera se inactivaba al "enhancer" (Fig. 5A). Sin embargo, actualmente se tienen varias evidencias experimentales que van en contra de este modelo.

Experimentos realizados en transfecciones transitorias con plásmidos circulares y lineales mostraron que los "insulators" no alteran en ninguna forma a los "enhancers" y que su actividad de bloqueo es dependiente de su posición y requieren de proteínas específicas para realizarla. Aunque estos resultados se inclinan mayoritariamente a que el "enhancer" realiza su función por medio del "tracking", ya que el "insulator" bloquea esta función de manera mecánica, a la fecha no existe un estudio que favorezca a uno de los modelos en particular.

Se mencionó que la segunda característica funcional de los "insulators" es proteger a un transgen del efecto de posición cuando se encuentran flanqueándolo. En 1998 Pikaart y colaboradores demostraron que el "insulator" 5' del locus  $\beta$ -globina de pollo es capaz de proteger a un transgen del efecto de posición y de la desacetilación y la metilación del locus que son sinónimo de represión (18). Basados en estos datos y en los de los "insulators" de *Drosophila*, se propusieron, entre otros, dos modelos sobre cómo el "insulator" podría funcionar para proteger a un dominio del efecto de posición: en el primero, sugieren que el "insulator" funciona como una barrera para el acceso de los complejos de metilación y desacetilación, quizás secuestrando



**Figura 5.** Modelos propuestos para el funcionamiento de los "insulators". A) Modelos para la actividad de bloqueo de "enhancer". B) Modelos para la actividad de protección contra el efecto de posición. Referirse al texto para mayores detalles.

desacetilasas de histonas e impidiendo que actúen sobre el dominio; en el segundo modelo, podrían reclutar acetilasas de histonas en mayor número que desacetilasas y así vencer las reacciones de desacetilación que pudieran llevarse a cabo en el *locus* al cual estén flanqueando (Fig. 5B). Anteriormente se mencionó que al integrar un transgen en el genoma, su expresión está supeditada al entorno cromatínico en el cual se integre. Si

el transgen se integra en una región de heterocromatina, la expresión del gen es silenciada, ya que la heterocromatina invade al transgen, creando una estructura represora la cual impide su transcripción (Fig 5B). De este concepto se desprende otra actividad de los "insulators", que pueden actuar como una barrera física que impida la invasión de regiones de heterocromatina al transgen (19). Salta a la vista que dadas las propiedades de los "insulators" se

está ante un novedoso tipo de elementos de regulación que pueden contribuir a evitar la variabilidad y extinción en la expresión de genes con fines terapéuticos (20) (Tabla II).

En conclusión, los "insulators" son elementos que ayudan a organizar el genoma en dominios cromosomales al servir como límites entre genes o *loci* que se expresan diferencialmente durante el desarrollo, protegiéndolos de los efectos de la cromatina adyacente (ya sea heterocromatina o elementos de regulación en *loci* adyacentes) y ayudando a mantener los dominios abiertos para que puedan ser transcritos.

Otros elementos que delimitan los dominios, son las llamadas secuencias de unión a la matriz nuclear ("Matrix Attachment Regions o Scaffold Attachment Regions, MARs o SARs"), las cuales son elementos en *cis* que están involucrados en la organización de los dominios estructurales asociados con sitios de hipersensibilidad y que se encuentran enmarcando regiones que presentan una sensibilidad generalizada al corte por la DNasa I. A diferencia de los "insulators", no todas las MARs confieren protección contra el efecto de posición, ni tienen una expresión dependiente del número de copias y la gran mayoría no tiene función de bloqueo de "enhancer". Estas secuencias son ricas en A y T, no tienen una secuencia consenso y se identificaron como sitios de unión a la matriz nuclear; el ADN de estas regiones tiene un surco menor más angosto, es propensa a desenrollarse y se dobla con mayor facilidad. Estas secuencias generalmente son reconocidas por la histona H1 y por la topoisomerasa II. Estas dos proteínas contribuyen a compactar la fibra de cromatina en "loops" o bien forman

heterocromatina, por el contrario, otras proteínas como MATH-20 y HMG-IFI, se unen cooperativamente a estas regiones y pueden desplazar a las proteínas que inducen la compactación y de esta manera crear accesibilidad. Tres excepciones a esta regla son las encontradas en la MAR del *locus* de lisozima de pollo, llamadas elementos A, las cuales se encuentran en 5' y 3' delimitando el *locus*. Estas MARs son capaces de proteger contra el efecto de posición y tienen actividad de bloqueo de "enhancer". Otro ejemplo clásico es el de las MARs 5' y 3' del gene humano de la apolipoproteína B (apo B), éstas solamente tienen actividad de protección contra el efecto de posición, sin embargo no son capaces de bloquear a un "enhancer". Existen otros elementos MAR en el gen  $\mu$  de la inmunoglobulina, éstos fueron encontrados flanqueando el "enhancer" y el promotor del gene. Posteriormente se estudió el papel que jugaban estos elementos en la expresión del gene  $\mu$  de la inmunoglobulina. Se generaron ratones transgénicos los cuales llevaban construcciones del gene  $\mu$  con o sin las secuencias MARs, además, analizaron el efecto de la metilación del ADN previo a la transfección (en células en cultivo) en la actividad del gen  $\mu$  con o sin las MARs. Encontraron que el "enhancer"  $\mu$  es capaz de activar a un promotor distal en transfecciones estables en células en cultivo, pero no en ratones transgénicos; sin embargo, al estar en combinación con las MARs, el "enhancer" fue capaz de activar al gen también en los ratones transgénicos. Además, el "enhancer" en combinación con una MAR fue capaz de inducir accesibilidad a una región cromatínica. Las secuencias MARs fueron capaces, además, de localizar al transgen a la

matriz nuclear y finalmente obtuvieron un dato significativo e inédito en el cual demostraron que las secuencias MARs podían revertir la metilación del ADN, lo que evitaba la represión genética. Sin embargo, las MARs por sí solas son incapaces de ejercer estas funciones, por lo que se propuso que la acción de las MARs era dependiente del "enhancer" (21).

Además de las MARs, se ha demostrado que en el genoma ocurren naturalmente secuencias repetidas naturales en A-T, las cuales presentan algunas propiedades que pudieran estar involucradas en la definición de dominios y a las cuales resulta difícil atribuirles un actividad funcional. Por otra parte, también existen secuencias repetitivas como la familia CRI que ha sido localizada en los límites de algunos dominios como en los extremos 5' y 3' del *locus*  $\beta$ -globina de pollo, y se ha propuesto que las repetidas de A y T junto con las CRI pudieran formar una estructura diferente de cromatina que contribuyera a delimitar el dominio.

Al analizar las propiedades de las secuencias delimitadoras como "insulators" o las secuencias MARs, se hace evidente la necesidad de seguir estudiando estas secuencias ya que forman parte fundamental de la organización de los genes en dominios transcripcionalmente activos. A continuación se revisará lo relacionado con las interacciones y la cooperación de estos límites para que los dominios puedan ser regulados.

### TERRITORIOS

Hace algún tiempo se pensaba que el núcleo era una estructura donde la transcripción, la duplicación y todas las actividades de procesamiento del ADN, se llevaban a cabo en cualquier lugar de éste. Actualmen-

te se entiende que esto no puede ser cierto ya que se sabe que el núcleo tiene una estructura tridimensional y que esta estructura se adapta a distintas circunstancias, como por ejemplo en respuesta a diferentes tipos de estrés o durante las distintas etapas del ciclo celular para facilitar las funciones celulares, nucleares y de segregación.

En un núcleo en interfase, los cromosomas están organizados en parches llamados territorios cromosomales, los cuales, basados en modelos propuestos por varios grupos, sugieren que se encuentran separados por canales llamados dominios intercromosomales. En esta etapa del ciclo celular, los cromosomas son altamente dinámicos y se ha visto que los genes activos transcripcionalmente se encuentran más cerca de los canales intercromosomales aledaños a la superficie de los territorios y con una mayor superficie de contacto con éstos. Se ha comprobado que los cromosomas se mueven dentro de su territorio y se ha propuesto que la explicación a este movimiento está determinada por la abundancia de factores involucrados en el procesamiento, la transcripción y la duplicación en focos localizados aledaños a estos territorios. Para explicar este concepto, existe un ejemplo clásico que demuestra el movimiento o reclutamiento de genes a una región en las células B de ratón que es rica en una proteína llamada Ikaros. Esta proteína es capaz de inactivar a los genes a través de su relocalización a una región centromérica de heterocromatina. Mientras que los genes que necesitan ser transcritos permanecen en zonas favorables del nucleoplasma, Ikaros recluta a los genes que deben ser reprimidos a sitios de heterocromatina para inactivarlos (22). Esta represión involu-

era el movimiento de *loci* específicos a un compartimiento nuclear donde son silenciados activamente.

En levaduras se ha observado que los complejos que participan en el silenciamiento genético, particularmente la familia de las proteínas SIR, se localizan en los telómeros cerca de la periferia nuclear donde se encuentran en compartimentos. Se han identificado por lo menos dos proteínas que ayudan a que esto ocurra, una de ellas es el complejo Ku, el cual al ser mutado en cualquiera de sus dos subunidades impide la posición perinuclear del telómero. Además, se ha demostrado que la represión de genes reporteros baja a medida que se alejan de los telómeros (donde se encuentran los complejos represores). Estas evidencias apoyan y asocian el concepto de territorios y dominios. En *Drosophila*, se ha descrito un modelo en el cual se asocian los "insulators" y los elementos de respuesta Polycomb (Pc-G), estos últimos involucrados en la represión de genes pueden actuar en *trans* para mediar la represión y lo hacen a través de interacciones proteína-ADN y proteína-proteína con miembros de la familia Polycomb en zonas definidas dentro del núcleo. Se ha propuesto que la organización y actividad de los "insulators" es influenciada por los elementos Pc-G al establecer dominios subnucleares. Corces y colaboradores han demostrado una clara correlación entre la presencia del "insulator" *gypsy* y una localización subnuclear cercana y en la periferia del núcleo. Las evidencias van más allá dado que demostraron que Su(Hw), el componente central de la acción del "insulator" *gypsy*, colocaliza con una alta frecuencia con la proteína lamina, componente estructural de la membrana nuclear (23).

Las evidencias mencionadas demuestran que en el núcleo debe existir una organización tal que permita la correcta regulación de *loci* independientes, facilitando el remodelaje de la cromatina, ya sea para que se active o se reprima la transcripción en un momento determinado del desarrollo y que los límites de los dominios son elementos importantes para que esta organización se lleve a cabo de manera más eficiente. Además, los nuevos datos surgidos de estudios citológicos apoyan la idea de la necesidad de tener elementos estructurales, ya sean membrana o matriz nuclear, para el anclaje de secuencias que favorezca una topología óptima y así llevar a cabo la expresión de los genes contenidos en un dominio. Actualmente se han hecho estudios en los cuales se ha logrado comprobar que además de los "insulators", los "enhancers" también contribuyen a la organización de los dominios dentro de un contexto cromosómico relocalizándolos lejos de regiones altamente represoras. A continuación se presentan las evidencias.

### RELOCALIZACIÓN

Recientemente, dos grupos de investigación independientes analizaron la posibilidad de que los "enhancers" y LCRs contribuyan también a la regulación del dominio al relocalizarlo lejos de regiones de heterocromatina en el interior del núcleo. En el primer trabajo se tomó a uno de los sitios de hipersensibilidad, específicamente el 5' HS2 del *locus*  $\beta$ -globina humano, el cual anteriormente había sido caracterizado como un "enhancer" y en el que se identificaron varios sitios de unión a factores transcripcionales como GATA-1, NF-E2, AP1 y EKLF, y se realizaron mutaciones en sitios específicos para la unión a estos fac-

tores, aboliendo su unión. La meta de este estudio era tratar de contestar dos preguntas: primero, si la interrupción de la unión de factores abolía la actividad del "enhancer" y segundo, si estas interrupciones afectaban la localización subnuclear de un transgen.

Se valieron de una técnica de recombinación homóloga para integrar a un transgen cuya expresión era dirigida por el HS2 completo o con las diversas mutaciones en los sitios de unión a factores, actuando en una región cercana al centrómero. Visualizaron al transgen en todos los casos mediante una técnica de FISH (hibridación *in situ* fluorescente) y comprobaron su expresión; cuando el "enhancer" se encontraba intacto, de una manera estadísticamente significativa, demostraron que el transgen era relocalizado lejos del centrómero y se expresaba adecuadamente, lo que no pasaba en presencia del "enhancer" mutado. Concluyeron que se requiere un "enhancer" funcional tanto para suprimir el silenciamiento genético como para relocalizar a un gen lejos de la heterocromatina centromérica durante la interfase, y que la supresión del silenciamiento requiere de la relocalización del transgen lejos de regiones de heterocromatina (24).

Basados en estos resultados se propuso un modelo en el cual los factores que se unen al "enhancer", posiblemente actúan reclutando y translocando al transgen a compartimentos nucleares o territorios donde existen factores y elementos basales de la transcripción en grandes concentraciones, por sus interacciones con éstos. A su vez no se descarta la posibilidad de que alguno de los factores del "enhancer" reclute a complejos de acetilación, metilación o de remodelaje de la



cromatina completando el ciclo regulador que lleve a la expresión regulada de un gen.

En el segundo trabajo se analizó la posibilidad de que el LCR también estuviese involucrado en la apertura de la cromatina y en la activación transcripcional. El grupo de Groudine se valió de líneas celulares que tenían por un lado al LCR completo del *locus*  $\beta$ -globina (con los 5 sitios de hipersensibilidad) y por otro la supresión presente en la talasemia hispánica  $\Delta$ HS1-HS5, y una línea celular donde ellos generaron una supresión más pequeña que abarcaba  $\Delta$ HS2-5. Por medio de FISH analizaron si la ausencia del LCR afectaba a la localización del *locus* y encontraron que la localización en las líneas con el *locus* silvestre y las que tenían la supresión  $\Delta$ HS2-5, se localizaban lejos del centrómero y tenían a las histonas H3 y H4 hiperacetiladas a lo largo del *locus* (25). Estos datos indicaban que en realidad el LCR no era necesario para la localización del *locus* lejos de la heterocromatina centromérica, ya que la línea celular con la supresión parcial del LCR ( $\Delta$ HS2-5) también se localizaba lejos del centrómero y además la sensibilidad al corte por la DNasa I se mantenía. Por otra parte, la hiperacetilación de las histonas H3 y H4 es característica de un *locus*  $\beta$ -globina que está abierto. Anteriormente, se había pensado que ésta era una característica que dependía del LCR, pero con los datos anteriores se demostró que el LCR no es necesario para esta acetilación, por lo que infirieron que para la localización del *locus* lejos de la heterocromatina centromérica mediaba una estructura más abierta y una hiperacetilación generalizadas del *locus* y que estas actividades eran independientes del LCR. Por otro

lado, la acetilación de H3 cerca del LCR en el *locus* silvestre, que es transcripcionalmente activo, es mucho mayor cuando el LCR se encontraba intacto. Estos datos sugirieron que el LCR no actúa directamente en mantener una estructura abierta de la cromatina, ni ayuda a alejar al *locus* de entornos heterocromáticos y por lo tanto represores, más sin embargo podría actuar a otro nivel para regular la expresión del *locus*.

Esto crea una controversia en cuanto a lo que se conoce del LCR, ya que en estudios realizados por Proudfoot (1997) y de manera independiente por Fraser y colaboradores (2000), demostraron la presencia de transcritos intergénicos los cuales son dependientes de la integridad del LCR (11, 12). Estos dos grupos han propuesto que los transcritos intergénicos contribuyen a la apertura de la cromatina, ya que se encuentran presentes a lo largo del *locus*, son estadio específicos y sólo se encuentran en células donde el *locus* es expresado, es decir, células eritroides. Una de las explicaciones de lo resultados que indican que el LCR no es necesario para la apertura del dominio, es que la regulación de la expresión del *locus* se da en varias etapas y posiblemente niveles, y que la combinación de los diferentes factores juega un papel importante para la regulación del *locus*. Actualmente se están realizando diversos estudios para resolver esta controversia.

#### MODELOS DE APERTURA Y CERRADO DE DOMINIOS

Se ha establecido que la cromatina no solamente funciona para ayudar a contener al genoma dentro del núcleo, sino también para reprimir la actividad genética, se ha mencionado que esta estructura represora

debe remodelarse para permitir la expresión regulada de los genes. Se ha visto que los elementos de regulación en *cis* contribuyen a la formación de una estructura más abierta de la cromatina, muchos de estos elementos en *cis*, interactúan con elementos proteicos para lograr este objetivo. En esta parte de la revisión se tratará de explicar cómo los complejos de remodelaje, la acetilación y la desacetilación de los residuos amino-terminales de las histonas, además de la metilación, juegan papeles importantes para la apertura y el cerrado de la cromatina.

Los complejos de remodelaje están compuestos por conjuntos de proteínas capaces de desplazar o disociar parcialmente un nucleosoma, para que la región donde se encuentren posicionados pueda ser más accesible. Entre estos complejos de remodelaje existen los que son dependientes de ATP, entre estos el mejor descrito es el complejo SWI/SNF, el cual fue inicialmente descrito en levadura y que actualmente se sabe que tiene homólogos en mamíferos; contiene una subunidad llamada SWI2/SNF2 con actividades de ATPasa y helicasa de ADN. Un modelo propuesto para la acción de SWI/SNF es que este complejo puede encarrilarse en el ADN y moverse a lo largo de éste y así desplazar a los nucleosomas, aunque esta propuesta no está del todo aceptada. Un segundo modelo propone que las histonas H2A y H2B se desestabilizan, se tiene evidencia que esta desestabilización facilita el acceso a factores de transcripción al ADN nucleosomal. Se ha encontrado que este complejo es capaz de participar en la unión de derivados de un activador transcripcional GAL4 a sus sitios específicos en un mo-

nonucleosoma reconstituido, de una manera dependiente de ATP, también alterar el patrón de digestión de un nucleosoma y cooperar para interrumpir un arreglo de nucleosomas preensamblados.

Otros complejos que utilizan ATP son: RSC y NURF (del inglés "Nucleosome Remodeling Factor"). RSC contiene varias subunidades homólogas a SWI/SNF y se ha visto que contiene una actividad dependiente de ATPasa capaz de alterar la estructura nucleosomal *in vitro*. CHRAC ("CHRomatin Accessibility Complex") moviliza un arreglo nucleosomal y arregla los nucleosomas en una espaciación periódica. Se sabe que muchos de estos complejos de remodelaje comparten subunidades, lo que implicaría un ahorro de energía para la célula, además se ha visto que existen homologías de las actividades muy conservadas entre estos complejos en los diferentes organismos, lo que indica su importancia.

Además de los complejos de remodelaje dependientes de ATP, existen los que reclutan una actividad de desacetilasas o acetilasas de histonas. Anteriormente se discutió que los extremos amino-terminales de las histonas quedan por fuera del octámero rodeando a la fibra de ADN y estabilizando la unión de ésta con el octámero. Estos extremos amino-terminales de las histonas son pequeños, varían entre 11 y 15 residuos y sufren diferentes modificaciones como acetilación, fosforilación, metilación y ribosilación que correlacionan con una estructura más abierta o cerrada del dominio, entre otras funciones. Los extremos amino-terminales de las histonas del octámero pueden ser modificados reversiblemente por la acetilación de sus residuos de lisina, los cuales

interaccionan iónicamente con los fosfatos cargados negativamente de la estructura del ADN. La acetilación de las histonas resulta en una neutralización de la carga y como consecuencia una separación del ADN de las histonas, lo que le permite al ADN nucleosomal ser más accesible a los factores de transcripción. La hiperacetilación de histonas en un dominio, se ha correlacionado con regiones activas del genoma, mientras que las regiones inactivas están hipoacetiladas. Los niveles de acetilación dependen de complejos que tienen actividad de acetiltransferasas o de desacetilasas. Algunos de los activadores y coactivadores que tienen actividad de acetiltransferasa, son GCN5, p300/CBP y la TAF<sub>II</sub>250. GCN5 acetila la histona H3 en la lisina 14 y a la histona H4 en las lisinas 8 y 16 pero sorprendentemente se descubrió que GCN5 sólo acetila histonas libres. Por su parte, P/CAF (p300/CBP associated factor) interacciona con coactivadores transcripcionales como p300/CBP (CREB-binding protein) que es una acetiltransferasa, la cual es capaz de acetilar a las histonas libres o en el octámero. P/CAF acetila a las histonas libres (H3 y H4) y a la histona H3 en el octámero. Además, se sabe que p300 y CBP interaccionan con numerosos factores de transcripción como CREB, c-jun/v-jun y Fos. La TAF<sub>II</sub>250 es parte del complejo de TFIID y tiene una actividad de acetiltransferasa, que acetila a H3 y a H4 en sus extremos amino-terminales y se propone que gracias a esta subunidad TFIID es capaz de desestabilizar a los nucleosomas que están en el promotor, lo que confiere al complejo de transcripción la habilidad de abrir una región de la cromatina sobre la cual

está transcribiendo (26). Dentro de los complejos de desacetilasas de histonas que se han identificado está la HDAC-1 en humanos, Rpd3 en levadura y su homólogo en *Drosophila*. Estos complejos se asocian a represores que controlan la proliferación y diferenciación celular como Mad(mx1)/Max, Sin3 y YY1.

Otro sistema que es utilizado para controlar el cerrado de un dominio o de una zona amplia del genoma, es la metilación del ADN en dinucleótidos CpG, la metilación de citosinas en el carbono 5 de los dinucleótidos es una característica del genoma de eucariotes. En vertebrados, 60-90% de los CpG's están metilados y los CpG's que no están metilados (cerca del 15% de todos los CpG's del genoma) se encuentran en islas CpG. Se ha comprobado que la metilación de regiones del genoma provoca la pérdida de sitios de hipersensibilidad, y le confiere al ADN resistencia a la digestión por nucleasas. Este mecanismo es independiente de secuencias específicas y es un sistema de expresión global. El mecanismo molecular por el que la metilación del ADN reprime la transcripción no es del todo conocido, sin embargo existen reportes que relacionan a la metilación con la desacetilación de histonas. Una proteína que se une a regiones metiladas es MeCP2, que es capaz de interactuar con la cromatina de una manera dependiente de la metilación y recientemente se ha comprobado que puede unirse a componentes del complejo mSin3A/HDAC. La asociación entre MeCP2 y este complejo es muy clara, por lo que se ha propuesto que una región del genoma que esté metilada puede ser silenciada por desacetilación dependiente de MeCP2 generando un estado permanente de represión

a nivel de la estructura de la cromatina. El efecto del silenciamiento mediado por MeCP2, puede ser parcialmente revertido al utilizar un inhibidor de las desacetilasas como la tricostatina A (TSA), lo que indica que la acetilación puede contrarrestar, al menos en parte, el silenciamiento de un gen. Otra forma de revertir el silenciamiento por metilación sería el reclutamiento de desmetilasas; sin embargo, hasta la fecha no se han encontrado evidencias contundentes de su existencia. La metilación del ADN no es necesariamente la primera causa de inactivación de un gene, podría ser un evento secundario o haber más de una manera de inactivación, lo que queda claro es que estos diferentes mecanismos (activadores y represores) actúan de una manera cuidadosamente regulada para permitir la apertura o cerrado de dominios en la cromatina.

Recientemente, los modelos que se han propuesto acerca de la manera cómo los límites de los dominios regulan la apertura y cerrado de los mismos, relacionan a los "insulators" y MAR's con complejos de remodelaje, metilación y acetilación/desacetilación. Felsenfeld propuso que los límites actuarían reclutando acetilasas y complejos de remodelaje para crear una estructura más accesible del dominio. Recientemente se han encontrado evidencias en el "insulator" 5'HS4  $\beta$ -globina de pollo que indican la presencia de una actividad de acetilasa de histonas, la cual al parecer tiene direccionalidad. Además, Felsenfeld propuso que los límites eran capaces de bloquear la acción de desacetilasas y metilasas para impedir que éstas actuaran sobre el dominio y mantenerlo abierto. Por otra parte, al reunir todos los estudios acerca de los límites de los dominios (en particular de

los "insulators"), existen evidencias que demuestran que las funciones de estos límites dependen del contexto cromatínico en el cual se encuentren, es decir, la actividad de bloqueo de "enhancer" está destinada a evitar la promiscuidad entre los elementos de regulación de dominios adyacente como sucede en el caso del "insulator" 3' del *locus*  $\beta$ -globina de pollo. Este "insulator" sólo presenta actividad de bloqueo de "enhancer"; por el contrario, el "insulator" 5'HS4 del mismo *locus* presenta ambas actividades (bloqueo de "enhancer" y protección contra el efecto de posición). Se ha propuesto que esto se debe a que río arriba del "insulator" se encuentra una región de heterocromatina y por lo tanto debe de evitar que esta estructura represora invada el dominio (Fig. 4). Pero no en todos los casos los "insulators" se han encontrado flanqueando un *locus*, también puede haber "insulators" que se encuentren de un solo lado del *locus* o entre dos genes como el "insulator" del RMD del *locus* *H19/Igf2*, o incluso *locus* en los cuales no sería necesaria su presencia.

Todas estas evidencias muestran que el futuro del estudio de la regulación de la expresión genética se verá íntimamente relacionado con aspectos tanto topológicos, a través de la formación de dominios, cómo con distintas localizaciones y relocalizaciones de los dominios en nuevos compartimientos subnucleares.

#### CONCLUSIONES Y PERSPECTIVAS

El objetivo de esta revisión fue definir las características de un dominio transcripcional, sus elementos estructurales y de regulación y la manera cómo éstos colaboran para la expresión específica de los genes

contenidos en ellos. El hecho de que la cromatina sea una estructura altamente represora, da relevancia a la existencia de los elementos que son capaces de remodelarla para poder activar y regular la transcripción genética. Entre los elementos que regulan la transcripción revisamos a los LCRs, "enhancers" y los "insulators". Estos últimos tienen especial y novedosa importancia ya que son los encargados de delimitar a los dominios, de cooperar con distintos elementos para mantener a los dominios accesibles y protegerlos de los efectos de la cromatina adyacente. La organización del genoma en dominios independientes y la cooperación de los elementos inherentes a ellos para su regulación específica ha llevado a proponer a los dominios como un primer nivel de regulación de la expresión genética.

Al parecer, las investigaciones a futuro tienden a centrarse en dos tipos de enfoques: en estudios citológicos y estudios aplicables como la terapia génica. En los primeros se tendrá que dilucidar la relación entre los elementos de regulación de los dominios y los territorios cromosomales, al parecer se puede imaginar una relación entre los "insulators", MARs, "enhancers" y LCRs con la relocalización de los dominios a sitios donde pueden ser transcritos con mayor eficiencia y topológicamente, lejos o cerca (según sea requerido) de regiones potencialmente represoras dentro del núcleo. Esto ayudaría a entender mejor lo que hasta ahora ha permanecido como una caja negra: la manera como la arquitectura del núcleo favorece la formación de dominios y a tener una idea más clara de lo que ocurre realmente *in vivo* en cuanto a la dinámica de la regulación de la expresión genética.

Dentro de los estudios aplicables estaría el análisis de los "insulators" y secuencias MARs como elementos con una importancia fundamental en la terapia génica. Como se mencionó, el "insulator" 5'HS4 del locus  $\beta$ -globina de pollo está siendo utilizado para flanquear transgenes con resultados favorables. Se puede decir que éste es un claro ejemplo de cómo los estudios realizados en ciencia básica apoyan

a la ciencia aplicada en medicina.

Por último, si se supone que la formación de un dominio es importante, queda claro que no basta solamente con conocer las secuencias codificantes para entender el funcionamiento y regulación del genoma, y que esto debe ser tomado en cuenta al realizar e interpretar los resultados de proyectos tan ambiciosos como la secuenciación del genoma humano.

## AGRADECIMIENTOS

Agradecemos a los Drs. Marietta Tuena y Alejandro Zentella y a los miembros del laboratorio del Dr. Félix Recillas por sus sugerencias y lectura crítica de este manuscrito. Este trabajo fue apoyado por los donativos de la Dirección General de Asuntos del Personal Académico-UNAM (IN203200) y del Consejo Nacional de Ciencia y Tecnología-CONACyT (33863-N).

## REFERENCIAS

1. Luger K, Mader A, Richmond R, Sargent D y Richmond T (1997) Crystal structure of the nucleosome core particle at 2.8Å resolution. *Nature* 389: 251-260.
2. Widom J (1998) Chromatin Structure: linking structure to function with histone H1. *Curr Biol* 8: R788-R791.
3. Orphanides G y Reinberg D (2000) RNA polymerase II elongation through chromatin. *Nature* 407: 471-475.
4. Wu C, Wong Y C y Elgin S C (1979) The chromatin structure of specific genes: II Disruption of chromatin structure during gene activity. *Cell* 16: 807-814.
5. Boyes J y Felsenfeld G (1996) Tissue-specific factors additively increase the probability of the all-or-none formation of a hypersensitive site. *EMBO J* 15: 2496-2507.
6. Blackwood E y Kadonaga J (1998) Going the distance: A current view of enhancer action. *Science* 281: 60-63.
7. Kamakaka R (1997) Silencers and locus control regions: opposite sides of the same coin. *Trends Biochem Sci* 22: 124-128.
8. Martin D, Fiering S, Groudine M (1996) Regulation of  $\beta$ -globin gene expression: straightening out the locus. *Curr Op Gen Dev* 6: 488-495.
9. Fraser P y Grosveld F (1998) Locus control regions, chromatin activation and transcription. *Curr Op Gen Dev* 10: 361-365.
10. Bulger M y Groudine M (1999) Looping versus linking: toward a model for long-distance gene activation. *Genes Dev* 13: 2465-2477.
11. Ashe H, Monks J, Wijgerde M, Fraser P y Proudfoot N (1997) Intergenic transcription and transduction of the human  $\beta$ -globin locus. *Genes Dev* 11: 2494-2509.
12. Gribnau J, Diderich K, Pruzina S, Calzolari R y Fraser P (2000) Intergenic transcription and developmental remodeling of chromatin subdomains in the human  $\beta$ -globin locus. *Mol Cell* 5: 377-386.
13. Kellum R y Schedl P (1991) A position-effect assay for boundaries of higher order chromosomal domains. *Cell* 64: 941-950.
14. Gdula D, Gerasimova T y Corces V (1996) Genetic and molecular analysis of the gypsy chromatin insulator of *Drosophila*. *Proc Natl Acad Sci USA* 93: 9378-9383.
15. Chung J, Bell A y Felsenfeld G (1997) Characterization of the chicken  $\beta$ -globin insulator. *Proc Natl Acad Sci USA* 94: 575-580.
16. Bell A, West A y Felsenfeld G (1999) The protein CTCF is required for the enhancer-blocking activity of the vertebrate "insulators". *Cell* 98: 387-396.
17. Bell A y Felsenfeld G (2000) Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature* 405: 482-485.
18. Pikaart M J, Recillas Targa F y Felsenfeld G (1998) Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by "insulators". *Genes Dev* 12: 2852-2862.
19. Bell A y Felsenfeld G (1999) Stopped at the border: boundaries and "insulators". *Curr Opin Gen Dev* 9: 191-198.
20. Emery D, Yannaki E, Tubb J y Stamatoyannopoulos G (2000) A chromatin insulator protects retrovirus vectors from chromosomal position effects. *Proc Natl Acad Sci USA* 97: 9150-9155.

21. Forrester W, Fernández L y Grosschedl R (1999) Nuclear matrix attachment regions antagonize methylation-dependent repression of long-range enhancer-promoter interactions. *Genes Dev* 13: 3003-3014.
22. Lamond A y Earnshaw W (1998) Structure and function in the nucleus. *Science* 280: 547-553.
23. Gerasimova T, Byrd K y Corces V (2000) A chromatin insulator determines the nuclear localization of DNA. *Mol Cell* 6: 1025-1035.
24. Francastel C, Walters M, Groudine M y Martin D (1999) A functional enhancer suppresses silencing of a transgene and prevents its localization close to centromeric heterochromatin. *Cell* 99: 259-269.
25. Schübeler D, Francastel C, Cimbora D, Reik A, Martin D y Groudine M (2000) Nuclear localization and histone acetylation: a pathway for chromatin opening and transcriptional activation of the human  $\beta$ -globin locus. *Genes Dev* 14: 940-950.
26. Wu C (1997) Chromatin remodeling and the control of gene expression. *J Biol Chem* 272: 28171-18174.

Prospects and implications of using chromatin insulators in gene therapy and transgenesis

Félix Recillas-Targa<sup>1\*</sup>, Viviana Valadez-Graham<sup>1</sup> and Catherine M. Farrell<sup>2</sup>

<sup>1</sup>Departamento de Genética Molecular, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, Apartado Postal 70-242, México D.F. <sup>2</sup>National Institutes of Health, Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, Maryland 20892.

\* Corresponding author: Félix Recillas-Targa  
Instituto de Fisiología Celular  
Universidad Nacional Autónoma de México  
Departamento de Genética Molecular  
Apartado Postal 70-242  
México, 04510, D.F.  
Tel: (52 55) 56 22 56 74  
FAX: (52 55) 56 22 56 30  
Email: [frecilla@ifc.unam.mx](mailto:frecilla@ifc.unam.mx)  
<http://www.ifsiol.unam.mx/>

Abbreviations: Dnmt, DNA methyltransferase; HDAC, histone deacetylase; SUV39H1, histone methyltransferase; HP1, heterochromatin protein 1; LCR, locus control region; MAR, matrix attachment region; SAR, scaffold attachment region; cHS4, chicken 5'  $\beta$ -globin insulator; Mi-2, member of the SNF2 family of chromatin remodeling complexes; MBD, methyl-CpG-binding domain; AAV, adeno-associated virus; LTR, long terminal repeat; HS, DNase I hypersensitive site; MSCV, murine stem cell virus; GFP, green fluorescent protein.

## Summary

Gene therapy has emerged from the idea of inserting a wild-type copy of a gene in order to restore the proper expression and function of a damaged gene. Initial efforts have focused on finding the proper vector and delivery method to introduce a corrected gene to the affected tissue or cell type. Even though these first attempts are clearly promising, several problems remain unsolved. A major problem is the influence of chromatin structure on transgene expression. To overcome chromatin-dependent repressive transgenic states, researchers have begun to use chromatin regulatory elements to drive transgene expression. Insulators or chromatin boundaries are able to protect a transgene against chromatin position effects at their genomic integration sites, and they are able to maintain transgene expression for long periods of time. Therefore, these elements may be a very useful tool in gene therapy applications for ensuring high-level and stable expression of transgenes.

## **Introduction**

For the past fifteen years, gene therapy has become an important and useful tool with the potential to correct diseases.<sup>(1,2)</sup> Originally defined to contribute to the treatment or prevention of diseases, gene transfer has now found its limitations.<sup>(2)</sup> Based on continued progress in the field of gene expression and chromatin structure, we are now in a better position to understand integration processes and the defects and virtues of gene therapy vectors.

One of the critical aspects of introducing a transgene into the genome is the great variability or even a total lack of gene expression, events mainly attributed to the chromatin environment at the genomic insertion site. But this is not the only difficulty; for example, interference by viral regulatory elements, particularly repetitive sequences and silencers, can jeopardize transgene expression.<sup>(2)</sup> In recent years a novel type of chromatin regulatory element, named an insulator or boundary element, has emerged as a promising candidate to suppress transgene expression obstacles, particularly by its ability to block suppressive position effects in randomly selected integration sites.

In the present review we mainly focus our attention on this promising type of chromatin element, which we can seriously begin to consider as a candidate for use in different types of gene therapy vectors or transgenes.

### **Regulatory elements that counteract position effects**

For quite a long time it was assumed that all regulatory elements capable of counteracting position effects may be divided into two groups: dominant positive regulatory elements that impose an active chromatin configuration even in heterochromatic surroundings, and barrier elements that can physically block the passage of signals (either positive or negative) from the



surrounding chromatin. Locus control regions (LCRs) and strong enhancers constitute the first group of regulatory elements, while matrix attachment region elements and chromatin insulators belong to the second group.

The discovery of the LCR upstream of the human  $\beta$ -globin genes, with its ability to confer copy number-dependent gene expression independent of the integration site, originally made this element a promising candidate for incorporation in therapeutic vectors.<sup>(3-4)</sup> However, LCR regulatory elements have been constrained in their use, since these are tissue-specific, and in the case of retroviral vectors, their adverse effects on viral stability and titers are well documented.<sup>(4-7)</sup> Several alternative options have been tested with varied results, including core-LCR elements, or the  $\alpha$ -globin HS-40 element.<sup>(8-10)</sup> Despite this optimistic scenario, position effects and silencing of viral elements cannot be overcome in all cases,<sup>(11)</sup> since LCR-containing transgenes can be subject to position effects in stably transfected cell lines and in mice.<sup>(11,12)</sup>

Matrix- or scaffold-attachment regions (MARs/SARs) have also been used as alternatives for protection against position effects in vertebrates, with even more successful examples of their use seen in plants.<sup>(13-21)</sup> In the case of in the immunoglobulin  $\mu$  heavy chain locus, the enhancer needs to be flanked by MAR sequences for proper enhancer-promoter distant communication.<sup>(14)</sup> Further analysis demonstrated that the MAR sequences protect the enhancer against DNA methylation,<sup>(14)</sup> similar to the case of the chicken  $\beta$ -globin cHS4 insulator, where protection of regulatory elements against DNA methylation seems to be one of the strategies to allow sustained transgene expression (see below and Fig. 1A).<sup>(22,23)</sup> For the same MAR sequences, local attraction of histone acetylation has been demonstrated<sup>(24)</sup>, a property once more consistent with the cHS4  $\beta$ -globin insulator.<sup>(22,25,26)</sup>

In addition, an Alu element, a short interspersed repetitive sequence found in the 5' non-coding region of the human keratin 18 (K18) gene, has also the ability to protect a transgene

against position effects.<sup>(27)</sup> While this Alu element, LCRs, MARs/SARs and insulators are all capable of counteracting position effects, the only apparent concordance between them is that all are sequence elements located in non-coding regions.<sup>(28,29)</sup>

### **Insulators and transgene expression**

The barrier function of an insulator can protect a chromatin domain from the non-specific effects of the surrounding chromatin, by blocking the passage of regulatory signals from adjacent loci and chromatin domains.<sup>(29-31)</sup> Felsenfeld and collaborators have found that by flanking a transgene with the chicken  $\beta$ -globin insulator it is possible to protect a transgene from the position effects of the surrounding chromatin and from extinction of expression during long-term propagation in culture in the absence of drug selection (Fig. 1A).<sup>(32,33)</sup> An important characteristic of this particular insulator is that it functions in a vast number of vertebrate cell types and organisms tested, including human cells.<sup>(30)</sup>

Insulators are attractive chromatin components from both basic and practical perspectives. As far as gene therapy vectors are concerned, the obvious application of insulators is to create a mini-domain that is protected from the regulatory signals of the host cell genome (Fig. 1B). In this mini-domain, any necessary regulatory sequences could be incorporated along with the transgene itself. In the absence of an insulator, transgene expression is dominantly influenced by the surrounding heterochromatin, causing position effect variegation through heterochromatin spreading over a transgene.<sup>(34,35)</sup> This progressive heterochromatinization is associated with silencing epigenetic marks,<sup>(34)</sup> including histone deacetylation and DNA methylation, which have been clearly shown in silenced non-insulated transgenes.<sup>(22,23,32)</sup> Experimental data shows that sooner or later the great majority of non-insulated transgenes will extinguish their expression, although insulators are not needed in all cases, since strong promoters or enhancers can efficiently counteract progressive transgene silencing (see below).<sup>(36)</sup>

Thus, insulators seem to have mechanisms to counteract, at least partially, the silencing effects of heterochromatin. However, it should be mentioned that to date there is no direct demonstration that vertebrate insulators are able to block heterochromatin silencing. In addition, the position effect studies in the presence of insulators have mostly compared transgenes at different integration sites, and it will be necessary to analyze insulator activity at the same insertion sites, as well as integrants that are close to centromeric and telomeric sequences.

In addition to their ability to guard against position effects, insulators appear to have other favorable properties applicable to gene therapy and transgenesis. One possibility, though not well explored, is that insulators may contribute to the integration event to make this process more efficient. Insulators may also stabilize the transgene, avoiding deletions or recombination. This may have been a relevant factor in a transgenic study by O'Malley and collaborators, where no founder mice were obtained for an inducible transgene in the absence of insulators (Table 1).<sup>(37)</sup> In contrast, when the same transgene was shielded with two copies of the 1.2 kb DNA fragment of the cHS4 insulator (Fig. 1A), 4 founder mice were obtained, with 3 of them showing high and constant levels of expression (Table 1).<sup>(37)</sup>

Not only may the barrier function of an insulator be relevant in transgenesis, but the enhancer-blocking function may also have its applications. In the case of the cHS4 insulator, and in practically all vertebrate insulators found to date, the enhancer-blocking activity is attributable to the CTCF transcription factor (Fig. 1A).<sup>(28-31,38)</sup> Although it has been shown that the barrier and enhancer-blocking activities of the cHS4 insulator are separable, and the enhancer-blocking activity is not required for position effect protection in stably transfected cell lines (Fig. 1A),<sup>(33)</sup> there are instances where the enhancer-blocking function could potentially be applied. For example, in gene therapy, an important application of this property could be its potential to prevent insertional activation of adjacent oncogenes after integration in the genome, a problem that caused a leukemia-like disorder in a recent gene therapy protocol against an

immune disease.<sup>(2,39)</sup> Another contribution of the enhancer-blocking property arises from the fact that selectable markers incorporated in the same vector can influence transgene expression.<sup>(40)</sup> Thus, if the experimental transgene is flanked by insulators, the influence of these selectable gene regulatory elements could be prevented.<sup>(41)</sup> However, it should be noted that the inclusion of selectable markers creates a bias towards favorable chromatin integration sites, and only clones in which the marker gene is active are selected, thus preventing the analysis of clones in which the transgene and its associated marker gene have been completely silenced.<sup>(36)</sup>

An unexplored topic is the relocation of transgenes during gene expression extinction in vertebrate cells. Since it has been suggested that insulators can anchor the active domains to the nuclear periphery close to nuclear pores, in the absence of insulators the loss of such cytolocalization during transgene silencing may not be too unlikely.<sup>(42-45)</sup>

How does an insulator protect a transgene from chromosomal position effects and prevent extinction of its expression? Several studies using the chicken *CHS4*  $\beta$ -globin insulator and the *Drosophila gypsy* insulator, among others, have partially answered these questions, leading to three main models of insulator function.<sup>(28-31)</sup> All of these models are based on the supposition that, in contrast to the traditional view, insulators play an active role in establishing chromatin domains (Fig. 1B). These models are: 1) the insulator actively recruits histone acetylases (and perhaps other activities associated with active chromatin), maintaining a histone acetylated (open) chromatin configuration within the domain; 2) the insulator captures or titrates DNA methyltransferases and/or histone deacetylases and methylases, avoiding gene silencing within the protected domain and thus preventing heterochromatin-mediated domain repression; and 3) the insulator relocates the gene away from heterochromatic regions to an active sub-nuclear compartment, perhaps in association with the nuclear periphery and nuclear pore components.<sup>(42,43)</sup>

Detailed studies have recently addressed the mechanisms of transgene expression extinction and the mechanisms of protection conferred by the  $\beta$ -globin insulator. In these studies the DNA methylation and histone acetylation status of the transgene and its promoter were analyzed in insulated and non-insulated stably transformed cell lines.<sup>(22,23)</sup> In non-insulated lines the transgene body and the promoter region were DNA methylated, and the histones had low levels of acetylation. In contrast, in insulated lines the transgene maintained high levels of acetylation and the promoter was protected from DNA methylation in a region where transcription factors are known to bind, allowing transgene expression despite methylation of the gene body.<sup>(22)</sup> Furthermore, it was demonstrated that Mi-2 and a methyl-binding protein, MBD3, bind to silenced transgenes only in non-insulated and silenced cell lines,<sup>(22)</sup> suggesting that in addition to histone deacetylation, specific promoter methylation contributes to transgene silencing. Furthermore, in non-functional but insulated transgenes, acetylation by the insulator prevented high levels of specific promoter DNA methylation.<sup>(22)</sup> This notion is also consistent with a recent time-course study analyzing the order of events during silencing of the same transgene in the absence of insulators, where it was shown that specific promoter DNA methylation was acquired at a later stage than the loss of histone acetylation and histone H3-K4 methylation, which constitute primary events during silencing of this transgene.<sup>(23)</sup> Thus, taken together, these results gave further information about insulator function, suggesting that acetylation by the insulator prevents promoter methylation in a dominant way.<sup>(22,23)</sup> It has also been shown that this insulator is highly enriched in histone acetylation and histone H3-K4 methylation at its natural genomic location at the 5' end of the chicken  $\beta$ -globin domain, which is adjacent to a condensed chromatin region (Fig. 1A),<sup>(25,26)</sup> again supporting the model that the cHS4 insulator prevents transgene silencing by dominant maintenance of active chromatin marks.

### **Dominant enhancer-promoter elements as insulators**

There is not always a strict requirement for the use of insulators in transgene vectors, since the dominant action of a particular set of regulatory elements, such as certain promoters and enhancers, are sufficient to overcome position effects.<sup>(36)</sup> One possible explanation for the counteraction of position effects by dominant enhancer expression may be the particular cellular differentiation stage.<sup>(36)</sup> When transgene expression encounters the optimal differentiation context, the transcription activating elements will counteract the repressive position effects with some ease, even allowing continuous levels of expression over time. The dominant regulatory elements which overcome the need for insulator elements may not only be enhancer elements, but also can be certain cellular or viral promoters.<sup>(36)</sup> Obviously, a thorough study of the factors that bind to these elements will be necessary to understand how position effect protection is achieved, not only by these elements but by the classical vertebrate insulator elements as well.

In the presence of a strong enhancer it has been shown that the additional assistance of the chicken cHS4 insulator does not improve gene expression and position effect protection.<sup>(38,46)</sup> This observation is not surprising since part of the attractive aspect of insulators is their neutral action on the associated transgene, i.e., overall they lack enhancer or silencer activities and they preserve the intrinsic activity of the transgene regulatory elements.<sup>(38,46)</sup>

### **Insulators, transgenesis and gene therapy vectors**

Since 1998 there have been several studies of transgene expression where the transgene was flanked by the  $\beta$ -globin insulator (Table 1). This insulator has not only been used to improve transgene expression, but also to generate reproducible transgenic animal phenotypes, with promising results for future studies of human diseases. The impact of insulator use in transgenic animals is to reduce the number of transgenic founders required in order to obtain animals with an appropriate expression level and tissue distribution of the transgene.<sup>(47-50)</sup>

An alternative use of insulators is related to the improvement of inducible transgene expression, which can be easily explained by the fact that insulator sequences contribute to the generation of an open chromatin context that allows much faster and direct regulatory factor recognition and action, with no necessity for extensive chromatin remodeling.<sup>(40,51)</sup> This becomes relevant, particularly for inducible systems in which gene expression needs to be highly controlled, but most importantly in the scenario where the gene product is toxic. Thus, shielding those kinds of transgenes with insulators assures minimal basal expression. To this extent the chicken  $\beta$ -globin insulator has already been offered by Invitrogen in the commercial vector pBC1 with the aim to express genes of interest in the mammary glands of transgenic animals.

#### **Viral vectors and insulators**

Distinct types of viruses have been engineered to become highly specific and efficient for gene delivery to particular cell types or tissues.<sup>(2)</sup> Several types of viruses, including retroviruses, adenoviruses, adeno-associated viruses (AAV), herpes simplex viruses, and more recently lentiviruses, have been modified in the laboratory for their potential use in future gene therapy applications.<sup>(2)</sup> One could think that the use of insulator sequences would be restricted to the integrated types of viral vectors like retroviral or lentiviral vectors. Interestingly, an insulator contribution to episomally derived viral vectors is also relevant.<sup>(51-53)</sup>

Depending on the cell type and reporter gene used, complementary modifications of viral vectors are required in addition to the use of insulator sequences, with the aim to ameliorate transgene expression. These modifications include codon usage optimization and modifications of the control elements, as well as elimination of putative mammalian RNA processing signals to significantly improve transgene expression.<sup>(8,54)</sup> Such modifications can even contemplate mutations at CpG dinucleotides to avoid progressive and permanent DNA methylation.

With regard to using insulators in viral vectors, a major drawback is that insulators can affect vector titers and the stability of the viral particles.<sup>(55)</sup> In general, the reduction in vector titers roughly correlates with the size of the inserts.<sup>(56)</sup> Thus, the size of the insulator sequence to be employed should be taken into account and shorter DNA sequences may help to avoid vector titer problems (Fig. 1A). To this end, it has been demonstrated that two copies of the core cHS4 insulator (Fig. 1A), equivalent to 500 bp, are as efficient in protecting a transgene against position effects as are two copies of the original 1.2 kb DNA fragment.<sup>(33,57,58)</sup>

#### Studies with retroviral vectors:

In a study using recombinant murine retroviruses, it has been shown that the presence of the chicken cHS4 insulator does not affect viral titers and stability.<sup>(55)</sup> Interestingly, in the same viral system the 1.2 kb version of the cHS4 insulator was inserted into the 5' and 3' LTRs in both orientations, and in their native orientation (5'-3') the insulation was more than twice as efficient than in the opposite orientation (Fig. 2A).<sup>(55)</sup> Curiously, in other investigations using murine retroviral vectors, the orientation of the same insulator element did not interfere with viral transduction and expression levels, and the only difference from the previously mentioned work is that here two copies of the 1.2 kb DNA fragment were used as recommended originally (Fig. 1A).<sup>(32,33,59)</sup> In addition, in stable cell lines insulated retrovirus expression was 10-fold higher compared to that seen with a non-insulated vector, clearly showing the ability of the insulator to elevate and enable homogeneous virally transduced gene expression.<sup>(54)</sup> However, in contrast to this report, another report shows that vector titers can be affected by this insulator.<sup>(60)</sup>

Viral vectors are not only subject to the selective pressure of the chromatin site of integration, but in addition they are influenced positively, though mainly negatively, by viral regulatory and structural elements, like the viral long terminal repeats (LTRs) that are responsible for provirus silencing *in vivo*.<sup>(2,7)</sup> In the case of retroviral vectors, there are retroviral



regulatory elements that can induce silencing with a progressive increase in DNA methylation and methyl-CpG-binding protein recruitment, along with the associated histone deacetylation and methylation.<sup>(61)</sup> As a means of combating these negative viral elements, it has recently been demonstrated that the cHS4 insulator is able to block retroviral silencing in a murine stem cell virus (MSCV) vector in transgenic mice.<sup>(60)</sup> Interestingly, the results obtained with this MSCV retrovirus are in agreement with the data obtained previously in chicken stably transfected cell lines.<sup>(33,60)</sup> First, when two copies of the cHS4 core element are placed on each side of an LCR- $\beta$ -globin transgene (Fig. 3), the highest protection against silencing is obtained, an effect that is translated to the most elevated copy number and with homogeneous gene expression, again confirming that two copies of the 250 bp core insulator are as good as the two copies of the original 1.2 kb element (Fig. 3).<sup>(32,33,60)</sup> Second, deletion of FII corresponding to the CTCF DNA-binding sequence of the core insulator (Fig. 1A), which is responsible for the enhancer-blocking activity, does not perturb the protection against viral silencing, indicating that the enhancer- and silencer-blocking components of this insulator are separable, again consistent with the previous report (Fig. 3).<sup>(33)</sup> Third, when the minimal enhancer-blocking element containing FII and FIII was tested in quadruple, there was no transgene expression from the MSCV retrovirus vector, demonstrating that CTCF and the FIII-associated factor(s) accentuate silencing (Fig. 3). In summary, these results demonstrate that the topological conformation of protein-protein and protein-DNA interactions at the 250 bp core insulator are critical for the co-habitation of the enhancer-blocking, silencer-blocking and position effect protection activities, and all of these components, with the exception of the enhancer-blocking activity, are required to overcome retroviral silencing.

In yet another study with retroviral vectors, Rivella and collaborators demonstrated that viral transgene expression generated in mouse erythroleukemia cells in the absence of drug selection is improved by the presence of insulators, which can reduce the *de novo* methylation

originating at the 5' LTR of the retroviral vector (Fig. 2B).<sup>(59)</sup> This observation could imply, based on data recently obtained in *Neurospora crassa* and *Arabidopsis thaliana* and more recently at mammalian centromeric regions, that transcripts could be responsible for heterochromatin formation through histone H3-lys9 methylation and the incorporation of HP1.<sup>(62,63)</sup> Thus, in such a scenario the insulator may interfere with heterochromatin formation, possibly through active and continuous histone acetylation, as discussed above.<sup>(22,25)</sup> However, unexpectedly, when the same transgene was used in embryonic stem cells, the retroviral vectors were not protected against DNA methylation by the cHS4 insulator.<sup>(59)</sup> One possible explanation has to do with differentiation timing and *de novo* compared to maintenance DNA methylation, i.e., in stem cells insulators may not protect transgenes against *de novo* DNA methylation through the action of the Dnmt1 enzyme. In contrast, in differentiated cell types, Dnmt3a and Dnmt3b are the DNA methyl-transferases responsible for transgene silencing, in which case their activity could be counteracted by the cHS4 insulator.<sup>(63,64)</sup> An alternative explanation is that in stem cells there is a specific silencer factor interacting with the LTR and internal retroviral sequences, and the location of the insulator within the LTR (Fig. 2B) does not permit it to block such a dominant silencing effect, independently of the DNA methylation status. Nevertheless, no matter what the reason for the failure of the insulator to protect against DNA methylation in embryonic stem cells, these results underscore the fact that insulators may not function in all cell types, and this may be a limiting factor when a transgene needs to go through a developmental program before reaching its desired effect.

In an independent series of experiments with oncoretroviruses, the cHS4 insulator was used in combination with the  $\alpha$ -globin LCR ( $\alpha$ HS-40) to drive expression of the human  $\gamma$ -globin gene in a mouse bone marrow transduction and transplantation assay. In these studies there was a clear insulator effect, ranging from 2-5% expression for the non-insulated oncoretrovirus vector to 49% for the insulated vector (Fig. 2D).<sup>(8)</sup> Such an improvement represents a reasonable

increase in transgene expression that may allow us to reach therapeutic levels in future hemoglobinopathy protocols. Thus, the use of regulatory elements complementary to insulators, such as the  $\alpha$ -globin LCR element used in the above study, seems to be critical. An alternative aspect of insulator use in viral vectors could arise from selection or enrichment of the transduced cells. For example, stem or progenitor cells in which an integrated viral vector expresses the transgene efficiently can be pooled and enriched before transplantation. Such a strategy has already been attempted in retrovirally transduced bone marrow, yielding more than 9 months of continuous and constant human  $\beta$ -globin gene expression.<sup>(65)</sup> Thus, the incorporation of insulators may facilitate the initial selection process, in addition to their ability to maintain expression of the transgene.

#### Studies with adeno-associated viral vectors:

Adeno-associated virus vectors have also been tested for their improved capacity for gene expression in the presence of the cHS4 insulator (Fig. 2C).<sup>(57)</sup> In one study using this vector system, the incorporation of the LCR HS2-4 core sequences from the human  $\beta$ -globin locus in the absence of an insulator was not sufficient to overcome position effects. However, the addition of two copies of the 250 bp core cHS4 insulator (Fig. 2C) remarkably reduced gene expression variability,<sup>(57)</sup> again confirming that two copies of the 250 bp core element can maintain the same properties as two copies of the original 1.2 kb element.<sup>(33,60)</sup> Interestingly, fluorescence *in situ* hybridization in metaphase chromosomes has demonstrated single and random integration of the recombinant adeno-associated virus vector containing the core cHS4 insulator, strongly suggesting that the insulator is not forcing or driving the viral vector to a particular and favorable chromatin environment.<sup>(57)</sup> Such an observation reinforces the idea that the shielded

transgene is not being influenced by the integration site, and the pattern of expression reflects the intrinsic activity of the regulatory elements incorporated in it.

Other insulators, notably invertebrate insulators, have also been tested in AAV vector systems. In one such study, adenovirus-mediated vectors with heterologous promoters unsuccessfully attempted to use the *Drosophila gypsy* insulator, demonstrating the limitation of using *Drosophila* insulators in mammalian and viral vectors.<sup>(51)</sup> However, the *gypsy* insulator was able to stimulate expression levels from a retroviral vector in another study.<sup>(60)</sup> Interestingly, an insulator has recently been identified within a DNase I hypersensitive site of adeno-associated virus itself, where a 336 bp DNA fragment possesses both enhancer-blocking activity and the ability to protect an integrated transgene against chromosomal position effects in transfected mammalian cells.<sup>(66)</sup>

#### Studies with lentiviral vectors:

In recent years a viral alternative has emerged with lentiviruses based on their genetic stability, but they are still subject to chromosomal position effects.<sup>(7,67,68)</sup> In one study where a self-inactivating lentivirus expresses the GFP protein under the control of the human elongation 1 $\alpha$  promoter (Fig. 2E),<sup>(69)</sup> low frequency integration events are observed, and GFP expression is apparently maintained for over 60 days in culture. In this series of experiments the scaffold attachment sequence from the human interferon- $\beta$  gene locus and the chicken cHS4 insulator were incorporated in the lentiviral vector, with these elements placed downstream of the GFP transgene, either individually or in tandem (Fig. 2E).<sup>(69)</sup> If GFP expression was quantified in terms of GFP expressing cells, there were no clear differences between the presence and absence of the SAR and/or the cHS4 insulator. However, in the presence of both the SAR and cHS4 elements, but not the cHS4 insulator alone, there was an order of magnitude improvement in the

average of GFP fluorescence intensity when evaluated in long-term culture (until 60 days) in the absence of selection.<sup>(69)</sup> These results demonstrate that the SAR and cHS4 elements do not quantitatively improve transgene expression, but instead they qualitatively ameliorate GFP expression, supporting the notion that the lentivirus transgene is not increasing the number of expressing cells, but in contrast is counteracting chromosomal position effects, and in this case probably variegation. In a second example, similar lentiviral vectors with the human interferon- $\beta$  SAR sequence and the cHS4 insulator were tested in transduced cells of the human CD34<sup>+</sup> hematopoietic progenitor line KG1 and in primary CD34<sup>+</sup> cord blood cells differentiated *in vitro* into monocytes.<sup>(56)</sup> The results clearly show that incorporation of the SAR/cHS4 combination consistently yielded elevated and homogeneous levels of GFP expression for periods of up to 4 months, even with three different promoters.<sup>(56)</sup> Thus, the SAR/cHS4 sequences seem to be a combination that positively contributes to the protection against position effects in lentiviral vectors.

This prompts us to question how such a combination of elements may act to improve lentiviral transgene expression. A possible contribution of the human interferon- $\beta$  SAR element may be its ability to avoid *de novo* DNA methylation,<sup>(70)</sup> which may complement the role of the cHS4 insulator in methylation avoidance. Another possibility may lie in the nature of the SAR/MAR elements, which are defined as AT-rich sequences with no particular consensus, and they can interact with the nuclear matrix, contributing to the formation of loop attachment sites that are arranged in helical spirals along the axis of the metaphase scaffold. SARs may also include topoisomerase II consensus binding sequences, which may enable relaxation of supercoiled DNA molecules.<sup>(17)</sup> In this context, the presence of the SAR/cHS4 elements in lentiviral vectors could alter the superhelical tension of the DNA, favoring enhancer-promoter communication through the formation of a local and optimal topological conformation.

Another clue as to how the SAR/cHS4 combination works may come from studies of the *Drosophila* Su(Hw) insulator. When an enhancer is flanked by the Su(Hw) insulator, or if one copy is placed between an enhancer and promoter, the enhancer-blocking activity of this insulator is accentuated,<sup>(71,72)</sup> in agreement with the insertion of the cHS4 insulator in the U3 region of the 3' LTR of the lentiviral vector, which caused decreased transgene expression.<sup>(56)</sup> However, when two copies of the Su(Hw) insulator are placed between an enhancer and promoter, the enhancer activity is favored, probably by chromatin loop formation.<sup>(71,72)</sup> These observations are similar to the placement of the SAR/cHS4 elements together in the lentiviral vector (Fig. 2E), which consistently yielded elevated levels of transgene expression.<sup>(56)</sup> In such a scenario, possible loop formation, together with the protein complexes that interact with the insulator and SAR sequences, as well as the distance between the enhancer and promoter, should represent critical parameters for successful transgene activation.

In conclusion, one of the aspects emerging from these series of studies is that one chromatin component, being either a matrix/scaffold or an insulator element, is not always sufficient for homogeneous and sustained expression levels in viral vectors, but that combinations of these elements could contribute more effectively, with each element having its individual virtues for improvement of transgene expression.

## **Conclusions**

In recent years chromatin insulators or boundary elements have emerged as favorable candidates to improve transgene expression. The use of insulators, particularly vertebrate insulators, along with strategies such as naked DNA injection and the use of cationic lipids or viral vectors will certainly ameliorate gene expression in a particular target cell or tissue. The main goal is to establish, or at least attempt to estimate, appropriate gene transfer levels of the

therapeutic gene, and of course to reach adequate expression levels to achieve curative or corrective effects. For some applications reproducibility represents a key factor. Such an event can probably be reached if the insulators supply constant expression independently of the integration site.

However, several unsolved questions about the use of insulators remain. For instance, we do not know whether insulator activity is linked to cell cycle variations. If it is then it is possible that cell cycle variations could explain differences in insulation from cell type to cell type, or in particular cellular differentiation stages. As mentioned above, this may be a critical issue in transgene expression, especially in cases where the transgene needs to go through a differentiation program. The stability of the viral vectors is also another problem to be taken into consideration, since the presence of repetitive or identical DNA sequences shielding a transgene may favor undesirable recombination events in some cases. Thus, to improve vector stability, we may consider the incorporation of different insulators or insulators in combination with matrix or scaffold attachment sequences.<sup>(56)</sup> Another unsolved question is whether insulators are able to counteract variegated gene expression, and how this impacts transgene expression. To answer these questions, among others, it is certain that further basic studies on insulators and their mechanisms of action will be necessary before they can be successfully employed in gene therapy protocols. It is also likely that the separable functions of insulators (barrier activity and enhancer- or silencer-blocking activities) will have to be exploited, together with the use of complementary chromatin elements (e.g., MARs and LCRs), in order to ensure maximal benefits in transgenesis. In addition, it is likely that the beneficial components to be employed will vary for different transgenes and vectors, and various issues, such as where to place the insulator and/or other chromatin regulatory elements, will have to be addressed.

Another key issue is the optimization of vector design in order to ensure a successful therapeutic effect. From the data published it has become clear that a viral vector

requires a minute design that involves: i) knowledge of the internal regulatory and structural viral elements like enhancers, silencers, and LTRs; ii) based on such a distribution, the location where the insulator sequence will be incorporated becomes relevant; iii) the choice of transgene regulatory elements and the avoidance of elements detrimental to transgene expression, such as the avoidance of CpG-rich sequences to minimize DNA methylation; iv) the size and combination of insulators to be used to favor elevated titer production and improved vector stability; v) the insulators would probably need to have both properties, enhancer- or silencer-blocking and position effect protection activities, to ensure maximal action. At this point it is important to recall that an optimal vector design is not the ultimate solution, since there are other considerations beyond the vector and the transgene. One such consideration is the immune response, which has been one of the cornerstones of gene therapy protocols. As a consequence we probably need to consider engineering new vectors, not only with therapeutic genes, but in the same vector to incorporate genes that could modulate or diminish the immune response.<sup>(2)</sup> In this regard, it is encouraging that lentiviruses and adeno-associated viruses do not seem to induce an inflammatory or immune response.<sup>(2)</sup>

At the present time there is an intense debate concerning the risks of "insertional mutagenesis" of viral and non-viral vectors as a cause of cancer or a mis-reaction in a gene therapy protocol. One possible alternative is to direct the insertion of vectors to a "secure" non-coding area of the genome. We would like to speculate two possibilities for achieving this: the first one is by using the homologous recombination Cre-Lox system in combination with insulators, and the second option is to drive the insertion event directly into heterochromatin in an attempt to avoid integration in coding sequences.<sup>(61,73)</sup> Obviously, this strategy would require the presence of insulators in the transgenes, and it would also require some certainty of the effectiveness of the insulator shielding. Clearly, further studies of insulator function in repressive chromatin environments will be necessary to employ such a strategy.



Another important consideration in gene therapy is the type of cells the therapeutic vector is delivered to and how these cells could subsequently be delivered to a patient to ultimately cure a human genetic defect. Obviously, this subject becomes entangled with many ethical considerations, such as the debate over the use human embryonic stem cells. Among the various options, stem cells derived from adult bone marrow, including hematopoietic and mesenchymal stem cells, have become promising candidates for use in human gene therapy.<sup>(74)</sup> Nonetheless, no matter what the route chosen for a human gene therapy protocol, in terms of using insulator elements it is still necessary to ensure that the insulators are effective in the cells they are delivered to and at the appropriate developmental stage for correcting the genetic defect.

Thus, this is one of the first reviews that summarizes transgenes, gene therapy vectors and the use of chromatin insulators. Our goal was to outline and critically reconsider all advantages objectively, but at the same time we consider the limitations of using such chromatin elements. Despite these limitations and the fact that we do not fully understand how they function, insulator elements (perhaps in combination with other chromatin elements) will probably contribute to the next generation of gene therapy vectors. In any case, these elements give way to a new approach for the design of gene therapy vectors that will increase the potential of gene therapy protocols and the way that gene therapy is seen nowadays.

### **Acknowledgments**

We thank James Ellis and Sergey V. Razin for helpful discussions and comments on the manuscript.

This work was supported by the Dirección General de Asuntos del Personal Académico-UNAM (IN203200), the Consejo Nacional de Ciencia y Tecnología, CONACyT (33863-N), the Third World Academy of Sciences (TWAS, Grant 01-055 RG/BIO/LA), and the Fundación Miguel Alemán.

## References

1. Rivella S, Sadelain M. Genetic treatment of severe hemoglobinopathies: the combat against transgene variegation and transgene silencing. *Sem Hematol* 1998; 35: 112-125.
2. Thomas CE, Ehrhardt A, Kay MA. Progress and problems with the use of viral vectors for gene therapy. *Nat Rev Genet* 2003; 4: 346-358.
3. Grosveld F, van Assendelft GB, Greaves DR, Kollias G. Position-independent, high-level expression of the human  $\beta$ -globin gene in transgenic mice. *Cell* 1987; 51: 975-985.
4. Sadelain M, Wang CH, Antoniou M, Grosveld F, Mulligan RC. Generation of a high-titer retroviral vector capable of expressing high levels of the human  $\beta$ -globin gene. *Proc Natl Acad Sci USA* 1995; 92: 6728-6732.
5. Leboulch P, Huang GM, Hamphries RK, Oh YH, Eaves CJ, Juan DY, London IM. Mutagenesis of retroviral vectors transducing human  $\beta$ -globin gene and  $\beta$ -globin locus control region derivatives results in stable transmission of an active transcriptional structure. *EMBO J* 1994; 13: 3065-3076.
6. Emery DW, Chen H, Li Q, Stamatoyannopoulos G. Development of a condensed locus control region cassette and testing in retrovirus vectors for  $\gamma$ -globin. *Blood Cells Mol Dis* 1998; 24: 322-339.
7. Pannell D, Ellis J (2001) Silencing of gene expression: implications for design of retrovirus vectors. *Rev Med Virol* 2001; 11: 205-217.
8. Emery DW, Yannaki E, Tubb J, Nishino T, Li Q, Stamatoyannopoulos G. Development of virus vectors for gene therapy of Cain hemoglobinopathies: flanking with a chromatin insulator reduces  $\gamma$ -globin gene silencing in vivo. *Blood* 2002; 100: 2012-2019.
9. Festenstein R, Tolaini M, Corbella P, Mamalaki C, Parrington J, Fox M, Miliou A, Jones M, Kioussis D. Locus control region function and heterochromatin-induced position effect variegation. *Science* 1996; 271: 1123-1125.
10. Festenstein R, Kioussis D (2000) Locus control regions and epigenetic chromatin modifiers. *Curr Opin Genet Dev* 2000; 10: 199-203.

11. Ellis J, Pannell D. The  $\beta$ -globin locus control region versus gene therapy vectors: a struggle for expression. *Clin Genet* 2001; 59: 17-24.
12. Alami R, Grealley JM, Tanimoto K, Hwang S, Feng YQ, Engel JD, Fiering S, Bouhassira EE.  $\beta$ -globin YAC transgenes exhibit uniform expression levels but position effect variegation in mice. *Hum Mol Genet* 2000; 9: 631-636.
13. Phi-Van L, Strätling WH. The matrix attachment regions of the chicken lysozyme gene co-map with the boundaries of the chromatin domain. *EMBO J* 1988; 7: 655-664.
14. Forrester WC, Fernández LA, Grosschedl R. Nuclear matrix attachment regions antagonize methylation-dependent repression of long-range enhancer-promoter interactions. *Genes Dev* 1999; 13: 3003-3014.
15. Lutzki C, Senaheera D, Skelton D, Petersen D, Kohn DB. Lentivirus vectors incorporating the immunoglobulin heavy chain enhancer and matrix attachment regions provide position-independent expression in B lymphocytes. *J Virol* 2003; 77: 7341-7351.
16. Mankin SL, Allen GC, Phelan T, Spiker A, Thompson WF. Elevation of transgene expression level by flanking matrix attachment regions (MAR) is promoter dependent: a study of the interactions of six promoters with the RB7 3'MAR. *Trans Res* 2003; 12: 3-12.
17. Hart CM, Laemmli UK. Facilitation of chromatin dynamics by SARs. *Curr Opin Genet Develop* 1998; 8: 519-525.
18. Namciu SJ, Blochlinger KB, Fournier RE. Human matrix attachment regions insulate transgene expression from chromosomal position effects in *Drosophila melanogaster*. *Mol Cell Biol* 1998; 18: 2382-2391.
19. Allen GC, Spiker S, Thompson WF. Use of matrix attachment regions (MARs) to minimize transgene silencing. *Plant Mol Biol* 2000; 43: 361-376.
20. Petersen K, Leah R, Knudsen S, Cameron-Mills V. Matrix attachment regions (MARs) enhance transformation frequencies and reduce variance of transgene expression in barley. *Plant Mol Biol* 2002; 49: 45-58.
21. Fukuda Y, Nishikawa S. Matrix attachment regions enhance transcription of a downstream transgene and the accessibility of its promoter region to micrococcal nuclease. *Plant Mol Biol* 2003; 51: 665-675.
22. Mutskov VJ, Farrell CM, Wade PA, Wolffe AP, Felsenfeld G. The barrier function of an insulator couples high histone acetylation levels with specific protection of promoter DNA from methylation. *Genes Dev* 2002; 16: 1540-1554.
23. Mutskov V, Felsenfeld G. Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9. *EMBO J* 2004; 23: 138-149.

24. Fernández LA, Winkler M, Grosschedl R. Matrix attachment region-dependent function of the immunoglobulin  $\mu$  enhancer involves histone acetylation at a distance without changes in enhancer occupancy. *Mol Cell Biol* 2001; 21: 196-298.
25. Litt MD, Simpson M, Recillas-Targa F, Prioleau M-N, Felsenfeld G. Transitions in histone acetylation reveal boundaries of three separately regulated neighboring loci. *EMBO J* 2001; 20: 2224-2235.
26. Litt MD, Simpson M, Gaszner M, Allis CD, Felsenfeld G. Correlation between histone lysine methylation and developmental changes at the chicken  $\beta$ -globin locus. *Science* 2001; 293: 2453-2455.
27. Willoughby DA, Vilalta A, Oshima RG. An Alu element from the K18 gene confers position-independent expression in transgenic mice. *J Biol Chem* 2000; 275: 759-768.
28. Recillas-Targa F, Razin SV. Chromatin domains and regulation of gene expression: familiar and enigmatic clusters of chicken globin genes. *Crit Rev Eukaryot Gene Expr* 2001; 11: 227-242.
29. Razin SV, Farrell CM, Recillas-Targa F. Genomic domains and regulatory elements operating at the domain level. *Int Rev Cytol* 2003; 226: 63-125.
30. West AG, Gaszner M, Felsenfeld G. Insulators: many functions, many mechanisms. *Genes Dev* 2002; 16: 271-288.
31. Burgess-Beusse B, Farrell C, Gaszner M, Litt M, Mutskov V, Recillas-Targa F, Simpson M, West A, Felsenfeld G. The insulation of genes from external enhancers and silencing chromatin. *Proc Natl Acad Sci USA*, 2002; 99, suppl.4: 16433-16437.
32. Pikaart MJ, Recillas-Targa F, Felsenfeld G. Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators. *Genes Dev* 1998; 12: 2852-2862.
33. Recillas-Targa F, Pikaart MJ, Burgess-Beusse B, Bell AC, Litt MD, West AG, Gaszner M, Felsenfeld G. Position-effect protection and enhancer blocking by the chicken  $\beta$ -globin insulator are separable activities. *Proc Natl Acad Sci USA* 2002; 99, 6883-6888.
34. Ayyanathan K, Lechner MS, Bell P, Maul GG, Schultz DC, Yamada Y, Tanaka K, Torigoe K, Rauscher III FJ. Regulated recruitment of HP1 to a euchromatic gene induces mitotically heritable, epigenetic gene silencing: a mammalian cell culture model of gene variegation. *Genes Dev* 2003; 17: 1855-1869.
35. Donze D, Kamakaka RT. Braking the silence: how heterochromatic gene repression is stopped in its tracks. *BioEssays* 2002; 24: 344-349.
36. Sutter NB, Scalzo D, Fiering S, Groudine M, Martin DIK. Chromatin insulation by a transcriptional activator. *Proc Natl Acad Sci USA* 2003; 100: 1105-1110.

37. Wang Y, DeMayo FJ, Tsai SY, O'Malley BW. Ligand-inducible and liver-specific target gene expression in transgenic mice. *Nat Biotechnol* 1997; 15: 239-243.
38. Bell AC, West AG, Felsenfeld G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* 1999; 98: 387-396.
39. Hacein-Bey-Abina S, von Kalle C, Schmidt M, Le Deist F, Wulffraat N, McIntyre E, Radford I, Villeral JL, Fraser CC, Cavazzana-Calvo M et al. A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. *N Engl J Med* 2003; 348: 255-256.
40. Anastatssiadis K, Kim J, Daigle N, Sprengel R, Schöler HR, Stewart AF. A predictable ligand regulated expression strategy for stably integrated transgenes in mammalian cells in culture. *Gene* 2002; 298: 159-172.
41. Hasegawa K, Nakatsuji N. Insulators prevent transcriptional interference between two promoters in a double gene construct for transgenesis. *FEBS Lett* 2002; 520: 47-52.
42. Gerasimova TI, Byrd K, Corces VG. A chromatin insulator determines the nuclear localization of DNA. *Mol Cell* 2000; 6: 1025-1035.
43. Ishii K, Arib G, Lin C, Van Houwe G, Lemmli UK. Chromatin boundaries in budding yeast: the nuclear pore connection. *Cell* 2002; 109: 551-562.
44. Francastel C, Magis W, Groudine M. Nuclear relocation of a transactivator subunit precedes target gene activation. *Proc Natl Acad Sci USA* 2001; 98, 12120-12125.
45. Yusufzai TM, Tagami H, Nakatani Y, Felsenfeld G. CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. *Mol Cell* 2004; 13: 291-298.
46. Recillas-Targa F, Bell AC, Felsenfeld G. Positional enhancer-blocking activity of the chicken  $\beta$ -globin insulator in transiently transfected cells. *Proc Natl Acad Sci USA* 1999; 96:14354-14359.
47. Taboit-Dameron F, Malassagne B, Viglietta C, Puissant C, Leroux-Coyau M, Chéreau Ch, Attal J, Weill B, Houbedine L-M. Association of the 5'HS4 sequence of the chicken  $\beta$ -globin locus control region with human EF1 $\alpha$  gene promoter induces ubiquitous and high expression of human CD55 and CD59 cDNAs in transgenic rabbits. *Transg Res* 1999; 8: 223-235.
48. Potts W, Tucker D, Wood H, Martin C. Chicken  $\beta$ -globin 5'HS4 insulators function to reduce variability in transgenic founder mice. *Biochem Biophys Res Commun* 2000; 273: 1015-1018.
49. Guglielmi L, LeBert M, Truffinet V, Cogne M, Denizot Y. Insulators to improve expression of a 3' IgH LCR-driven reporter gene in transgenic mouse models. *Biochem Biophys Res Commun* 2003; 307: 466-471.
50. Frazar TF, Weisbein JL, Anderson SM, Cline AP, Garrett LJ, Felsenfeld G, Gallagher PG, Bodine DM. Variegated expression from the murine band 3 (AE1) promoter in transgenic

- mice is associated with mRNA transcript initiation at upstream start sites and can be suppressed by the addition of the chicken  $\beta$ -globin 5'HS4 insulator element. *Mol Cell Biol* 2003; 23: 4753-4763.
51. Steinwaerder DA, Lieber A. Insulation from viral transcriptional regulatory elements improves inducible transgene expression from adenovirus vectors in vitro and in vivo. *Gene Ther* 2000; 7: 556-567.
  52. Vassaux G, Hurst HC, Lemoine NR. Insulation of a conditionally expressed transgene in an adenoviral vector. *Gene Ther* 1999; 6: 1192-1197.
  53. Ye X, Liang M, Meng X, Ren XW, Chen H, Li Z-Y, Ni S, Lieber A, Hu F. Insulation from viral transcriptional regulatory elements enables improvement to hepatoma-specific gene expression from adenovirus vectors. *Biochem Biophys Res Commun* 2003; 307: 759-764.
  54. Wells KD, Foster JA, Moore K, Pursel VG, Wall RJ. Codon optimization, genetic insulation and an rTA reporter improve performance of the tetracycline switch. *Transg Res* 1999; 8: 371-381.
  55. Emery DW, Yannaki E, Tubb J, Stamatoyannopoulos G. A chromatin insulator protects retrovirus vectors from chromosomal position effects. *Proc Natl Acad Sci USA* 2000; 97: 9150-9155.
  56. Ramezani A, Hawley TS, Hawley RG. Performance- and safety-enhanced lentiviral vectors containing the human interferon- $\beta$  scaffold attachment region and the chicken  $\beta$ -globin insulator. *Blood* 2003; 101: 4717-4724.
  57. Inoue T, Yamaza H, Sakai Y, Mizuno S, Ohno M, Hamasaki N, Fukumaki Y. Position-independent human  $\beta$ -globin gene expression mediated by a recombinant adeno-associated virus vector carrying the chicken  $\beta$ -globin insulator. *J Hum Genet* 1999; 44: 152-162.
  58. Qu Z, Thottassery JV, Ginkel SV, Manuvakhova M, Westbrook L, Roland-Lazenby C, Hays S, Kern FG. Homogeneity and long-term stability of tetracycline-regulated gene expression with low basal activity by using the rTA2S-M2 transactivator and insulator-flanked reporter vectors. *Gene* 2004; in press.
  59. Rivella S, Callergari JA, May C, Tan CW, Sadelain M. The cHS4 insulator increases the probability of retroviral expression at random chromosomal integration sites. *J Virol* 2000; 74: 4679-4687.
  60. Yao S, Osborne CS, Bharadwaj RR, Pasceri P, Sukonnik T, Pannell D, Recillas-Targa F, West AG, Ellis J. Retrovirus silencer blocking by the cHS4 insulators is CTCF independent. *Nucleic Acids Res* 2003; 31: 5317-5323.
  61. Lorincz MC, Schübeler D, Hutchinson SR, Dickerson DR, Groudine M. DNA methylation density influences the stability of an epigenetic imprint and Dnmt3a/b-independent *de novo* methylation. *Mol Cell Biol* 2000; 22: 7572-7580.

62. Lachner M, Jenuwein T. The many faces of histone lysine methylation. *Curr Opin Cell Biol* 2002; 14: 286-298.
63. Santos F, Zakhartchenko V, Stojkovic M, Peters A, Jenuwein T, Wolf E, Reik W, Dean W. Epigenetic marking correlates with developmental potential in cloned bovine preimplantation embryos. *Curr Biol* 2003; 13: 1116-1121.
64. Lehnertz B, Ueda Y, Derijck AA, Braunschweig U, Perez-Burgos L, Kubicek S, Chen T, Li E, Jenuwein T, Peters AH. Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr Biol* 2003; 13: 1192-1200.
65. Kalberer CP, Pawliuk R, Imren S, Bachelot T, Takekoshi KJ, Fabry M, Eaves CJ, London IM, Humphries RK, Leboulch P. Preselection of retrovirally transduced bone marrow avoids subsequent stem cell gene silencing and age-dependent extinction of human  $\beta$ -globin in engrafted mice. *Proc Natl Acad Sci USA* 2000; 97: 5411-5415.
66. Ogata T, Kozuka T, Kanda T. Identification of an insulator in AAVS1, a preferred region for integration of adeno-associated virus DNA. *J Virol* 2003; 77: 9000-9007.
67. May C, Rivella S, Callegari J, Heller G, Gaensler KM, Luzzatto L, Sadelain M. Therapeutic haemoglobin synthesis in  $\beta$ -thalassaemic mice expressing lentivirus-encoded human  $\beta$ -globin. *Nature* 2000; 406: 82-86.
68. Pawliuk R, Westerman KA, Fabry ME, Payen E, Tighe R, Bouhassira EE, Acharya SA, Ellis J, London IM, Eaves CJ, et al. Correction of sickle cell disease in transgenic mouse model by gene therapy. *Science* 2001; 294: 2368-2371.
69. Ma Y, Ramazani A, Lewis R, Hawley RG, Thomson JA. High-level sustained transgene expression in human embryonic stem cells using lentiviral vectors. *Stem Cells* 2003; 21: 111-117.
70. Dang Q, Auten J, Plavec I. Human  $\beta$  interferon scaffold attachment region inhibits de novo methylation and confers long-term, copy number-dependent expression to a retroviral vector. *J Virol* 2000; 74: 2671-2678.
71. Cai HN, Shen P. Effects of cis arrangement of chromatin insulators on enhancer-blocking activity. *Science* 2001; 291: 493-495.
72. Muravyova E, Golovnin A, Gracheva E, Parshikov A, Belenkaya T, Pirrotta V, Georgiev P. Loss of insulator activity by paired Su(Hw) chromatin insulators. *Science* 2001; 291: 495-498.
73. Feng YQ, Seibler J, Alami R, Eisen A, Westerman KA, Leboulch P, Fiering S, Bouhassira EE. Site-specific chromosomal integration in mammalian cells: highly efficient CRE recombinase-mediated cassette exchange. *J Mol Biol* 1999; 292: 775-785.

74. Prockop DJ, Gregory CA, Spees JL. One strategy for cell and gene therapy: harnessing the power of adult stem cells to repair tissues. *Proc Natl Acad Sci USA* 2003; 100 (suppl. 1): 11917-11923.
75. Song A, Winer S, Tasi H, Sampson A, Pasceri P, Ellis J, Elliot JF, Dosch H-M. Deviation of islet autoreactivity to cryptic epitopes protects NOD mice from diabetes. *Eur J Immunol* 2003; 33: 546-555.
76. Guglielmi L, LeBert M, Truffinet V, Cogné M, Denizot Y. Insulators to improve expression of a 3' IgH LCR-driven reporter gene in transgenic mouse models. *Biochem Biophys Res Commun* 2003; 307: 466-471.
77. Guglielmi L, Truffinet V, Cogné M, Denizot Y. The  $\beta$  globin HS4 insulator confers copy-number expression of IgH regulatory elements in stable B cell transfectants. *Immunol Letters* 2003; 89: 19-123.
78. Rival-Grevier S, Pantano T, Viglietta C, Prince S, Attal J, Jovilet G, Houdebine L-M. The insulator effect of the 5'HS4 region of the  $\beta$ -globin chicken locus on the rabbit *WAP* gene promoter activity in transgenic mice. *Transg Res* 2003; 12: 723-730.
79. Izumi M, Gilbert DM. Homogeneous tetracycline-regulatable gene expression in mammalian fibroblasts. *J Cell Biochem* 1999; 76: 280-289.
80. Johansen J, Tornoe J, Moller A, Johansen TE. Increase *in vitro* and *in vivo* transgene expression levels mediated through *cis*-acting elements. *J Gene Med* 2003; 5: 1080-1089.



**Table 1.** Use of the chicken 5'cHS4  $\beta$ -globin insulator in transgene expression

<b>Organism, cells and vectors</b>	<b>Regulatory elements and transgenes</b>	<b>Observations</b>	<b>Reference</b>
Transgenic mouse	Human transthyretin promoter driving a chimeric Gi-VP liver-specific <i>trans</i> -activator.	No transgenic mice from 10 founders with no insulator. With 2 x 1.2 kb insulator, 3 of 4 mice with high expression.	37
Trangenic mouse	Mouse tyrosinase promoter regulating the mouse tyrosinase minigene.	Homogenous pigmentation distribution with 2 x 1.2 kb cHS4 insulator.	48
Transgenic mouse	Anion exchange protein 1 (AE1) promoter/human $\gamma$ -globin transgene.	Position-independent and copy-number-dependent expression. Insulated transgene favors correct mRNA transcription initiation site.	50
Transgenic mouse	Tetracycline-regulated tet07 minimal promoter ICA69 transgene to generate a NOD mouse. Model system for human type I diabetes.	Prevents leakage and dramatically enhances transgene expression. 2 x 1.2 kb insulator used.	75
Transgenic mouse	pV <sub>H</sub> promoter and the 3' IgH-LCR fused to GFP.	B cell-specific and sustained GFP expression. In some transgenic lines there is progressive silencing with animal age that seems to be associated with DNA methylation.	76
Trangenic rabbit	Human elongation factor 1 $\alpha$ promoter expressing CD55 and CD59.	Enhancement of CD55 and CD59 mRNA concentrations in different tissues.	47
Transgenic mouse	Rabbit whey acidic protein WAP gene promoter regulating the luciferase gene.	All 6 lines generated in the presence of the cHS4 insulator expressed the transgene.	78

		Two out of 8 expressed the non-insulated transgene. Lack of copy-number-dependence and ectopic variability in expression observed.	
Human breast cancer ML20 cells derived from MCF-7	TRE-reporter vector, responsive to rTA25-M2 <i>trans</i> -activation, directing the expression of humanized <i>Renilla</i> green fluorescent protein.	The use of the double core 2 x 250 bp cHS4 insulator reduced the level of expression in the uninduced state. Does not affect maximal expression levels and allows homogeneous induction and low doses of doxycycline.	58
Murine pre-B and B cell lines	pV <sub>H</sub> promoter-fused to GFP linked to 3' or 5' murine IgH regulatory elements.	IgH elements now behave like a real LCR when the transgene is insulated.	77
Chinese Hamster ovary, CHO cell line	tTA and hygromycin resistance gene were co-expressed. Ptet promoter responsive to tTA drives the expression of GFP.	Both reporter vectors are flanked by 2 x 1.2 kb cHS4 insulator. Results show only partial protection against position effects.	79
Human embryonic stem cells transduced with lentiviral vectors	Human elongation factor 1 $\alpha$ (EF1 $\alpha$ ) using GFP as transgene.	Incorporation of the human interferon- $\beta$ MAR and cHS4 insulator drastically reduces GFP expression variability.	69
Rat HiB5 stem cell line and graft to the rat striatum	Human UbC promoter fused to GFP and WPRE, a regulatory element from Woodchuck hepatitis virus.	Flanking the transgene with two copies of the 1.2 kb cHS4 insulator significantly increased protein and mRNA expression levels.	80
Human CD34+ hematopoietic progenitor line KG1a and primary CD34+ cord blood cells differentiated <i>in vitro</i> to monocytes; transduced with lentiviral vectors	Human elongation factor 1 $\alpha$ (EF1 $\alpha$ ) using GFP as transgene.	Incorporation of the human interferon- $\beta$ MAR and cHS4 insulator shows higher and homogeneous transgene expression when both elements are incorporated.	56
Retrovirus	GFP gene transcribed from the LTR promoter.	Significant increase in the fraction of transduced cells.	55
Retrovirus	Cell surface marker NTP in recombinant retrovirus derived from murine leukemia virus.	Infection performed without any selection. cHS4 insulator prevents silencing and delays DNA methylation of the LTR.	59
Murine stem cell virus (MSCV) in transgenic mouse	Human HS2-4-LCR driving expression of the $\beta$ -globin gene.	CTCF-independent retrovirus silencer-blocking activity with the cHS4 insulator. When two insulator dimers are used retroviral vector titers are compromised.	60

## Figure Legends

**Figure 1.** The chicken  $\beta$ -globin domain and genomic organization. **A:** The chromatin domain boundaries have been determined and the *in vivo* interaction of CTCF demonstrated.<sup>(31,33)</sup> The chicken  $\beta$ -globin 5'cHS4 insulator is located within a constitutive DNase I hypersensitive site and represents the transition between an open chromatin conformation and 16 kb of condensed chromatin.<sup>(25,31,33)</sup> The cHS4 insulator possesses separable enhancer-blocking and position effect (PE) protection properties.<sup>(33)</sup> The core element contains at least five protein-binding sites termed FI to FV.<sup>(31)</sup> The enhancer-blocking activity is mainly dependent on the 11-zinc finger CTCF protein.<sup>(31,38)</sup> In contrast, the CTCF site is dispensable for protection against position effects.<sup>(31,33)</sup> **B:** Theoretically, insulators had been proposed to shield a gene or group of genes (arrows) against the spreading of chromatin modifications and to avoid promiscuity between regulatory signals from distinct or unrelated domains (red line). Such properties contribute to position effect protection when the transgene is surrounded by insulators.<sup>(30,31,33)</sup>

**Figure 2.** Viral vectors employing the chicken cHS4  $\beta$ -globin insulator. **A:** MSCV-based vector named MGPN2 using one copy of the 1.2 kb cHS4 insulator (Ins 5' cHS4) expressing the green fluorescence protein (GFP, green) from the 5'LTR promoter (blue); Enh, enhancer; pA, polyadenylation site.<sup>(55)</sup> The GFP activities were determined by flow cytometric analysis.<sup>(55)</sup> **B:**

Moloney MuLC (Mo-MuLV) derived virus encodes a cell surface marker named NTP.<sup>(59)</sup> Mouse erythroleukemic MEL cells were injected and the viral vector was designed with the incorporation of one copy of the 1.2 kb cHS4 insulator.<sup>(59)</sup> Interestingly, the presence of the insulator decreases the level of DNA methylation originating from both the 5' and 3' LTRs.<sup>(59)</sup> C: Recombinant adeno-associated virus (rAAV). The herpes virus thymidine kinase (tk) promoter drives expression of the neomycin-resistance gene (*Neo<sup>r</sup>*). The human  $\beta$ -globin gene is present along with the LCR hypersensitive sites (HS2, HS3 and HS4).<sup>(57)</sup> This is the first time the core (2 copies of the 250 bp fragment, see Fig. 1A) cHS4 insulator is used successfully in a viral vector to protect a transgene against position effects in infected MEL cells.<sup>(57)</sup> D: MLV-based oncoretrovirus vector containing coding elements from the human  $\gamma$ -globin gene linked to the  $\alpha$ -globin LCR  $\alpha$ HS-40.<sup>(8)</sup> One copy of the chicken 1.2 kb cHS4 insulator was placed upstream of the 5'LTR and between the  $\alpha$ HS-40 element and the 3'LTR.<sup>(8)</sup> E: Self-inactivating lentiviral vector with the human elongation factor 1 $\alpha$  (EF1 $\alpha$ ) promoter driving the expression of the *GFP* gene.<sup>(69)</sup> For optimal ES cell transduction the lentivirus vector incorporates the human interferon- $\beta$  SAR element and one copy of the 1.2 kb cHS4 insulator inserted in U3 region of the 3'LTR.<sup>(69)</sup>

**Figure 3.** Silencer-blocking activity of diverse DNA fragments of the chicken cHS4  $\beta$ -globin insulator. The MSCV retrovirus vector actively silences the human LCR- $\beta$ -globin reporter gene in transgenic mice.<sup>(60)</sup> Vectors containing 2 copies of the 1.2 kb cHS4, two copies of the 250 bp core cHS4,<sup>(31,33,38,46)</sup> four copies of the 90 bp DNA fragment including FII and FIII (termed Quad), and 2 copies of the core FII (the CTCF binding site) are shown.<sup>(33,38,60)</sup> Deletion of FII does not affect position effect protection and silencer-blocking activities.<sup>(33,60)</sup> Instead, the Quad

consistently reinforces silencing in stably transformed cell culture lines<sup>(33)</sup> and in transgenic mice generated by MSCV retrovirus infection.<sup>(60)</sup>

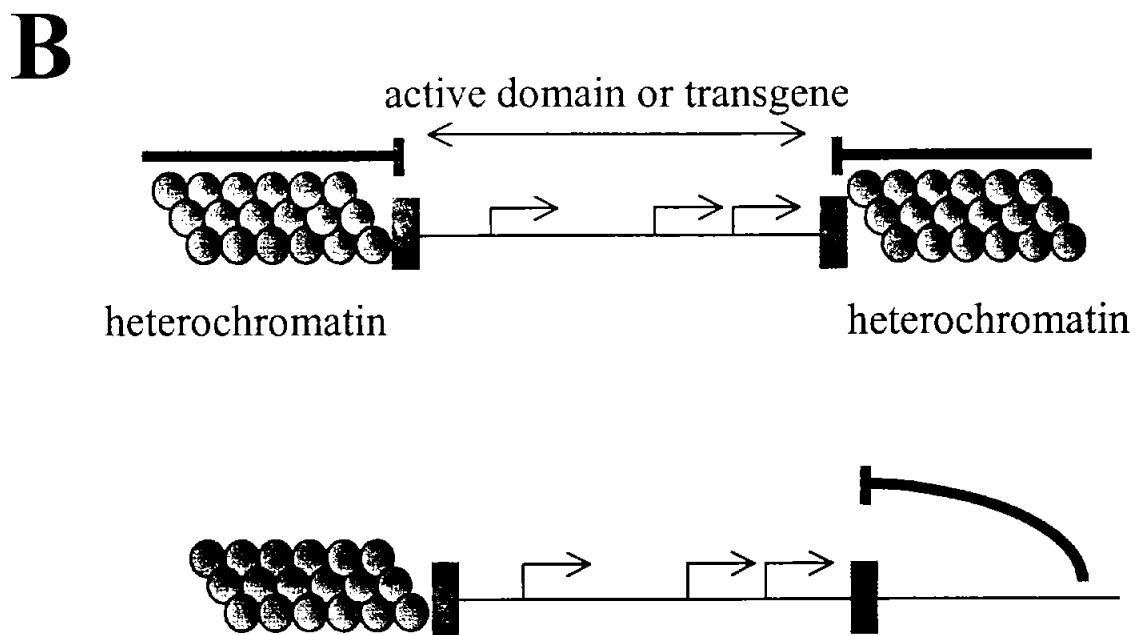
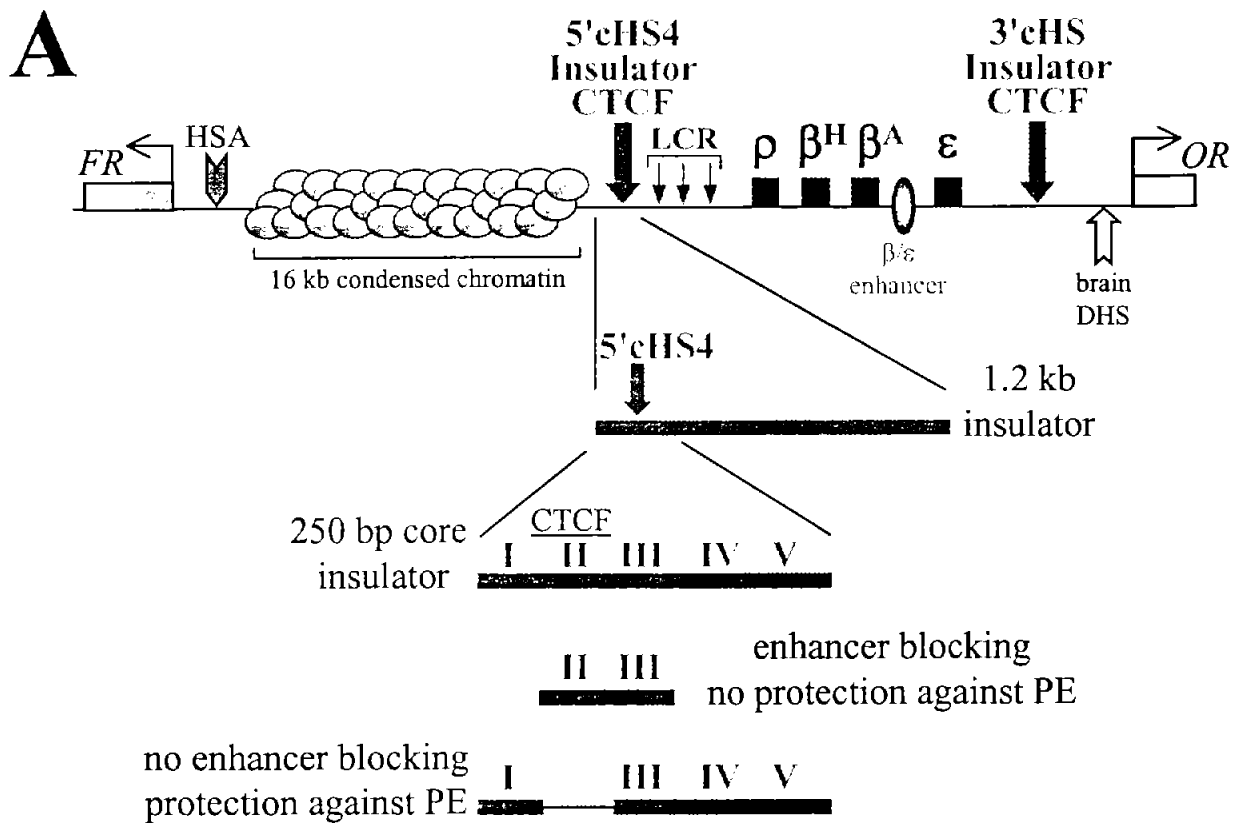


Figure 1

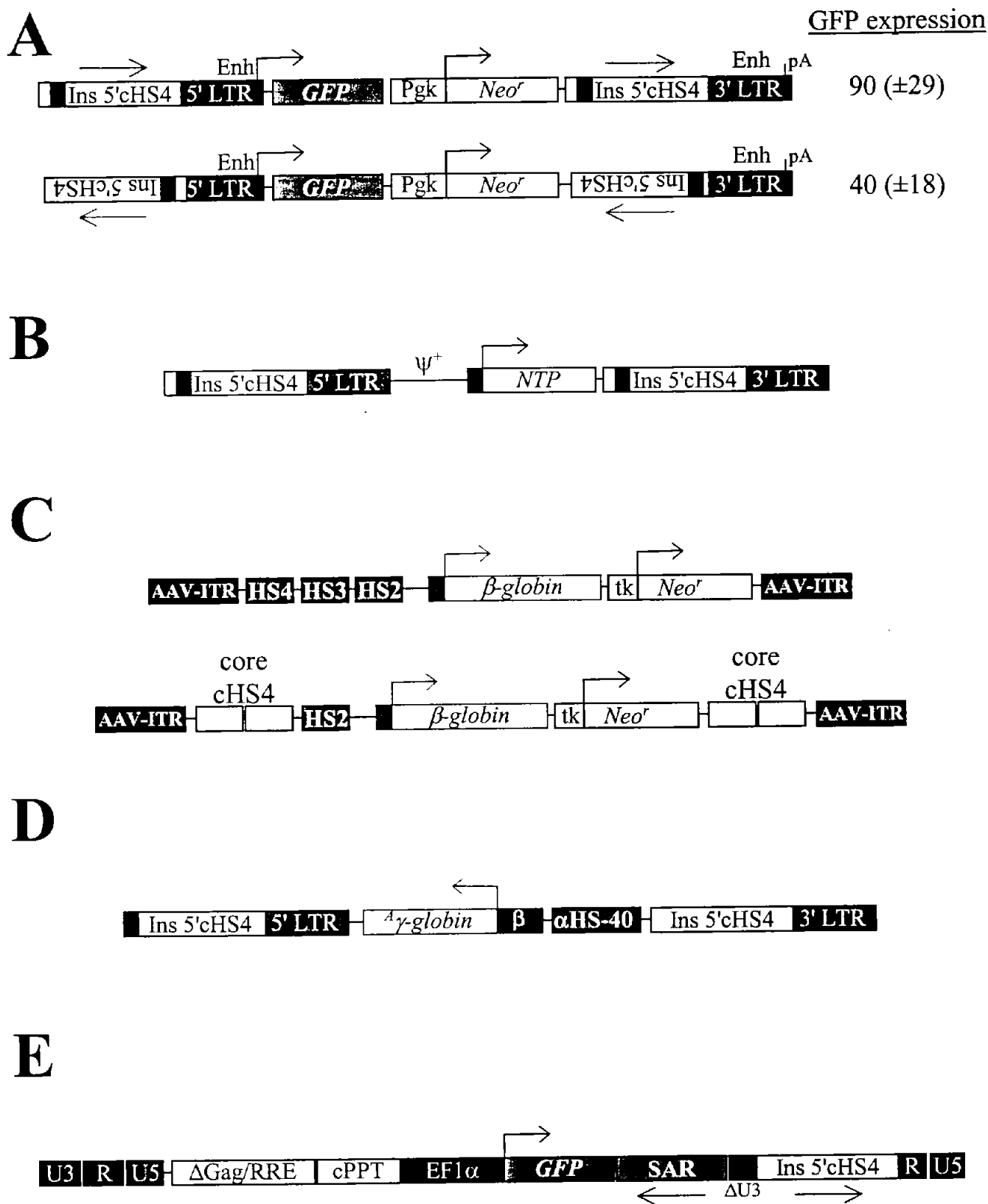


Figure 2

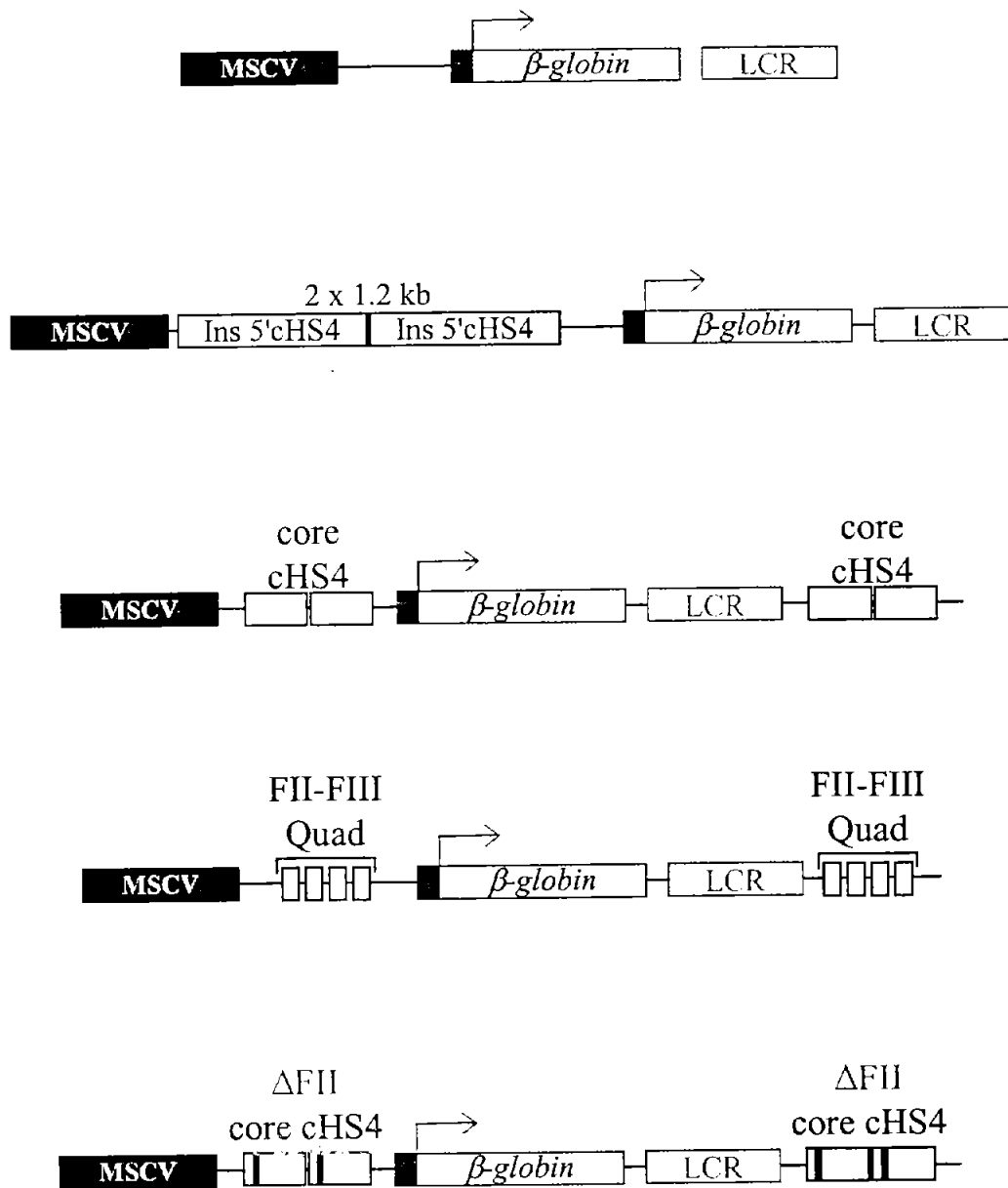


Figure 3



## ANEXO 1

### **Modulación de la actividad del enhancer localizado en el costado 3' del dominio $\alpha$ -globina de pollo durante la diferenciación eritroide (Rincón-Arano *et al* 2004).**

En el laboratorio estamos interesados en estudiar la relación que existe entre la estructura de la cromatina y la regulación de la expresión genética.

Para que exista una correcta expresión génica en tiempo y espacio, la célula se vale de distintos niveles de regulación, a nivel de la secuencia del ADN, de los promotores, de elementos de regulación a grandes distancias como enhancers y silenciadores, del remodelaje de la cromatina y actualmente la compartimentalización nuclear. Existen diversos modelos sobre cómo actúa un enhancer, se sabe que su organización es modular, es decir que tienen sitios de unión para múltiples factores y que sus funciones están determinadas por la unión de esos factores y a sus combinatorias. Sin embargo hasta la fecha, se desconoce el mecanismo de acción de estos elementos.

En el costado 3' del dominio  $\alpha$ -globina de pollo, se han encontrado un silenciador y un enhancer que regulan la transcripción de los genes del dominio. El enhancer fue descrito por el grupo de Felsenfeld en 1989 y de manera independiente por Scherrer y colaboradores, al hacer un ensayo de digestión a la DNasa I encontraron que existían 3 sitios de unión al factor GATA-1 (antes llamado EryfI) y mediante ensayos funcionales se determinó que el conjunto de esos sitios correspondían a un elemento tipo enhancer.

Se sabe que los elementos de regulación de los genes globina son muy conservados entre las diferentes especies. Con el objeto de encontrar más elementos de regulación en el costado 3' del dominio, se hizo una comparación entre las secuencias localizadas río abajo a partir del último gen del dominio entre el pato y el pollo. Posteriormente al realizar ensayos funcionales en transfecciones transitorias, se determinó que estas secuencias corresponden a un silenciador y a un enhancer, adicionalmente en el extremo 3' del enhancer se identificó un fragmento de ADN de 120 bp que corresponde a una región altamente conservada entre el dominio de pato y el de pollo (Fig. 1 del artículo).

En el silenciador, se determinó mediante ensayos de "footprint" *in vitro*, la protección de tres secuencias, es decir regiones de interacción proteína-ADN, a las cuáles se les nombró SF1, SF2 y SF3 (SF= "Silencing Factor"), y las cuáles corresponden a factores todavía desconocidos. Al hacer un footprint del fragmento de 120 bp y competencias con oligonucleótidos fríos con sitios de unión a diferentes factores eritroides, se determinó que existen 3 factores involucrados en la unión a esa zona: GATA-1, EKLF y NF-E2 además de los 3 sitios de unión a GATA-1. A partir de estos datos se propuso que en el fragmento de ADN de 120 bp podría estar formándose un complejo multipéptido que pudiera contribuir a la actividad del enhancer. Hasta la fecha se desconoce el mecanismo de acción de este enhancer y el papel que tienen las diversas proteínas que se asocian a él.

En un esfuerzo por entender como se llevan a cabo los procesos de regulación, mediados por el enhancer  $\alpha$ -globina de pollo, en este proyecto lo que se pretende es entender la función de este elemento y su contribución a la transcripción en distintas etapas del desarrollo eritroide. estudiando que factores se encuentran asociados al enhancer en cada estadio y además cómo contribuye el fragmento de 120 bp a la actividad del enhancer.

## **Discusión de Resultados**

En este trabajo hemos demostrado que el fragmento de ADN de 120 bp encontrado por análisis de homología entre las secuencias de ADN del extremo 3' del dominio entre pato y pollo, es parte integral del enhancer, y se comporta como un modulador de la actividad del mismo, dependiendo del estadio de diferenciación eritroide.

### **Los factores Sp1 y YY1**

Se encontró la presencia de dos nuevos factores, el factor Sp1 en el core (Fig. 5C y 8F) y el factor YY1 ("Ying Yang 1") en el fragmento de 120 bp (Fig. 8). El factor Sp1 es un factor ubicuo y al parecer está jugando un papel importante en estadios tardíos de la diferenciación eritroide en donde los genes globina se encuentran activos. Recientemente se ha descrito que existen interacciones entre el factor GATA-1 y los factores EKLF y Sp1, estas interacciones cooperan para llevar a cabo la actividad de elementos reguladores de la transcripción eritroide. El factor YY1 es un factor nuclear el cuál puede actuar como co-activador o como co-represor, dependiendo del contexto en el que se encuentre. La participación de este factor parece ser, al igual que la de Sp1 dependiente del estadio de diferenciación

eritroide. Este factor podría actuar como co-represor en estadios de diferenciación tempranos, aunque se encontró la presencia de este factor como parte del complejo de unión al fragmento de ADN de 120 bp tanto en células 6C2 como en RBC, sin embargo, los resultados son consistentes ya que al parecer la presencia de YY1 es más abundante en el complejo formado con extractos de células 6C2 (Fig. 8).

Un dato interesante es que los niveles de YY1 se mantienen constantes durante la diferenciación eritroide (demostrado aquí por Western-Blot), esto apoya un modelo en el cuál la regulación no está siendo llevada a cabo por medio de la concentración del factor, y en el cuál los niveles de la proteína no varían, así que se sugieren dos modelos: uno de competencia por el sitio de unión, en el que se propone que los factores YY1 y GATA-1 están compitiendo por el sitio de unión al fragmento de ADN de 120 bp, lo cuál está apoyado por los ensayos de *trans*-activación en los que se demuestra que al sobre-expresar el factor GATA-1 en el contexto del enhancer completo, core + 120 bp, suben los niveles de expresión del gen reportero, sin embargo estos niveles no se comparan con la *trans*-activación utilizando únicamente al "core" del enhancer; esto podría explicarse por el segundo modelo (Fig. 7). El segundo modelo involucra modificaciones post-traduccionales de los factores, ya que estos (YY1 y GATA-1) sufren modificaciones post-traduccionales como la acetilación y la desacetilación por parte de HAT's y HDAC's y es posible que estas modificaciones tengan un efecto sobre la afinidad de estas proteínas por sus secuencias blanco.

### **Los factores EKLF y NF-E2**

Al estudiar la participación de los factores que se unen al enhancer completo en un contexto cromatínico, es decir integrado al genoma de la célula, se ha determinado que el factor NF-E2 juega un papel mayoritario (Fig. 3), ya que su delección afecta de manera importante la actividad del gen reportero. Este factor es capaz de interactuar con la acetil-transferasa de histonas CBP, y probablemente tenga un papel preponderante en las modificaciones de cromatina y quizás también al modificar post-traduccionamente a los otros factores involucrados, ya que, por ejemplo, YY1 puede ser acetilado por CBP.

El factor EKLF es capaz de reclutar actividades remodeladoras de la cromatina, se sabe que es capaz de reclutar al complejo de remodelaje SWI/SNF y además es capaz de interactuar con el factor GATA-1, aunque en el contexto del este enhancer no se ha determinado si se reclutan estas actividades ni si este factor está interactuando con GATA-1.

### **El factor GATA-1**

Este factor es un regulador maestro de la diferenciación eritroide. Este factor es capaz de ejercer varias funciones, se sabe que GATA-1 puede interactuar con EKLf y Sp1, y que estas interacciones cooperan para la actividad de varios elementos de regulación eritroide. Además este factor puede reclutar actividades de acetil-transferasas de histonas (como por ejemplo a CBP) así que se ha propuesto que participa en funciones remodeladoras de la cromatina.

Se ha determinado que la concentración de esta proteína en células eritroides no se mantiene constante a lo largo de la diferenciación eritroide, de hecho se sabe que los niveles de esta proteína disminuyen en etapas tardías de diferenciación y datos en la literatura indican que una proteína homóloga a GATA-1, llamada GATA-2 es preponderante en los estadios de diferenciación primitivos y esto nos permite proponer que quizás es este factor GATA-2 el que tiene un papel importante en el enhancer en los pre-eritroblastos y eritroblastos de pollo.

En conclusión en este estudio se ha profundizado en el entendimiento de la actividad funcional y de la contribución de estos factores a la actividad neta del enhancer; además hemos encontrado datos interesantes acerca de los factores que se encuentran unidos a esta zona a lo largo de la diferenciación eritroide, muchos de ellos pueden actuar como co-activadores o como co-represores, se ha determinado que el factor NF-E2 es un componente importante del enhancer y que tiene un papel en el contexto cromatínico. Todos estos datos nos dan una nueva visión acerca de la organización del enhancer  $\alpha$ -globina de pollo y nos ha permitido proponer nuevos modelos a estudiar para explicar su modo de acción sobre los genes  $\alpha$ -globina de pollo.

### **Perspectivas**

Los resultados obtenidos en este artículo nos permiten proponer una serie de experimentos con la finalidad de entender las interacciones a nivel molecular de los diferentes factores involucrados en el enhancer  $\alpha$ -globina de pollo y que nos ayudarán a entender la contribución individual de los mismos durante las distintas etapas de diferenciación eritroide para la regulación de los genes del dominio.

Para este fin se propone clonar el ADNc del factor YY1 de pollo, la clonación permitirá producir cantidades suficientes de la proteína, las cuáles podrían utilizarse para generar anticuerpos específicos contra este factor y realizar ensayos de coimmunoprecipitación. Además el contar con el cDNA de YY1 permitiría realizar experimentos de *trans*-activación o *trans*-represión que ayudarían a comprobar o refutar alguno de los modelos propuestos.

Siguiendo con la línea de saber las interacciones específicas entre los factores actualmente nos encontramos generando anticuerpos contra las regiones amino y carboxilo terminales del factor GATA-1, expresando estos extremos en el sistema de baculovirus en colaboración con el grupo del Dr. Luis Vaca. Este sistema permite obtener grandes cantidades de proteína, la cuál se utilizará para hacer ensayos de co-inmunoprecipitación, la ventaja de hacer los anticuerpos contra los extremos amino y carboxilo terminales, es que seríamos capaces de discernir las interacciones que ocurren en estos diferentes dominios de la proteína, y cuáles son importantes para la unión con los diversos factores.

Cómo parte de las perspectivas a futuro en otro proyecto realizado en el laboratorio por el Químico Martín Escamilla se está estudiando la estructura de la cromatina del enhancer y del silenciador en los estadios de diferenciación eritroide, por experimentos de accesibilidad a la nucleasa micrococcal. En estos experimentos se ha observado que existe un remodelaje en la zona del enhancer. se propone hacer un vector de RNA de interferencia (RNAi) contra el factor GATA-1, hacer transfecciones con el RNAi de las células en los estadios de diferenciación eritroides, esperando observar un cambio en el patrón nucleosomal. Esto apoyaría o descartaría que el factor GATA-1 tenga un papel preponderante en el reclutamiento de actividades remodeladoras de la cromatina. En este proyecto también se ha propuesto estudiar las modificaciones de las histonas en la zona del enhancer.

En base a los datos de la literatura que indican que acetiltransferasas de histonas específicas están actuando en conjunto con los factores encontrados y con los datos de las modificaciones post-traduccionales de las histonas, se podrá inferir cuáles acetil-transferasas de histonas ( como CBP) o desacetilasas de histonas ( como HDAC1) están realizando estas modificaciones, y con anticuepos específicos se corroborará su presencia o ausencia a lo largo de la diferenciación.

**Modulation of the chicken 3'-side  $\alpha$ -globin enhancer activity during erythroid differentiation\***

**Héctor de Jesús Rincón-Arano<sup>§</sup>, Viviana Valadez-Graham<sup>§</sup>, Georgina Guerrero  
and Félix Recillas-Targa<sup>‡</sup>**

*Instituto de Fisiología Celular, Departamento de Genética Molecular, Universidad Nacional Autónoma de México, Apartado Postal 70-242, México D.F. 04510, México.*

*Running Title: The chicken  $\alpha$ -globin enhancer*

\*This work was supported by the Dirección General de Asuntos del Personal Académico-UNAM (IN203200), the Consejo Nacional de Ciencia y Tecnología, CONACyT (33863-N), the Third World Academy of Sciences (TWAS, Grant 01-055 RG/BIO/LA), and Fundación Miguel Alemán.

<sup>‡</sup>Corresponding author: Félix Recillas-Targa  
Instituto de Fisiología Celular  
Departamento de Genética Molecular  
Universidad Nacional Autónoma de México  
Apartado Postal 70-242  
México, D.F. 04510  
MEXICO  
Tel: (52 55) 56 22 56 74  
FAX: (52 55) 56 22 56 30  
Email: [frecilla@ifc.unam.mx](mailto:frecilla@ifc.unam.mx)  
<http://www.ifsiol.unam.mx/gmdepto.html>

<sup>§</sup>These authors contributed equally to the present work

The abbreviations used are: LCR, Locus Control Region; EKLF, Erythroid Krüppel-like Factor; YY1, Ying Yang 1 nuclear factor; FACS, fluorescence activated cell sorting; CAT, chloramphenicol acetyltransferase; RBC, red blood cells; Cytes, terminal differentiated erythrocytes.

**SUMMARY**

**Studying differential regulation of globin genes provides endless information about developmental gene expression and chromatin structure. We have previously identified contiguous silencer-enhancer elements located at the 3' of the chicken  $\alpha$ -globin domain. To better characterize these control elements, we decided to perform a systematic functional analysis and re-define the ubiquitous and stage-specific transcriptional regulators that interacting at this region. In contrast with previous reports, we found that the enhancer is composed of the core element that includes three GATA-1 binding sites and a 120 base pair DNA fragment where an EKLF, NF-E2 and a fourth GATA-1 factor interact. At this 120 base pair region we found the interaction of Ying Yang 1, and we suggest that its contribution to the enhancer activity is based on differential chromatin modifications at distinct erythroid stages. Functional experiments show that the enhancer activity over the adult  $\alpha^D$  promoter is differentially dependent on the erythroid differentiation state. Our observations show that the originally defined enhancer possesses a more complex composition and includes the set of nuclear factors that is found in other globin regulatory elements, like locus controls regions, that seems to modulate enhancer activity.**

## INTRODUCTION

Eukaryotic gene expression requires a perfect coordination between tissue- and stage-specific regulatory factors in the nuclear and chromatin structural environments. Long distance regulatory elements have been an attractive research topic over the years. In particular, the  $\alpha$ - and  $\beta$ -globin group of genes and their regulatory elements, including enhancers, silencers, locus control regions (LCRs) and insulators, have been studied (1-5). The interest in the study of gene families is based on differential gene expression with physicochemical properties specific to each developmental stage (2).

Locus control regions and enhancers are thought to regulate gene expression following two basic models: The binary and the progressive models (6). The binary model, also known as the "on or off" model, predicts that enhancers are responsible for increasing the probability of gene transcription without varying the transcription level of individual cells. In contrast, in the progressive model enhancers increase the rate of transcription in all cells. In any case, at the present time both models are supported by experimental evidence. Some particular features of enhancers are their capacity to act over long distances, with a modular organization and independently of their location. Less is known about enhancer influence on chromatin structure. What seems to be true is that an active enhancer is found in an open chromatin configuration probably attributed to the binding of nuclear factors and their associated chromatin remodeling activities (7-9). Recently, two new experimental strategies demonstrated that the LCR mechanism is based on chromatin loop formation with multiple physical contacts in close proximity between individual promoters and the DNase I hypersensitive sites that are located at long range (10, 11).

As in the human and mouse  $\alpha$ -globin loci, the chicken  $\alpha$ -globin domain is located in a constitutive open chromatin configuration (2). Such lack of domain shielding, like that seen in the chicken  $\beta$ -globin domain, particularly by specialized chromatin structures like insulators, may probably



### *The chicken $\alpha$ -globin enhancer*

require tighter control for developmental regulation of gene expression (3, 5). Based on such a prediction, we decided to further characterize the 3' silencer-enhancer elements (Fig. 1). In this work we are particularly interested in functional characterization of the 3' chicken  $\alpha$ -globin enhancer during distinct erythroid differentiation stages.

Previous work has identified silencer-enhancer elements in the 3' of the chicken  $\alpha$ -globin locus, around 450 base pair (bp) downstream of the adult  $\alpha^A$  globin gene (12, 13). Knezetic and Felsenfeld identified three GATA family binding sites, as sufficient for the enhancer activity (12). Alignment of duck and chicken homologous sequences enabled us to also identify a highly conserved 120 bp DNA fragment with binding sequences for Sp1 family, NF-E2 a small Maf family member and GATA family (Fig. 1; 13, 14). With the aim of further characterizing this region, we decided to address two main questions. The first is to determine the contribution of the 120 bp DNA fragment to the full enhancer activity, and the second is to determine which nuclear factors are involved in the enhancer function during distinct erythroid differentiation stages. Using the chicken  $\alpha^D$  globin promoter we were able to demonstrate that the 120 bp DNA fragment constitutes an integral component of the enhancer. Interestingly, by itself the 120 bp fragment does not possess any enhancer activity, and when MARE and GATA sites are mutated in the context of the entire enhancer, the activity reduced around 50% in transient transfection assay. In contrast, NF-E2 site mutation in stably integrated environment is sufficient to decrease approximately 50% of the enhancer function, consistent with the suggested contribution of NF-E2 to gene activation, locus relocation and protection against position effects (15, 16). Additionally, we identify a novel interaction of the Ying Yang 1 (YY1) nuclear factor in the proximity of the fourth GATA-1 sequence. We propose that YY1 is able to modulate the chicken  $\alpha$ -globin 3' enhancer activity probably by changing chromatin structure.

## EXPERIMENTAL PROCEDURES

*Cell culture and transfections-* The 6C2 cell line corresponds to an avian erythroblastosis transformed chicken bone marrow cell line arrested at the CFU-E stage and considered to be pre-erythroblast (17, 18). 6C2 cells were grown in minimal essential medium ( $\alpha$ -MEM; GIBCO) supplemented with 10% fetal calf serum, 2% chicken serum, 1 mM HEPES pH 7.2 and 50  $\mu$ M 2-mercaptoethanol. The avian erythroblastosis virus-transformed and temperature-sensitive chicken erythroblast cell line LSCCHD3 (HD3) was grown at 37°C in DMEM medium (GIBCO) supplemented with 8% fetal calf serum, 2% chicken serum and antibiotics (17, 18).

*Transient and stable transfections-* Ten-day-old chicken embryonic erythrocyte transient transfections were performed essentially as described previously (19-21). Typically, 6  $\mu$ g of the pC $\alpha$ 3' derived plasmids and 6  $\mu$ g of pTk- $\beta$ gal were co-transfected.  $10^8$  cells were electroporated in a 0.4 mm cuvette under the following conditions: 500  $\mu$ Fa, 450 Volts using the BioRad GenePulser II apparatus. Transfected cells were cultured with 1.4 ml of complete L-15 medium at 37°C for 24-48 hrs. CAT activity assays were performed as described (21) and the  $\beta$ -galactosidase activity for normalization was obtained using a  $\beta$ -galactosidase kit (Promega). Dual luciferase assay transfections and measurements were carried out as indicated by the manufacturer (Promega) in a TD-20/20 Turner BioSystems luminometer.

*The chicken  $\alpha$ -globin enhancer*

*Plasmid constructs*- DNA sub-fragments contained in the genomic *Bam*HI-*Bam*HI restriction sites (13) were PCR amplified from 10-day-old embryonic erythrocytes using genomic DNA previously fractionated by *Eco*RI enzymatic digestion, with the following primers: ASIL1, 5'-CGCGGA TCCCATGCACTCCTTACCCCATG-3' ; ASIL2, 5CGCGGATCCCAGCACAGCTTCACCC TGCG; AE1,CGCGGATCCGCGCAGGGTGAAGCTGTGCTG-3' ; A120P4, 5CGCGGAT CCGACCGTGGTAACCCAGGCCG-3' ; A120P3, 5CGCGGATCCGACGTGGGCAGCAGATA GCCTCG-3' ; EE2, 5CGCGGATCCGAGGCTATCTGCTGCCACG-3' ; A7, 5CGCGGAT CCTGCAGCAGGTTGAGCAG-3' . All synthetic oligonucleotides contain *Bam*HI restriction site. DNA binding motif mutation on the EKLF, NF-E2, the fourth GATA-1 and the double mutant NF-E2/GATA-1 sites were generated in the context of the full enhancer, i.e., including the original three GATA-1 binding sites (12, 13). The two step PCR overlapping method was used to generate two and three base pair changes using the following primers (with their corresponding complementary strands): MABekl02, 5'-GCAGATAGCCTCGttTtGGTAAAACCTGC-3' ; MABnfe02, 5CCTGCCAGG TGCTagGTAAGTGCAGTCTG-3' and MABgata02, 5GATCCTGCCAGGTaagTTCATGC AGAGATC-3' . For the double mutant NF-E2/GATA-1 sites, the E4- $\Delta$ GATA-1 construct was used as a PCR template. All the mutated DNA fragments were directly cloned in the pCAT $\alpha^D$  reporter vector (22) and the DNA was sequenced using the primer GGGGAGGTGTGGGAGGTTT. For the series of experiments where we used the dual-luciferase reporter system (Promega), we generated the pG $\alpha^D$  vector by cloning the  $\alpha^D$  globin promoter into the *Bgl*III site of the pGL3-basic poly-linker (Promega). All the DNA sub-fragments to be tested were sub-cloned into the *Bam*HI site downstream of the firefly reporter gene.

*The chicken  $\alpha$ -globin enhancer*

*Electrophoretic Mobility Shift Analysis-* Synthetic oligonucleotides were radioactively labeled with [ $\gamma$ - $^{32}$ P]dATP, annealed and acrylamide gel purified. Double-stranded oligonucleotides were prepared by annealing of the reverse strand. Mixtures of two complementary strands were first heated for 10 minutes at 90°C and then incubated at room temperature for 10 minutes in 2X annealing buffer (20 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA). For binding reactions 0.5 pmol (20,000 cpm each) of radiolabeled oligonucleotide were incubated with 20  $\mu$ g of different nuclear extracts in 2X GS buffer (40 mM HEPES pH 7.9, 100 mM KCl, 10 mM MgCl<sub>2</sub>, 10% glycerol, plus 1  $\mu$ g of poly dI-dC as a non-specific competitor and 1 mM DTT) in a 20  $\mu$ l final volume for 15 min on ice and then 15 min at room temperature. Competitions were carried out with 20 to 40 pmol of cold oligonucleotides; incubation with competitors was carried out before adding the labeled oligonucleotide, for 30 min on ice and 30 min at room temperature. Samples were then analyzed on a 5 or 6% polyacrylamide gel. After electrophoresis, gels were dried and exposed to X-ray film. For supershift reactions, anti-YY1 polyclonal antibody (H-414; Santa Cruz Biotechnology) was preincubated with nuclear extracts overnight with agitation at 4°C. The following oligonucleotides (all purified by polyacrylamide gel electrophoresis) were used in different mobility shift experiments (in all cases only one of the complementary chains is shown): A1G (fourth GATA-1), 5'-CGCGGATCCTGCCAGGTATCTT CATGCAGAGATG-3' ; AYY1A (YY1), 5'-GATTGCCAGGTATCTTtcgGCAGAGATG-3' ; YY1A (YY1 consensus), 5'-CGCTCCGCGGCCATCTTGGCGGCTGGT-3' (23); Sp1cu (Sp1 consensus), 5' ATTCGATCGGGGCGGGGCGAGC-3' (24, 25); Sp1a, 5CAGCAGCAGGGCCTGGGATGGC CATAG-3' ; Sp1b, 5CCAGCTCACGGGTGTGGTGAGGGGAC-3' .

*In vitro DNase I footprinting-* Two DNA fragments were analyzed: a 282 bp fragment including the full enhancer and a 154 bp fragment including the >120 bp DNA fragment (Fig. 1) were generated by

*The chicken  $\alpha$ -globin enhancer*

*Hin*pI-*Hin*fI and *Alu*I-*Hin*fI restriction enzyme digestions, respectively. Both DNA sub-fragments were end-labelled with [ $\alpha$ -<sup>32</sup>P]dATP using Klenow polymerase I. Labelled probes were incubated for 10 min on ice with 1  $\mu$ g of poly(dI-dC) (Pharmacia) as a non-specific competitor, and increasing concentrations of nuclear extracts. DNase I dilution mixture was added (to the naked DNA) for 90 seconds to 3 min (in the presence of nuclear extracts) to the sample on ice. The reaction was stopped with 50  $\mu$ l of STOP buffer (100 mM Tris-HCl pH 7.5, 10 mM EDTA, 100 mM NaCl and 0.1% SDS). DNA samples were incubated with proteinase K and RNase and phenol/chloroform extracted three times. The DNA pellets were resuspended in formamide-dye, heated to 90°C for 3 min, and loaded onto an 8% polyacrylamide-10M Urea sequencing gel. Gels were dried and exposed to an X-ray film at -70°C for 24-48 hrs. The oligonucleotides used for competition are: NF-E2, 5'-CTAGTGATGA GTCAGCCGGATC-3' (26) and GATA-1, 5'-CACGCGGATAAGATAAGGCCGGGCT-3' (27).

*Western blot*- Cells were harvested with lysis buffer containing 20 mM HEPES pH 7.9, 400 mM KCl, 20% glycerol, 2 mM DTT, plus protease inhibitors; freeze-thawed two times, and proteins were quantified by the Bradford assay (BioRad). One hundred  $\mu$ g of protein extracts were separated by SDS-polyacrylamide gel electrophoresis and the resolved proteins were transferred to PVDF membranes (Pharmacia). Membranes were blocked with 5% non-fat milk for 2 hours and then incubated with rabbit polyclonal YY1 (H-414) antibody (Santa Cruz Biotechnology) for different time intervals (2 hours to overnight) at 4°C. After washing, the blots were incubated for 1 h with an appropriate secondary antibody (Santa Cruz Biotechnology), developed by a chemiluminiscent kit (PIERCE), washed and exposed to X-ray film (Kodak-Bio-Max).

*The chicken  $\alpha$ -globin enhancer*

*Chromatin Immunoprecipitation*- The ChIP assay was performed as described by Farnham and collaborators (28) with slight modifications. After formaldehyde crosslinking and sonication, chromatin corresponding to  $2 \times 10^7$  cells was incubated overnight at  $4^\circ\text{C}$  with a secondary antibody, goat anti mouse IgG (Santa Cruz Biotechnology) as a negative control and  $10 \mu\text{g}$  of anti-GATA-1 (H-200), anti-Sp1 (PEP2) and anti-YY1 (H-414) (Santa Cruz Biotechnology). Next day,  $20 \mu\text{l}$  of protein A/G plus agarose (Santa Cruz Biotechnology) was added per reaction and incubated for 5 hours at  $4^\circ\text{C}$ . The mixture was centrifuged and the supernatant was saved as the UNBOUND fraction. The immunoprecipitated fraction was washed as in the original protocol (28).  $0.5 \mu\text{g}$  of RNase A was added to each sample followed by a 30 min incubation at  $37^\circ\text{C}$ , then samples were incubated at  $67^\circ\text{C}$  overnight to reverse formaldehyde crosslinking, samples were precipitated by adding 2 volumes of 100% ethanol and incubated overnight at  $-20^\circ\text{C}$ . The next day samples were centrifuged at  $4^\circ\text{C}$ , dissolved in TE buffer containing  $60 \mu\text{g}$  of proteinase K (Sigma), and incubated at  $45^\circ\text{C}$  for 2 hr. DNA was extracted twice with phenol:chloroform and once with chloroform, then samples were precipitated by adding 1/10 vol. of 5 M NaCl,  $20 \mu\text{g}$  of glycogen and 3 volumes of 100% ethanol. The DNA was dissolved in  $40 \mu\text{l}$  of deionized water. Radioactive PCR reactions were then performed. Two pairs of oligonucleotides were designed. The first one amplifies a 169 bp DNA fragment of the chicken  $\alpha$ -globin 3'-side enhancer containing the 120 bp element, and the oligonucleotide sequences were as follows: A120P3, 5'-GACGTGGGCAGCAGATAGCCTCG-3' and A120P4, 5GCCGGACCC CAATGGTGCCAG-3' . The second pair of oligonucleotides, used as control, amplifies a 301 bp sequence containing a portion of the chicken  $\beta$ -globin enhancer (110-558 bp after the poly-A signal of the  $\beta^A$  gene; 19): 0026 5'-CTGGGTGGGGGCAGGT-3', 0027 5'-CTGCTTTTGCTGCCCTGTG-3'.

*The chicken  $\alpha$ -globin enhancer*

Polymerase chain reactions were performed at an annealing temperature of 56.3°C for 20 cycles, run on a 6% polyacrylamide gel at 150 Volts in 0.5X TBE. Gels were dried and exposed to an x-ray film (Kodaz Bio-Max) for 24 to 72 hours and developed.

## RESULTS

We have previously identified silencer-enhancer elements located around 450 bp downstream of the adult  $\alpha^A$  gene at the 3'-side of the chicken  $\alpha$ -globin domain (Fig. 1; 13). By sequence alignment between the duck and chicken, we also identified a 120 bp DNA fragment with putative DNA binding sites for erythroid factors (13, 29). Further characterization identified a large footprint over 60 bp including EKLF, NF-E2 and a fourth GATA-1 binding sites (Fig. 5; 13, 29). Our original series of functional assays were carried out with the SV40 promoter driving the expression of the *CAT* reporter gene (13). We have recently noticed that SV40 promoter-enhancer control elements are not efficiently functional in erythroid-cells (21). Therefore we changed the strategy for studying the silencer-enhancer in different erythroid differentiation stages.

*Transient and stable transfections of the silencer-enhancer in the 6C2 pre-erythroblast cell line-* The effect of these regulatory elements was first analyzed in the virally transformed chicken erythroid 6C2 cell line that is arrested at the colony forming unit erythrocytes stage (CFU-E; 30, 31). Instead of the SV40 control elements used before, we now used the chicken adult  $\alpha^D$ -promoter driving the expression of the *CAT* reporter gene in order to test the function of the silencer-enhancer elements (Fig. 2A; construct 1). The silencer-enhancer was tested in different combinations in transient transfections using 6C2 cells, and no significant activity in any construct was observed (Fig. 2A). We consider two possibilities to explain these results. The first one could come from low transfection efficiencies and the second from a real lack of activity of the promoter and the silencer-enhancer in any combination in this pre-erythroblast cell line. To address each possibility we first transfected the pGE $\alpha^D$  and pGE $\alpha^D$ E4 plasmids corresponding to the  $\alpha^D$  promoter alone and the  $\alpha^D$  promoter in the presence of the entire enhancer (E4), respectively, directing the expression of the green fluorescent protein followed by



### *The chicken $\alpha$ -globin enhancer*

evaluation of transfection efficiencies by fluorescence cytometry (FACS) and using the viral CMV promoter-enhancer as a positive control (data not shown). We clearly conclude that the lack of activity is not due to the transfection efficiency since we reach high levels of *GFP* expression using the pGCMV plasmid. To confirm our observations we asked whether in chromatin environment with the same set of constructs will give a different result in 6C2 cells (Fig. 2B). Cells were transfected with the pGE $\alpha$ D plasmid carrying the neomycin resistance cassette (*Neo<sup>r</sup>*). After selection, stable pools were split in three and the results of CAT activities are shown in figure 2B. Consistent with our previous data from transient transfection experiments we found that the  $\alpha^D$  promoter and the 3' control elements are not functional in this cell line. Previously our group has shown that endogenous globin genes are expressed at basal levels in 6C2 cells (31; Guerrero and Recillas-Targa, *unpublished observations*). Interestingly, in the case of the chicken adult  $\beta^A$  promoter and the  $\beta/\epsilon$  enhancer there is considerable transcription activation of a reporter gene suggesting that the pools of transcription factors are not limiting in this case (32). In summary, in a pre-erythroblast transformed cell line the  $\alpha^D$  promoter alone and the 3' silencer-enhancer are not active, consistent with the fact that chicken  $\alpha$ -globin genes are expressed at low levels in 6C2 cells.

*Transient and stable transfections of the 3'-side  $\alpha$ -globin enhancer in HD3 cells-* To test the enhancer activity in more matured stage we transfected the same constructs into the HD3 avian transformed cell line that corresponds to the erythroblast stage (18). For these experiments we used luciferase assay which is more sensitive than the CAT assay. We observe only a modest increase in luciferase activity for the core and the entire enhancer, latter of which also contains the 120 pb DNA fragment, compared to the promoter alone (Fig. 3A; constructs 1, 2 and 3). When the 120 pb DNA fragment is tested alone there is even a reduction in the reporter activity (Fig. 3A; compare constructs 1 and 4). We decided to

### *The chicken $\alpha$ -globin enhancer*

include in our study the silencer and the silencer-enhancer with the 120 bp DNA fragment and we confirm our previously observations showing that the 360 bp DNA elements 5' to the enhancer correspond to a silencer element (Fig. 3A; constructs 5 and 6; 13, 29).

To further investigate the enhancer function in a chromatin context, stable cell lines were generated and individual clones isolated in a semisolid agar and monitored green fluorescent gene expression by FACS. Interestingly, we found that the highest intensities of fluorescence are obtained with the promoter and the core enhancer containing the three GATA-1 sites (Fig. 3B,  $\alpha^D E$ ). In contrast, lower and homogeneous GFP expression was found when the EKLF and the fourth GATA-1 binding sites are mutated in the context of the entire enhancer (Fig. 3B;  $\alpha^D E \Delta EKLF$  and  $\alpha^D E \Delta GATA-1$ ). However, when mutation of the NF-E2 site is tested we observed the greatest loss in gene expression shown by a decrease in fluorescence intensity (Fig. 3B;  $\alpha^D E \Delta NF-E2$ ).

These results demonstrate that the silencer and enhancer elements are functional in HD3 cells, and the 120 bp DNA fragment is not able to activate reporter expression. We have also found that deletion of the NF-E2 binding site in the context of the entire enhancer is responsible for a significant decrease in reporter activation, supporting the contribution of NF-E2 to the enhancer activity and suggesting a possible role in chromatin remodeling.

#### *Transient transfections in 10-day-old chicken embryonic erythrocytes and enhancer blocking assays-*

The same constructs tested above were then transiently transfected in chicken 10-day-old embryonic erythrocytes (RBC). For the circular plasmids we observed a robust activation by the core enhancer composed of three GATA-1 binding sites (Fig. 4; constructs 1 and 2; 12-14). These results support the

### *The chicken $\alpha$ -globin enhancer*

idea that the 3' chicken  $\alpha$ -globin enhancer is not only composed of the core three GATA-1 sites as previously defined, but in addition requires the 120 bp and their associated nuclear factors (12).

In addition, we asked if the silencer could act as an enhancer blocking element on linearized plasmids (Fig. 1; 21). For the enhancer blocking assay we placed the silencer between the  $\alpha^D$  promoter and the enhancer (Fig. 4; construct 5), and also located the enhancer between the  $\alpha^D$  promoter and the silencer (Fig. 4; construct 6). These two constructs were linearized to test the positional enhancer blocking activity of the silencer element (21). The enhancer blocking assay shows that the silencer reduced enhancer activity 4.7-fold regardless of the position (Fig.4; constructs 5 and 6). Since there is no difference in activity between two constructs, this suggests that the silencer has no insulating activity.

In summary, our results argue in favor of the presence of two regulatory elements, one positive and the other negative side by side, where the enhancer activity is regulated during the course of distinct erythroid differentiation stages.

*DNA-protein interactions in the 120 bp DNA fragment-* Previous results demonstrated the interaction of the erythroid specific GATA-1 factor with three DNA consensus sequences at the core enhancer (Fig. 5; 12, 29). As mentioned before and based on our sequence alignment, we were able to define a highly conserved 120 bp DNA sequence immediately downstream of the core enhancer (Fig. 5A; 13, 14). We also demonstrated, using nuclear extracts from HD3 cells and terminally differentiated erythrocytes from adult chicken, the appearance of a large footprint covering around 60 bp, coincident with the presence of an EKLF, NF-E2 and a fourth GATA-1 consensus binding site (Fig. 5A; 13, 14, 29). We have obtained the first overview of the entire enhancer that includes the core element and the 120 bp DNA fragment (Fig. 5A). We observed a large protected area covering around 250 bp which includes three GATA-1 sites and the 120 pb DNA fragment (Fig. 5B). This is the first time in which the

### *The chicken $\alpha$ -globin enhancer*

complete enhancer is analyzed in terms of which potential nuclear factors are involved. In addition, a large complex appears to be formed at the 120 bp DNA fragment that includes Sp1 family, GATA family binding sites and MARE (Fig. 5B, D). Thus, the core enhancer is not only composed of GATA-1 binding sites as previously suggested, but also of other erythroid and non-erythroid factors (Fig. 5A, C; 12). In addition, there are Sp1/EKLF interactions between the GATA-1 sites and a large complex formed by the binding of EKLF, NF-E2 and a fourth GATA-1 site in the 120 bp DNA fragment (Fig. 5A).

We further characterized the 120 bp DNA fragment with sequence specific competitions. Using HD3 nuclear extracts we carried out an *in vitro* DNase I footprinting experiment with 100-fold and 250-fold molar excess of DNA fragments containing NF-E2 and GATA-1 consensus sequences (Fig. 5D; 26, 27). We confirm our previous results showing a large protected area of 60 bp and in addition we discover a prominent DNase I hypersensitive site downstream of the protected area (Fig. 5D; lanes 3-5 and arrow). When the same reaction is then done with increased concentrations of cold NF-E2 sequence competitor, we observe the reappearance of specific bands that match exactly with the NF-E2 sequence in the 120 bp DNA fragment (Fig. 5D; lines 6 and 7). Interestingly, for NF-E2 competition the DNase I hypersensitive site disappears. Our interpretation of these results is that, there are multiple DNA-protein interactions including co-factor interactions responsible for the formation of a large complex in the 120 bp DNA fragment. Such interactions could induce changes in DNA conformation, creating DNase I hypersensitive site a few base pairs downstream (Fig. 5D). Such a site is lost when the NF-E2 sequence is missing, suggesting that this factor may be playing a role in anchoring other interactions, conclusions which are further supported by our mutational studies (Fig. 5D; compare lanes 3-5 and 6-7). In examining competition with the GATA-1 site of the chicken adult  $\beta^A$ -globin gene promoter, we found that with a 100-fold excess of competitor we observed 60 bp footprint but when

### *The chicken $\alpha$ -globin enhancer*

this was further increased to 250-fold the majority of the bands including GATA-1 site reappeared (Fig. 5D, lanes 10-11). These results not only confirm the interaction of the GATA-1 factor, but in addition suggest that its binding is also critical for the complex formation, even though there is no loss of the downstream DNase I hypersensitive site.

In summary, the 3' chicken  $\alpha$ -globin enhancer is not only restricted to the core enhancer and three GATA-1 binding sites, but in addition its architecture is complemented with other well known erythroid-specific partners. Interestingly, there is an exact match between the number and factors interacting with the human  $\alpha$ -globin -40HS upstream regulatory element and the 3'-side chicken  $\alpha$ -globin enhancer (see discussion; 33).

*Mutational analysis of the DNA-binding sites of the 120 bp DNA fragment-* To better estimate the contribution of the 120 bp DNA fragment to the overall enhancer activity, we designed point mutations of the following individual sites: EKLF, NF-E2, GATA-1 of the 120 bp DNA fragment and a double mutant NF-E2/GATA-1 (Fig. 6; 14). Such mutations were tested in the context of the entire enhancer, i.e., the core element and the 120 bp DNA fragment (Fig. 5A, 6). In addition, we carried out a comparative study by transiently transfecting the constructs in HD3 cells and chicken 10-day-old embryonic erythrocytes. As we have already shown, the full enhancer action is reached in the presence of the core and the 120 bp DNA fragment (Fig. 6A, B; constructs 2 and 3). In both scenarios the 120 bp DNA fragment alone does not have any significant activity even though EKLF, NF-E2 and GATA-1 nuclear factors interact with this sequence (Fig. 6A, B; constructs 4). Indeed, our results strongly support the idea that the 120 bp DNA fragment is part of the enhancer, and we suggest that it could modulate its activity in distinct differentiation conditions (see below). In any case, the *in vitro* footprinting results suggest that multiple nuclear factor interactions with accentuated interaction in

### *The chicken $\alpha$ -globin enhancer*

erythrocytes when globin genes are actively transcribed compared to primitive or terminally differentiated erythroid cells (Fig. 6A, B; 13). One of the most relevant observations from this series of experiments is the drastic difference in reporter activity levels of the transiently transfected plasmids between HD3 cells when compared to 10-day-old embryonic chicken erythrocytes. In RBC the overall activity is significantly higher, suggesting that the  $\alpha^D$  promoter and the 3' enhancer coupling is probably optimal in RBC. This confirms that the capacity of the enhancer is differentially regulated in distinct erythroid differentiation stages.

We found a slight decrease in activation levels in transient transfections in HD3 cells for when NF-E2 and GATA-1 sites were mutated (Fig. 6A; constructs 6 and 7). Surprisingly, this pattern is not maintained in 10-day-old embryonic erythrocytes when the EKLF, NF-E2 and the GATA-1 sites were mutated (Fig. 6B; constructs 5, 6 and 7). The most dramatic loss of enhancer activity is seen for the NF-E2/GATA-1 double mutant in RBC, which loses around 50% of the activity (Fig. 6B; construct 8). These results are consistent with the suggested relevance of the NF-E2 and the fourth GATA-1 sites, and also support the contribution of the 120 bp DNA fragment to the enhancer activity. However, they are not in total agreement with the HD3 stable lines where the NF-E2 mutation is the only one that affects significantly enhancer activity (see Fig. 3B). This may seem to be a clear example of the relevance of the chromatin context and supports the notion that NF-E2 is critical for the chicken  $\alpha$ -globin enhancer, and that its activity is intimately associated with the chromatin context or even its nuclear and sub-nuclear location (see discussion; 15, 16).

*Trans-activation of the 3'  $\alpha$ -globin enhancer in 6C2 cells by GATA-1 cDNA* - It has been previously demonstrated that the transcription level and protein level of GATA-1 is elevated in the early stages of erythroid differentiation, and decreases drastically when red cells are terminally differentiated (34).

### *The chicken $\alpha$ -globin enhancer*

Based on such background and the contribution of the 120 bp DNA fragment to the overall enhancer activity, we decided to test the capacity of the chicken GATA-1 cDNA to *trans*-activate the non-active enhancer-promoter in 6C2 cells with the core enhancer (E) and the entire enhancer (E+120). As expected, increased concentrations of chicken GATA-1 cDNA are able to stimulate reporter activity (Fig. 7). Interestingly, a clear difference in the *trans*-activation capacity of GATA-1 is observed between the core enhancer, composed of three GATA-1 binding sites ( $\alpha^D E$ ) and the full enhancer ( $\alpha^D E+120$ ) formed by the core enhancer with the 120 bp DNA fragment. Higher reporter activity is reached when the core enhancer is present in the construct. In contrast, when the full enhancer is tested there is just partial *trans*-activation. Interestingly, the presence of only the 120 bp DNA fragment is sufficient to abate enhancer activity in half (Fig. 7; compare  $\alpha^D E$  to  $\alpha^D E+120$  and Fig. 2A; constructs 1 and 3). We interpret this result from the perspective of the 120 bp element where we have seen the binding of additional transcription factors including a fourth GATA-1 and a novel YY1 interaction (see below). Hence, the 120 bp DNA sequence seems to contain regulatory components that can modulate the activity of the enhancer in distinct erythroid differentiation stages.

*The Ying Yang 1 factor interacts with the 120 bp DNA fragment of the enhancer-* Previous observations in terminally differentiated erythroblasts showed two retarded complexes when the GATA-1 site is used in the 120 bp DNA fragment as a probe in a gel shift assay (29; Fig. 8A). These data suggest the interaction of an additional factor in close proximity or overlapping the GATA-1 motif in the 120 bp DNA fragment (Fig. 5A). In the present study we decided to use pre-erythroblast 6C2 cells, where the globin genes are not expressed, as a source of nuclear factors (30, 31). With the fourth GATA-1 sequence as a probe we found two retarded complexes, confirming the binding of GATA-1 and a second nuclear factor (Fig. 8A). Detailed sequence analysis revealed the presence of the TCATG motif

*The chicken  $\alpha$ -globin enhancer*

adjacent to the fourth GATA-1 site in the 120 bp DNA fragment (Fig. 8B). Further characterization showed that the TCAT sequence that was previously identified in the human  $\epsilon$  silencer interacts with the multifunctional Ying Yang 1 factor adjacent to a GATA-1 binding site (35, 36). In addition, a more recent report defines the same TCAT motif as an YY1 binding site in the murine beta interferon promoter with activation and repressive effects on transcription (37). All of this evidence prompted us to ask if YY1 could interact with the 120 bp DNA fragment, in proximity to the fourth GATA-1 binding site. We then carried out a gel shift assay using RBC nuclear extracts with the fourth GATA-1 sequence as a probe (Fig. 8B). We include in this experiment probes with the GATA-1 site mutated, a YY1 consensus sequence, the fourth GATA sequence with the putative YY1 site mutated, and the double mutant for the fourth GATA-1 and YY1 sequences (Fig. 8B). When the mutation of the GATA-1 sequence ( $\Delta$ GATA-1) is used as a probe, we observe less GATA-1 bound and the appearance of an additional slow migrating complex (Fig. 8B; lane 6). To demonstrate the presence of YY1 in RBC nuclear extracts, we carried out a control with YY1 consensus motif (Fig. 8B; lanes 8 and 9). We next mutated three bases of the predicted YY1 motif ( $\Delta$ YY1) (Fig. 8B; lanes 11 and 12). Consistent with the previous results, we found that the double GATA/YY1 mutations avoid the binding of the two corresponding factors. An additional slow migrating complex is now seen to allow for the possibility of the presence of a third factor or a degradation product (Fig. 8B; lanes 14 and 15). To further characterize YY1 interaction around the fourth GATA-1 site of the 120 bp DNA fragment, we performed a gel shift assay using a polyclonal antibody against human YY1 (Fig. 8C). Since we do not know if the YY1 antibody recognizes this factor in chicken erythroid nuclear extracts we first incubated the YY1 antibody with labeled YY1 consensus sequence and with 10-day old chicken embryonic erythrocytes and human HeLa nuclear extracts (Fig. 8C). A clear supershift complex is seen in each case (Fig. 8C; lanes 3 and 5). We next repeated the experiment using the GATA-1 sequence of



### *The chicken $\alpha$ -globin enhancer*

the 120 bp DNA fragment and RBC nuclear extracts. A specific supershift is seen when the YY1 antibody is used, strongly supporting the presence of YY1 on this sequence (Fig. 8C; lane 8). We then carried out an experiment using the YY1 consensus sequence as a probe, with RBC and 6C2 cells as sources of nuclear extracts and specific competitors containing YY1 and GATA-1 mutations over the GATA-1 sequence (Fig. 8D). The results are consistent with the previous data since YY1 interaction with the consensus sequence is only competed when the GATA-1 sequence is mutated, leaving access to the YY1 binding sequence. Interestingly, when the wild-type GATA-1 of the 120 bp DNA sequence is used as a cold competitor, there is no competition in RBC nuclear extracts (Fig. 8D; lane 5). In contrast, in 6C2 cells using increasing concentrations of cold competitors, slight competition is seen (Fig. 8D; lanes 13 and 14). This result shows that YY1 cannot bind to its recognition sequence when GATA-1 is present in RBC. However, in 6C2 cells partial competition may reflect a difference in GATA-1 and YY1 abundance. This idea seems to be further supported by the fact that the same nuclear extract concentrations are used (Fig. 8D; compare lanes 2 and 7). Finally, further confirmation of YY1 interaction with the fourth GATA-1 sequence is evident by competition with the GATA-1 mutant sequence (Fig. 8D; lanes 6 and 17). Previous reports have shown that GATA-1 varies in its messenger RNA and peptide concentrations during development, being at its highest peak around 4 days of chicken embryonic development, then decreasing gradually after day 8 (34). To our knowledge no data is available for YY1 in chicken cells. Thus, we performed western blots using HD24 (one of the most primitive avian erythroid cell lines), 6C2, HD3, RBC and the human erytholeukemic K562 cells as sources of protein extracts (Fig. 8E). As can be seen, there are no significant variations in YY1 relative abundance in the chicken cell types tested, thus one possibility is that GATA-1 is responsible for modulating such differential interactions. To finally confirm the *in vivo* YY1 interaction with the fourth GATA-1 sequence, chromatin immunoprecipitation experiments were done (Fig. 8F). Using 6C2 and HD3 as a source of erythroid cells and the chicken lymphoma cell line DT40 as a source of non-

### *The chicken $\alpha$ -globin enhancer*

erythroid cells, we confirm the interaction of YY1 and GATA-1 with primers amplifying the 120 bp DNA fragment (Fig. 8F). No amplification product is seen in DT40 cells, and interestingly, we observe an amplification product from the YY1 and GATA-1 immunoprecipitated samples in our  $\beta$ -globin control corresponding to the chicken  $\beta^A/\epsilon$  enhancer. This suggests the possibility that YY1 is also interacting with the  $\beta$ -globin enhancer, data not reported previously, although there is a putative YY1 site in the  $\beta^A/\epsilon$  enhancer defined by TRANSFAC analysis (data not shown). In addition we carried out a ChIP assay using an antibody against Sp1. We found that Sp1 is mainly present in 10 days RBC and absent in 6C2 and HD3 cells. This result is consistent with the fact that the enhancer is fully functional in 10 days RBC suggesting that Sp1 is, at least partially, contributing to such activity (Fig. 5C, 8F).

The results presented here support the interaction of YY1 with the GATA-1 sequence included in the 120 bp DNA fragment (Fig. 1, 5A), overlapping or in close proximity to the GATA-1 factor binding site (Fig. 8B). Variations in GATA-1 concentration could dictate the ability of YY1 to act in a regulated way, and this is an idea that is currently being tested. This is based on the fact that a competition between these two factors for their DNA binding sequences occurs. Based on the GATA-1 *trans*-activation experiments and the apparent enrichment of YY1 in primitive and terminally differentiated erythroid cells, we suggest that YY1 may negatively modulate the enhancer function in those stages of differentiation. Further experiments are required to address such a model.

## DISCUSSION

Over the years the study of globin genes from the perspective of chromatin structure and erythroid gene expression has become an excellent model for understanding mechanisms governing tissue- and developmental-stage specific gene regulation. Differential regulation of gene expression falls in the center of many biological processes, including hemoglobinopathies. The understanding of how regulatory elements work is a difficult but at the same time a fascinating task. In this particular line of thinking, it is important to recall that actually there is still no clear idea of how enhancer and silencers work (6). Even though recent results support the notion that long-distance contacts occur by multiple loop formation, the coordinated action of enhancers throughout cell differentiation remains to be clarified. In addition, the regulation of globin gene expression during erythroid development involves the formation of chromatin domains, local chromatin remodeling and the determination of erythroid characters (2, 3). With this in mind we carried out a detailed study of the chicken  $\alpha$ -globin 3' enhancer. We characterized the enhancer activity over different erythroid differentiation stages, showing the highest activation rates in 10-day-old embryonic chicken erythrocytes. Deletional and mutational analysis show that the 120 bp DNA fragment is an integral part of the enhancer and we suggest that it is responsible for modulating the overall enhancer function. In addition, we identify novel DNA-binding proteins interacting with this region, particularly the presence of the YY1 factor.

Based on our observations, the 3' chicken enhancer contains binding sites for a classical set of nuclear factors: GATA-1, NF-E2, EKLF, Sp1 and YY1 (Fig. 5A, 8F). The present knowledge for the activity of each of these factors has evolved and likely involves chromatin remodeling associated activities. In the case of GATA-1, it is well known that its recognition sequence is found in globin promoters, enhancers and LCRs. GATA-1 is an essential erythroid transcription factor and the GATA-motif distribution over regulatory elements seems to be critical for several reasons. We usually find

### *The chicken $\alpha$ -globin enhancer*

more than one binding site that is not far away from another (33). Even in some cases the sites are distributed in tandem, apparently increasing their affinity to DNA (30). Related to the present study there is clear evidence in the literature showing that GATA-1 can interact with the erythroid Krüppel-like factor (EKLF) and Sp1, and that such interactions cooperate in the activity of several erythroid-specific control elements (38, 39). Interestingly, the distribution of GATA-1, EKLF and Sp1 sites over the 3'  $\alpha$ -globin enhancer favors the idea of multiple contacts between these transcription factors, generating a highly specific combinatorial complex. More recently, it has been demonstrated that GATA-1 is able to interact with the histone acetyltransferase CBP *in vitro* and *in vivo*, strongly supporting the influence of GATA-1 in chromatin remodeling (7, 40, 41). To further complicate the scenario, there are two alternative pathways for GATA-1 function. Detailed studies on GATA-1 interactions revealed the protein-protein contact with a recently defined protein called "Friend of GATA" or FOG. FOG has been shown to be a co-factor of GATA-1, positively and negatively modulating its response during erythroid and megakaryocyte differentiation (42). The other pathway involves a GATA-1 homologue that is GATA-2 in distinct and some times overlapping in some stages of erythroid differentiation. It is generally accepted that GATA-2 action is found in earlier stages of differentiation (43-45). In contrast, GATA-1 is expressed and more abundant in later stages of erythroid maturation, and decreases in concentration when terminal differentiation is reached (34). In the 3' enhancer there is no significant influence of GATA-1 over the  $\alpha^D$  promoter in 6C2 cells. We can speculate that GATA-2 and not GATA-1 is playing a role in such a pre-erythroblast stage, where the globin locus is prepared not only transcriptionally but also at the level of chromatin domain structure to initiate regulated globin gene expression (7, 31, 46).

Another relevant erythroid factor is NF-E2 that forms an erythroid-specific heterodimer that binds to the TGCTGAGTCA sequence motif (47). NF-E2 is a member of the Maf transcription factors

### *The chicken $\alpha$ -globin enhancer*

family composed of the erythroid-specific large subunit p45 and the more general p18 small subunit (also known as MafK; 47). Like GATA-1, NF-E2 can also interact with the histone acetyltransferase CBP, probably involving transcription regulation and chromatin modification (48). Concerning the effects of NF-E2 on globin regulation, it has been demonstrated that interference of NF-E2 p45 produces a drastic decrease in globin gene expression (49). In contrast, over-expression of NF-E2 p45 induces erythroid terminal differentiation (49, 50). This observation argues in favor of the relevance of NF-E2 in globin gene expression and chromatin modifications. It is interesting to mention that the great majority of the studies focus their attention on NF-E2 interactions with the LCRs and gene promoters. In fact, a recent study shows that the p45 subunit of NF-E2 contributes to the histone hyperacetylation at the mouse adult  $\beta$ -globin gene promoter and not at the LCR (51). These data raise the question about the function of NF-E2 at HS2, HS3 and HS4 of the LCR.

In a more recent publication, Groudine and collaborators demonstrated that the p18 and p45 subunits of NF-E2 are located in distinct sub-nuclear compartments (16). Interestingly, in undifferentiated mouse erythroleukemic MEL cells, the p18 subunit of NF-E2 is largely located in close association with centromeric heterochromatin. In contrast, the p45 subunit of NF-E2 is mainly found in euchromatic compartments (16). Once differentiation is induced, there is heterodimerization of the p18 and p45 subunits in euchromatic compartments. These observations strongly suggest that erythroid differentiation can be in part modulated by the capacity of globin regulator nuclear factors to be relocated away from heterochromatin compartments, thus switching their capacity to interact with their target sequences in globin control elements. It has become attractive to define the contribution of NF-E2 to enhancer activity, particularly from the point of view of sub-nuclear relocation of the  $\alpha$ -globin locus. This may be even more relevant since we have found that mutation of the NF-E2 site has a significant effect on the enhancer activity in an integrated context, i.e., in the chromatin environment.

### *The chicken $\alpha$ -globin enhancer*

As mentioned before, the erythroid Krüppel-like factor, or EKLF, is an erythroid specific transcription factor that recognizes the CACCC sequence motif and is present in the great majority of globin control elements (52, 53). It has been suggested that EKLF participates in the switch transition from embryonic and fetal globin to adult gene expression (54). In chromatin reconstitution experiments, ELKF interacts with subunits of the SWI/SNF chromatin remodeling complex, supporting the notion that EKLF participates in chromatin modifications (55, 56). As mentioned before, EKLF also interacts with GATA-1 (39). At the present time it remains unclear what the contribution of EKLF is to the activity of globin control elements and its intersection or combinatorial interactions with Sp1.

Our results define the presence of the Ying Yang 1 nuclear factor, both *in vivo* and *in vitro*, close to the GATA-1 site in the 120 bp DNA fragment of the chicken 3'  $\alpha$ -globin enhancer. YY1 is involved in positive and negative activation based on co-factor recruitment, including histone acetyltransferases like CBP and p300, and on the other hand with histone deacetylases like HDAC1, HDAC2 and HDAC3 (36, 57, 58). In addition, in a recent publication it has been demonstrated that indeed YY1 can be acetylated by both p300 and PCAF, and deacetylated by HDAC1 and HDAC2 (59). The biological significance of such post-translational modifications of YY1 remains to be clarified, but it has been suggested, as for GATA-1, that those modifications could increase DNA-binding affinities to the DNA with regulatory consequences (59, 60). All of these observations intend to define the contribution of YY1 to the 3' chicken  $\alpha$ -globin enhancer. One possible speculation is related with the idea that YY1 function is modulating the activity or the contribution of some other factors to the enhancer activity in different developmental stages. Another possibility is that YY1 participates in the regulated chromatin remodeling of the enhancer. In addition, the data presented here do not discard the possibility of an interaction between GATA-1 and YY1 in the 120 bp DNA fragment of the enhancer. Further investigations are required to define the role of YY1 in the enhancer function.

### *The chicken $\alpha$ -globin enhancer*

In an artificially reconstituted erythroid system, the formation of a DNase I hypersensitive site is not dependent on GATA-1 and NF-E2 (8). In addition, it requires the incorporation of additional transcription factors, such as EKLF and notoriously, Sp1 (8). Even more, several globin regulatory elements in different organisms incorporate a varied and well define set of general and erythroid transcription factors. The most notorious is the human -40HS, which is constituted by the same group of factors presented here and in a similar organization (33). Thus, the combination of transcription factors are critical for the enhancer or silencer activity, but in addition those interactions recruit chromatin remodeling activities that act in concert to define the architecture of those elements that are capable of reaching the highest and regulated levels of gene regulation, particularly during distinct stages of erythroid differentiation.

The biological significance of the silencer-enhancer elements remains unclear. The present study shows a more complete characterization of the 3' chicken  $\alpha$ -globin enhancer and its differential activity over distinct erythroid differentiation stages. We are certainly now in a better position to ask novel questions that contribute to the mechanistic understanding of the mode of action of this couple of regulatory elements. The distribution of DNA-binding proteins over the region predicts a differential combinatorial action of those factors influenced by polypeptide concentrations, competition for DNA-binding motifs, and chromatin remodeling activities. We are currently analyzing the chromatin structure over the silencer-enhancer to in the near future further support the contribution of chromatin remodeling events to the silencer and enhancer activities.

*The chicken  $\alpha$ -globin enhancer*

*Acknowledgments-* We are indebted with Catherine M. Farrell for critical reading of the manuscript, suggestions and constant support. We also thank Tomoyuki Sawado and Jessica Halow for critical reading of the manuscript.



REFERENCES

1. Bulger, M., and Groudine, M. (1999) *Genes Dev.* **13**, 2465-2477
2. Recillas-Targa, F. (2000) *Cell. Mol. Biol. Lett.* **5**, 451-467
3. Recillas-Targa, F., and Razin, S. V. (2001) *Crit. Rev. Eukaryot. Gene Exp.* **11**, 227-242
4. Burgess-Beusse, B., Farrell, C., Gaszner, M., Litt, M., Mutskov, V., Recillas-Targa, F., Simpson, M., West, A., and Felsenfeld, G. (2002) *Proc. Natl. Acad. Sci. USA* **99**, suppl. 4, 16433-16437
5. Razin, S. V., Farrell, C. M., and Recillas-Targa, F. (2003) *Inter. Rev. Cytol.* **226**, 63-125
6. Blackwood, E. M., and Kadonaga, J. T. (1998) *Science* **281**, 60-63
7. Boyes, J., and Felsenfeld, G. (1996) *EMBO J.* **15**, 2496-2507
8. Goodwin, A. J., McInerney, J. M., Glander, M. A., Pomerantz, O., and Lowrey, C. H. (2001) *J. Biol. Chem.* **276**, 26883-26892
9. Sutter, N. B., Scalzo, D., Fiering, S., Groudine, M., and Martin, D.I. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 1105-1110
10. Carter, D., Chakalova, L., Osborne, C. S., Dai, Y.-F., and Fraser, P. (2002) *Nature Genet.* **32**, 1-4
11. Tolhuis, B., Palstra, R.-J., Splinter, E., Grosveld, F., and de Laat, W. (2002) *Mol. Cell* **10**, 1453-1465
12. Knezetic, J. A., and Felsenfeld, G. (1989) *Mol. Cell. Biol.* **9**, 893-901
13. Recillas Targa, F., De Moura Gallo, C. V., Huesca, M., Scherrer, K., and Marcaud, L. (1993) *Gene* **129**, 229-237
14. Recillas Targa, F., De Moura Gallo, C. V., and Scherrer, K. (1993) *Biochem. Biophys. Res. Commun.* **190**, 1163-1172
15. Sawado, T., Igarashi, K., and Groudine, M. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 10226-10231
16. Francastel, C., Magis, W., and Groudine, M. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 12120-12125
17. Beug, H., von Kirchbachl, A., Döderlein, G., Conscience, J.-F., and Graf, T. (1979) *Cell* **18**, 375-390
18. Beug, H., Palmieris, S., Freudenstein, C., Zentgraf, H., and Graf, T. (1982) *Cell* **28**, 907-919

37. Weill, L., Shestakova, E., and Bonnefoy, E. (2003) *J. Virol.* **77**, 2903-2914
38. Zon, L. I., Youssoufian, H., Mather, C., Lodish, H. F., and Orkin, S. H. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 10638-10641
39. Merika, M., and Orkin, S. H. (1995) *Mol. Cell. Biol.* **15**, 2437-2447
40. Blobel, G. A., Kakajima, T., Eckner, R., Montminy, R., and Orkin, S. H. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 2061-2066
41. Letting, D. L., Rakowski, C., Weiss, M. J., and Blobel, G. A. (2003) *Mol. Cell. Biol.* **23**, 1334-1340
42. Tsang, A. P., Visvader, J. E., Turner, C.A., Fujiware, Y., Yu, C., Weiss, M. J., Crossley, M., and Orkin, S. H. (1997) *Cell* **90**, 109-119
43. Leonard, M. W., Brice, M., Engel, J. D., and Papayannopoulou, T. (1993) *Blood* **82**, 1-9
44. Leonard, M. W., Lim, K. C., and Engel, J. D. (1993) *Development* **119**, 519-531
45. Ikonomi, P., Noguchi, C. T., Miller, W., Kassahun, H., Hardisson, R., and Schechter, A. N. (2000) *Gene* **261**, 277-287
46. Litt, M. D., Simpson, M., Recillas-Targa, F., Prioleau, M.-N., and Felsenfeld, G. (2001) *EMBO J.* **20**, 2224-2235
47. Blank, V., and Andrews, N. C. (1997) *Trends Biochem. Sci.* **22**, 437-441
48. Hung, H.-L., Kim, A. Y., Hong, W., Rakowski, C., and Blobel, G. A. (2001) *J. Biol. Chem.* **276**, 10715-10721
49. Francastel, C., Poindessous-Jazat, V., Augery-Bourget, Y., and Robert-Lezenes, J. (1997) *Leukemia* **11**, 273-280
50. Igarashi, K., Itoh, K., Motohashi, H., Hayashi, N., Matuzaki, Y., Nakauchi, H., Nishizawa, M., and Yamamoto, M. (1995) *J. Biol. Chem.* **270**, 7615-7624
51. Johnson, K. D., Christensen, H. M., Zhao, B., and Bresnick, E. H. (2001) *Mol. Cell* **8**, 465-471
52. Dang, D. T., Pevsner, J., and Yang, V. W. (2000) *Int. J. Biochem. Cell Biol.* **32**, 1103-1121
53. Perry, C., and Soreq, H. (2002) *Eur. J. Biochem.* **269**, 3607-3618
54. Wijerde, M., Gribnau, J., Trimbom, T., Nuez, B., Philipsen, S., Grosveld, F., and Fraser, P. (1996) *Genes Dev.* **10**, 2894-2902
55. Armstrong, J. A., Bieker, J. J., and Emerson, B. M. (1998) *Cell* **95**, 93-104

*The chicken  $\alpha$ -globin enhancer*

56. Brown, R. C., Pattison, S., van Ree, J., Coghill, E., Perkins, A., Jane, S. M., and Cunningham, J. M. (2002) *Mol. Cell. Biol.* **22**, 161-170
57. Lee, J. S., Galvin, K. M., See, R. H., Eckner, R., Livingston, D., Moran, E., and Shi, Y. (1995) *Genes Dev.* **9**, 1188-1198
58. Yang, W. M., Inouye, C., Zeng, Y., Bearss, D., and Seto, E. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 12845-12850
59. Yao, Y. L., Yang, W. M., and Seto, E. (2001) *Mol. Cell. Biol.* **21**, 5979-5991
60. Boyes, J., Byfield, P., Nakatani, Y., and Ogryzto, V. (1998) *Nature* **396**, 594-598

FIGURES LEGENDS

**FIG. 1. Scheme of the chicken  $\alpha$ -globin gene domain.** The silencer-enhancer control elements are located around 450 bp downstream of the adult  $\alpha^A$  gene. The enhancer is located adjacent to the silencer where at least three DNA-binding sequences for unknown factors have been previously defined (SF1, SF2 and SF3; 13). The first characterization of the enhancer determined the presence of three GATA-1 binding sites (12, 13). From a sequence alignment of the duck and chicken homologous sequences, we identified a highly conserved 120 bp DNA fragment with binding sites for EKLF, NF-E2 and GATA-1 (13, 14).

**FIG. 2. Functional characterization of the 3' chicken  $\alpha$ -globin silencer-enhancer in pre-erythroblasts.** *A*, Different silencer and enhancer sub-fragments were transiently transfected using the *CAT* reporter gene under the control of the chicken adult  $\alpha^D$  promoter. Plasmids were transfected in the transformed chicken pre-erythroblast 6C2 cell line. *B*, The same constructs were stably co-transfected with a plasmid containing the *Neo* resistance gene and pools of stable lines were analyzed. Transfection efficiencies were normalized with protein concentrations.

**FIG. 3. Functional characterization of the 3' chicken  $\alpha$ -globin silencer-enhancer in erythroblasts.** *A*, Different silencer and enhancer sub-fragments were transiently transfected using the *CAT* reporter gene under the control of the chicken adult  $\alpha^D$  promoter. Plasmids were transfected in the transformed chicken pre-erythroblast 6C2 cell line. *B*, Stable transfections in HD3 cells of constructs expressing the GFP protein. Data is analyzed in terms of distinct levels of intensity of fluorescence (M1 to M4) shown at the top. The percentage of each region is shown on the graph by the 4 bars. Data is representative of at least 10 independent clones. At the bottom a representative flow cytometry graph is presented for

*The chicken  $\alpha$ -globin enhancer*

each construct tested. Constructs  $\alpha^D$ :promoter alone,  $\alpha^{DE}$ :promoter+core enhancer,  $\alpha^D+120$ :promoter+120 bp DNA fragment alone. The EKLF, NF-E2 and GATA-1 mutations are tested in the context on the entire enhancer.

**FIG. 4. Functional characterization of the 3' chicken  $\alpha$ -globin enhancer and enhancer blocking assay in 10-day-old chicken embryonic erythrocytes.** The core enhancer (e) corresponds to the three GATA-1 binding sites. The silencer (s) represents a 360 bp DNA fragment. Insulator enhancer blocking activity is defined by the capacity of the insulator element to block or reduce enhancer-promoter communication only when the element is placed between the enhancer and the promoter (4, 21). Based on the insulator definition, we carried out an enhancer blocking assay in transiently transfected cells (RBC) with linearized plasmids where the silencer is located in its native location "between" the enhancer and the promoter (construct 5). The inversion of the insert places the silencer "outside" in relation to the entire enhancer and the promoter (construct 6).

**FIG. 5. DNA-protein interactions over the 3' chicken  $\alpha$ -globin enhancer.** *A*, Scheme of the DNA-protein interactions over the 3' chicken  $\alpha$ -globin enhancer. The DNA sequence includes the three GATA-1 sites (12, 13) and the 120 bp DNA fragment. Several GGTGG and putative Sp1/EKLF sites are found interspaced between the GATA-1 sites. EKLF and NF-E2 are shown in italics and the YY1 binding site is outlined. *B*, A *Hin*pI-*Hin*fI (285 bp) DNA fragment containing the full enhancer was labeled, incubated with increasing concentrations of HD3 nuclear extracts and *in vitro* digested with DNase I nuclease. Interestingly, several footprints are seen in addition to the GATA-1 protected sequences. In particular, several GGTGG sequences and two Sp1-like motifs are protected (open

*The chicken  $\alpha$ -globin enhancer*

boxes). Notice the large footprinted area starting at the second GATA-1 site that includes the 120 bp DNA fragment. *C*, Electrophoretic mobility shift assay using the Sp1a and the Sp1b sequences as probes (see A for location). A clear retarded complex is formed in lanes 2 and 8 corresponding to a putative Sp1 protein that can be competed with 25-, 50- and 100-fold molar excesses of a consensus binding sequence for Sp1 (lanes 3-5 and 9-11; 25). *D*, To further characterize the 120 bp DNA fragment previously seen as a large 60 bp protected area (14, 29), we carried out competitions with known sequences that bind NF-E2 from the human porphobilinogen deaminase erythroid promoter (26) and GATA-1 from the chicken  $\beta$ -globin enhancer (27). Using increasing concentrations of HD3 nuclear extracts we once again observe the 60 bp footprint (lanes 5, 6 and 7). Notice the formation of an intense DNase I hypersensitive site (HS) 40 bp downstream of the protected area. When competing with two concentrations of cold NF-E2 binding sequence, we observe the re-appearance of a band located just in the middle of the NF-E2 recognition sequence in the 120 bp DNA fragment (see small arrow on lanes 8 and 9). Interestingly, this competition causes the loss of the HS suggesting that NF-E2 is at least in part responsible for the *in vitro* induction of a particular DNA conformation in conjunction with the other factors interacting near this site. For the GATA-1 competition we observe the re-appearance of the majority of the bands seen in the naked DNA digestion (see lanes 10 and 11). The HS is maintained and we speculate that GATA-1 could mediate a pivotal role for the formation of multiple interactions over the 120 bp DNA fragment.

**FIG. 6. Mutational study of the DNA-binding sequences of the 120 bp DNA fragment.** *A*, Transient transfections in HD3 cells. Mutations of the putative EKLF, NF-E2, the fourth GATA-1 and the double mutant NF-E2/GATA-1 sites are analyzed in the context of the entire enhancer. *B*, The same constructs were transiently transfected in 10-day-old chicken embryonic erythrocytes. Compare the levels of

*The chicken  $\alpha$ -globin enhancer*

normalized luciferase activities between HD3 and RBC cells. Notice the differences in the graph scale between the two sets of experiments.

**FIG. 7. *Trans*-activation of the chicken 3'  $\alpha$ -globin enhancer by GATA-1 over-expression.** The chicken cDNA for GATA-1 under the control of the CMV promoter was co-transfected with the core enhancer composed of the three GATA-1 sites (Fig. 1;  $\alpha^{\text{DE}}$ ) and with the full enhancer that includes the 120 bp DNA fragment. Co-transfections were done in 6C2 cells and transcriptional activation was measured with the dual-luciferase assay.

**FIG. 8. The Ying Yang 1 nuclear factor interacts *in vitro* and *in vivo* with the 120 bp DNA fragment of the 3' chicken  $\alpha$ -globin enhancer.** *A*, Gel mobility shift assays were carried out using nuclear extracts from the pre-erythroblast 6C2 cell line. Two retarded complexes are seen as previously observed with adult chicken terminally differentiated nuclear extracts (29). Based on our previous results, the fast migrating complex corresponds to the interaction of GATA-1 and the slow migrating complex to an undetermined factor. *B*, Sequence analysis shows a putative binding site (TCAG) for the YY1 factor. Interestingly, there is striking similarity between the fourth GATA-1 sequence and the adjacent TCATG motif with the human  $\epsilon$  silencer where GATA-1 and YY1 binding sites are located side by side (35). Based on such analysis, we carried out a gel shift assay using different probes, including the YY1 consensus, a fourth GATA-1 ( $\Delta$ GATA-1) and YY1 ( $\Delta$ YY1) mutations using increasing concentrations (5 and 10  $\mu$ g) of RBC nuclear extracts. *C*, On the left panel the consensus YY1 sequence was used as a probe in a supershift assay and the reaction was incubated in the presence of a polyclonal YY1 antibody with RBC and HeLa cell nuclear extracts. On the right panel the fourth GATA-1 motif was radioactively labeled and a supershift is seen using RBC nuclear extracts. The

*The chicken  $\alpha$ -globin enhancer*

GATA-1/YY1 complex and the supershift were efficiently and specifically self competed with 100-fold molar excess of cold GATA-1 sequence. *D*, To confirm the interaction of the YY1 factor with the fourth GATA-1 sequence, we carried out a gel shift assay using the consensus YY1 and the fourth GATA-1 sequences as probes and 6C2 and RBC cells as sources of nuclear extracts. Cold competitions are shown where YY1 corresponds to the self competition,  $\Delta$ YY1 is mutant oligonucleotide of the newly defined YY1 binding site, GATA-1 is the intact fourth GATA-1 sequence and  $\Delta$ GATA-1 is the fourth GATA-1 site mutated leaving the YY1 binding site intact. Notice the differences in retarded complexes abundance for 6C2 and RBC (compare lanes 2 and 8). In each set of experiments, 10  $\mu$ g of nuclear extracts were used. *E*, Western blots showing no differences in YY1 abundance in chicken cells from distinct differentiation stages. A human erythroleukemic K562 cell extract was included as a control. *F*, Chromatin immunoprecipitation assays using two different sets of primers which amplifies: in the chicken  $\alpha$ -globin enhancer the 120 bp DNA fragment (120 bp) generating a 169 bp amplification product and the chicken  $\beta$ -globin enhancer ( $\beta^A/\epsilon$ ) a 301 bp product. Input (i), bound (b) and unbound (ub) amplification products are shown. An irrelevant antibody ( $\alpha$ -X) was used and no amplification product was observed. Erythroid (6C2 and HD3) and non-erythroid cells (DT40) were used in these experiments. This is a representative result from two independent experiments.



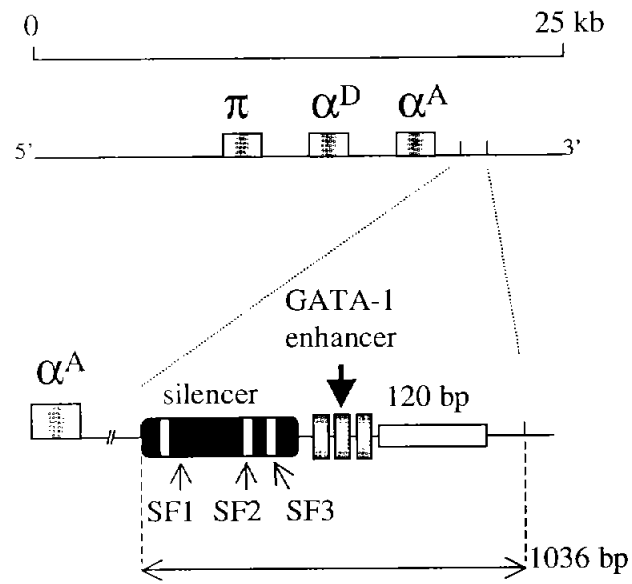


Figure 1

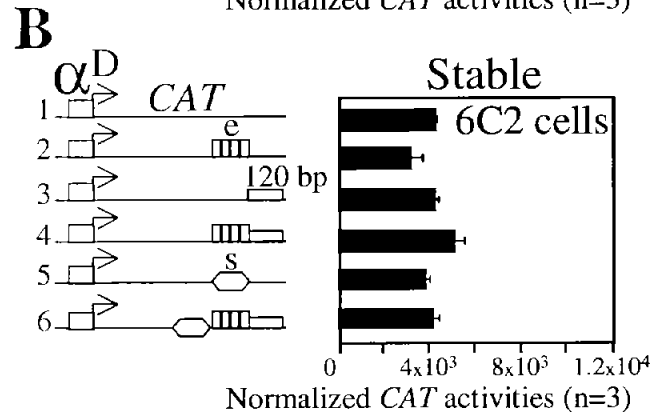
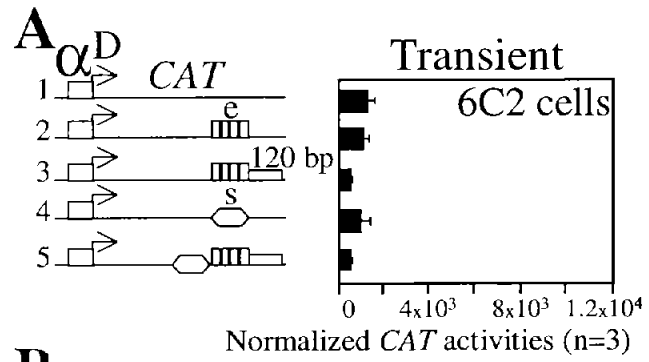


Figure 2

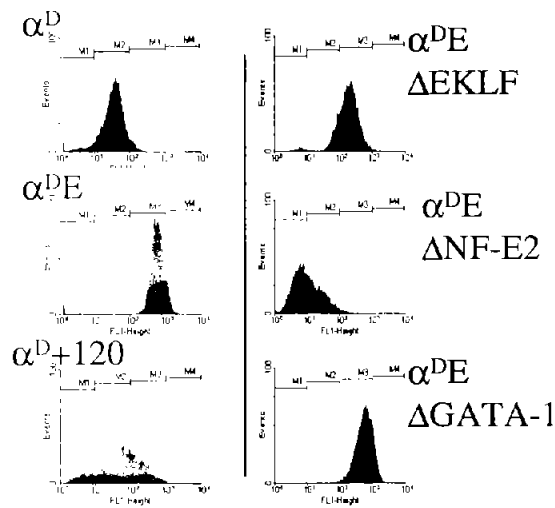
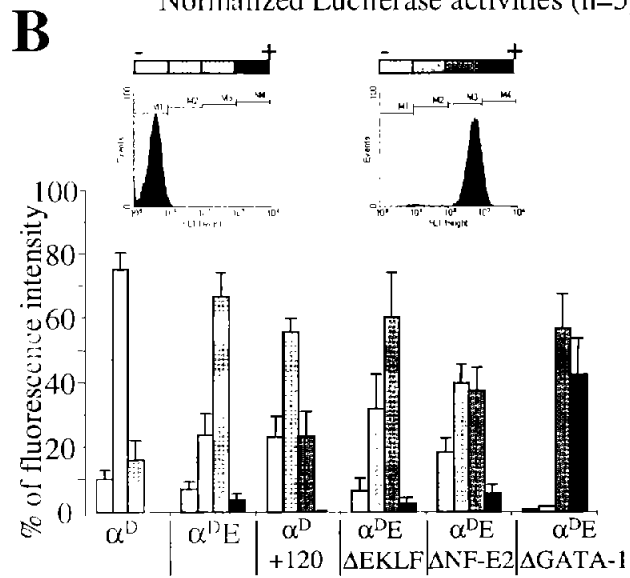
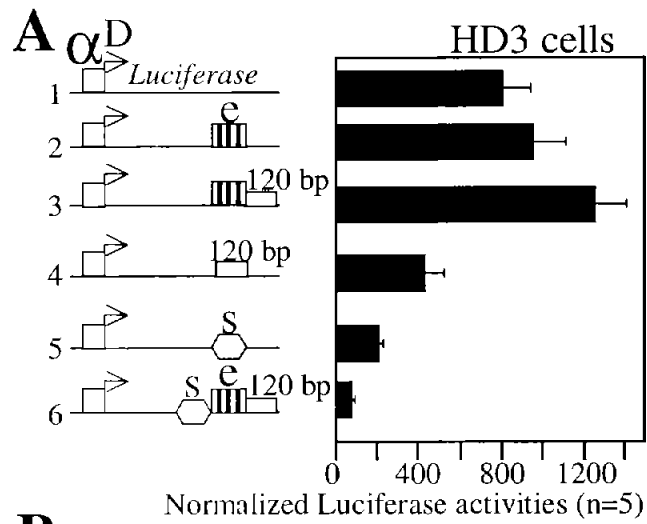


Figure 3

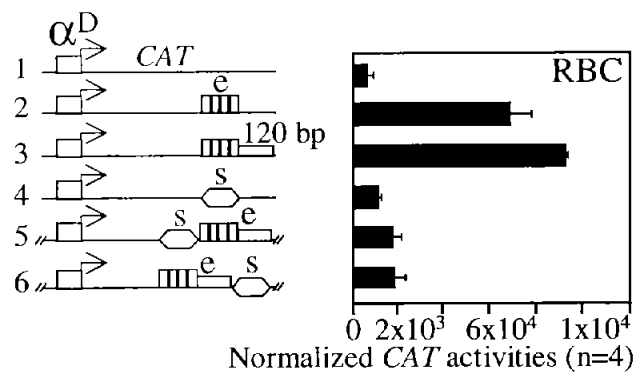


Figure 4

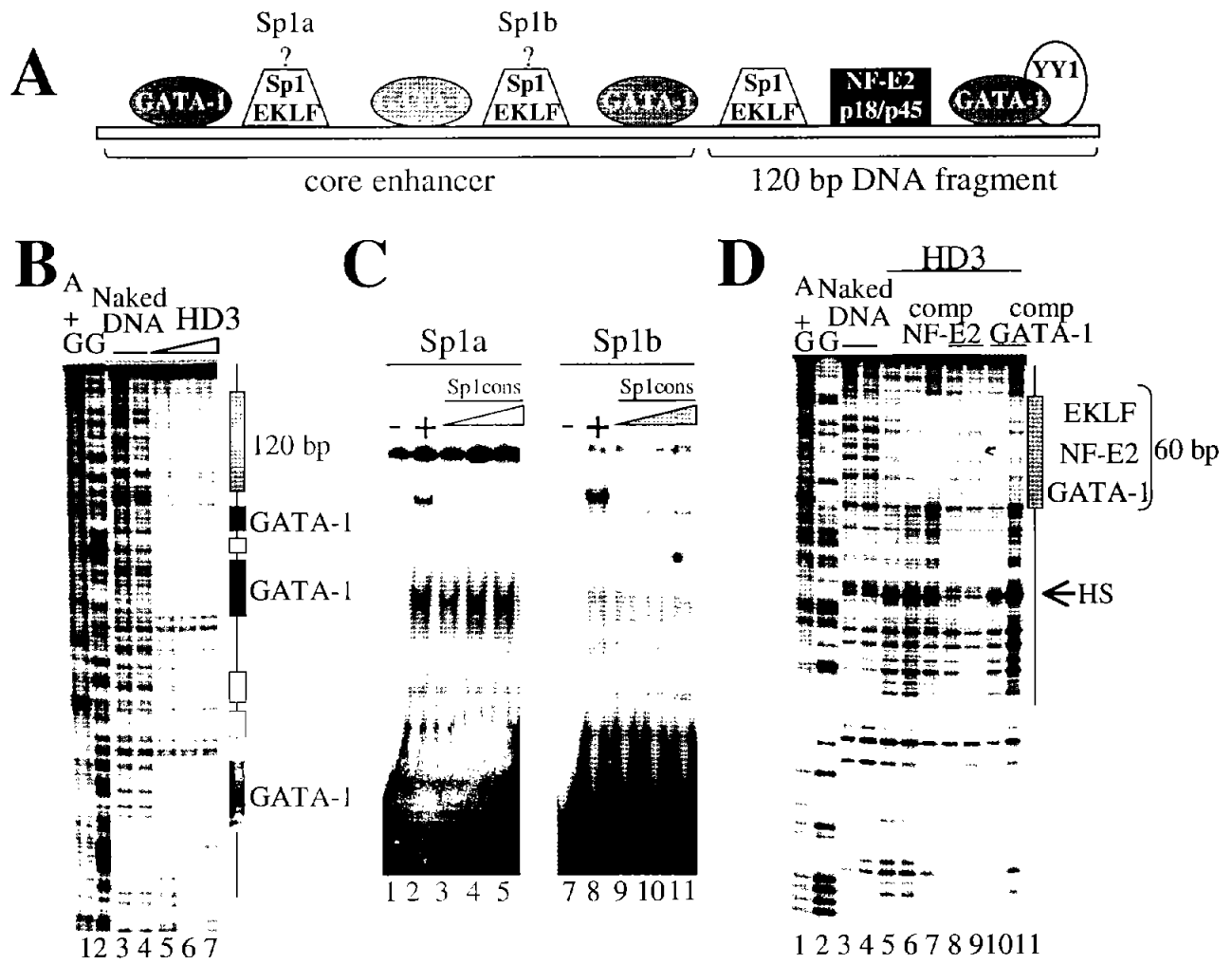


Figure 5

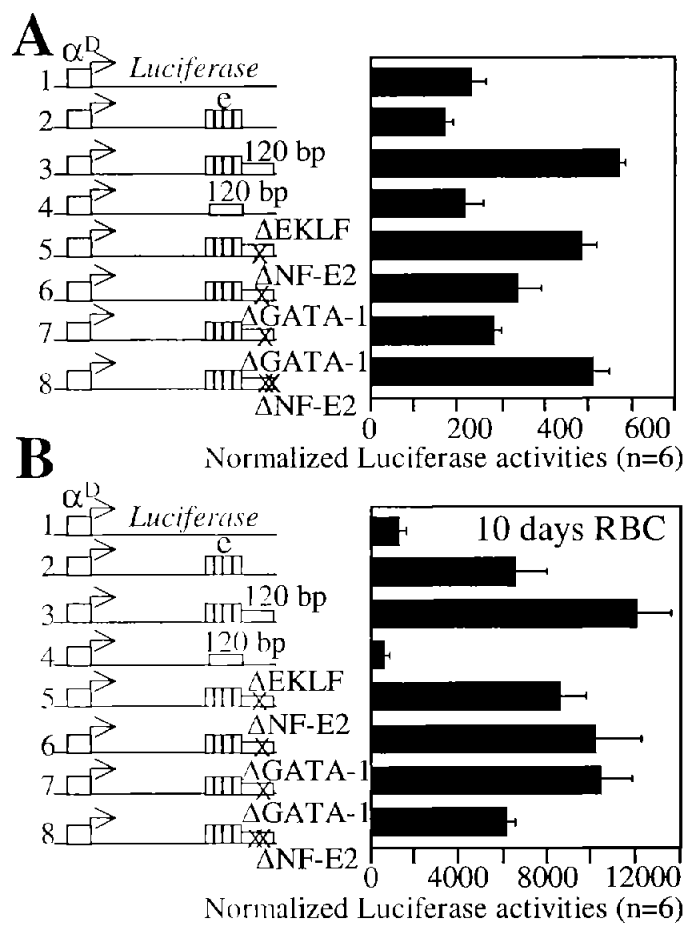


Figure 6

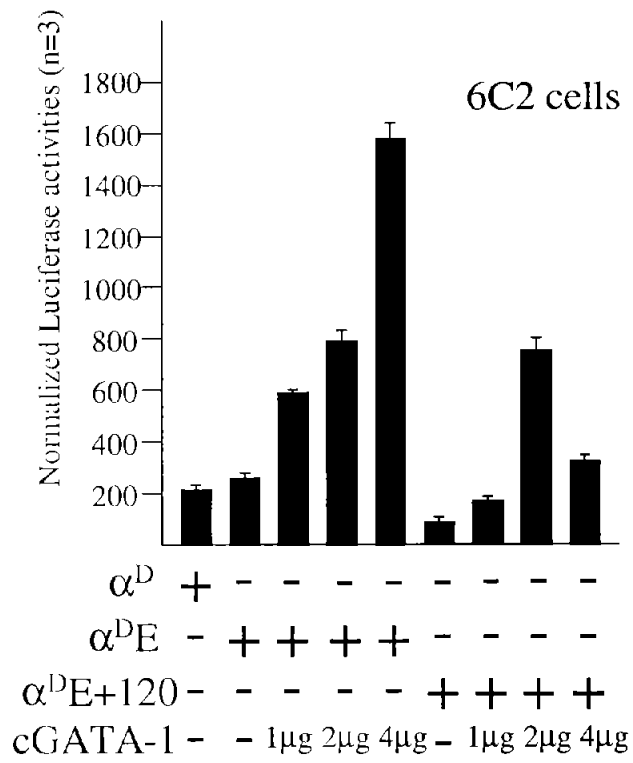


Figure 7





## APENDICE

### Reactivos

#### Cultivo:

Suero de Pollo GIBCO BRL cat # 16110-082

Suero Fetal Bovino GIBCO BRL cat # 16000-044

D-MEM GIBCO-BRL cat # 12800-017

L-15 GIBCO-BRL cat # 11415-064

Penicilina/Estreptomicina GIBCO-BRL cat # 15140-148

Geneticina GIBCO-BRL cat # 10131-035

$\alpha$ -MEM GIBCO-BRL cat # 11900-016

Buffer HEPES 1M Sigma cat # H0887

$\beta$ -mercaptoetanol GIBCO-BRL cat # 21985-023

#### Otros Reactivos:

Dual Luciferase Kit PROMEGA cat # E1910

TnT Kit PROMEGA cat # L1170

Glicina GIBCO BRL cat # 15527-013

Deoxycholate FLUKA cat # 30970

Glicógeno ROCHE cat # 901393

Poly dI-dC AMERSHAM cat # 27-7880

Proteinasa K Sigma cat # P2308

Rnasa A Sigma cat # R6315

Membranas PVDF MILLIPORE cat # IPVH 00010

$^3$ H Acetil CoA AMERSHAM cat # TRK 688

$\beta$ -galactosidase kit PROMEGA cat # E2000

SuperSignal Kit PIERCE cat # 34080

$\alpha$ -CTCF N17 SANTA CRUZ cat # sc-5916