



Análisis bioinformático de los reguladores epigenéticos

Sergio Lois Olmo

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tdx.cat) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tdx.cat) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tdx.cat) service has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized neither its spreading and availability from a site foreign to the TDX service. Introducing its content in a window or frame foreign to the TDX service is not authorized (framing). This rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

TESIS DOCTORAL

DEPARTAMENTO DE BIOQUÍMICA Y BIOLOGÍA MOLECULAR

Facultad de Biología de la Universidad de Barcelona

Programa de doctorado en Biomedicina

Bienio 2003 - 2005

ANÁLISIS BIOINFORMÁTICO DE LOS REGULADORES EPIGENÉTICOS

**Memoria presentada por el Licenciado Sergio Lois Olmo
para optar al Grado de Doctor por la Universidad de Barcelona**

La presente tesis doctoral ha sido realizada bajo la dirección del Dr. Xavier de la Cruz Montserrat, del Laboratorio de Biología Computacional y Bioinformática del Parc Científic de Barcelona, y la co-dirección de la Dra. Marian Martínez-Balbás del Instituto de Biología Molecular de Barcelona (CSIC).

Dr. Xavier de la Cruz

Director

Dra. Marian Martínez-Balbás

Co-Directora

Dr. Josep L. Gelpí

Tutor

Sergio Lois Olmo

Autor



Barcelona, Abril 2012.

AGRADECIMIENTOS

*“Cuando emprendas el viaje a Ítaca
ruega que sea largo el camino,
lleno de aventuras, lleno de experiencias”.*

“El Viaje a Ítaca” – **Konstantino Kavafis**

Seguramente, citar una estrofa de “El viaje a Ítaca” no sea la forma más original de afrontar un texto como este. Emplear la idea de un largo viaje como alegoría de una tesis doctoral es algo bastante recurrente. De todas formas, si mi intención fuera simplemente destacar que la realización de una tesis supone un largo recorrido con algún que otro infortunio, bastaría con remitirme a los famosos relatos de Julio Verne. Sin embargo, con esta breve alusión al poema de Kavafis me gustaría destacar la importancia del camino independientemente de cual sea el destino. El camino hacia nuestros sueños está repleto de experiencias enriquecedoras. Aprender a disfrutar del camino puede llegar a ser más satisfactorio que cumplir el ansiado sueño. *“Ítaca te ha dado el bello viaje”*. Lo importante no era simplemente escribir esta tesis, sino aprender lo que he aprendido, conocer a quien he conocido, viajar donde he viajado y sentir lo que he sentido. Estas primeras líneas reservadas a los agradecimientos, están dedicadas a todas aquellas personas que han hecho que este camino haya sido aún más enriquecedor y, que en mayor o menor medida, me han facilitado la llegada a Ítaca sin temer a los Lestrigones, a los Cíclopes o, incluso, al fiero Poseidón. En cada una de estas páginas hay un poco de cada unos de ellos: sus consejos, sus enseñanzas, sus motivaciones, sus ánimos y sus experiencias.

La importancia en el orden de los agradecimientos la reservo únicamente para *papa* y *mama*, porque ellos se merecen el mayor agradecimiento posible por transmitirme parte de lo que hoy soy, por su apoyo incondicional, su amor eterno y su sacrificio extremo. Cualquier camino que decida tomar será más fácil siempre que me acompañen y sé que allí estarán. Cierro esta etapa con la certeza que me lo han dado todo.

A un amigo responsable de mi llegada a Ítaca, padre científico, y director de esta tesis, *Xavier de la Cruz*, que ha hecho posible que emprendiera este camino, y que este fuera lo más llano posible. Le agradezco todo lo que he aprendido a su lado (mucho y

no solo de ciencia), su comprensión, su dedicación y sus consejos. Para él toda mi admiración y agradecimiento.

A *Marian Martínez-Balbás*, por su contribución a mi trabajo en diferentes colaboraciones, por su dedicación y por introducirme al interesante mundo de la epigenética. A las chicas de su laboratorio, especialmente a *Naiara Akizu* y *Noemi Blanco*, por la ayuda prestada, su disposición y los momentos en las barbacoas en el Montseny.

A *David Piedra*, compañero de viaje y amigo, un alma gemela en la que sentirse reflejado cuando algo funciona y cuando no también. Con toda seguridad, el camino fue más alegre compartiéndolo con él. A *Rebeca García*, un golpe de aire fresco y gallego, por sus consejos, por reírme las gracias y por compartir conmigo momentos especiales. A *Montse Barbany*, por contribuir con su experiencia a mi trabajo y dar continuidad a los proyectos. Al resto del *Molecular Modeling and Bioinformatic Group*, por los buenos momentos vividos y la ayuda prestada, especialmente a *Dr. Modesto Orozco*, por la oportunidad ofrecida.

En este agradable paseo por los agradecimientos, también me gustaría hacer una necesaria parada en *Molins de Rei* para agradecer a mis amistades que me hayan acompañado en este viaje. Vuestra ayuda no queda plasmada en un artículo científico, en un punto de las conclusiones o en un párrafo de los resultados, pero vuestro apoyo es muy importante aunque no sepáis exactamente en que o para que, por lo que parte de esta tesis también os pertenece, amigos. Gracias a todos!

Por último, me gustaría agradecer al *Institut de Medicina Predictiva i Personalitzada del Càncer* (especialmente al MAPLAB) el hacerme más dulce si cabe está última etapa del viaje. Sin lugar a duda, conocerlos ha sido una de las experiencias más enriquecedoras del viaje y seguramente, hoy soy mejor compañero y amigo gracias a lo que me habéis transmitido cada uno de vosotros. Gracias también!

Cada vez que se mueve un profundo deseo, se emprende un nuevo viaje a Ítaca...

TABLA DE CONTENIDOS

INDICE DE TABLAS Y FIGURAS	7
ABREVIATURAS Y NOMENCLATURA	9
INTRODUCCIÓN.....	15
1. Base molecular de las diferencias fenotípicas	15
1.1. Estudios con gemelos monocigóticos. El papel del entorno.....	17
1.2. Estudios con animales genéticamente idénticos. El tercer componente.....	18
2. ¿Qué entendemos por epigenética?	19
2.1. Mecanismos epigenéticos de control de la expresión génica	19
2.1.1. Remodelación dependiente de ATP.....	20
2.1.2. Variantes de histonas	20
2.1.3. Metilación del ADN.....	21
2.1.4. Modificaciones covalentes de las histonas	22
3. Cromatina y procesos epigenéticos: el rol de las histonas y sus modificaciones. ...	23
3.1. Regulación de la estructura de la cromatina	24
3.1.1. Modificaciones covalentes de las histonas	25
3.1.2. Efectos de las modificaciones de las histonas sobre la estructura de la cromatina	26
3.1.3. Función de las modificaciones covalentes de las histonas	29
4. Reconocimiento de las modificaciones de las histonas	32
4.1. Dominios que reconocen marcas de acetilación	36
4.1.1. Bromodominio.....	36
4.1.2. PHD – Doble dominio PHD de la proteína DPF3b.....	38
4.2. Dominios que reconocen marcas de metilación.....	39
4.2.1. Cromodominio	40
4.2.2. Dominios chromo barrel y chromoshadow.....	43
4.2.3. TUDOR.....	43
4.2.4. PHD finger.....	46
4.2.5. MBT	47
4.2.6. PWWP	49
4.2.7. WD40.....	50
4.2.8. Características comunes del reconocimiento de las lisinas metiladas	51
4.3. Reconocimiento de otras modificaciones	52
4.3.1. Fosforilación.....	52
4.3.2. Metilación de las argininas.....	53
4.3.3. Histonas no modificadas	53
OBJETIVOS	57

DISCUSIÓN GLOBAL	61
1. Propiedades estructurales y patrón de distribución de los dominios de interacción con la cromatina	61
2. Modulación funcional mediante variaciones locales en el centro activo	63
2.1. Lugar de unión de los dominios de interacción de los reguladores epigenéticos	63
2.2. Péptido de histonas.....	66
2.2.1. Características estructurales del péptido de histona: patrones de contactos atómicos.....	66
3. <i>Splicing</i> alternativo como mecanismo de regulación de los reguladores epigenéticos	67
3.1. Reducciones drásticas de tamaño	68
3.2. Cambios en el dominio catalítico	68
3.3. Cambios en los dominios de interacción	68
CONCLUSIONES GENERALES	73
BIBLIOGRAFÍA.....	77
PUBLICACIONES	93
INFORME DEL DIRECTOR DEL FACTOR DE IMPACTO DE LOS ARTICULOS PUBLICADOS.....	94
Do protein motifs read the histone code?.....	95
The functional modulation of epigenetic regulators by alternative splicing	109
Characterization of structural variability sheds light on the specificity determinants of the interaction between effector domains and histone tails.....	125

INDICE DE TABLAS Y FIGURAS

TABLAS

T1. Modificaciones covalentes de las colas de las histonas y su función asociada.....	25
T2. Familias de enzimas modificadoras de la cromatina y residuos modificados relacionados.....	32
T3. Dominios de reconocimiento de modificaciones de histonas.....	35
T4. Proteínas humanas que presentan cromodominios en su secuencia.....	41
T5. Resumen de las modificaciones reconocidas por el dominio MBT	48

FIGURAS

F1. Capas de organización de la cromatina en el núcleo de las células de mamífero.....	20
F2. Estructura cristalográfica de la cromatina	22
F3. Organización del ADN en la estructura de la cromatina	24
F4. Modificaciones covalentes de las colas de las histonas	26
F5. Modelo de cooperación de las enzimas mod. de la cromatina y los dominios de reconocimiento	29
F6. Distribución de las mod. covalentes de las histonas en diferentes elementos funcionales del genoma	31
F7. Bromodominio de GCN5 en complejo con un péptido de histona H4 acetilado (H4K16Ac).....	37
F8. Doble PHD finger de la proteína humana DPF3 unido a un péptido de histona acetilado (H3K14Ac).....	38
F9. Reconocimiento de la modificación H3K14Ac mediante el dominio PHD1.....	39
F10. Familias de enzimas involucradas en la metilación de determinados residuos de la histona H3	40
F11. Estructura cristalográfica del cromodominio de la proteína Polycomb unido a H3K27me3.....	42
F12. Estructura del dominio Tudor de 53BP1 en complejo con el péptido H4K20.....	44
F13. Estructura del doble dominio Tudor	45
F14. Estructura cristalográfica del dominio PHD finger de la proteína BPTF humana unida a H3K4me2.....	46
F15. Estructura cristalográfica del dominio MBT de L3MBT1 unido al péptido de histona	47
F16. Estructura cristalográfica del dominio PWWP de HDGF2 en complejo con H3K79me3	49
F17. Estructura cristalográfica del dominio WD40 de la proteína WDR5 en complejo con H3K4me2.....	50

ABREVIATURAS Y NOMENCLATURA

53BP1	<i>p53 binding protein-1</i>
ADN	Ácido desoxirribonucleico (en inglés DNA)
ARN	Ácido ribonucleico (en inglés RNA)
AS	Del inglés, <i>Alternative Splicing</i>
ASH	<i>Absent, small and homeotic</i>
ATP	Trifosfato de Adenosina
BAF	<i>Brg/Brahma associated factors</i>
BPTF	<i>Bromodomain PHD Finger Transcription Factor</i>
BRCA	<i>Breast Cancer</i>
BRCT	<i>C-terminal portion of the BRCA-1</i>
BRG	<i>Brahma-related gene</i>
BS	Del inglés, <i>Binding Site</i>
CARM1	<i>Coactivator-Associated Arginine Methyltransferase</i>
CBP	<i>CREB-Binding Protein</i>
CBX	<i>Chromobox homolog</i>
CREB	<i>cAMP reponse element-binding</i>
CTCF	<i>CCCTC-binding factor</i>
CHD	<i>Chromo Helicase DNA-binding</i>
ChIP-chip	Inmunoprecipitación de cromatina combinada con tecnología de microarrays
ChIP-seq	Inmunoprecipitación de cromatina combinada con secuenciación de ADN
cADN	ADN complementario
DBD	<i>DNA-binding domain</i>
DPFB3	<i>Double PHD Finger protein B3</i>
DNMT	<i>DNA methyltransferase</i>
DOT	<i>Disruptor of telomeric silencing</i>
ESET	<i>ERG-associated protein with SET domain</i>
EZH	<i>Enhancer of Zeste</i>
GCN5	<i>General Control of amino-acid synthesis 5</i>
GLP	<i>G9a-like protein</i>
HAT	<i>Histone Acetyltransferase</i>
HCP	<i>High CpG content promoter</i>
HDAC	<i>Histone Deacetylase</i>

HDGF	<i>Hepatoma-derived growth factor</i>
HDM	<i>Histone Demethylase</i>
HMT	<i>Histone Methyltransferase</i>
HP1	<i>Heterochromatin Protein 1</i>
IFN	<i>Interferon</i>
IHGSC	<i>International Human Genome Sequencing Consortium</i>
ING2	<i>Inhibitor of Growth protein 2</i>
ISWI	<i>Imitation Switch</i>
JHDM	<i>JmJC-Containing Histone Demethylase protein</i>
JMJD	<i>Jumonji domain-containing protein</i>
L3MBTL1	<i>Lethal-(3) Malignant Brain Tumor repeat-Like protein 1</i>
LCP	<i>Low CpG content promoter</i>
LSD	<i>Lysine specific demethylase</i>
MAGE	<i>Melanoma Antigen Gene</i>
MBS	<i>Del inglés, modified binding site</i>
MBT	<i>Malignant Brain Tumor</i>
MDC1	<i>Mediator of DNA damage checkpoint protein 1</i>
MLL	<i>Mixed lineage leukemia protein</i>
MORF4L1	<i>Mortality Factor 4 like 1 protein</i>
MRG	<i>Melanocyte-specific gene-Related Gene</i>
MSL	<i>Male-specific Lethal complex</i>
MYST	<i>Moz, Ybf2/Sas3, Sas2 and Tip60 family</i>
mARN	<i>Ácido Ribonucleico mensajero</i>
NSD1	<i>Nuclear receptor binding SET domain protein 1</i>
NURF	<i>Nucleosome Remodelling Factor</i>
NuA4	<i>Nucleosome acetylating H4</i>
NuRD	<i>Nucleosome Remodelling and deacetylase</i>
PB1	<i>Polybromo-1 protein</i>
PC	<i>Polycomb protein</i>
PCAF	<i>p300/CBP-associated factor</i>
PDB	<i>Protein Data Bank</i>
PHD	<i>Plant Homeo domain</i>
PRKDC	<i>Protein kinase, DNA-activated, catalytic polypeptide</i>
PRMT1	<i>Protein Arginine Methyltransferase 1</i>
PTBP1	<i>Polypyrimidine tract-binding protein 1</i>

PWWP	<i>Proline-Tryptophan-Tryptophan-Proline Motif</i>
p300/CBP	<i>p300/CREB binding protein complex</i>
RBBP	<i>Retinoblastoma binding protein</i>
RBP2	<i>Retinoblastoma binding protein 2</i>
RIZ	<i>Retinoblastoma protein-Interacting Zinc finger gene</i>
RMN	<i>Resonancia Magnética Nuclear</i>
RMSD	<i>Root Mean Square Deviation</i>
RNAPII	<i>Del inglés, RNA polymerase II</i>
RPD3	<i>Reduced Potassium Dependency 3</i>
RSC	<i>Chromatin Structure Remodelling Complex</i>
SAGA	<i>Spt-Ada-Gcn5-Acetyltransferase complex</i>
SANT	<i>Swi3, Ada2, N-CoR y TFIIIB</i>
SCM	<i>Sex combs on midleg</i>
SCML	<i>Sex combs on midleg-like</i>
SERPIN	<i>Serpine Protease Inhibitor</i>
SET	<i>Su(var)39, E(z) and trx</i>
SETDB1	<i>Set domain, bifurcated 1</i>
SFMBT	<i>Scm-related gene containing four mbt domains</i>
SMN	<i>Survival Motor Neuron</i>
SMYD1	<i>SET and MYND domain containing 1</i>
SUV39H	<i>Suppressor of variegation 3-9 homologue</i>
SWIRM	<i>SWI3, RSC8 and MOIRA</i>
SWI/SNF	<i>SWitch/Sucrose Non Fermentable (Complejo Remodelante)</i>
TAF	<i>TBP-associated factor</i>
TBP	<i>TATA-binding protein</i>
TIP60	<i>Tat Interactive protein, 60 kDa</i>
TF	<i>TATA-box binding protein</i>
TFIID	<i>Transcription Factor IID</i>
UTX	<i>Ubiquitously transcribed tetratricopeptide repeat, X chromosome</i>
WD40	<i>Tryptophan-aspartic acid 40 amino acids motif</i>
WDR5	<i>WD repeat-containing protein 5</i>
WHSC1	<i>Wolf-Hirschhorn syndrome candidate 1</i>

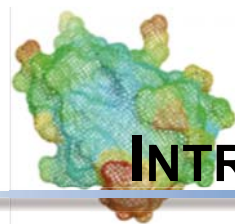
La nomenclatura común para las modificaciones de las histonas es:

- a) El nombre de la histona (H1, H2A, H2B, H3, H4, y variantes)
- b) Abreviatura del aminoácido modificado siguiendo el código IUPAC de una letra.
- c) La posición en la que se produce la modificación desde el extremo N-terminal.
- d) El tipo de modificación (me: metil, ph: Fosfato, ac: Acetil, ub: Ubiquitina)
- e) En el caso de las metilaciones, representa el número de grupos metilo añadidos (1: monometilación, 2: dimetilación, 3: trimetilación)

Por ejemplo:

H₃_(a)K_(b)4_(c)me_(d)1_(e)

H3K4me1 denota una monometilación de la cuarta lisina de la histona H3.



INTRODUCCIÓN

INTRODUCCIÓN

*“No encontraras trabas en el camino
si se mantiene elevado tu pensamiento y es exquisita
la emoción que toca el espíritu y el cuerpo”.*

1. Base molecular de las diferencias fenotípicas

En Febrero de 2001, el *International Human Genome Sequencing Consortium* (IHGSC) publicaba los primeros resultados generales del proyecto de secuenciación masiva del genoma humano en la revista *Nature*¹. Los datos presentados recogían casi 3.000 millones de pares de bases de ADN que cubrían más del 90% del genoma humano. Dos años más tarde, el IHGSC anunciaba la consecución del Proyecto Genoma Humano y publicaba una secuencia que alcanzaba una cobertura del 99% del genoma humano con un 99.99% de exactitud². Comenzaba así, de forma “oficial”, la denominada era post-genómica, cuyo objetivo declarado era obtener los frutos de los proyectos genoma desarrollados hasta el momento. A la luz de este objetivo surgió la genómica comparativa, una estrategia poderosa en la que la bioinformática jugaba un papel vital. Dicha estrategia se basa en el análisis y la comparación de genomas de diferentes organismos, con el objetivo de conocer los fundamentos moleculares de la evolución de las especies y determinar así la función de los genes y regiones no codificantes que integran los genomas.

En aquel momento, autores como Chris P. Ponting creían que la creciente disponibilidad de genomas daría lugar a una nueva disciplina conocida como zoología genómica³, en la cual las diferencias entre genomas explicarían la aparente distinción tanto fisiológica como morfológica entre especies, especialmente si estas diferencias eran el resultado de una adaptación evolutiva. Sus estudios comparativos de los genomas de humano y ratón proporcionaron información sobre la evolución adaptativa de algunos grupos de genes involucrados en placentación, sistema inmunitario, olfato y reproducción. Aunque estos grupos de genes permiten explicar diferencias obvias en la fisiología humana y la del roedor, no explican las grandes diferencias fenotípicas entre ambas especies.

Durante los siguientes años se publicaron diferentes estudios comparativos entre el genoma humano y el de diferentes organismos modelo que han divergido notablemente respecto al linaje humano como: ratón⁴, rata⁵, gallina⁶ y pez⁷. El acceso a los diferentes genomas revolucionó nuestra comprensión de cómo se organiza la información genética, y cómo ésta ha evolucionado con el paso del tiempo, y mejoró sustancialmente la anotación funcional de los diferentes genomas. Globalmente, la mayoría de estos estudios reflejan que más del 95 % del genoma evoluciona de forma neutra, y que sólo un 5 % parece estar bajo presión selectiva, lo que indica que existen otros elementos genómicos (como regiones no traducidas, elementos reguladores, elementos estructurales cromosómicos, etc.) que podrían desempeñar una función biológica. La comparación entre el genoma humano y el de los organismos modelo citados anteriormente refleja que comparten la mayoría de sus genes, e incluso estos se encuentran en el mismo orden que en su ancestro común. Las diferencias más notables observadas entre genomas se deben a la expansión de ciertas familias de genes que afectan sistemas fisiológicos que han sido objeto de innovación en un determinado linaje⁴⁻⁷.

En Julio de 2005 *The Chimpanzee Sequencing and Analysis Consortium* publicó en la revista *Nature* una secuencia preliminar del genoma del chimpancé común⁸. Este hecho marcó un hito en los estudios comparativos entre especies, ya que permitía cotejar el genoma humano con el de su pariente vivo más cercano. A pesar de las diferencias acumuladas desde que humanos y chimpancés divergimos de nuestro ancestro común (unos treinta y cinco millones de cambios puntuales en nucleótidos, cinco millones de eventos de inserción/delección y varios reordenamientos cromosómicos) los patrones evolutivos de los genes codificantes en humano y chimpancé están fuertemente correlacionados y dominados por la fijación de alelos neutros o ligeramente deletéreos⁸. Es decir, la mayoría de cambios evolutivos se deben a la deriva neutra, y sólo una pequeña porción de la variación interespecífica total está causada por cambios adaptativos. Estos datos, junto con los estudios comparativos mencionados anteriormente, sugieren que no toda la variación fenotípica observada entre especies puede explicarse por la información contenida en el ADN. Ello nos lleva de forma natural a preguntarnos: ¿qué otros factores participan en las diferencias fenotípicas observadas entre el ser humano y los organismos modelo?

1.1. Estudios con gemelos monocigóticos. El papel del entorno.

Diferentes estudios muestran que además del mensaje contenido en el ADN, existen otras fuentes de regulación de la expresión génica que contribuyen a la variabilidad fenotípica⁹. Los estudios comparativos de gemelos monocigóticos son un ejemplo clásico en genética, ya que al tratarse de individuos genéticamente idénticos han sido utilizados frecuentemente para demostrar el papel que desempeñan los factores ambientales (como por ejemplo, la nutrición o agentes químicos/físicos) en la determinación del fenotipo^{10,11}. Aunque la herencia genética influye claramente en el riesgo de padecer una determinada enfermedad o en la adquisición de un determinado fenotipo, importantes discordancias entre gemelos monocigóticos reflejan que la secuencia de ADN por sí sola no puede determinar completamente el fenotipo¹². Por ejemplo, la mitad de gemelos monocigóticos, en el caso de la esquizofrenia, no comparten la enfermedad¹³.

Sin embargo, incluso las variaciones de entorno muestran limitaciones a la hora de explicar ciertas diferencias entre individuos genéticamente idénticos. Por ejemplo, en un estudio pionero realizado con 100 parejas de gemelos monocigóticos y dicigóticos¹⁰, se compararon las evaluaciones físicas y psicológicas de gemelos que habían crecido compartiendo el mismo entorno con aquellos que habían sido separados desde la infancia, con el objetivo de aislar la contribución del entorno en el fenotipo. Las correlaciones entre ambos tipos de gemelos monocigóticos para la mayoría de variables observadas fueron casi idénticas (patrones electroencefalográficos, presión sistólica, frecuencia cardíaca, respuesta electrodérmica, cuestionarios de personalidad, etc.). El hallazgo de dicho estudio es, en general, consistente con otros estudios en gemelos monocigóticos. Por ejemplo, la úlcera péptica, una enfermedad con un claro componente ambiental (exposición a *Helicobacter pylori*) ha sido objeto de estudio basado en gemelos monocigóticos. Los resultados reflejan que los factores genéticos son realmente importantes en la vulnerabilidad, sin embargo, la comparación de gemelos que han crecido separados con aquellos que han compartido el mismo entorno, sugiere que compartir el mismo entorno tienen un efecto leve en la susceptibilidad a sufrir úlcera péptica¹⁴.

1.2. Estudios con animales genéticamente idénticos. El tercer componente

En una serie de experimentos diseñados para explorar la contribución relativa de los genes, el ambiente y otros factores en la adquisición de un determinado fenotipo¹⁵, se pudo demostrar que la mayoría de variabilidad no genética no era debida al entorno. Las fuentes de variación genética fueron minimizadas utilizando animales consanguíneos, pero la reducción de la variabilidad genética no redujo la cantidad de variación observada en los fenotipos. El entorno controlado del laboratorio no tuvo mayor efecto en la variabilidad interindividual al compararlo con la variabilidad del entorno natural. Solo un 20-30% de la variabilidad podría atribuirse a factores ambientales, y el 70 – 80% restante de variación no genética se adjudicó a un “tercer componente” cuya base molecular se desconocía¹⁵.

El clonaje de mamíferos se ha podido llevar a cabo recientemente en varias especies¹⁶⁻¹⁸, y también supone una oportunidad para diferenciar los efectos genéticos del resto de factores que pueden influir en la adquisición de un determinado fenotipo. Los clones fruto de estos experimentos, a pesar de tener el mismo genoma que el donador, exhiben una amplia variedad de anomalías fenotípicas que, obviamente, no pueden atribuirse a causas genéticas y tampoco pueden atribuirse al entorno en el que viven, ya que estos primeros ejemplares de animales clonados son supervisados en entornos estrictamente controlados¹².

Normalmente, cuando se descartan las fuentes de variación genética, se considera que los factores ambientales son la fuente del resto de la variación. Sin embargo, en los ejemplos antes mencionados, la variación persiste a pesar de tratarse de ambientes invariantes. Entre el componente ambiental y el componente genético, existe un “tercer componente” epigenético, que explicaría cómo los factores ambientales pueden dar lugar a múltiples fenotipos partiendo de la misma base genética. Los estudios de Fraga y colaboradores¹¹ sugieren que los factores externos y/o internos pueden tener un impacto notable en el fenotipo al alterar la expresión génica mediante cambios en los perfiles epigenéticos.

Los mecanismos moleculares encargados de esta regulación epigenética de la expresión génica podrían ser uno de los factores que contribuyen a las diferencias fenotípicas entre especies, particularmente en aquellos casos en los que ambas

especies comparten una base genética muy similar, como en el caso del hombre y el chimpancé. A lo largo de esta introducción profundizaremos con algo más de detalle en las bases moleculares del control epigenético de la expresión génica.

2. ¿Qué entendemos por epigenética?

Aunque el término “epigenética” se ha utilizado mucho durante los últimos años, Conrad Hal Waddington introdujo el término a principios de los años 40¹⁹. Él definió epigenética como: “*the branch of biology which studies the causal interactions between genes and their products which bring the phenotype into being*”. En el sentido original de esta definición, la epigenética se refiere a todas las vías moleculares que modulan la expresión del genotipo dando lugar a un determinado fenotipo. Con el paso de los años, y el rápido crecimiento de la genética, el significado de la palabra se ha ido restringiendo.

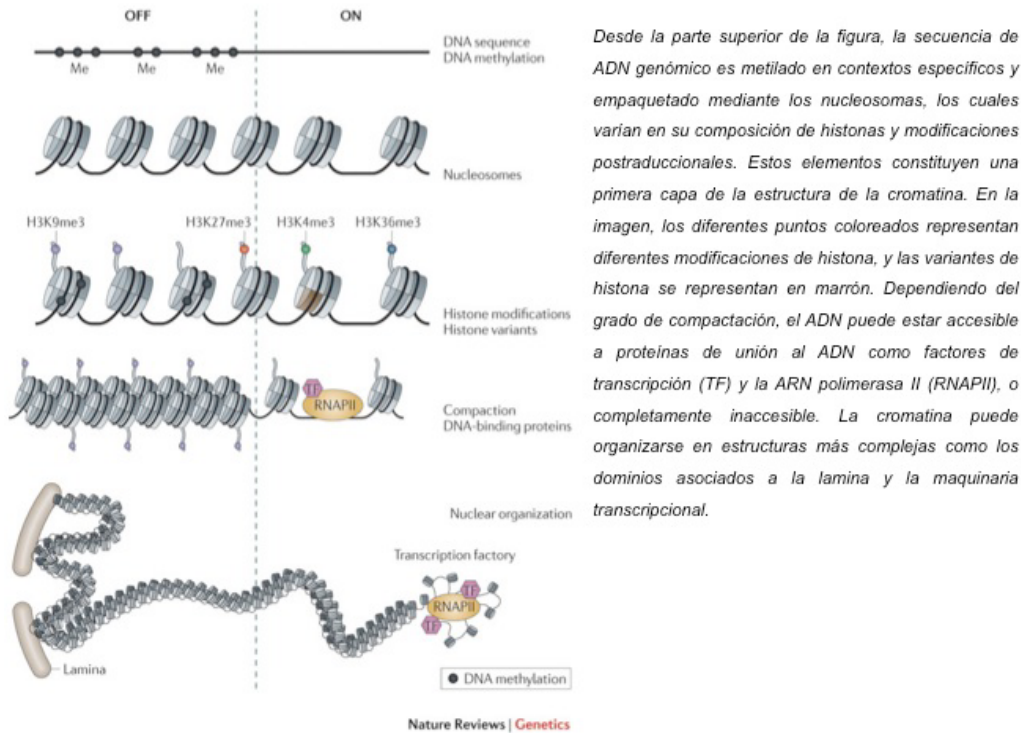
El uso contemporáneo más extendido del término epigenética, y el que se empleará a lo largo de la presente tesis, hace referencia al **estudio de los cambios heredables y transitorios en la expresión génica que no implican un cambio en la secuencia primaria de ADN**²⁰. Debido a la similitud con el término “genética”, se han generado algunos paralelismos muy extendidos y que con mayor o menor frecuencia también se emplearán a lo largo del texto. Por ejemplo, en relación al concepto de “genoma”, el epigenoma se refiere al estado epigenético global de una célula. Por otro lado, la frase “código genético” ha sido adaptada a “código epigenético”, y se emplea para referirse al conjunto de características epigenéticas que, en su combinación, generan diferentes fenotipos en diferentes células¹².

2.1. Mecanismos epigenéticos de control de la expresión génica

Existen distintos mecanismos que, bien de forma individual, bien en combinación, generan distintos niveles de regulación de la expresión génica. Por ejemplo, los **factores de represión** que reconocen y se unen a motivos de secuencia en *cis*, como es el caso de la proteína TBP que inhibe componentes de la maquinaria de transcripción²¹; o bien, los **tiempos de replicación del ADN**, que permiten distinguir entre regiones de replicación tardía, que correlacionan con la represión génica a nivel de genes individuales; o regiones de replicación temprana, donde se agrupan los genes *housekeeping*²². Desde el punto de vista epigenético, los principales

mecanismos de regulación de la actividad transcripcional de los genes son²³: (i) remodelación dependientes de ATP de la estructura de la cromatina, (ii) introducción de las variantes de histonas, (iii) metilación del ADN, y (iv) modificaciones covalentes de las histonas (**Figura 1**).

Figura 1 | Capas de organización de la cromatina en el núcleo de las células de mamífero



2.1.1. Remodelación dependiente de ATP

Algunos complejos, como el conocido *SWI/SNF ATP-dependent remodelling complex*, facilitan el movimiento de las histonas a lo largo de la molécula de ADN, lo que conlleva la compactación o descompactación de la cromatina. Estudios *in vitro* han demostrado que los complejos SWI/SNF producen un desensamblaje de la estructura del nucleosoma dependiente de ATP²⁴. La principal función de estos complejos es regular la transcripción génica, activándola o reprimiéndola dependiendo del grado de compactación de la cromatina; sin embargo, también están implicados en el ensamblaje de la cromatina²⁵.

2.1.2. Variantes de histonas

Hace años que se sabe de la existencia de las variantes de histonas H2A y H3. Sin embargo, ha sido durante los últimos años cuando se han relacionado con estados alternativos de la cromatina. Estas variantes se depositan durante la replicación y sustituyen a las histonas canónicas, distinguiendo diferentes estados de la cromatina

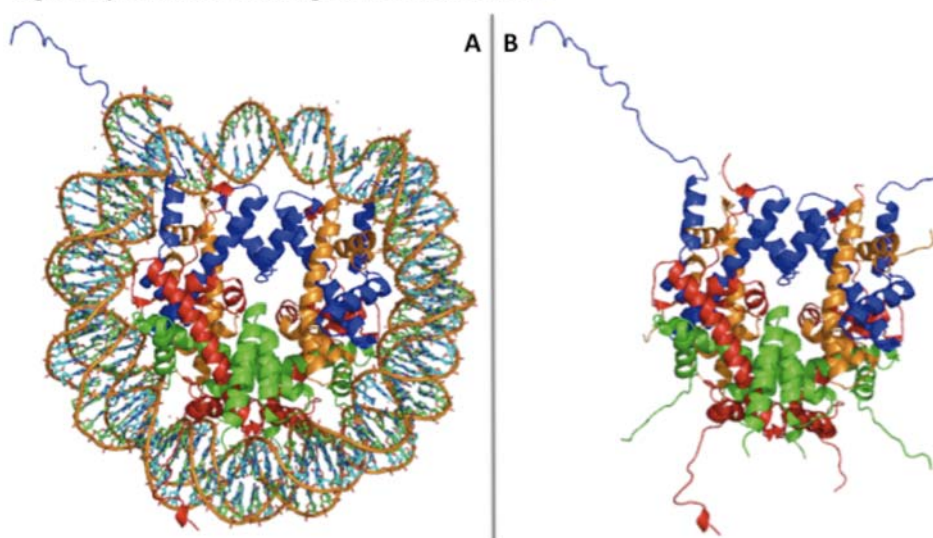
en los centrómeros, en el cromosoma X inactivo en mamíferos y en *loci* transcripcionalmente activos²⁶. Recientemente, se ha observado que la cromatina activa está enriquecida en ciertas variantes de histonas. Por ejemplo, la variante de histona H3.3 es un marcador de cromatina activa, y el desplazamiento de la histona H3 por la variante H3.3 proporciona un mecanismo dinámico para la rápida activación de la cromatina, por ejemplo, cuando ha sido previamente inactivada por metilación de la lisina 9²⁷.

2.1.3. Metilación del ADN

La metilación del ADN es una de las modificaciones epigenéticas mejor caracterizada, que desempeña un papel crítico en el control de la actividad génica y en la arquitectura del núcleo de la célula. La metilación del ADN se establece y se mantiene mediante la familia de las metiltransferasas de histona (DNMTs). En humanos, se produce en las citosinas que preceden a guaninas -el comúnmente denominado dinucleótido CpG-, se establece durante las primeras etapas del desarrollo y es una marca epigenética heredable²⁸. Dicho dinucleótido no está distribuido de forma aleatoria en el genoma, sino que existen regiones ricas en él, conocidas como islas CpG. Aproximadamente el 60% de los genes tienen su región promotora ubicada en estas islas²⁹ que normalmente están desmetiladas, excepto en subgrupos particulares de genes cuyas islas CpG aparecen metiladas en células tumorales^{30,31}. La metilación del ADN está relacionada con el silenciamiento génico e introduce un control adicional de ciertos genes específicos de tejido, inhibiendo su expresión de forma diferencial, dependiendo del tipo celular o tejido³². Por ejemplo, los genes SERPIN, miembros de la familia de inhibidores de proteasas, o los genes MAGE de la línea germinal, están silenciados en la mayoría de tejidos excepto en tumores malignos³³. La impronta genómica asegura la expresión monoalélica de un gen o dominio genómico determinado mediante metilación del ADN³⁴, y una reducción de dosis génica similar se produce en la inactivación del cromosoma X en hembras³⁵. La hipermetilación de secuencias repetitivas del genoma probablemente previene la inestabilidad cromosómica, limitando las translocaciones y disrupciones génicas originadas por la reactivación de secuencias de ADN transponibles³⁶. Células con defectos espontáneos en las DNMTs³⁷ o disrupciones experimentales de dichas enzimas presentan anomalías nucleares³⁸.

La relación entre la metilación del ADN y la estructura de la cromatina se establece a través de determinadas modificaciones covalentes de las histonas, concretamente mediante la deacetilación de las histonas y la metilación de H3K9. Existen evidencias de que la metilación del ADN influye en el patrón de modificaciones de las histonas. Por ejemplo, los componentes de la maquinaria de la metilación del ADN reclutan complejos represores que contienen deacetilasas de histonas²³.

Figura 2 | Estructura cristalográfica del nucleosoma



A. OCTÁMERO DE HISTONAS + 146 bp DE ADN. La estructura cristalográfica del nucleosoma determinada mediante rayos X (código PDB: 1AOI) muestra como se ensambla el octámero de histonas y como la superhélice de 146 bases de ADN se organizan alrededor de este. Un tetrámero de histonas formado por dos histonas H3 (azul) y dos histonas H4 (naranja), junto dos dímeros de las histonas H2A (rojo) y las histona H2B (verde) integran el octámero de histonas. **B. OCTÁMERO DE HISTONAS.** En esta figura se ha retirado la superhélice de ADN para facilitar la visualización de las colas de las histonas, y como su estructura desordenada pasa sobre y entre los giros de ADN para quedar expuestas a las diferentes maquinarias biológicas

2.1.4. Modificaciones covalentes de las histonas

Las histonas están sujetas a una amplia variedad de modificaciones del extremo N-terminal y, con menor frecuencia, del extremo C-terminal y el dominio globular. La presente tesis se centra exclusivamente en el estudio de diferentes aspectos de este mecanismo epigenético, y por lo tanto se dedicará un mayor esfuerzo a su descripción. Para ello primero repasaremos los aspectos estructurales y funcionales de las histonas, para después describir los diferentes tipos de modificaciones que pueden experimentar.

3. Cromatina y procesos epigenéticos: el rol de las histonas y sus modificaciones.

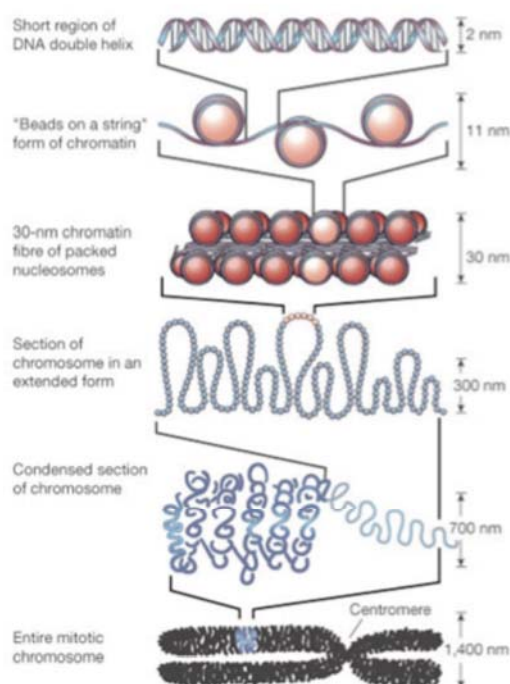
El núcleo de la célula eucariota, de apenas 10 μm de diámetro, contiene genomas cuyo tamaño puede ir de los 12 millones de pares de bases en *Saccharomyces cerevisiae* hasta los 6.000 millones de pares de bases en *Homo sapiens* (cuya fibra de ADN se extiende más allá de los 2 metros³⁹). Ello es posible debido a que los genomas eucariotas están compactados y organizados en el interior del núcleo celular, en una estructura nucleoproteica estable conocida como cromatina, cuya unidad básica es el nucleosoma (**Figura 2**).

El nucleosoma está formado por un octámero de histonas (un tetrámero de histonas H3 y H4, flanqueado por dos dímeros de histonas H2A y H2B) sobre el cual se enrollan 146 pares de bases de ADN⁴⁰ (**Figura 2A**). Las histonas son pequeñas proteínas básicas que se encuentran en el núcleo de la célula eucariota, con una región central altamente estructurada que forma el corazón del nucleosoma, conocida como dominio de plegamiento de la histona; sus extremos N y C terminales sobresalen de la superficie del nucleosoma, tal y como muestra la estructura cristalográfica del nucleosoma⁴¹ (**Figura 2B**) y no adoptan una estructura definida. La principal función de las histonas es estructural ya que permiten reducir el volumen que ocupa el ADN para poderlo ubicar en el interior del núcleo celular, además de reforzar y aumentar su manejabilidad durante la metafase.

El nucleosoma es el primer nivel de compactación de la cromatina; en presencia de histonas H1, o *linker*, u otras proteínas asociadas a la cromatina, los nucleosomas pueden condensarse en fibras de 30 nm de diámetro⁴². El sistema mediante el cual la fibra de 30 nm se empaqueta hasta llegar al nivel máximo de compactación, el cromosoma, es aún poco conocido⁴³ (**Figura 3**). Dependiendo del grado de compactación, podemos distinguir varios tipos de cromatina que se corresponden con diferentes niveles de actividad transcripcional⁴⁴: (i) **euromatina**, o también denominada cromatina activa, que coincide con regiones menos compactadas del genoma y que son, por lo tanto, transcripcionalmente activas, y (ii) **heterocromatina**, que debido a su elevada compactación corresponde a regiones transcripcionalmente inactivas, por lo que también recibe el nombre de cromatina inactiva.

Desde el punto de vista de la función génica, una consecuencia muy importante de la compactación de la cromatina es que impide el acceso de las maquinarias biológicas encargadas de procesos tales como la replicación, la transcripción, la reparación y la recombinación del ADN^{45,46}. Por lo tanto, para que todos estos procesos ocurran en el lugar y momento adecuados, se han desarrollado una serie de mecanismos moleculares que actúan sobre la estructura de la cromatina, regulando su grado de compactación.

Figura 3 | Organización del ADN en la estructura de la cromatina



La subunidad fundamental de la cromatina es el nucleosoma, que consiste en 147 pares de bases de ADN enrolladas alrededor del octámero de histonas. Cada nucleosoma está conectado con su vecino a través de un pequeño fragmento de ADN (10–80 bp) y esta cadena de polinucleosomas se pliega formando una fibra de 30 nm de diámetro. Esta fibra queda estabilizada mediante la unión de la histona H1 a cada nucleosoma y al ADN linker. Finalmente, el nivel superior de compactación del ADN es el cromosoma mitótico. Los mecanismos que permiten que estas fibras alcancen niveles mayores de compactación en el interior del núcleo celular son desconocidos

3.1. Regulación de la estructura de la cromatina

Son varios los mecanismos centrados en las histonas que inducen cambios locales y transitorios en la compactación de la cromatina, facilitando así el acceso al ADN, y contrarrestan el efecto represivo que ejerce la cromatina: (i) modificaciones covalentes de la región amino terminal de las histonas⁴⁷⁻⁵⁰, (ii) alteración de la estructura nucleosomal mediante enzimas capaces de utilizar la energía obtenida de la hidrólisis del ATP²⁴ e (iii) incorporación de variantes de histonas⁵¹. La presente tesis se centra en el primero de ellos, sobre el que a continuación proporcionamos información más detallada.

3.1.1. Modificaciones covalentes de las histonas

Estas modificaciones se concentran principalmente en el dominio N-terminal de las histonas H2A, H2B, H3 y H4, también conocido como cola de la histona. Este dominio es un fragmento de secuencia de unos 40 aminoácidos que sobresale considerablemente de la superficie del nucleosoma (**Figura 2B**). A diferencia del dominio globular, no adopta una estructura tridimensional definida⁴¹, y por ello, a pesar de comprender aproximadamente el 25% de la masa de la histona⁵² no es visible en la estructura cristalográfica del nucleosoma⁵³. Por este motivo, el dominio N-terminal de la histona no parece contribuir a la formación del núcleo del nucleosoma⁴¹, pero sí proporciona una superficie expuesta que facilita la interacción con otras proteínas^{41,54}. Todo ello le convierte en una diana ideal para el almacenamiento de la información epigenética. Esta importancia funcional viene confirmada indirectamente por la elevada conservación de su secuencia aminoácida entre diferentes especies⁵⁵ y, directamente, por estudios en los que se muestra que la delección del extremo N-terminal en las histonas H3 y H4 resulta letal⁵⁶.

Tabla 1 | Modificaciones covalentes de las colas de las histonas y su función asociada

Modificación	Residuo	FUNCIÓN ASOCIADA			
		Transcripción	Reparación	Replicación	Condensación
Acetilación	K-ac	X	X	X	X
Metilación	K-me1	X	X		
	K-me2				
Metilación	K-me3	X			
	R-me1				
	R-me2a				
Fosforilación	Rme2s	X	X		X
	S-ph, T-ph				
Ubiquitinación	K-Ub	X	X		
Sumoilación	K-Su	X			
Ribosilación de ADP	E-Ar	X			
Deiminación	R -> Cit	X			
Isomerización de Prolinas	P-cis -> P-trans	X			

Hasta el momento se han identificado 8 tipos de modificaciones covalentes de las histonas (**Tabla 1**), entre las que acetilación, metilación y fosforilación son las más estudiadas y mejor caracterizadas. Tales modificaciones se han correlacionado con varias actividades nucleares, incluyendo replicación, ensamblaje de la cromatina y

3.1.2.1. Efectos directos

Los efectos directos de las modificaciones se refieren a aquellos que alteran las propiedades físicas del nucleosoma, como podrían ser los contactos con el ADN, la movilidad, el tamaño, la conformación o la estabilidad, o bien, alteran la capacidad del nucleosoma para formar estructuras de orden superior, a través de la modulación de los contactos internucleosomales⁶³. Más concretamente, se cree que determinadas modificaciones de los residuos de la histona podrían afectar al grado de compactación de la cromatina⁶⁴⁻⁶⁶ mediante su efecto sobre las interacciones electrostáticas entre la cola de la histona y el ADN^{52,67-69}. Estas interacciones son debidas a la basicidad de la cola de la histona, que permite una estrecha asociación al esqueleto ácido del ADN. Pero ciertas modificaciones, como por ejemplo la acetilación de las lisinas, reducirían la carga global del extremo N-terminal de la histona, debilitando su interacción con el ADN. A su vez, este cambio de afinidad entre el ADN y las histonas se traduciría en un cambio en la conformación del nucleosoma que facilitaría el paso a las diferentes maquinarias moleculares. El modelo alternativo⁷⁰ postula que la acetilación de la histona podría modificar la conformación de la cromatina, al aumentar el contenido de α -hélices del extremo N-terminal. Ello permitiría que esta modificación regulase el estado conformacional de la cromatina.

Hay que señalar que los mecanismos directos no gozan de una aceptación total, ya que estudios biofísicos de nucleosomas completos indican que estos mecanismos producen efectos mínimos en la conformación de la cromatina⁷¹⁻⁷³. Además, algunos estudios indican que la acetilación de la histona no tiene un efecto inhibitor en la interacción histona-ADN⁷⁴ ni un efecto estimulante en el acceso de los factores de transcripción⁷⁵.

3.1.2.2. Efectos indirectos: el código de histonas

Las evidencias señaladas al final de la sección anterior, junto con un número creciente de estudios publicados en los últimos años, sugieren que la cola de las histonas funcionaría como plataforma de señalización para el reclutamiento de otros reguladores de la cromatina involucrados en diferentes eventos celulares.

A principios de la década de los noventa se propuso que las modificaciones covalentes de las histonas podían actuar como lugares de unión, capaces de reclutar ciertas proteínas involucradas en definir la eucromatina o la heterocromatina^{76,77}. Este

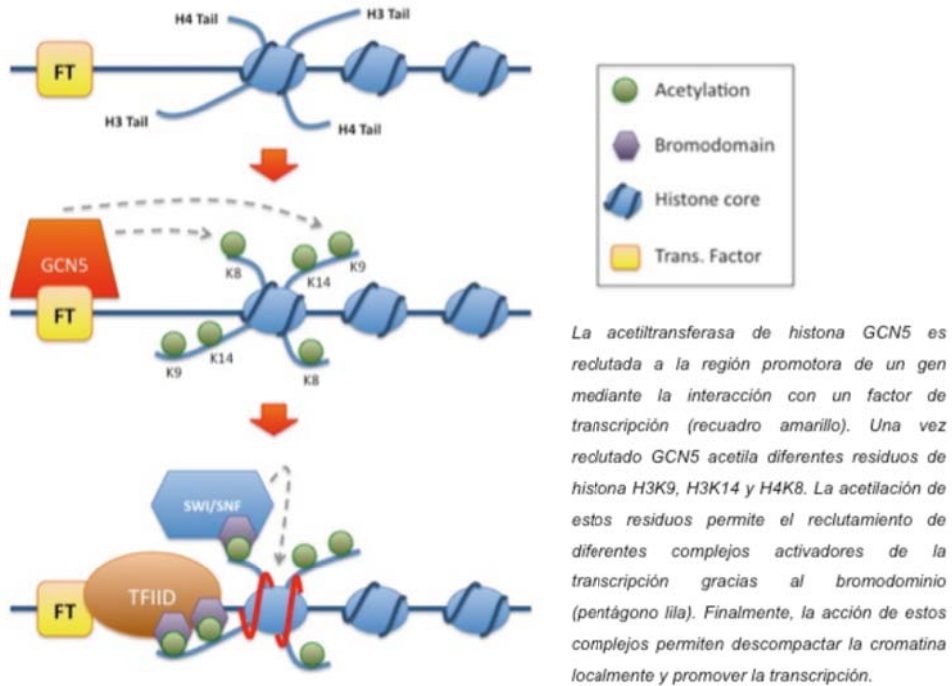
modelo defiende que las modificaciones de las histonas son “leídas” por una serie de módulos proteicos denominados efectores, que facilitan una serie de eventos celulares mediante el reclutamiento de la maquinaria biológica a la cromatina^{78,79} (**Figura 5**). Diferentes estudios apoyan este modelo, al desvelar la existencia de una gran diversidad de módulos proteicos con la capacidad de reconocer diferentes modificaciones post-traduccionales^{60,71,79}. Procesos celulares tan variados como transcripción⁸⁰, replicación⁸¹, pluripotencia de las células madre⁸², silenciamiento génico^{80,83}, inactivación del cromosoma X⁸⁴, reparación del ADN⁸⁵, apoptosis^{86,87}, carcinogénesis^{88,89}, herencia epigenética⁹⁰ y programas de expresión génica durante el desarrollo^{91,92}, requieren la interacción de un determinado módulo efector con la correspondiente modificación de la cromatina.

Bryan Turner fue el primero en postular que la acetilación de la cola de las histonas, podría actuar como una marca epigenética mediante la cual el grado de actividad génica se mantiene entre divisiones celulares⁹³. Posteriormente, Strahl & Allis llevaron dicha hipótesis un paso más allá, sugiriendo que distintos patrones de modificaciones actúan secuencialmente, o en combinación para formar un “código de histonas” leído por otras proteínas encargadas de desencadenar distintos eventos celulares⁶². La interpretación de dicho código permitiría la remodelación del nucleosoma o de las fibras de cromatina, bien directamente, o a través del reclutamiento de otras proteínas. Estos cambios en la estructura de la cromatina o en la acción de los complejos multiproteicos reclutados acabarían determinando el estado transcripcional de los genes^{61,62,94} (**Figura 5**).

En los últimos años, un número creciente de datos experimentales han contribuido a comprobar y mejorar diferentes aspectos de la hipótesis del código de histonas^{55,95}. Se ha confirmado que el código de histonas involucra tanto modificaciones independientes, como combinaciones de ellas. Por ejemplo, la modificación H3K9me produciría silenciamiento^{96,97}, mientras que la misma modificación acompañada de H3K4me y H3K20me actuaría como una señal de activación⁹⁸. Otro aspecto que ha cobrado importancia sería la idea de que las modificaciones tanto en una como en diferentes histonas pueden ser interdependientes. Dicho de otra forma, modificaciones en un residuo pueden ocasionar la modificación de otro residuo en *cis* o, más sorprendentemente, en *trans*. Por ejemplo, se ha demostrado que la metilación en

H3K4 impide el anclaje del NURD e inhibe la metilación de H3K9, impidiendo la incorporación de señales de silenciamiento⁹⁹.

Figura 5 | Modelo de cooperación de las enzimas modificadoras de la cromatina y los dominios de reconocimiento (ilustrado en el caso del bromodominio)



3.1.3. Función de las modificaciones covalentes de las histonas

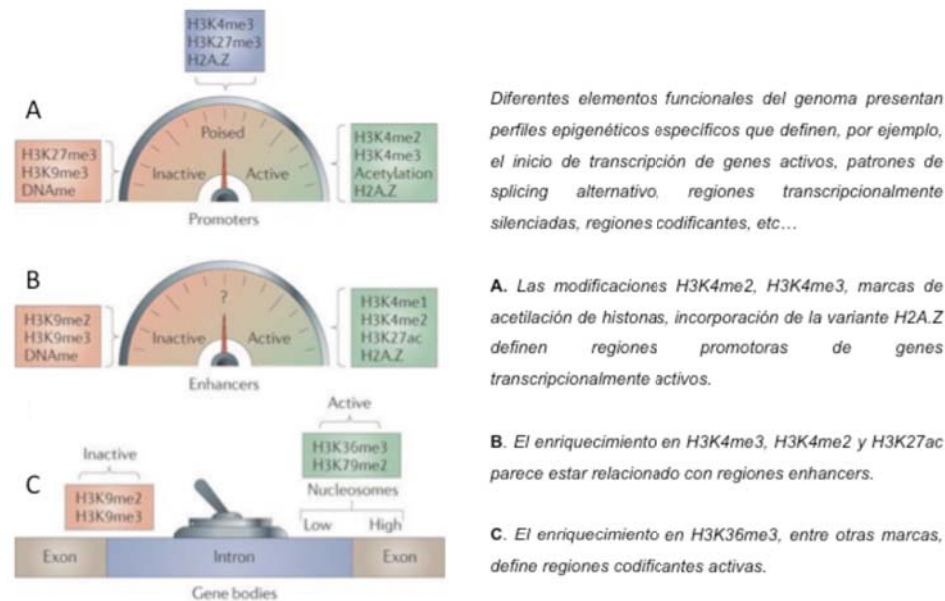
Actualmente, existe un vacío en la comprensión mecánica de cómo las diferentes modificaciones de las histonas se traducen en una determinada función biológica como la transcripción por ejemplo, o cómo pueden ser utilizadas para propagar la información contenida en el genoma de una generación a la siguiente, como en el caso de la herencia epigenética. Sin embargo, estudios a nivel del genoma completo¹⁰⁰, combinando ensayos de inmunoprecipitación de la cromatina con análisis de microarrays (ChIP-chip) o secuenciación (ChIP-seq), han permitido observar que las diferentes modificaciones de las histonas no se distribuyen de forma aleatoria a lo largo del genoma, sino que existen diferentes elementos funcionales enriquecidos en estas marcas covalentes. Éstos se enumeran a continuación.

3.1.3.1. Promotores

Los promotores de los mamíferos pueden ser clasificados en base a características de su secuencia. La mayoría de ellos coinciden con regiones con un alto contenido CG y del dinucleótido CpG, las denominadas islas CpG. Estos han sido denominados promotores con un alto contenido CpG (HCPs), en contraste con los promotores con un bajo contenido en CpG (LCPs). Estas dos clases de promotores presentan diferentes patrones de modificaciones covalentes y diferentes modos de regulación¹⁰¹. Estudios iniciales de ChIP-chip en células de mamíferos asociaron la modificación de histona H3K4me3 con el inicio de transcripción de genes activos^{82,102}. Sin embargo, otros estudios revelaron que H3K4me3 es una modificación que se localiza en todos los promotores HCP, independientemente de su estado transcripcional^{101,103}. Los *loci* que presentan la modificación H3K4me3 están acompañados de otras características que definen un estado más accesible de la cromatina como: (i) la presencia de marcas de acetilación en las histonas, (ii) incorporación de las variantes de histona H3.3 y (iii) H2A.Z e hipometilación del ADN^{31,104-108} (**Figura 6A**). Algunos investigadores han explicado la relación entre la modificación H3K4me3 y los promotores HCP mediante el reconocimiento físico de dinucleótidos CpG demetilados por el dominio CXXC en complejos metiltransferasa H3K4¹⁰⁹, por ejemplo el complejo SET1.

A diferencia de los promotores HCP, los LCP parecen inactivos por defecto y, la mayoría de ellos, carecen de las marcas H3K4me3 (o H3K4me2) en células madre embrionarias y en varias líneas celulares diferenciadas^{31,101}. Algunos estudios sugieren que algunos promotores LCP, portadores de H3K4me2, permanecen generalmente inactivos pero son inducidos durante la diferenciación celular, cuando se produce un cambio en el estado de metilación de H3K4me2 a H3K4me3¹¹⁰. Los promotores reprimidos muestran patrones únicos de modificaciones epigenéticas que reflejan diferentes modos de silenciamiento génico: H3K27me3, la modificación asociada a los represores Polycomb; H3K9me3, asociada a heterocromatina constitutiva; y metilación del ADN (**Figura 6A**). En células madre embrionarias, aproximadamente un 20% de promotores HCP presentan la modificación H3K27me3. Estos promotores, que también son portadores de la modificación H3K4me3, han sido denominados bivalentes ya que muestran propiedades de cromatina activa e inactiva¹¹¹. Estos promotores controlan genes que permanecen silenciados en células pluripotentes, pero que pueden inducirse rápidamente o quedar silenciados de forma estable, dependiendo de la fase de desarrollo.

Figura 6 | Distribución de las modificaciones covalentes de las histonas en diferentes elementos funcionales del genoma



3.1.3.2. Genes

Estudios recientes han puesto de manifiesto que los patrones de la cromatina pueden distinguir entre exones e intrones, e incluso desempeñar un papel en determinar los patrones de *splicing*. Las principales modificaciones observadas en regiones transcritas son H3K36me3^{101,112} y H3K79me2⁸⁰. También se ha puesto de manifiesto que exones expresados tienen un fuerte enriquecimiento en H3K36me3^{113,114} comparado con los intrones, además de H2BK5me1, H4K20me1 y H3K79me1¹⁰⁶ (**Figura 6C**). Han sido propuestos diferentes mecanismos que relacionan estas modificaciones de las histonas con diferentes patrones de *splicing* alternativo. Algunos autores han especulado que los nucleosomas posicionados sobre exones podrían facilitar los eventos de *splicing* al ralentizar el avance de la RNAPII¹¹⁵ o incluso que existe una relación directa entre las modificaciones de las histonas y la maquinaria de *splicing*¹¹⁶.

Este último estudio muestra que las modificaciones de histonas H3K36me3, H3K4me3, H3K4me1 y H3K27me3 varían a lo largo del gen FGFR2 en células epiteliales y mesenquimales. Los autores de dicho estudio han sugerido un modelo en el que las modificaciones de las histonas son interpretadas por la maquinaria de *splicing* mediante la proteína MORF4L1 y el regulador de *splicing* PTBP1. Hay que indicar, sin embargo, que estudios recientes de regulación génica en células únicas

muestran que el *splicing* alternativo no es forzosamente co-transcripcional¹¹⁷, en cuyo caso el papel de las histonas y sus modificaciones quedaría un tanto en entredicho.

3.1.3.3. Enhancers

Los *enhancers* son elementos de ADN que reclutan factores de transcripción, RNAPII y reguladores de la cromatina que estimulan la transcripción en promotores distales¹¹⁸. Los perfiles de modificaciones de las histonas parecen jugar un papel relevante en la identificación de dichos elementos. Además de modificaciones específicas, los *enhancers* están ocupados preferentemente por proteínas de unión a secuencias específicas de ADN⁸¹ y coactivadores como p300¹¹⁹. Heintzman y colaboradores ha identificado un enriquecimiento de H3K4me y la eliminación de H3K4me3 como una marca para la señalización de *enhancers* en células humanas¹²⁰ (**Figura 6B**). Los *enhancers* también parecen estar relacionados con enriquecimiento en H3K27ac, H2BK5me1, H3K4me2, H3K9me1, H3K27me1 y H3K36me1¹⁰⁵. Estos perfiles de modificaciones de las histonas podrían contribuir a proporcionar accesibilidad al genoma o reflejarían la proximidad física de los *enhancers* con la maquinaria que mantiene la cromatina activa en zonas promotoras¹¹⁸.

4. Reconocimiento de las modificaciones de las histonas

Como hemos visto en secciones anteriores, la hipótesis del código de histonas proporciona una visión mecanística del rol funcional de las modificaciones de las histonas⁶². Las responsables de realizar estas modificaciones son unas enzimas (también conocidas como reguladores epigenéticos) con la capacidad de introducir o eliminar modificaciones covalentes en el extremo N-terminal de las histonas¹²¹.

Tabla 2 | Familias de enzimas modificadoras de la cromatina y residuos modificados relacionados

Familia	Enzima	Residuo Modificado
Acetiltransferasas (HAT)	HAT1	H4K5, H4K12
	CBP/P300	H3K14, H3K18, H4K5, H4K8, H2AK5, H2BK12, H2BK15
	PCAF/GCN5	H3K9, H3K14, H3K18
	TIP60	H4K5, H4K8, H4K12, H4K16, H3K14
	HB01 (ScESA1, SpMST1)	H4K5, H4K8, H4K12
	ScSAS3	H3K14, H3K23
	ScSAS2 (SpMST2)	H4K16

	ScRTT109	H3K56	
Deacetilasas (HDAC)	SirT2 (ScSir2)	H4K16	
Metiltransferasas (HMT)	SUV39H1	H3K9	
	SUV39H2	H3K9	
	G9a	H3K9	
	ESET/SETDB1	H3K9	
	EuHMTasa/GLP	H3K9	
	CLL8	H3K9	
	SpClr4	H3K9	
	MLL1	H3K4	
	MLL2	H3K4	
	MLL3	H3K4	
	MLL4	H3K4	
	MLL5	H3K4	
	SET1A	H3K4	
	SET1B	H3K4	
	ASH1	H3K4	
	Sc/Sp SET1	H3K4	
	SET2 (Sc/Sp SET2)	H3K36	
	NSD1	H3K36	
	SYMD2	H3K36	
	DOT1	H3K79	
	Sc/Sp DOT1	H3K79	
	Pr-SET 7/8	H4K20	
	SUV420H1	H4K20	
	SUV420H2	H4K20	
	SpSet9	H4K20	
	EZH2	H3K27	
	RIZ1	H3K9	
	Desmetilasas lisina	LSD1/BHC110	H3K4
		JHDM1a	H3K36
		JHDM1b	H3K36
JHDM2a		H3K9	
JHDM2b		H3K9	
JMJD2A/JHDM3A		H3K9, H3K36	
JMJD2B		H3K9	
JMJD2C/GASC1		H3K9, H3K36	
Metiltransferasas Arginina	JMJD2D	H3K9	
	CARM1	H3R2, H3R17, H3R26	
	PRMT4	H4R3	
	PRMT5	H3R8, H4R3	
Kinasas (Ser/Thr)	Haspin	H3T3	
	MSK1	H3S28	
	MSK2	H3S28	
	CKII	H4S1	
Ubiquitinasas	Mst1	H2BS14	
	Bmi/Ring1A	H2AK119	
	RNF20/RNF40	H2BK120	
Isomerasas de Prolinas	ScFPR4	H3P30, H3P38	

En 1996, dos grupos descubrieron enzimas modificadoras de la cromatina relacionadas por homología de secuencia con reguladores transcripcionales. Uno de ellos aisló una deacetilasa de histonas (HDAC) de mamífero que compartía un 60% de

identidad de secuencia con Rpd3, un represor transcripcional de levadura¹²². Simultáneamente, el otro grupo purificó una acetiltransferasa de histona (HAT) de *Tetrahymena thermophila* con una elevada homología con el adaptador transcripcional de levadura, Gcn5¹²³. La identificación de estas enzimas supuso un hito en la comprensión de las funciones biológicas asociadas a las modificaciones de las histonas, ya que fue la primera evidencia directa de una relación entre éstas y la regulación transcripcional. El descubrimiento de Gcn5 y Rpd3 fue el punto de partida para la posterior identificación y caracterización de otras familias de HATs y HDACs, así como otras clases de enzimas modificadoras de la cromatina, incluyendo quinasas^{124,125}, metiltransferasas específicas de lisinas y argininas^{48,126,127} (HMTs), arginina deaminasas^{128,129}, ubiquitinasas¹³⁰, deubiquitinasas¹³¹⁻¹³³, desmetilasas específicas de lisinas y argininas (HDMs)¹³⁴⁻¹³⁶ (**Tabla 2**). Estudios previos habían implicado a algunas de estas proteínas en la regulación de la transcripción, o de otras funciones biológicas, resaltando aún más la correlación existente entre las modificaciones de las histonas y los procesos dependientes de la cromatina.

La identificación de enzimas modificadoras de la cromatina ha centrado mucho interés científico durante la pasada década. La mayoría de las modificaciones que éstas introducen son dinámicas, y se ha identificado la correspondiente enzima que elimina la modificación introducida. De esta forma, podemos distinguir dos tipos de enzimas modificadoras de la cromatina: (i) las encargadas de añadir modificaciones covalentes, los denominados escritores del código de histonas (*writers*), y (ii) las encargadas de retirar dichas modificaciones, los denominados borradores del código de histonas (*erasers*).

La mayoría de estas actividades reguladoras de la cromatina están integradas por subunidades proteicas que presentan módulos enzimáticos y no-enzimáticos⁵⁵. Estos dominios participan en la incorporación, eliminación o reconocimiento de las modificaciones covalentes⁵⁵. El resto de esta sección se dedicará a la descripción de los módulos no enzimáticos en torno a los que gira principalmente esta tesis.

Los dominios de interacción con la cromatina desempeñan un importante papel en la interpretación biológica de las modificaciones de las histonas, al permitir el reclutamiento selectivo de las proteínas efectoras a determinados *loci*. En los últimos años, muchas líneas de investigación se han centrado en el estudio de la interacción

histona-dominio efector⁷¹ por lo que actualmente tenemos a nuestra disposición un número apreciable de estructuras de dominios en sus formas apo y/o holo, en las que los ligandos son normalmente péptidos de histona portadoras de diferentes modificaciones covalentes. Los resultados de estos estudios han vertido luz en diferentes aspectos de la interacción histona-dominio efector^{137,138}.

Tabla 3 | Dominios de reconocimiento de modificaciones de histonas

Dominio		Modificación
ROYAL FAMILY	Bromodominio	LYS acetiladas (H3 y H4)
	Cromodominio	H3K9me2/3, H3K27me2/3
	Barril Cromo	H3K36me2/3
	Tudor	Rme2s
	Doble/Tándem Tudor	H3K4me3, H4K20me1/2/3, Kme2
	MBT	H4K20me1/2, H1K26me1/2, H3K4me1, H3K9me1/2
	PWWP	H4K20me
PHD finger	H3K4me2/3, H3K9me3, H3K36me3, H3K14Ac	
WD40	H3R2/K4me2, (R, Sph, Tph)	
14-3-3	H3S10ph, H3S28ph (Sph, Tph)	
BRCT	H2AX S139ph (Sph, Tph)	
SANT	Colas de histona sin modificar	
SWIRM	ADN	

Como hemos visto anteriormente, las modificaciones más estudiadas de las histonas son la acetilación y la metilación, por lo que las familias de dominios mejor caracterizadas y que suelen encontrarse en proteínas asociadas a la cromatina son la familia de los bromodominios¹³⁹ y la de los cromodominios¹⁴⁰, involucradas, respectivamente, en el reconocimiento de lisinas acetiladas y metiladas. Sin embargo, la capacidad de reconocer residuos de histona modificados no es exclusiva de estos dominios. Existen otros dominios comunes en factores asociados a la cromatina con dicha propiedad: Tudor¹⁴¹, PHD¹⁴², SANT¹⁴³, SWIRM¹⁴⁴, MBT¹⁴⁵, WD40¹⁴⁶, PWWP¹⁴⁷, 14-3-3 y BRCT¹⁴⁸ (**Tabla 3**).

Una de las publicaciones de esta tesis (**publicación 1**) revisa de forma exclusiva los bromodominios, cromodominios y dominios SANT. Debido a la fecha de su publicación y al rápido crecimiento de la investigación en este área, sentimos la necesidad de actualizar su contenido. Así, los siguientes apartados contienen material

complementario a dicha publicación, actualizando la información sobre los dominios descritos en ella, y añadiendo aquellos nuevos dominios descritos desde la fecha de su aparición y más relevantes para la comprensión del resto de la presente tesis.

En las siguientes secciones se describen los diferentes dominios, agrupados de acuerdo con el residuo modificado que reconocen, distinguiendo básicamente entre acetilación y metilación. Al final se añade un grupo heterogéneo de dominios que reconocen diferentes modificaciones, y que están peor caracterizados. A fin de reducir la numeración de los epígrafes o secciones, la descripción de los dominios seguirá (cuando sea posible) una estructura común no numerada para todos ellos: una introducción con datos sobre su descubrimiento/caracterización/extensión, y otro sobre su estructura y mecanismos de reconocimiento de la histona.

4.1. Dominios que reconocen marcas de acetilación

Hasta hace poco se pensaba que los dominios encargados del reconocimiento de lisinas acetiladas eran únicamente los bromodominios. Sin embargo, recientemente se ha sugerido que el doble dominio PHD *finger* (PHD12) de la proteína DPF3b tiene la capacidad de unirse a histonas H3 y H4 acetiladas¹⁴⁹.

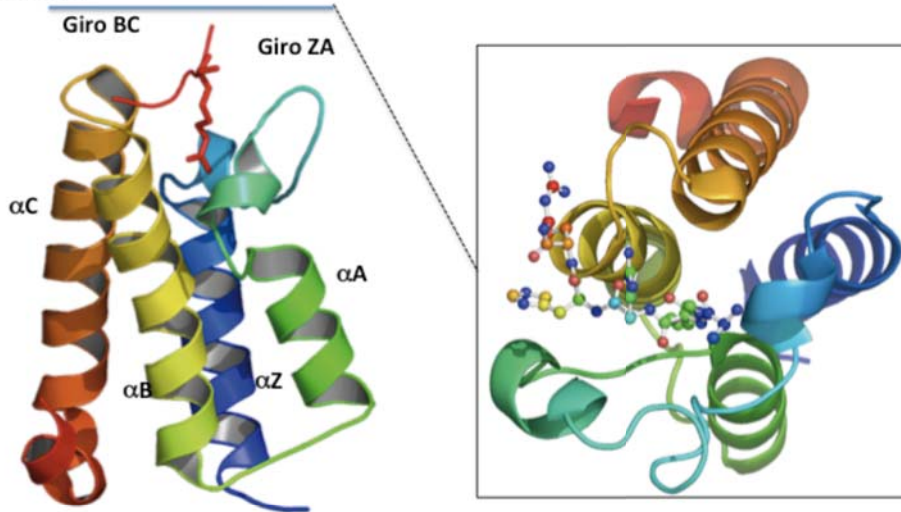
4.1.1. Bromodominio

El bromodominio fue el primer módulo proteico identificado que mostraba selectividad por lisinas acetiladas en el extremo amino terminal de las histonas H3 y H4¹⁵⁰. Fue identificado inicialmente en la proteína Brahma en *Drosophila melanogaster*¹⁵¹. Dicho dominio está presente en algunos reguladores transcripcionales asociados a la cromatina, incluyendo HATs nucleares como Gcn5p, p300/CBP y SAGA; complejos remodelantes de la cromatina como SWI/SNF y RSC; y algunos factores de transcripción como TAFII250¹⁵².

Actualmente, se conocen cuatro versiones de la estructura del bromodominio: el bromodominio del coactivador transcripcional PCAF determinado mediante RMN (el homólogo de *Saccharomyces cerevisiae* Gcn5p)¹⁵⁰, el bromodominio de la proteína humana GCN5 determinado mediante RMN¹³⁹, el doble bromodominio de TAF_{II}250 resuelto mediante cristalografía de rayos X¹⁵³ y el bromodominio de la proteína Gcn5

de levadura en complejo con la histona H4 acetilada determinada mediante cristalografía de rayos X¹⁵⁴ (Figura 7).

Figura 7 | Bromodominio de GCN5 en complejo con un péptido de histona H4 acetilado (H4K16ac)



A. La estructura del bromodominio consiste en un haz de cuatro hélices α (αZ , αA , αB , αC) conectados por dos giros (giro ZA y BC) que conforman un canal estrecho, profundo, de naturaleza hidrofóbica y revestido de residuos aromáticos muy conservados. Ahí el grupo acetilo de la histona H4 (H4K16Ac) es anclado mediante un puente de hidrógeno al grupo amida de una asparagina que está al final del canal. B. Vista superior del canal formado por las cuatro hélices α del bromodominio. Código PDB: 1E6I

Estudios de la estructura cristalográfica del bromodominio han sugerido que dos residuos de tirosina conservados (uno en el giro ZA y el otro en el extremo C-terminal de la hélice α_B) contribuyen a la cavidad hidrofóbica encontrada en la mayoría de los bromodominios¹⁵⁵, aunque no son necesariamente los determinantes del reconocimiento de las lisinas acetiladas.

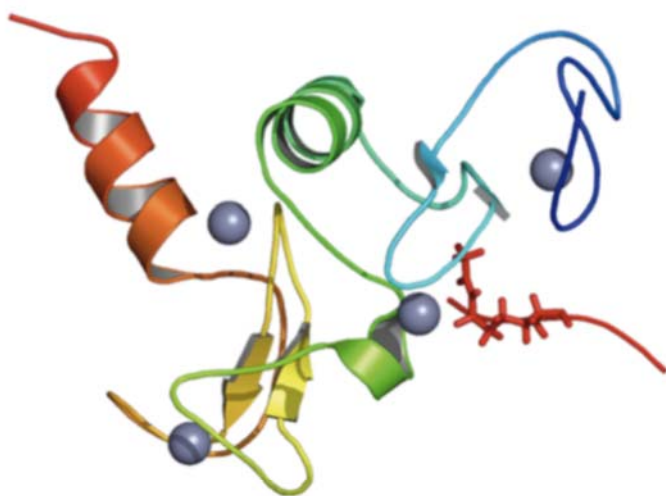
4.1.1.1. Doble bromodominio

Como ocurre con otros dominios encargados del reconocimiento de modificaciones de histonas, pueden encontrarse proteínas con múltiples copias del bromodominio. Por ejemplo, la estructura de TAF1 (TAF_I,250) muestra dos bromodominios dispuestos en forma de U¹⁵³. El plegamiento de los dominios se produce de forma independiente, y las cavidades que unen la lisina acetiladas quedan separadas por 25Å, lo que equivale a una distancia de 7-8 residuos en el péptido de histona. Estudios realizados con péptidos de histona H4 con diferentes residuos acetiladas revelan que el doble

bromodominio de TAF1 se une con mayor afinidad a péptidos di- o tetra- acetilados en K5/K12, K8/K16, K5/K8/K12/K16 que al péptido de histona H4 monoacetilado.

Otro ejemplo de múltiples copias del bromodominios lo encontramos en la proteína remodelante de la cromatina PB1 que contiene 6 copias en tándem en su extremo N-terminal¹⁵⁶. La presencia de múltiples copias del bromodominio permitiría reconocer patrones de acetilación específicos reclutando a PB1 en la cromatina¹⁵⁶.

Figura 8 | Doble PHD finger de la proteína humana DPF3 unido a un péptido de histona acetilado (H3K14ac)



La estructura del complejo muestra que cada uno de los módulos PHD presentan el plegamiento típico del dominio PHD finger: una lámina β formada por dos hebras β antiparalelas seguida por una pequeña hélice α en el extremo C-terminal. Este plegamiento queda estabilizado mediante dos átomos de zinc coordinados por el motivo Cys4-His-Cys3. En amarillo y cian se muestran las dos hebras β antiparalelas, y en naranja y verde las dos hélices α . Las esferas representan los átomos de zinc. En la imagen se muestra el reconocimiento del péptido de histona modificado (en rojo) por parte de uno de los dominios PHD. Código PDB: 2KWJ

4.1.2. PHD – Doble dominio PHD de la proteína DPF3b

Según los recientes análisis estructurales de Zeng y colaboradores¹⁴⁹, el dominio *PHD finger* en tándem descubierto en la proteína humana DPF3b se convierte en la primera alternativa al bromodominio en el reconocimiento de lisinas acetiladas. Concretamente, el dominio PHD1-PHD2 reconoce H3K14ac, aunque también reconoce preferentemente péptidos acetilados de la histona H4¹⁴⁹. La proteína DPF3b actúa asociada con el complejo remodelante de la cromatina BAF en el inicio de la transcripción durante el desarrollo muscular y del tejido cardíaco.

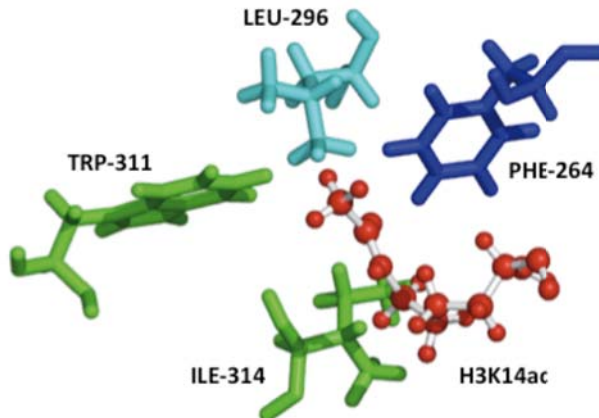
Según Zeng y colaboradores, el dominio en tándem PHD1-2 usa un mecanismo conservado como alternativa al reconocimiento de los grupos acetil por los

bromodominios. El PHD1-2 finger utiliza el motivo Asp263-Phe264 para el reconocimiento del grupo N-acetil de H3K14ac, mientras el bromodominio utiliza el motivo conservado Tyr-Asn para interactuar con H4K20Ac, por ejemplo, el bromodominio de CBP utiliza Tyr1167-Asn1168.

4.2. Dominios que reconocen marcas de metilación

Desde el descubrimiento de la primera metiltransferasa de histona (Suv39h) en el año 2000, la metilación de las histonas se ha transformado en una de las áreas de investigación más activa en la biología de la cromatina¹⁵⁷. A pesar de que tanto los residuos de arginina como los de lisina pueden ser metilados, poco se sabe acerca del reconocimiento de residuos de arginina metilados. Sin embargo, existe un gran número de publicaciones que describen módulos proteicos con capacidad de unirse a lisinas metiladas, entre los que el mejor caracterizado es el cromodominio.

Figura 9 | Reconocimiento de la modificación H3K14ac mediante el dominio PHD1



La modificación (K14Ac) así como algunos residuos colindantes, quedan intercalados a lo largo de un surco hidrofóbico situado en la lámina β de PHD1, donde la modificación queda envuelta por los residuos Phe264, Leu296, Trp311 y Ile314, una característica estructural común entre los dominios que reconocen lisinas metiladas. Los colores de los residuos implicados en el reconocimiento de K14Ac, se corresponden con las estructuras representadas en la figura anterior.

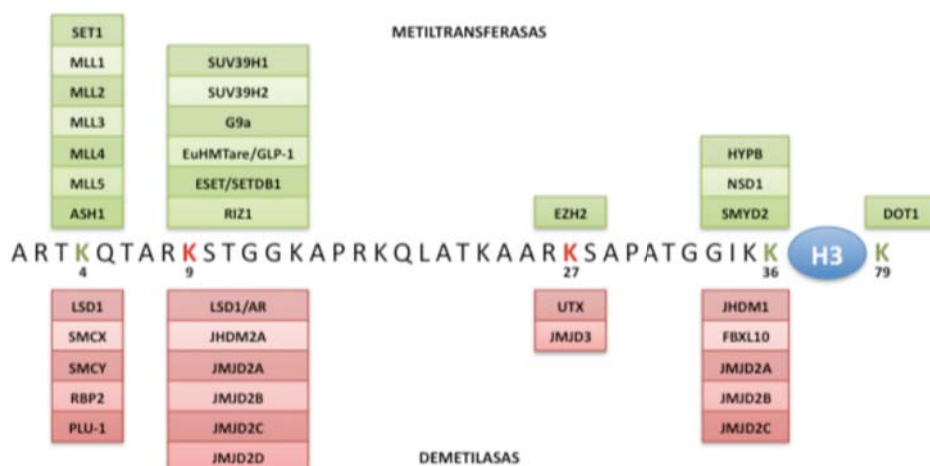
Básicamente, existen dos grandes grupos estructurales que han convergido evolutivamente hacia la unión de lisinas metiladas: la superfamilia denominada *Royal Family*¹⁵⁸ (que comprende los dominios Cromo, Tudor, MBT y PWWP) y los dominios PHD *finger*^{159,160}. Los distintos plegamientos de las proteínas que pertenecen a la *Royal Family* descienden de un ancestro común con la capacidad conservada de unirse a sustratos metilados¹⁵⁸. Estudios estructurales del reconocimiento de lisinas

metiladas sugieren que los módulos encargados del reconocimiento de esta modificación (Cromodominio, Tudor, MBT, PHD, etc.) utilizan mecanismos generales que les confieren especificidad hacia distintos niveles de metilación. A continuación, se describen dichos dominios, y posteriormente se enumeran brevemente las características comunes entre ellos.

4.2.1. Cromodominio

En 1991, Paro y colaboradores, identificaron una pequeña región común entre la proteína HP1 y PC de *Drosophila melanogaster*. Ambas proteínas juegan un papel importante en la modificación de la estructura de la cromatina y, por este motivo, a la región identificada se la denominó *chromatin organization modifier domain (chromodomain)*¹⁶¹.

Figura 10 | Familias de enzimas involucradas en la metilación de determinados residuos de la histona H3



Actividades enzimáticas involucradas en la metilación y desmetilación de los principales residuos modificados de la histona H3 (K4, K9, K27, K36 y K79). En color verde se representan las marcas de histonas activadoras mientras que los residuos de color rojo representan marcas represivas. En la parte superior se muestran las metiltransferasas de histona y en la parte inferior las demetilasas de histona.

En la figura 10, se muestran las principales familias de enzimas que participan en la metilación y desmetilación de determinados residuos de la histona H3. La metilación de las lisinas 9 y 27 de la histona H3 genera lugares de unión para los cromodominios de las proteínas HP1 y Polycomb, respectivamente. Por ejemplo, en la célula eucariota, la metilasa de histona SUV39H1 y la proteína HP1 interaccionan funcionalmente para reprimir la transcripción en regiones heterocromáticas¹⁶². H3K9 es metilada por SUV39H1 y sus homólogos⁴⁵, creando así un lugar de unión para el

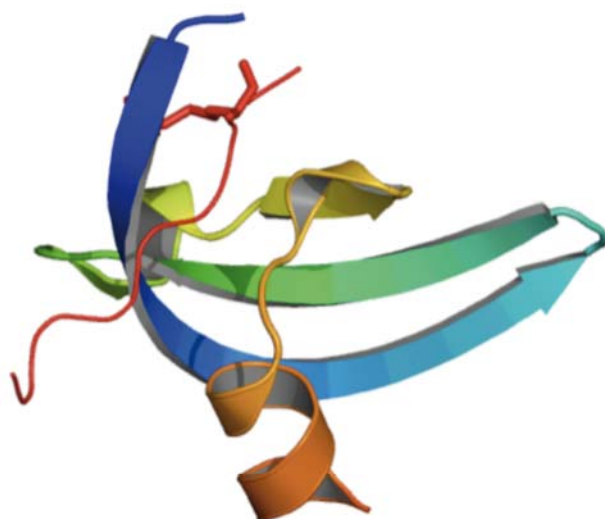
cromodominio de HP1^{93,94}, que muestra una gran especificidad por esta modificación⁹³. Experimentos de inmunofluorescencia *in situ* muestran la co-localización de H3K9me y HP1 en regiones de heterocromatina de *Drosophila melanogaster*¹⁶³. De forma análoga, H3K27 es metilada por un complejo multiproteico integrado por *Enhancer of Zeste* y *Extra Sex Combs* (ESC) en *Drosophila melanogaster*, o EZH2 y EED en humano. Dicha modificación es reconocida por el cromodominio de PC y parece estar involucrada en la represión de genes homeóticos¹⁶⁴⁻¹⁶⁷ y en la inactivación del cromosoma X^{168,169}.

Tabla 4 | Proteínas humanas que presentan cromodominios en su secuencia¹⁷³

Familia	Función	Proteína
Chromodomain-Helicase-DNA-Binding Family (CHD)	Reguladores positivos y negativos de la transcripción	Chd1
		Chd2
	Remodelamiento del nucleosoma dependiente de ATP Regulador negativo de la transcripción Autoantígeno para dermatomiositis (CHD4)	CHD3 (Mi2a)
		CHD4 (Mi2b)
Histone Methyltransferase Family (HMT)	Podría regular genes involucrados en desarrollo	Chd5
	Regula el silenciamiento génico mediado por el establecimiento de la heterocromatina constitutiva en pericentro y telómero	SUV39H1
		SUV39H2
Heterochromatin protein 1 Family (HP1)	Represión de la transcripción en regiones de heterocromatina	HP1a
	Represión de la transcripción en heterocromatina y eucromatina	HP1b
	Represión de la transcripción en eucromatina	HP1g
Polycomb family (PC)	Forman parte de complejos multiproteicos y mantienen la represión epigenética de los genes en la eucromatina	Cbx2
		Cbx4
		Cbx6
		Cbx7
		Cbx8
Homólogos Msl-3	Involucrados en la regulación de la transcripción mediante la acetilación de la histona y la modificación de la estructura de la cromatina	Msl-3 like
		MRG-15
Histone Acetyltransferase	Activación de la transcripción	MYST1
		Tip60
Retinoblastoma Binding Protein-1 (RBBP1)	Juega un papel importante en la represión de los promotores dependientes de E2F mediante la interacción con la proteína Retinoblastoma.	RBBP1

La determinación de las estructuras del cromodominio de HP1 en complejo con H3K9me^{170,171} y el cromodominio de Polycomb unido a H3K27me¹⁷² (**Figura 11**) supuso un avance importante para la comprensión de la interacción entre la cola de la histona y el cromodominio. En concordancia con su elevada homología de secuencia, la estructura global de los cromodominios de HP1 y PC es muy similar¹⁷⁰ (los esqueletos de C- α de ambos dominios pueden alinearse con una RMSD de 1.1Å). Muchas características de la interacción entre HP1 y H3K9me³ son muy parecidas a las de la interacción entre PC y H3K27me³¹⁷². A pesar de ello los cromodominios de HP1 y PC se unen específicamente a diferentes *loci* del genoma¹⁷², este efecto también se da en cromodominios de diferentes proteínas (**Tabla 4**) que presentan la capacidad de interactuar de forma selectiva con lisinas metiladas específicas, e incluso, muestran especificidad por diferentes grados de metilación.

Figura 11 | Estructura cristalográfica del cromodominio de la proteína Polycomb unido a H3K27me₃



La estructura del cromodominio consiste en tres hebras β (β 1- β 3) que forman una lámina β antiparalela. Una hélice α (α A) en el extremo C-terminal se sitúa sobre uno de los bordes de la lámina β cerca de β 1. El péptido de la histona H3 (en rojo) es reconocido por una hendidura formada entre el giro que conecta β 3 y α A, y el segmento N-terminal hasta β 1. Código PDB: 1PDQ

4.2.1.1. Doble cromodominio

Algunos complejos represivos remodelantes de la cromatina contienen componentes como la subunidad ATPasa Mi-2/CHD del complejo NuRD¹⁷⁴, la cual presenta dos cromodominios. A diferencia de HP1 y Polycomb que utilizan un único cromodominio para unirse a sus respectivos residuos de histona H3 metilados, la disposición de dos cromodominios en tándem favorece su cooperación para interactuar con las colas de las histona dimetiladas¹⁷⁵.

Las proteínas CHD están compuestas por un doble cromodominio en la región N-terminal, una helicasa SWI2/SNF2 central y un dominio de unión a ADN en su región C-terminal, y regulan el ensamblaje de los nucleosomas en un proceso dependiente de ATP y su movilización a lugares transcripcionalmente activos¹⁷⁶. El doble cromodominio de Chd1p fue asociado originalmente al reconocimiento de la modificación H3K4me mediante experimentos realizados en levadura¹⁷⁷. Los detalles moleculares de la interacción entre H3K4me2/3 y el ortólogo humano de dicha proteína (CHD1) han sido caracterizados en estudios estructurales y biofísicos¹⁷⁵.

4.2.2. Dominios chromo barrel y chromoshadow

El dominio *chromo barrel* es similar al cromodominio de HP1/CBX pero sin la hélice α de la región C-terminal y con dos hebras β adicionales contribuyendo a la lámina β . Los dominios *chromo barrel* de MRG15 y Eaf3 (reguladores de la acetilación de histonas global) interaccionan con H3K36me2 y H3K36me3 con una afinidad moderada-baja mediante los tres residuos aromáticos conservados del cromodominio¹⁷⁸, a pesar de la presencia de la hebra β adicional que podría ocupar la posición del péptido. Como ejemplo de la importancia de los residuos cercanos al lugar de unión en la especificidad, un triptófano adyacente al tercer residuo aromático de la jaula del cromodominio es importante para la unión de la lisina metilada¹⁷⁹. Dicho triptófano se corresponde con un residuo aromático en las proteínas CBX, y con un aspartato o histidina en las proteínas HP1.

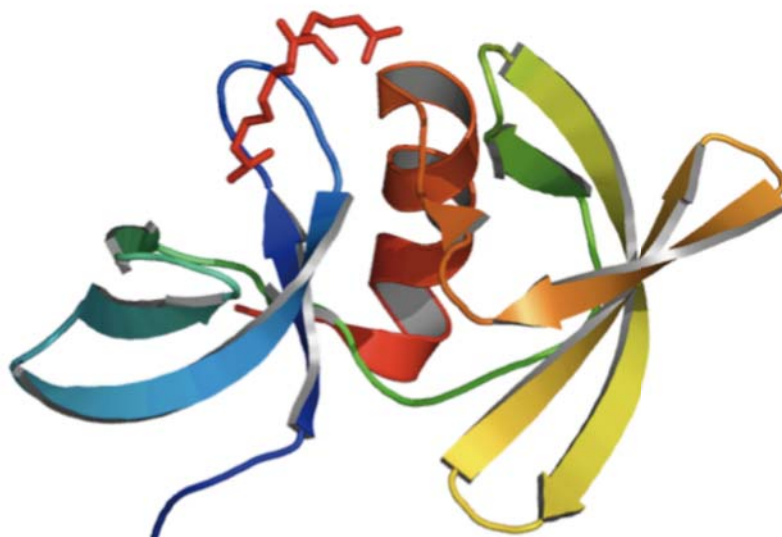
El dominio *chromoshadow*, parecido al cromodominio de HP1/CBX pero con una hélice adicional precediendo la hélice en la región C-terminal, y localizado en el lado C-terminal del cromodominio en proteínas HP1, carece de dos de los tres residuos aromáticos que coordinan la lisina metilada en el cromodominio, y no se une a histonas. *Chromoshadow* actúa como un dominio de dimerización de HP1 y, como tal, es responsable de la interacción proteína-proteína¹⁸⁰.

4.2.3. TUDOR

El dominio Tudor fue identificado por primera vez en la proteína Tud de *Drosophila melanogaster*¹⁸¹, que contenía 11 dominios Tudor. Inicialmente, se consideró que este dominio era el responsable de la unión al ARN¹⁴¹. Sin embargo, tras resolver y analizar su estructura tridimensional, se concluyó que, tanto las características estructurales,

como los ensayos de unión *in vitro*, sugerían que se trataba de un dominio de interacción con proteínas¹⁸². Concretamente, se observó la capacidad del dominio Tudor de unirse a pequeñas ribonucleoproteínas nucleares que mostraban residuos modificados post-traduccionamente y que contenían dimetilargininas¹⁸³.

Figura 12 | Estructura del dominio Tudor de 53BP1 en complejo con el péptido H4K20

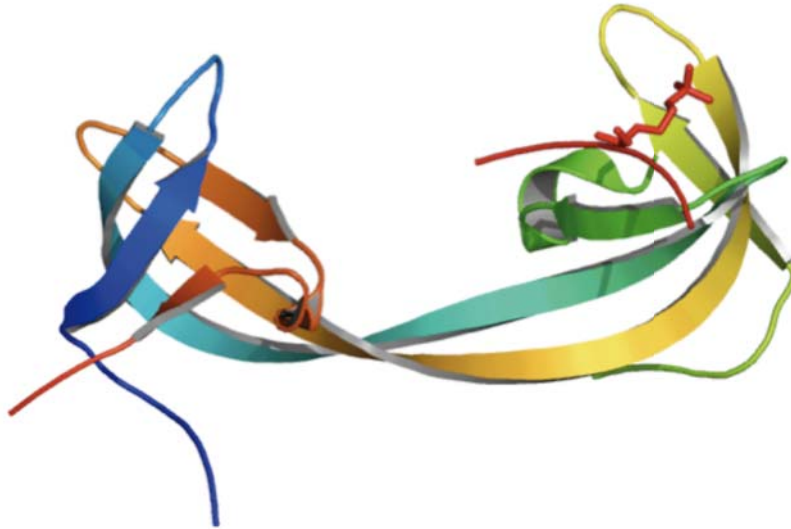


La estructura del dominio Tudor, muestra una lámina β antiparalela fuertemente doblada, formada por cuatro o cinco hebras β (β 1-5) dispuestas en un plegamiento similar a un barril β , cuyo borde inferior está formado por una larga y curvada hebra β (β 5) y cerrado por una interacción entre β 1 y la leu-142 de la hebra β 5. Las hebras β (β 1-4) están conectadas por giros cortos, mientras que las hebras β 4 y β 5 están conectadas mediante una pequeña hélice formando un ángulo de $\sim 90^\circ$. Código PDB: 2IG0

La interacción de los dominios Tudor de 53BP1 (**Figura 12**) y JMJD2A (**Figura 13**) con péptidos de histona metilados han sido las mejor caracterizadas¹⁸⁴, y son un claro ejemplo del papel de las modificaciones de las histonas en la reparación del ADN^{185,186} y en la represión de la transcripción¹⁸⁷, respectivamente. 53BP1 contiene dos dominios Tudor en tándem que se unen preferentemente a H4K20me2 y H4K20me1, pero no a péptidos sin modificar o H4K20me3¹⁸⁸. La estructura de los dominios Tudor en tándem de 53BP1 (**Figura 12**), en su estado apo- o en complejo con el péptido H4K20me2, revela las bases moleculares de este reconocimiento específico de bajos niveles de metilación de H4K20. El doble Tudor de 53BP1 forma dominios independientemente plegados, de los cuales el primero es el encargado del reconocimiento de las lisinas metiladas. El grupo metil de H4K20me2 queda posicionado en una jaula formada por cuatro residuos aromáticos y un aspartato, mientras que la arginina (H4R19) adyacente forma una interacción catión- π con el anillo de una tirosina adyacente y

también contacta con una fenilalanina del segundo dominio Tudor. Tal y como, se ha descrito anteriormente, la especificidad por lisinas con bajos niveles de metilación es atribuible a la formación de un puente de hidrógeno intermolecular entre H4K20me2 y el aspartato de la caja aromática, sumado a la repulsión estérica de la trilisina metilada.

Figura 13 | Estructura del doble dominio Tudor

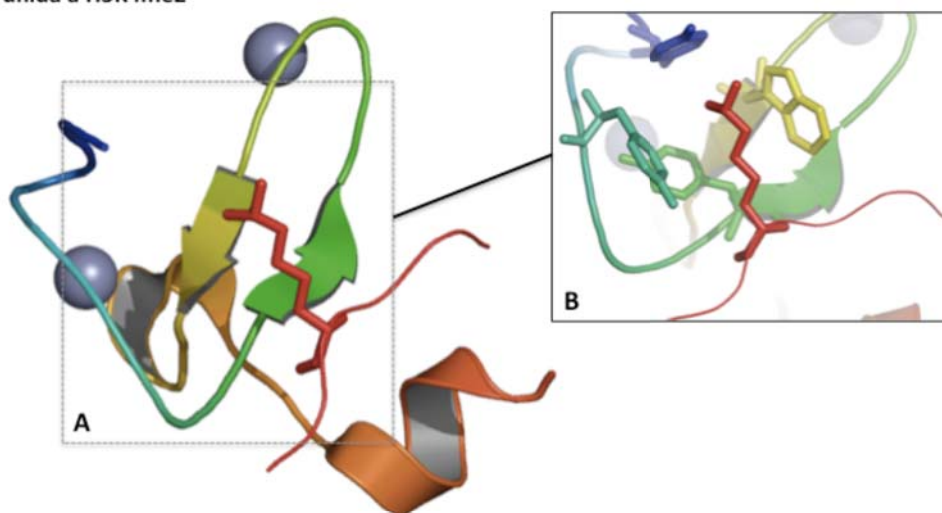


Los dos dominios Tudor quedan interdigitados, formando una estructura similar a una silla de montar, donde cada uno de los lóbulos se asemeja al dominio Tudor canónico de SMN. El dominio Tudor híbrido se forma por el intercambio de las hebras β 3 y 4, respecto al dominio Tudor canónico, mientras que las 2 largas hebras β unen los dos dominios individuales. Código PDB: 2GFA

JMJD2A es miembro de la superfamilia de las desmetilasas que presentan los dominios JmjC y JmjN, necesarios para su actividad desmetilasa, y dos dominios Tudor¹⁸⁹⁻¹⁹¹ (**Figura 13**). Experimentos de espectroscopia de masas aplicados al estudio de la metilación en las histonas, muestran que JMJD2A demetila H3K9me3 y H3K36me3^{189,191} (**Figura 10**).

La estructura descrita del doble Tudor de JMJD2A presenta la capacidad de unirse con la misma afinidad a H3K4me3 y H4K20me3. Esta interacción se produce con el segundo de los dominios Tudor, y la lisina metilada de cada péptido queda envuelta por residuos aromáticos^{192,193}. En el caso del péptido H3K4me3 y H3K420me3, la interacción queda estabilizada mediante la interacción entre un residuo de arginina de la cola de la histona y un aspártico del dominio Tudor. Por otro lado, los residuos N940, Y942 y T968 del dominio Tudor de JMJD2A forman puentes de hidrógeno intermoleculares de forma selectiva con algunos residuos de los péptidos H3K4me3 y H4K20me3¹⁸⁷.

Figura 14 | Estructura cristalográfica del dominio PHD finger de la proteína BPTF humana unida a H3K4me2



A. El dominio PHD finger está formado por dos hebras β que conectan los dos motivos zinc finger, y una pequeña hélice α ($\alpha 1$) en su extremo C-terminal. El residuo modificado se ancla en la superficie del dominio PHD sobre la lámina β antiparalela. B. El dominio PHD finger reconoce el grupo trimetilamonio de H3K4me3 mediante cuatro residuos aromáticos (dos de ellos situados en las dos hebras β) y la interacción queda estabilizada mediante interacciones π -catión. Código PDB: 2FSA

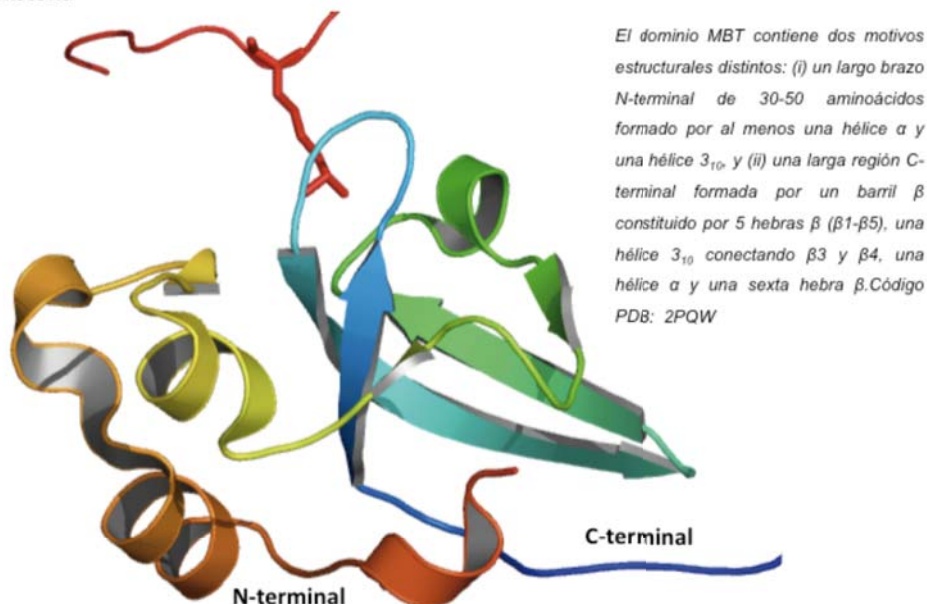
4.2.4. PHD finger

Estudios ChIP-chip del genoma han permitido establecer que las di- y trimetilación de H3K4 están asociadas con nucleosomas cercanos al promotor y a regiones codificantes de genes activos, respectivamente^{97,109}. Durante un tiempo, la conexión entre estas marcas de las histonas y la transcripción era una incógnita, hasta que evidencias estructurales identificaron al dominio PHD *finger* como el responsable del reconocimiento de H3K4me3, permitiendo el reclutamiento o la estabilización de determinados complejos^{159,160,194}. Recientemente, un estudio del dominio PHD *finger* de la subunidad TAF3 del factor de transcripción TFIID, ha permitido mejorar nuestro conocimiento acerca de la conexión entre activación transcripcional y la di- y trimetilación de H3K4¹⁹⁵. Los dominios PHD *finger* son pequeños módulos proteicos con pocos elementos de estructura secundaria formados por el segmento Cys₄-His-Cys₃ coordinado mediante dos iones zinc, y se encuentran en diferentes proteínas asociadas a la cromatina¹⁹⁶, como por ejemplo, en BPTF (**Figura 14**), Yng1p y ING2.

BPTF es la mayor subunidad del factor remodelante del nucleosoma (NURF), el cual estimula la transcripción *in vitro*¹⁹⁷. En su región N-terminal la proteína BPTF humana contiene un PHD *finger*, próximo a un bromodominio, con la capacidad de reconocer

H3K4me2 ($K_d = 5.0 \mu\text{M}$) y H3K4me3 ($K_d = 2.7 \mu\text{M}$) pero no la monometilación ni H3K4 sin modificar^{159,198}. La estructura del dominio PHD *finger* de BPTF en su forma libre y unido a H3K4me3 ha sido resuelta mediante RMN y cristalografía de rayos X.

Figura 15 | Estructura cristalográfica del dominio MBT de L3MBT1 unido al péptido de histona



4.2.5. MBT

Cuando Wismar y colaboradores clonaron el cDNA codificado en el gen L3MBT1 de *Drosophila melanogaster*, se dieron cuenta que contenía tres repeticiones de un motivo desconocido de ~100 aminoácidos¹⁹⁹. En *Drosophila* existen tres proteínas que contienen este motivo: Scm, una proteína Polycomb que participa en la represión de genes Hox, L(3)mbt y Sfmbt^{145,200,201}. Estas proteínas pertenecen al grupo Polycomb y están involucradas en el establecimiento y mantenimiento de un estado transcripcional reprimido de los genes de control del desarrollo, como los genes *Hox*.

Tan pronto como el dominio MBT (**Figura 15**) se relacionó con el control del desarrollo mediante proteínas PcG y la supresión de tumores a través de la proteína L(3)MBT, la estructura del dominio pasó a ser una pieza de vital importancia, lo que explica que actualmente dispongamos de 30 estructuras atómicas en la base de datos *Protein Data Bank* (PDB) que contienen el dominio MBT (**Tabla 5**).

Estudios *in vitro*, han demostrado que los dominios MBT tienen la capacidad de unirse a péptidos de histona con lisinas mono- y desmetiladas, discriminando aquellas lisinas que, o bien no están modificadas, o presentan trimetilaciones. En estructuras del dominio MBT junto con su ligando natural, se ha podido determinar que sólo una de las repeticiones del dominios MBT en una proteína es la encargada de acomodar la lisina metilada en su cavidad.

Tabla 5 | Resumen de las modificaciones reconocidas por el dominio MBT (adaptada de Bonasio, Lecona y Reinberg 2010)

Proteína	Ligando	Método	Referencia
hSCML2 (2MBT)	H3K4me ^a	NMR	Santiveri et al. <i>J Mol Biol</i> (2008)
	H3K9me ^a		
	H3K27me ^a		
	H3K36me ^a		
	H4K20me ^a		
hL3MBTL1 (3MBT)	H4K20me1/2	ITC	Min et al. <i>Nat Struc Mol Biol</i> (2007)
	H3K4me1/2		
	H3K9me1/2		
	H3K27me1/2		
	H3K36me1/2	FP	Li et al. <i>Mol Cell</i> (2007)
H4K20me1/2	ITC	Min et al. <i>Nat Struc Mol Biol</i> (2007)	
hL3MBT2 (4MBT)	H3K4me1	ITC	Guo et al. <i>NAR</i> (2009)
	H3K9me1/2		
	H3K27me1/2		
	H4K20me1/2		

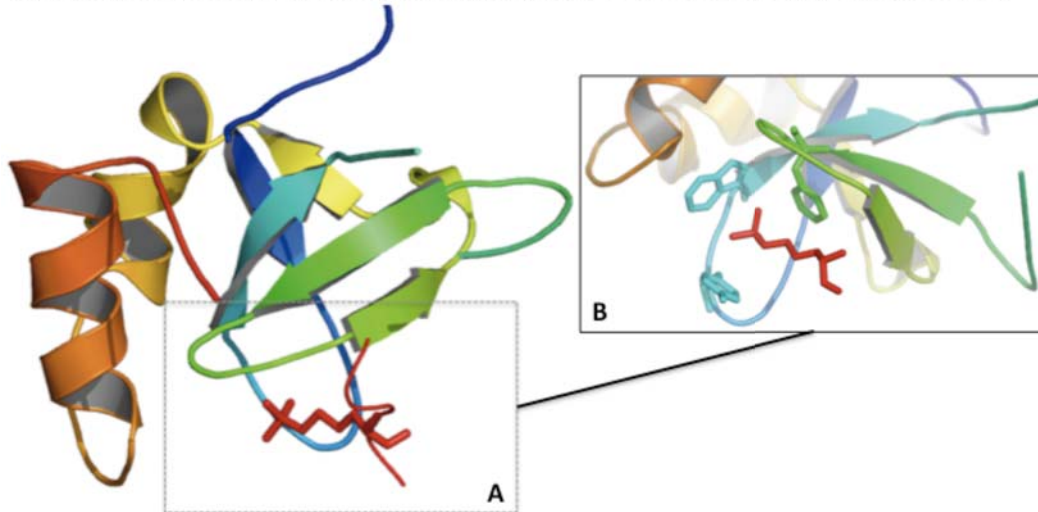
^a El autor no ha especificado el grado de metilación
 ITC = *Isothermal Titration Calorimetry*
 FP = *Fluorescence Polarization*

Estructuralmente, el dominio MBT es más parecido al dominio Tudor (**Figura 12**), y muestra diferencias con el Cromodominio (**Figura 11**). El reconocimiento del Cromodominio, permite que encajen mejor residuos adyacentes a la lisina metilada, adquiriendo mayor dependencia del contexto en el que se ubica la modificación, lo que permite discriminar, por ejemplo, entre H3K9me3 y H3K27me3¹³⁷. Además, los residuos de la cola de la histona adyacentes a la lisina metilada, no forman demasiados contactos con la superficie del dominio MBT, lo que explica la falta de especificidad del dominio *in vitro*²⁰².

4.2.6. PWWP

El dominio PWWP, nombrado así porque presenta un motivo conservado Pro-Trp-Trp-Pro, fue identificado por primera vez en el gen WHSC1²⁰³. Se trata de un módulo de ~70 aminoácidos que está presente en todos los eucariotas, y ha sido identificado en más de 60 proteínas involucradas en procesos tales como, la regulación transcripcional, reparación y metilación del ADN. Entre las diferentes familias de proteínas que contienen el dominio PWWP cabe destacar la familia de proteínas relacionadas con HDGF ya que siempre presentan un dominio PWWP (**Figura 16**), y la familia de proteínas asociadas a la cromatina.

Figura 16 | Estructura cristalográfica del dominio PWWP de HDGF2 en complejo con H3K79me3



A. La estructura de alta resolución del dominio PWWP de HDGF muestra dos partes diferenciadas. Una región muy conservada (a la derecha de la imagen) formada por un barril β integrado por 5 hebras β (β1-β5), β4 y β5 están conectadas mediante una hélice 3₁₀ (α1) y una región helicoidal (a la izquierda de la imagen) formada por una pequeña hélice α (α2) y una hélice α (α3) de 10 residuos. Los extremos N- y C-terminal se ubican en el mismo lado de la molécula. El motivo Pro-Trp-Trp-Pro se encuentra al principio de la hebra β2. B. Reconocimiento de la lisina metilada mediante la caja aromática de PWWP. Código PDB: 3QJ6

Según un estudio publicado el año 2004 por Ge y colaboradores²⁰⁴, además de su similitud con otros miembros de la *Royal Family* (Cromodominio y el dominio Tudor)¹⁸⁴, sugiere que el dominio PWWP podría tener la capacidad de unirse a ligandos metilados. Concretamente, Ge y colaboradores sugieren que podría ser un dominio esencial para ubicar las metiltransferasas *de novo* en la cromatina. Mutaciones en este dominio generan anomalías similares a las que se producen por falta de metilación en el ADN satélite, como inestabilidad centromérica.

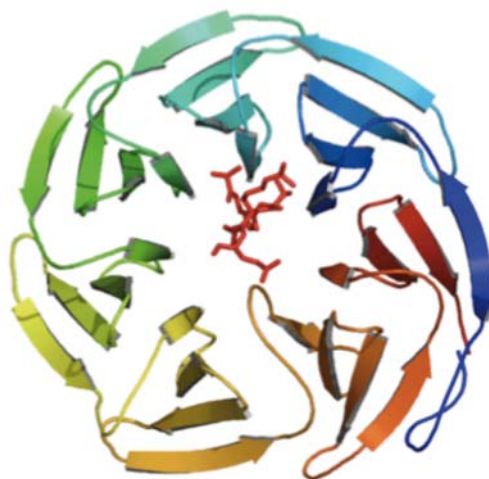
La proteína de levadura Pdp1 contiene un dominio PWWP que reconoce H4K20me1, pero no H4K20me2 o H4K20me3, mediante al menos dos de los residuos aromáticos

conservados equivalentes a aquellos que constituyen el lugar de unión del dominio Tudor²⁰⁵.

4.2.7. WD40

WDR5 es una subunidad de varios complejos con actividad metiltransferasa de H3K4 como MLL1, MLL2 y SET1. Esta subunidad es imprescindible para la unión de los complejos a H3K4 di- o trimetiladas²⁰⁶. La proteína WDR5 contiene 7 copias del dominio WD40, que representa otra clase de dominios de reconocimiento de lisinas metiladas diferente al cromodominio y otros miembros de la *Royal Family* (**Figura 17**).

Figura 17 | Estructura cristalográfica del dominio WD40 de la proteína WDR5 en complejo con H3K4me2



Cada copia de WD40 da lugar a una lámina β formada por 4 hebras β antiparalelas. Las siete copias de WD40 dispuestas circularmente forman un β propeller de siete caras, con un canal estrecho central. La sexta cara del β propeller queda ligeramente distorsionada en comparación con el resto de copias de WD40, debido al bucle que une la quinta y la sexta cara, y que permite la interacción con la histona H3. Una red de puentes de hidrógeno y contactos de Van der Waals anclan el extremo N-terminal de la histona H3 en la cavidad de WDR5. Código PDB: 2G9A

La gran familia de proteínas que contienen el dominio WD40 está presente en todos los eucariotas y parece implicada en diversas funciones como la transducción de señal, tráfico vesicular, ensamblaje del citoesqueleto, control del ciclo celular, apoptosis y regulación de la transcripción²⁰⁷.

Según el análisis estructural de Couture y colaboradores, el dominio formado por las copias de WD40 en WDR5 no es capaz de discriminar entre diferentes grados de metilación de la lisina 4 y la metilación de esta no es un prerrequisito para su unión. La región N-terminal de WDR5 carece de residuos aromáticos y, por lo tanto, no puede participar en las interacciones π -catión con K4me2, una característica común entre los motivos de unión a lisinas metiladas, como los dominios Tudor y Cromodominios. Sin

embargo, otros análisis estructurales revelan que K4me2 podría quedar estabilizada principalmente por un par de puentes de hidrógeno poco convencionales entre los dos grupos metil de K4 y el E322 de WDR5²⁰⁸. En contraste con su falta de especificidad por K4, WDR5 participa en múltiples interacciones con los residuos A1, R2 y T3 de la histona H3, siendo R2 y T3 los factores determinantes para el reconocimiento de la histona H3 por WD40 en el caso de la proteína WDR5²⁰⁹.

4.2.8. Características comunes del reconocimiento de las lisinas metiladas

Tal y como se mencionó anteriormente los dominios efectores encargados del reconocimiento de lisinas metiladas presentan unas características comunes que son relativamente independientes de la estructura global del dominio¹⁴⁸. A continuación se describen las más destacadas.

4.2.8.1. Modos de reconocimiento.

Considerando las estructuras conocidas de los complejos histona-efector, se han identificado, principalmente, dos modos de unión de las lisinas metiladas relacionados con la especificidad de la interacción: *cavity-insertion* y *surface-groove*⁶⁸.

En el modo *cavity-insertion*, el grupo metilamónio de la cadena lateral de la lisina metilada queda enterrada en el interior de una cavidad profunda y estrecha con la capacidad de filtrar el ligando por tamaño. Por ejemplo, el dominio Tudor en tándem de 53BP1¹⁸⁸ (**Figura 12**) y el dominio MBT de la proteína L3MBTL1²¹⁰ (**Figura 15**) reconocen bajos niveles de metilación mediante este mecanismo de reconocimiento.

Por lo contrario, en el modo *surface-groove*, las cavidades son más amplias y accesibles, de forma que la cadena lateral de la lisina metilada, sujeta a restricciones de tamaño menos estrictas, se ubica a lo largo de un surco formado en la superficie de la proteína. Como ejemplos de complejos histona-efector que siguen este mecanismo de reconocimiento tenemos: el Cromodominio de HP1^{170,172}, el doble Cromodominio de CHD1¹⁷⁵ y el doble dominio Tudor de JMJD2A¹⁹² (**Figura 13**).

4.2.8.2. Jaula aromática

Una característica común de los dominios que reconocen residuos de lisina metilados es la ubicación del grupo metilamónio en el interior de una jaula aromática (*aromatic cage*) formada por entre dos y cuatro residuos aromáticos, en ocasiones complementados por uno o más residuos ácidos. (**Figuras. 9, 14B, 16B**) El reconocimiento de la lisina metilada queda estabilizado electrostáticamente mediante interacciones π -catión y en menor medida, por contactos hidrofóbicos^{211,212}.

4.2.8.3. Cavidades de reconocimiento estáticas

Mientras los péptidos de histona pueden sufrir cambios conformacionales inducidos por la unión, la proteína efectora no suele mostrar perturbaciones estructurales apreciables. La naturaleza compacta y rígida del plegamiento de los dominios de reconocimiento descarta movimientos substanciales como consecuencia de la unión con el péptido de histona¹⁴⁸. Concretamente, la jaula aromática en complejo con el péptido de histona es muy estática en relación a su forma libre. Existen algunas excepciones en los casos de 53BP1¹⁸⁸ y WDR5²¹³, en los que existe movimientos de la cadena lateral tras la unión con el péptido de histona.

4.3. Reconocimiento de otras modificaciones

Además de las acetilaciones y las metilaciones, existen otras modificaciones post-traduccionales de las histonas para las que todavía no se conoce la proteína efectora asociada, o no está tan caracterizada como los casos anteriores. Estas modificaciones se describen a continuación.

4.3.1. Fosforilación

Algunos residuos de serina y treonina de las histonas pueden ser fosforilados, y podrían estar implicados en el control del ciclo celular, la reparación del ADN y la regulación de la transcripción^{148,184}. La fosforilación suele ocurrir tanto en la cola, como en la región central de las histonas, y se produce principalmente en residuos de serina. A pesar del gran número de dominios implicados en el reconocimiento de residuos fosforilados en proteínas que no son histonas, nuestro conocimiento de los módulos que reconocen residuos de histona fosforilados se limita a: (i) la familia de proteínas 14-3-3 involucradas en el reconocimiento de H3S10ph en el contexto Ala-

Arg-Lys-Ser, y (ii) el dominio en tándem BRCT de MDC1 involucrado en el reconocimiento de la fosfoserinas 139 (S139ph) de la variante de histona γ H2AX^{148,184}. En este contexto, las proteínas que contienen copias del dominio WD40 se han caracterizado estructuralmente como efectores con la capacidad de unirse a fosfoserinas, además de a lisinas metiladas, y a lisinas y argininas sin modificar¹⁴⁸. Actualmente, un desafío en el campo de la biología de la cromatina es identificar más proteínas efectoras asociadas a la cromatina que pueden interpretar la información codificada en la fosforilación de las histonas.

4.3.2. Metilación de las argininas

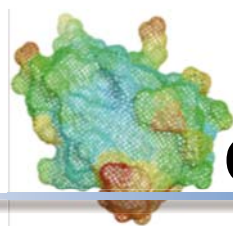
Aunque el interés en la metilación de argininas de las histonas ha quedado eclipsado por el interés en las lisinas metiladas, se han identificado dos co-activadores transcripcionales, CARM1 y PRMT1, que son reclutados por las modificaciones H3R17 y H4R3, respectivamente¹⁴⁸. Se sabe también que este mecanismo de señalización puede ser revertido mediante ciertas actividades desmetilasas y desmetilimididasas^{128,129}.

Diversos estudios han establecido la relación entre las argininas metiladas y diferentes dominios previamente identificados y asociados a otras modificaciones de las colas de las histonas¹⁸⁴. Por ejemplo, el dominio Tudor permite el reclutamiento de la proteína efectora a residuos de arginina metilados tanto asimétrica como simétricamente, y en este reconocimiento parece estar implicada la caja aromática del mismo modo que en el reconocimiento de lisina metiladas²¹⁴. Otro estudio centrado en la interacción entre los dos bromodominios de CHD1 y H3K4me3 sugiere que H3R2 es importante para la interacción, y la metilación asimétrica de esta posición disminuye la afinidad por el péptido¹⁷⁵. Otro dominio que también ha sido relacionado con el reconocimiento de argininas metiladas es el dominio WD40²⁰⁹ y PHD finger¹⁸⁴.

4.3.3. Histonas no modificadas

El dominio SANT es un dominio estructuralmente similar al dominio de unión al ADN (DBD) de Myb, que se ha identificado en algunas proteínas de complejos que actúan como corepresores o coactivadores transcripcionales, como por ejemplo, complejo SAGA con actividad HAT, SMRT, ISWI. A pesar de que su función aun no parece

demasiado clara, análisis bioquímicos de los dominios SANT de Ada2, Iswi y SMRT parecen indicar que estaría involucrado en el reconocimiento de colas de histonas sin modificar²¹⁵.



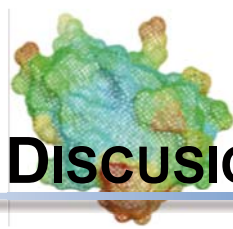
OBJETIVOS

OBJETIVOS

*“Siempre en la mente has de tener a Ítaca.
Llegar allá es tu destino,
pero no apresures el viaje”.*

De forma general, el objetivo de esta tesis es la caracterización de diferentes aspectos estructurales y funcionales de los reguladores epigenéticos y sus consecuencias para la interpretación biológica de las modificaciones de las histonas. A efectos prácticos, dicho objetivo se divide en las tres partes siguientes:

- a) **Análisis de la distribución de los dominios de interacción mejor caracterizados entre las diferentes enzimas modificadoras de la cromatina.** Los dominios de interacción son los responsables del reclutamiento de las enzimas modificadoras de la cromatina en determinadas regiones del genoma. Este objetivo se centra en determinar la asociación entre las diferentes actividades enzimáticas involucradas en la regulación epigenética (HATs, HDACs, HMTs, etc.) y los diferentes dominios de interacción (Bromodominio, Cromodominio y Dominio SANT).
- b) **Caracterización de la variabilidad estructural de los dominios de interacción de las enzimas modificadoras de la cromatina.** Este objetivo se centra en describir y clasificar la variabilidad estructural de los componentes fundamentales de la interacción entre la histona y la proteína efectora y, posteriormente, extraer las posibles consecuencias para la comprensión de la especificidad de unión a las diferentes modificaciones.
- c) **Caracterización de las propiedades moleculares y del papel regulador del *splicing* alternativo de las enzimas modificadoras de la cromatina.** La finalidad de este objetivo es la de caracterizar de forma exhaustiva los patrones de *splicing* alternativo de los reguladores epigenéticos, y clasificarlos en categorías de acuerdo con su acción reguladora potencial de la función génica.



DISCUSIÓN GLOBAL



DISCUSIÓN GLOBAL

*“Es mejor que dure muchos años
y que ya viejo llegues a la isla,
rico de todo lo que hayas guardado en el camino
sin esperar que Ítaca te de riquezas”.*

En este trabajo se presentan tres estudios que abordan aspectos complementarios de la función de los reguladores epigenéticos y su regulación: (i) propiedades estructurales y patrón de distribución de los dominios de interacción con la cromatina; (ii) la modulación funcional mediante variaciones locales en el centro activo y en la transición desorden–orden experimentada por la cola de las histonas tras su unión a los dominios efectores; y (iii) los patrones de *splicing* alternativo de los reguladores epigenéticos y su posible impacto funcional. A continuación se discuten las implicaciones mayores de los resultados obtenidos.

1. Propiedades estructurales y patrón de distribución de los dominios de interacción con la cromatina

Los dominios efectores y sus características estructurales y fisicoquímicas juegan un papel fundamental en la interacción entre los reguladores epigenéticos y la cromatina. Está claro que las distintas modificaciones de las histonas tienen distintos efectos funcionales, y que estas modificaciones son añadidas y reconocidas por diferentes dominios efectores; no parece tan claro, sin embargo, cómo se coordinan las modificaciones de las histonas y los dominios efectores para desencadenar un determinado efecto funcional.

Autores como Matthew J. Bottomley¹⁵⁷ y Ronen Marmorstein⁵², han hecho un esfuerzo por reconciliar el estado del arte sobre los dominios efectores implicados en la regulación de la cromatina y el código de histonas, y sugerir un marco conceptual adecuado para comprender cómo las modificaciones covalentes de las histonas regulan la expresión génica mediante dominios efectores conservados en los reguladores epigenéticos. De las publicaciones de estos autores se desprende que la composición y distribución de estos dominios juega un papel importante en la actividad de los reguladores epigenéticos.

En esta línea, y centrados en la importancia del reclutamiento como el origen de la selectividad de la función enzimática, en la **publicación 1** se analiza la distribución de los Bromodominios, Cromodominios y dominios SANT entre las enzimas implicadas en la regulación de los estados de la cromatina. También se describen las características estructurales y funcionales de dichos dominios, en el contexto de su posible papel en la interpretación del código de histonas.

En dicho trabajo se observa, de forma general, como el patrón de presencia/ausencia y la distribución desigual de los dominios efectores entre los diferentes reguladores epigenéticos indican que estos dominios efectores confieren propiedades específicas de unión a la cromatina a las diferentes familias de reguladores epigenéticos. Por ejemplo, la interacción entre el Bromodominio de BRG1, un componente del complejo SWI/SNF, y H4K8Ac, favorece el reclutamiento de dicho complejo a la región promotora de IFN- β ²¹⁶. Del mismo modo, el doble Bromodominio de TAF_{II}250, un factor de transcripción con actividad HAT, es reclutado al promotor de IFN- β mediante la acetilación de los residuos K9 y K14 de la histona H3, estabilizando la unión de SWI/SNF²¹⁶.

Los datos de la **Publicación 1 (Tabla 1)** muestran que las enzimas modificadoras de la cromatina sólo contienen un tipo de dominio, aunque su número de copias puede variar. Esta duplicación del dominio podría contribuir a la especificidad de unión, al incrementar la estabilidad de la unión en histonas que presenten varias modificaciones apropiadamente espaciadas. Este sería el caso del doble Bromodominio de TAF_{II}250 que se une a la histona H3 acetilada en las lisinas 9 y 14¹⁵³. El reclutamiento *in vivo* de TAF_{II}250 al promotor de IFN- β sólo ocurre cuando ambos residuos están acetilados; además, la mutación de uno de estos residuos cancela el reclutamiento²¹⁶. En estos datos, y otros similares, se basa la teoría de la multivalencia²¹⁷, según la cual las modificaciones de las histonas no operan solas, sino que actuarían coordinadas con otras modificaciones. Este complejo código de modificaciones epigenéticas no está necesariamente restringido a una única cola de histona, sino que podría involucrar a dos o más colas de un determinado nucleosoma, a nucleosomas adyacentes o a nucleosomas ubicados de forma discontinua en la secuencia de ADN primaria. De este modo, la combinación de módulos efectores podría incrementar la especificidad, afinidad y la dinámica de los complejos macromoleculares asociados a la cromatina²¹⁷.

Esta idea puede explicar cómo una determinada combinación de modificaciones de histonas y los dominios que las reconocen pueden regular la transcripción a corto plazo. Hay que decir, sin embargo, que no explica completamente cuál es la contribución de estos dominios al establecimiento y mantenimiento, e incluso la herencia, de los estados transcripcionales a largo plazo. Según Cosma y colaboradores²¹⁸, los dominios de unión a la cromatina pueden tener un papel central en el establecimiento y mantenimiento de dichos estados. Ello podría deberse a la habilidad de algunas enzimas para incorporar marcas de histonas que ellas mismas reconocen, como por ejemplo, enzimas HAT que unen preferentemente péptidos acetilados mediante el Bromodominio.

2. Modulación funcional mediante variaciones locales en el centro activo

En el contexto descrito, la especificidad de estos dominios de unión a la cromatina es un tema importante que requiere mención especial. En particular, una pregunta crítica es por qué estos dominios son capaces de discriminar entre residuos de lisina portadores de la misma modificación. La disponibilidad de información estructural para diferentes complejos histona-efector proporciona una excelente oportunidad para mejorar nuestra comprensión de las bases moleculares de la regulación epigenética, al aclarar las propiedades biofísicas y funcionales de esta interacción y sus componentes. En los siguientes apartados, se discuten cuáles son los posibles determinantes de la especificidad de los dominios efectores.

2.1. Lugar de unión de los dominios de interacción de los reguladores epigenéticos

Tal como se muestra en la **publicación 1 (Fig. 1D-F)** los lugares de unión de cada familia de dominios presentan un grado de similitud de secuencia considerable. Sin embargo, también es cierto que la similitud no es absoluta, lo cual podría estar asociado a las diferencias de especificidad al sustrato. Esto ha sido confirmado en el caso de los Cromodominios, mediante estudios de intercambio de dominios que demuestran que no existe una conservación uniforme de la función²¹⁹. Previamente, Taverna y colaboradores¹⁴⁸ habían descrito las principales características del reconocimiento molecular de las modificaciones de las histonas por diversas familias

de dominios efectores. En este amplio estudio sugieren que los diferentes complejos histona-efector pueden clasificarse en función del modo de reconocimiento empleado: *surface-groove* y *cavity-insertion*. En el primero, la cadena lateral de la lisina metilada descansa sobre un surco en la superficie del efector, mientras que en el segundo, la lisina metilada queda insertada en una cavidad estrecha y profunda.

En la **publicación 3**, partiendo de estos resultados, caracterizamos estructuralmente la variabilidad de la interacción entre los dominios efectores y colas de las histonas, con el objetivo de contribuir a la identificación de los determinantes de la especificidad de dicha interacción. Un aspecto interesante de nuestro trabajo es que no se observa una relación unívoca entre las clases que nosotros proponemos y aquellas sugeridas previamente por Taverna y colaboradores¹⁴⁸. Esta contradicción es sólo aparente, y se debe a que estos autores centran su trabajo en el lugar de unión del residuo de histona modificado (al que nos referiremos de ahora en adelante, y por consistencia con el texto de la correspondiente publicación, como MBS), mientras que en nuestro caso consideramos la totalidad del lugar de unión (al que nos referimos como BS). Concretamente, en nuestra clasificación su clase *surface-groove* queda subdividida en dos clases: *flat-groove* y *narrow-groove*. Para ilustrar la importancia de considerar el lugar de unión de forma amplia, tomaremos como ejemplo el caso de las familias Polycomb y HP1¹³⁷. En general, los Cromodominios de los miembros de la familia Polycomb muestran una preferencia por H3K27me3²²⁰, mientras que los Cromodominios de la familia HP1 la muestran por H3K9me3. Se dan, sin embargo, excepciones: por ejemplo, los miembros de la familia Polycomb Cbx4 y Cbx7 son capaces de unirse a H3K9me3, de hecho Cbx4 lo hace de forma preferente. Si analizamos el alineamiento múltiple de la familia Polycomb^{221,223}, vemos que fuera de la región MBS algunos residuos del lugar de unión de Cbx4 y Cbx7 son diferentes a los del resto de miembros de Polycomb. Concretamente, ambos Cromodominios presentan una valina en lugar de Ala-28 (de acuerdo con la numeración del Polycomb de *Drosophila*). Si ahora consultamos el residuo equivalente en HP1, vemos que es una valina (Val-26), y que además es un residuo que juega un papel importante en la unión con la histona¹⁶³.

La comparación entre MBS y BS proporciona una explicación para los cambios de especificidad entre dominios con el mismo MBS: dichos cambios serían debidos a diferencias en la parte restante del BS. Una serie de resultados experimentales

refuerzan esta idea. Por ejemplo, ensayos de mutagénesis para el segundo dominio de JMJD2A (PDB: 2GFA) muestran que las mutaciones de Asp945 (un residuo que está en contacto con el péptido de histona, aunque no con la lisina modificada) a Ala y Arg, reducen y eliminan el reconocimiento de H3K4me3, respectivamente¹⁹². Un caso parecido ocurre para la mutación Tyr1500Ala del dominio Tudor en tándem de la proteína 53BP1 (PDB:2IG0) que reduce la afinidad de unión por H4K20me2¹⁸⁸. Por su parte, la mutación Val26Met del cromodominio de HP1 de *Drosophila melanogaster* impide la unión con la histona H3¹⁶³. En estos dos casos, los residuos mutados pertenecen al lugar de unión de los dominios implicados, pero no están en contacto con los residuos de histona modificados. Ello confirma la contribución del resto del BS al reconocimiento de la histona, y a la especificidad de dicho reconocimiento.

Extender esta idea a los BS de la misma clase estructural resulta más difícil. Ello es debido por una parte al gran parecido global que tienen, y por otra al hecho de que a pesar de este parecido pueden tener especificidad hacia sustratos diferentes. Los resultados de la comparación de los BS de efectores homólogos (BS de los cromodominios de la clase *narrow-groove*) apuntan en la misma dirección. Podría parecer que debido a la gran similitud estructural de estos BS, que comparten el mismo MBS, tendrían que tener la misma especificidad de sustrato. Sin embargo, este no es siempre el caso, como muestran los cromodominios de HP1 y Polycomb. Al considerar su BS, estos dominios pertenecen a la misma clase, *narrow-groove*, y si consideramos exclusivamente el residuo de histona modificado (tal y como sugieren Taverna y colaboradores) pertenecen a la misma clase MBS, *surface-groove*, sin embargo, estos Cromodominios reconocen dos sustratos diferentes, H3K9me3 y H3K27me3, respectivamente¹³⁷. Fischle y colaboradores han mostrado que distintas características del lugar de unión fuera del MBS pueden jugar un papel importante en la determinación de las diferentes especificidades de sustrato de los cromodominios de Polycomb y HP1¹³⁷. Nuestros resultados de la comparación de los BS de efectores homólogos generalizan esta observación, al mostrar que la variabilidad estructural entre BS homólogos (descritos por la distribución de BS de efectores con el mismo plegamiento representada en la Fig. 4A - **Publicación 3**), es mayor que el error experimental.

2.2. Péptido de histonas

A pesar del enfoque de la tesis, centrada principalmente en los dominios efectores de los reguladores epigenéticos, hemos creído necesario añadir un apartado referente al péptido de histona a la luz de los resultados obtenidos en la **publicación 3**. Ello es debido a que una observación detallada de los péptidos unidos sugiere la existencia de un determinante adicional de la especificidad de la interacción entre el regulador epigenético y la histona, y refuerza la clasificación de los lugares de unión.

A continuación se discuten brevemente las principales características estructurales observadas en los péptidos de histona y su contribución a los contactos atómicos establecidos entre el péptido de histona y el dominio efector.

2.2.1. Características estructurales del péptido de histona: patrones de contactos atómicos.

Nuestro análisis de la estructura del péptido nos proporciona una visión consistente con la clasificación de los lugares de unión: encontramos un grado sustancial de variabilidad que surge de las diferencias entre los correspondientes efectores. Sin embargo, identificamos un motivo estructural, el motivo *hook-like*, presente en la mayoría de péptidos independientemente del modo de reconocimiento. Cuando la secuencia del motivo es RTK observamos una gran variabilidad en sus patrones de interacción, esta variabilidad está relacionada con las diferencias del lugar de unión de los efectores. Esto indica que la estructura atómica cercana, pero exterior al MBS, está involucrada en la determinación de la especificidad de la interacción. Por el contrario, cuando la secuencia es ARK su patrón de interacción está bien conservado. Todos los péptidos de histona que contienen el tripéptido ARK están unidos a los Cromodominios de la clase *narrow-groove*, lo que confirma que los determinantes de la especificidad de sustrato hacia diferentes péptidos de histona involucra átomos del efector en contacto con átomos de histona fuera de la secuencia de residuos conservados. Esta idea coincide con los resultados obtenidos por Fischle y colaboradores¹³⁷ en el caso de Polycomb y HP1. A pesar de estas diferencias, cabe destacar la identificación del motivo estructural *hook-like*, presente en la mayoría de péptidos estudiados, independientemente del modo de reconocimiento. Este motivo muestra un elevado grado de conservación estructural y podría jugar un papel importante en la afinidad de la unión del péptido al efector.

3. *Splicing* alternativo como mecanismo de regulación de los reguladores epigenéticos

Considerando la importancia de los dominios efectores en la función de los reguladores epigenéticos resulta natural pensar que, de forma similar a lo que ocurre en el caso de los factores de transcripción^{222,223}, la determinación de su presencia/ausencia mediante eventos de AS constituye un nivel adicional de regulación funcional. Es bien sabido que el AS es un fenómeno presente en la gran mayoría de los genes humanos^{224,225} y que da lugar a isoformas que por su estructura de dominios tienen un carácter regulador de la isoforma principal^{222,223,226}.

Previamente Tajul-Arifin y colaboradores¹⁷³, se centraron en la identificación y el análisis de proteínas codificadas en el transcriptoma de ratón que contenían el Cromodominio. Además de ampliar el catálogo de proteínas que presentan dicho dominio, los resultados de este estudio mostraban que algunas isoformas lo perdían total o parcialmente. Ello apuntaba a que estas proteínas podían tener alterada la especificidad a su sustrato o actuar como antagonista de la isoforma que conserva el Cromodominio. Este estudio señalaba al AS como una fuente de regulación funcional; sin embargo, considerando el gran número de dominios efectores conocidos, se trataba de un estudio parcial. Dado este contexto científico y los trabajos sobre AS realizados en el grupo, nos planteamos estudiar ampliamente los patrones de AS, analizando su tipo y frecuencia, considerando la totalidad de dominios, tanto los de interacción, como los catalíticos, e incluyendo todas las familias de reguladores epigenéticos anotadas hasta la fecha. **(Publicación 2)**.

Globalmente, los resultados de nuestro estudio confirman y refuerzan la idea de que el AS tiene un papel relevante en la modulación de las propiedades funcionales de los reguladores epigenéticos, eliminando o alterando su estructura modular de dominios. De acuerdo con la naturaleza del cambio y su localización en la proteína podemos definir tres clases: (i) reducciones drásticas de tamaño, (ii) cambios en el dominio catalítico y, finalmente, (iii) cambios en los dominios de interacción. A continuación se discute la interpretación funcional de dichas clases.

3.1. Reducciones drásticas de tamaño

La generación de isoformas inactivas constituye un mecanismo sencillo de regulación de la cantidad de proteína funcional presente en la célula²²⁷⁻²²⁹. Normalmente, las isoformas inactivas son versiones cortas de la proteína funcional en las que se han perdido la mayoría de sus dominios funcionales²²⁸. En nuestro estudio hemos identificado un conjunto de isoformas que presentan una importante reducción de tamaño relativo al de la proteína activa (entre el 35 y 95%), con la mayoría de los dominios funcionales eliminados o seriamente dañados. Estas isoformas encajan con la descripción de isoformas inactivas y, por lo tanto, su expresión correspondería a mecanismos de regulación de cantidad de proteína funcional presente en la célula (**Publicación 2. Tabla 4**). Por ejemplo, esta situación se da en la quinasa humana ATM, que pierde la mayoría de sus dominios catalíticos en una de las isoformas lo que hace inverosímil la conservación de la función.

3.2. Cambios en el dominio catalítico

Generalmente, las isoformas que han perdido el dominio catalítico se pueden comportar como reguladores dominantes negativos de la isoforma funcionalmente completa, tal y como se ha observado en el caso de los factores de transcripción²²³. En nuestro caso hemos identificado diferentes genes con isoformas que bien han perdido el dominio catalítico o éste ha perdido su capacidad funcional debido a las deleciones que ha experimentado (**Publicación 2. Tabla 2**). Por ejemplo, la quinasa PRKDC, muestra que los cambios introducidos no son estructuralmente neutros, lo que concuerda con la ligera actividad inhibidora descrita para la isoforma corta de PRKDC en relación a su isoforma de mayor tamaño²³⁰. Hay que señalar que en este caso, a pesar de no presentar actividad quinasa, la isoforma corta de PRKDC participa en procesos de reparación del ADN²³⁰. Por lo tanto, no podemos descartar la posibilidad de que en algunos casos, las isoformas sin unidad catalítica tengan otros papeles funcionales.

3.3. Cambios en los dominios de interacción

A *priori* el significado funcional de la pérdida de los dominios de interacción podría corresponder a una desregulación de la actividad enzimática a través de un mecanismo dependiente de la naturaleza de la interacción afectada. Si esta es necesaria para la formación de un complejo entre la enzima y otras subunidades

proteicas necesarias para la catálisis, la desregulación se produciría por la formación de un complejo inactivo. Probablemente, este es el caso de la isoforma corta de la acetiltransferasa de histona GCN5L2. Si el dominio de interacción es el responsable del reclutamiento de un regulador epigenético determinado, la desregulación sería una consecuencia de la incapacidad de la enzima para alcanzar su sustrato. Otra posibilidad sería que la falta del dominio de unión a la cromatina no permita la perpetuación de los dominios activos de la cromatina. Este efecto ha sido propuesto por diferentes autores para enzimas portadoras del Bromodominio^{231,232}. En nuestro estudio hemos identificado diferentes genes humanos cuyas isoformas han perdido total o parcialmente los dominios de interacción, por ejemplo, GCN5L2, MYST1 y MORF4L1 (**Publicación 2. Tabla 3**).

Los resultados de este trabajo constituyen un primer paso hacia la comprensión del impacto del AS como un nivel adicional de regulación de la expresión génica a través de las enzimas modificadoras de la cromatina. Debido a que los reguladores epigenéticos actúan específicamente sobre un determinado gen o globalmente sobre el genoma^{121,233}, los diferentes mecanismos de regulación funcional descritos podrían influir en la expresión de un gran conjunto de genes.

CONCLUSIONES GENERALES



CONCLUSIONES GENERALES

*“Sin ella no habrías aprendido el camino.
No tiene otra cosa que darte ya”.*

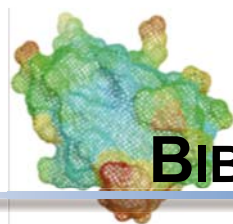
A continuación se resumen las principales conclusiones obtenidas en los diferentes estudios presentados en esta tesis:

A) De la caracterización funcional y estructural de los dominios de interacción presentes en las enzimas modificadoras/remodelantes de la cromatina, concluimos que:

- Existe una distribución desigual de los dominios de interacción entre las diferentes familias de enzimas modificadoras/remodelantes de la cromatina. Esta distribución desigual tiene el potencial de actuar como regulador de la especificidad de los reguladores epigenéticos.
- Los lugares de unión de los dominios de interacción pueden clasificarse en tres clases diferentes en función de características estructurales comunes: *flat-groove*, *narrow-groove* y *cavity-insertion*.
- Las variabilidades estructural y química globales constituyen un factor importante en el establecimiento de la especificidad hacia una determinada modificación de la histona.
- Existe un motivo conservado (*hook-like*) en los péptidos de histona participante en la interacción histona-dominio.
- Nuestros datos indican que al unirse el dominio efector, la cola de las histonas experimenta, en su lugar de unión, una transición desorden-orden que podría contribuir a la especificidad de la interacción histona-dominio

B) En relación a la variabilidad funcional introducida por los eventos de AS en las diferentes familias de reguladores epigenéticos estudiadas:

- AS se da en un número importante de reguladores epigenéticos.
- En general, los cambios introducidos por AS pueden ordenarse en las tres categorías siguientes: (i) reducciones drásticas del tamaño, (ii) cambios en el dominio catalítico y (iii) cambios en los dominios de interacción.
- A cada una de las categorías anteriores corresponde una interpretación funcional que sugiere un efecto regulador importante por parte del AS.



BIBLIOGRAFÍA

BIBLIOGRAFÍA

“ Ve a muchas ciudades egipcias
para que aprendas y aprendas de los sabios”.

1. Lander, E.S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
2. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931-945 (2004).
3. Emes, R.D., Goodstadt, L., Winter, E.E. & Ponting, C.P. Comparison of the genomes of human and mouse lays the foundation of genome zoology. *Hum. Mol. Genet* **12**, 701-709 (2003).
4. Waterston, R.H. et al. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520-562 (2002).
5. Gibbs, R.A. et al. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**, 493-521 (2004).
6. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695-716 (2004).
7. Jaillon, O. et al. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* **431**, 946-957 (2004).
8. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69-87 (2005).
9. Carroll, S.B. Evolution at Two Levels: On Genes and Form. *PLoS Biol* **3**, (2005).
10. Bouchard, T.J., Lykken, D.T., McGue, M., Segal, N.L. & Tellegen, A. Sources of human psychological differences: the Minnesota Study of Twins Reared Apart. *Science* **250**, 223-228 (1990).
11. Fraga, M.F. et al. Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl. Acad. Sci. U.S.A* **102**, 10604-10609 (2005).
12. Wong, A.H.C., Gottesman, I.I. & Petronis, A. Phenotypic differences in genetically identical organisms: the epigenetic perspective. *Hum. Mol. Genet* **14 Spec No 1**, R11-18 (2005).
13. Kendler, K.S. Twin Studies of Psychiatric Illness: An Update. *Arch Gen Psychiatry* **58**, 1005-1014 (2001).
14. Malaty, H.M., Graham, D.Y., Isaksson, I., Engstrand, L. & Pedersen, N.L. Are Genetic Influences on Peptic Ulcer Dependent or Independent of Genetic Influences for *Helicobacter pylori* Infection? *Arch Intern Med* **160**, 105-109 (2000).
15. Gärtner, K. A third component causing random variability beside environment and genotype. A reason for the limited success of a 30 year long effort to standardize laboratory animals? *Lab. Anim* **24**, 71-77 (1990).
16. Ross, P.J. & Cibelli, J.B. Bovine somatic cell nuclear transfer. *Methods Mol. Biol* **636**, 155-177 (2010).

17. Kishigami, S. & Wakayama, T. Somatic cell nuclear transfer in the mouse. *Methods Mol. Biol* **518**, 207-218 (2009).
18. Galli, C., Lagutina, I., Duchi, R., Colleoni, S. & Lazzari, G. Somatic cell nuclear transfer in horses. *Reprod. Domest. Anim* **43 Suppl 2**, 331-337 (2008).
19. WaddingtonCH The epigenotype. *Endeavour* **1**, 18-20 (1942).
20. Bird, A. Perceptions of epigenetics. *Nature* **447**, 396-398 (2007).
21. Thiel, G., Lietz, M. & Hohl, M. How mammalian transcriptional repressors work. *Eur. J. Biochem* **271**, 2855-2862 (2004).
22. Goldman, M.A., Holmquist, G.P., Gray, M.C., Caston, L.A. & Nag, A. Replication timing of genes and middle repetitive sequences. *Science* **224**, 686-692 (1984).
23. Hsieh, T.-F. & Fischer, R.L. BIOLOGY OF CHROMATIN DYNAMICS. *Annu. Rev. Plant. Biol.* **56**, 327-351 (2005).
24. Kingston, R.E. & Narlikar, G.J. ATP-dependent remodeling and acetylation as regulators of chromatin fluidity. *Genes Dev* **13**, 2339-2352 (1999).
25. Deuring, R. et al. The ISWI chromatin-remodeling protein is required for gene expression and the maintenance of higher order chromatin structure in vivo. *Mol. Cell* **5**, 355-365 (2000).
26. Henikoff, S., Furuyama, T. & Ahmad, K. Histone variants, nucleosome assembly and epigenetic inheritance. *Trends in Genetics* **20**, 320-326 (2004).
27. McKittrick, E., Gafken, P.R., Ahmad, K. & Henikoff, S. Histone H3.3 is enriched in covalent modifications associated with active chromatin. *Proc Natl Acad Sci U S A* **101**, 1525-1530 (2004).
28. Turker, M.S. The establishment and maintenance of DNA methylation patterns in mouse somatic cells. *Semin. Cancer Biol* **9**, 329-337 (1999).
29. Bestor, T.H., Gundersen, G., Kolstø, A.B. & Prydz, H. CpG islands in mammalian gene promoters are inherently resistant to de novo methylation. *Genet. Anal. Tech. Appl* **9**, 48-53 (1992).
30. Herman, J.G. & Baylin, S.B. Gene silencing in cancer in association with promoter hypermethylation. *N. Engl. J. Med* **349**, 2042-2054 (2003).
31. Weber, M. et al. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet* **39**, 457-466 (2007).
32. Ohgane, J., Yagi, S. & Shiota, K. Epigenetics: the DNA methylation profile of tissue-dependent and differentially methylated regions in cells. *Placenta* **29 Suppl A**, S29-35 (2008).
33. Bodey, B. Cancer-testis antigens: promising targets for antigen directed antineoplastic immunotherapy. *Expert Opin Biol Ther* **2**, 577-584 (2002).
34. Feinberg, A.P., Cui, H. & Ohlsson, R. DNA methylation and genomic imprinting: insights from cancer into epigenetic mechanisms. *Semin. Cancer Biol* **12**, 389-398 (2002).
35. Reik, W. & Lewis, A. Co-evolution of X-chromosome inactivation and imprinting in mammals. *Nat. Rev. Genet* **6**, 403-410 (2005).
36. Bestor, T.H. Transposons reanimated in mice. *Cell* **122**, 322-325 (2005).
37. Xu, G.L. et al. Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* **402**, 187-191 (1999).

38. Espada, J. et al. Epigenetic disruption of ribosomal RNA genes and nucleolar architecture in DNA methyltransferase 1 (Dnmt1) deficient cells. *Nucleic Acids Res* **35**, 2191-2198 (2007).
39. Redon, C. et al. Histone H2A variants H2AX and H2AZ. *Current Opinion in Genetics & Development* **12**, 162-169 (2002).
40. Bradbury, E.M. Current ideas on the structure of chromatin. *Trends in Biochemical Sciences* **1**, 7-9 (1976).
41. Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F. & Richmond, T.J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251-260 (1997).
42. Schalch, T., Duda, S., Sargent, D.F. & Richmond, T.J. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* **436**, 138-141 (2005).
43. Borland, L., Harauz, G., Bahr, G. & van Heel, M. Packing of the 30 nm chromatin fiber in the human metaphase chromosome. *Chromosoma* **97**, 159-163 (1988).
44. Bassett, A., Cooper, S., Wu, C. & Travers, A. The folding and unfolding of eukaryotic chromatin. *Curr. Opin. Genet. Dev* **19**, 159-165 (2009).
45. Marmorstein, R. & Trievel, R.C. Histone modifying enzymes: structures, mechanisms, and specificities. *Biochim. Biophys. Acta* **1789**, 58-68 (2009).
46. Kornberg, R.D. & Lorch, Y. Chromatin-modifying and -remodeling complexes. *Curr. Opin. Genet. Dev* **9**, 148-151 (1999).
47. Hebbes, T.R., Thorne, A.W. & Crane-Robinson, C. A direct link between core histone acetylation and transcriptionally active chromatin. *EMBO J* **7**, 1395-1402 (1988).
48. Rea, S. et al. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593-599 (2000).
49. Wu, J. & Grunstein, M. 25 years after the nucleosome model: chromatin modifications. *Trends Biochem. Sci* **25**, 619-623 (2000).
50. Berger, S.L. Histone modifications in transcriptional regulation. *Curr. Opin. Genet. Dev* **12**, 142-148 (2002).
51. Cheung, P. & Lau, P. Epigenetic Regulation by Histone Methylation and Histone Variants. *Mol Endocrinol* **19**, 563-573 (2005).
52. Wolffe, A.P. & Hayes, J.J. Chromatin disruption and modification. *Nucleic Acids Res* **27**, 711-720 (1999).
53. Zheng C, H.J., Hayes JJ Structures and interactions of the core histone tail domains. *Biopolymers* **68**, 539-46 (2003).
54. Arents, G. & Moudrianakis, E.N. The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proc. Natl. Acad. Sci. U.S.A* **92**, 11170-11174 (1995).
55. Marmorstein, R. Protein modules that manipulate histone tails for chromatin regulation. *Nat. Rev. Mol. Cell Biol* **2**, 422-432 (2001).
56. Ling, X., Harkness, T.A., Schultz, M.C., Fisher-Adams, G. & Grunstein, M. Yeast histone H3 and H4 amino termini are important for nucleosome assembly in vivo and in vitro: redundant and position-independent functions in assembly but not in gene regulation. *Genes Dev* **10**, 686-699 (1996).

57. Grunstein, M. Histone acetylation in chromatin structure and transcription. *Nature* **389**, 349-352 (1997).
58. Thompson, J.S., Ling, X. & Grunstein, M. Histone H3 amino terminus is required for telomeric and silent mating locus repression in yeast. *Nature* **369**, 245-247 (1994).
59. Durrin, L.K., Mann, R.K., Kayne, P.S. & Grunstein, M. Yeast histone H4 N-terminal sequence is required for promoter activation in vivo. *Cell* **65**, 1023-1031 (1991).
60. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693-705 (2007).
61. Jenuwein, T. & Allis, C.D. Translating the histone code. *Science* **293**, 1074-1080 (2001).
62. Strahl, B.D. & Allis, C.D. The language of covalent histone modifications. *Nature* **403**, 41-45 (2000).
63. Shogren-Knaak, M. et al. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311**, 844-847 (2006).
64. Cosgrove, M.S., Boeke, J.D. & Wolberger, C. Regulated nucleosome mobility and the histone code. *Nat. Struct. Mol. Biol* **11**, 1037-1043 (2004).
65. Garcia-Ramirez, M., Rocchini, C. & Ausio, J. Modulation of chromatin folding by histone acetylation. *J. Biol. Chem* **270**, 17923-17928 (1995).
66. Tse, C., Sera, T., Wolffe, A.P. & Hansen, J.C. Disruption of higher-order folding by core histone acetylation dramatically enhances transcription of nucleosomal arrays by RNA polymerase III. *Mol. Cell. Biol* **18**, 4629-4638 (1998).
67. Morales, V. & Richard-Foy, H. Role of histone N-terminal tails and their acetylation in nucleosome dynamics. *Mol. Cell. Biol* **20**, 7230-7237 (2000).
68. Hansen, J.C., Tse, C. & Wolffe, A.P. Structure and function of the core histone N-termini: more than meets the eye. *Biochemistry* **37**, 17637-17641 (1998).
69. Sivolob, A., De Lucia, F., Alilat, M. & Prunell, A. Nucleosome dynamics. VI. Histone tail regulation of tetrasome chiral transition. A relaxation study of tetrasomes on DNA minicircles. *J. Mol. Biol* **295**, 55-69 (2000).
70. Wang, X., Moore, S.C., Laszckzak, M. & Ausió, J. Acetylation increases the alpha-helical content of the histone tails of the nucleosome. *J. Biol. Chem* **275**, 35013-35020 (2000).
71. Ausio, J. & van Holde, K.E. Histone hyperacetylation: its effects on nucleosome conformation and stability. *Biochemistry* **25**, 1421-1428 (1986).
72. Libertini, L.J., Ausió, J., van Holde, K.E. & Small, E.W. Histone hyperacetylation. Its effects on nucleosome core particle transitions. *Biophys. J* **53**, 477-487 (1988).
73. Simpson, R.T. Structure of chromatin containing extensively acetylated H3 and H4. *Cell* **13**, 691-699 (1978).
74. Mutskov, V. et al. Persistent interactions of core histone tails with nucleosomal DNA following acetylation and transcription factor binding. *Mol. Cell. Biol* **18**, 6293-6304 (1998).
75. Polach, K.J., Lowary, P.T. & Widom, J. Effects of core histone tail domains on the equilibrium constants for dynamic DNA site accessibility in nucleosomes. *J. Mol. Biol* **298**, 211-223 (2000).
76. Turner, B.M. Decoding the nucleosome. *Cell* **75**, 5-8 (1993).

77. Turner, B.M., Birley, A.J. & Lavender, J. Histone H4 isoforms acetylated at specific lysine residues define individual chromosomes and chromatin domains in *Drosophila polytene nuclei*. *Cell* **69**, 375-384 (1992).
78. Seet, B.T., Dikic, I., Zhou, M.-M. & Pawson, T. Reading protein modifications with interaction domains. *Nat. Rev. Mol. Cell Biol* **7**, 473-483 (2006).
79. Ruthenburg, A.J., Allis, C.D. & Wysocka, J. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Mol. Cell* **25**, 15-30 (2007).
80. Li, B., Carey, M. & Workman, J.L. The role of chromatin during transcription. *Cell* **128**, 707-719 (2007).
81. Birney, E. et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799-816 (2007).
82. Kim, T.H. et al. A high-resolution map of active promoters in the human genome. *Nature* **436**, 876-880 (2005).
83. Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B. & Cavalli, G. Genome regulation by polycomb and trithorax proteins. *Cell* **128**, 735-745 (2007).
84. Brinkman, A.B. et al. Histone modification patterns associated with the human X chromosome. *EMBO Rep* **7**, 628-634 (2006).
85. van Attikum, H. & Gasser, S.M. The histone code at DNA breaks: a guide to repair? *Nat. Rev. Mol. Cell Biol* **6**, 757-765 (2005).
86. Ahn, S.-H. et al. Sterile 20 kinase phosphorylates histone H2B at serine 10 during hydrogen peroxide-induced apoptosis in *S. cerevisiae*. *Cell* **120**, 25-36 (2005).
87. Cheung, W.L. et al. Apoptotic phosphorylation of histone H2B is mediated by mammalian sterile twenty kinase. *Cell* **113**, 507-517 (2003).
88. Baylin, S.B. & Ohm, J.E. Epigenetic gene silencing in cancer - a mechanism for early oncogenic pathway addiction? *Nat. Rev. Cancer* **6**, 107-116 (2006).
89. Wang, G.G., Allis, C.D. & Chi, P. Chromatin remodeling and cancer, Part I: Covalent histone modifications. *Trends Mol Med* **13**, 363-372 (2007).
90. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet* **33 Suppl**, 245-254 (2003).
91. Lan, F. et al. A histone H3 lysine 27 demethylase regulates animal posterior development. *Nature* **449**, 689-694 (2007).
92. Swigut, T. & Wysocka, J. H3K27 demethylases, at long last. *Cell* **131**, 29-32 (2007).
93. Turner, B.M. Histone acetylation as an epigenetic determinant of long-term transcriptional competence. *Cell. Mol. Life Sci* **54**, 21-31 (1998).
94. Spotswood, H.T. & Turner, B.M. An increasingly complex code. *J. Clin. Invest* **110**, 577-582 (2002).
95. Turner, B.M. Cellular memory and the histone code. *Cell* **111**, 285-291 (2002).
96. Bannister, A.J. et al. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120-124 (2001).

97. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116-120 (2001).
98. Beisel, C., Imhof, A., Greene, J., Kremmer, E. & Sauer, F. Histone methylation by the *Drosophila* epigenetic transcriptional regulator Ash1. *Nature* **419**, 857-862 (2002).
99. Zegerman, P., Canas, B., Pappin, D. & Kouzarides, T. Histone H3 lysine 4 methylation disrupts binding of nucleosome remodeling and deacetylase (NuRD) repressor complex. *J. Biol. Chem* **277**, 11621-11624 (2002).
100. Pokholok, D.K. et al. Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122**, 517-527 (2005).
101. Mikkelsen, T.S. et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553-560 (2007).
102. Bernstein, B.E. et al. Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**, 169-181 (2005).
103. Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R. & Young, R.A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**, 77-88 (2007).
104. Wang, Z. et al. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet* **40**, 897-903 (2008).
105. Hon, G., Wang, W. & Ren, B. Discovery and annotation of functional chromatin signatures in the human genome. *PLoS Comput. Biol* **5**, e1000566 (2009).
106. Ernst, J. & Kellis, M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat. Biotechnol* **28**, 817-825 (2010).
107. Goldberg, A.D. et al. Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell* **140**, 678-691 (2010).
108. Meissner, A. et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**, 766-770 (2008).
109. Lee, J.-H. & Skalnik, D.G. CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. *J. Biol. Chem* **280**, 41725-41731 (2005).
110. Orford, K. et al. Differential H3K4 methylation identifies developmentally poised hematopoietic genes. *Dev. Cell* **14**, 798-809 (2008).
111. Bernstein, B.E. et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315-326 (2006).
112. Barski, A. et al. High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823-837 (2007).
113. Schwartz, S., Meshorer, E. & Ast, G. Chromatin organization marks exon-intron structure. *Nat. Struct. Mol. Biol* **16**, 990-995 (2009).
114. Andersson, R., Enroth, S., Rada-Iglesias, A., Wadelius, C. & Komorowski, J. Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Res* **19**, 1732-1741 (2009).

115. Kornblihtt, A.R., Schor, I.E., Allo, M. & Blencowe, B.J. When chromatin meets splicing. *Nat. Struct. Mol. Biol* **16**, 902-903 (2009).
116. Luco, R.F. et al. Regulation of alternative splicing by histone modifications. *Science* **327**, 996-1000 (2010).
117. Waks, Z., Klein, A.M. & Silver, P.A. Cell-to-cell variability of alternative RNA splicing. *Mol. Syst. Biol* **7**, 506 (2011).
118. Visel, A., Rubin, E.M. & Pennacchio, L.A. Genomic views of distant-acting enhancers. *Nature* **461**, 199-205 (2009).
119. Visel, A. et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854-858 (2009).
120. Heintzman, N.D. et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108-112 (2009).
121. Peterson, C.L. & Laniel, M.-A. Histones and histone modifications. *Curr. Biol* **14**, R546-551 (2004).
122. Taunton, J., Hassig, C.A. & Schreiber, S.L. A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science* **272**, 408-411 (1996).
123. Brownell, J.E. et al. Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell* **84**, 843-851 (1996).
124. Sassone-Corsi, P. et al. Requirement of Rsk-2 for epidermal growth factor-activated phosphorylation of histone H3. *Science* **285**, 886-891 (1999).
125. Thomson, S. et al. The nucleosomal response associated with immediate-early gene induction is mediated via alternative MAP kinase cascades: MSK1 as a potential histone H3/HMG-14 kinase. *EMBO J* **18**, 4779-4793 (1999).
126. Chen, D. et al. Regulation of transcription by a protein methyltransferase. *Science* **284**, 2174-2177 (1999).
127. Gary, J.D., Lin, W.J., Yang, M.C., Herschman, H.R. & Clarke, S. The predominant protein-arginine methyltransferase from *Saccharomyces cerevisiae*. *J. Biol. Chem* **271**, 12585-12594 (1996).
128. Cuthbert, G.L. et al. Histone deimination antagonizes arginine methylation. *Cell* **118**, 545-553 (2004).
129. Wang, Y. et al. Human PAD4 regulates histone arginine methylation levels via demethylimination. *Science* **306**, 279-283 (2004).
130. Robzyk, K., Recht, J. & Osley, M.A. Rad6-dependent ubiquitination of histone H2B in yeast. *Science* **287**, 501-504 (2000).
131. Emre, N.C.T. et al. Maintenance of low histone ubiquitylation by Ubp10 correlates with telomere-proximal Sir2 association and gene silencing. *Mol. Cell* **17**, 585-594 (2005).
132. Gardner, R.G., Nelson, Z.W. & Gottschling, D.E. Ubp10/Dot4p regulates the persistence of ubiquitinated histone H2B: distinct roles in telomeric silencing and general chromatin. *Mol. Cell. Biol* **25**, 6123-6139 (2005).

133. Henry, K.W. et al. Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. *Genes Dev* **17**, 2648-2663 (2003).
134. Chang, B., Chen, Y., Zhao, Y. & Bruick, R.K. JMJD6 is a histone arginine demethylase. *Science* **318**, 444-447 (2007).
135. Shi, Y. et al. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**, 941-953 (2004).
136. Tsukada, Y.-ichi et al. Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439**, 811-816 (2006).
137. Fischle, W. et al. Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev* **17**, 1870-1881 (2003).
138. Hansen, J.C., Lu, X., Ross, E.D. & Woody, R.W. Intrinsic protein disorder, amino acid composition, and histone terminal domains. *J. Biol. Chem* **281**, 1853-1856 (2006).
139. Hudson, B.P., Martinez-Yamout, M.A., Dyson, H.J. & Wright, P.E. Solution structure and acetyl-lysine binding activity of the GCN5 bromodomain. *J. Mol. Biol* **304**, 355-370 (2000).
140. Koonin, E.V., Zhou, S. & Lucchesi, J.C. The chromo superfamily: new members, duplication of the chromo domain and possible role in delivering transcription regulators to chromatin. *Nucl. Acids Res.* **23**, 4229-4233 (1995).
141. Ponting, C.P. Tudor domains in proteins that interact with RNA. *Trends Biochem. Sci* **22**, 51-52 (1997).
142. Aasland, R., Gibson, T.J. & Stewart, A.F. The PHD finger: implications for chromatin-mediated transcriptional regulation. *Trends Biochem. Sci* **20**, 56-59 (1995).
143. Aasland, R., Stewart, A.F. & Gibson, T. The SANT domain: a putative DNA-binding domain in the SWI-SNF and ADA complexes, the transcriptional co-repressor N-CoR and TFIIIB. *Trends Biochem. Sci* **21**, 87-88 (1996).
144. Da, G. et al. Structure and function of the SWIRM domain, a conserved protein module found in chromatin regulatory complexes. *Proc. Natl. Acad. Sci. U.S.A* **103**, 2057-2062 (2006).
145. Bornemann, D., Miller, E. & Simon, J. Expression and properties of wild-type and mutant forms of the *Drosophila* sex comb on midleg (SCM) repressor protein. *Genetics* **150**, 675-686 (1998).
146. Neer, E.J., Schmidt, C.J., Nambudripad, R. & Smith, T.F. The ancient regulatory-protein family of WD-repeat proteins. *Nature* **371**, 297-300 (1994).
147. Stec, I. et al. WHSC1, a 90 kb SET domain-containing gene, expressed in early development and homologous to a *Drosophila* dysmorphia gene maps in the Wolf-Hirschhorn syndrome critical region and is fused to IgH in t(4;14) multiple myeloma. *Hum. Mol. Genet* **7**, 1071-1082 (1998).
148. Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D. & Patel, D.J. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat. Struct. Mol. Biol* **14**, 1025-1040 (2007).
149. Zeng, L. et al. Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. *Nature* **466**, 258-262 (2010).

150. Dhalluin, C. et al. Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **399**, 491-496 (1999).
151. Tamkun, J.W. et al. brahma: A regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2. *Cell* **68**, 561-572 (1992).
152. Hassan, A.H. et al. Selective recognition of acetylated histones by bromodomains in transcriptional co-activators. *Biochem J* **402**, 125-133 (2007).
153. Jacobson, R.H., Ladurner, A.G., King, D.S. & Tjian, R. Structure and function of a human TAFII250 double bromodomain module. *Science* **288**, 1422-1425 (2000).
154. Owen, D.J. et al. The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase gcn5p. *EMBO J* **19**, 6141-6149 (2000).
155. Sanchez, R. & Zhou, M.-M. The role of human bromodomains in chromatin biology and gene transcription. *Curr Opin Drug Discov Devel* **12**, 659-665 (2009).
156. Thompson, M. Polybromo-1: the chromatin targeting subunit of the PBAF complex. *Biochimie* **91**, 309-319 (2009).
157. Bottomley, M.J. Structures of protein domains that create or recognize histone modifications. *EMBO Rep* **5**, 464-469 (2004).
158. Maurer-Stroh, S. et al. The Tudor domain «Royal Family»: Tudor, plant Agenet, Chromo, PWWP and MBT domains. *Trends Biochem. Sci* **28**, 69-74 (2003).
159. Li, H. et al. Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* **442**, 91-95 (2006).
160. Peña, P.V. et al. Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature* **442**, 100-103 (2006).
161. Paro, R. & Hogness, D.S. The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila. *Proc. Natl. Acad. Sci. U.S.A* **88**, 263-267 (1991).
162. Jones, D.O., Cowell, I.G. & Singh, P.B. Mammalian chromodomain proteins: their role in genome organisation and expression. *Bioessays* **22**, 124-137 (2000).
163. Jacobs, S.A. et al. Specificity of the HP1 chromo domain for the methylated N-terminus of histone H3. *EMBO J* **20**, 5232-5241 (2001).
164. Cao, R. et al. Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing. *Science* **298**, 1039-1043 (2002).
165. Czermin, B. et al. Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* **111**, 185-196 (2002).
166. Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P. & Reinberg, D. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes & Development* **16**, 2893-2905 (2002).
167. Müller, J. et al. Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell* **111**, 197-208 (2002).
168. Plath, K. et al. Role of Histone H3 Lysine 27 Methylation in X Inactivation. *Science* **300**, 131-135 (2003).

169. Silva, J. et al. Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. *Dev. Cell* **4**, 481-495 (2003).
170. Jacobs, S.A. & Khorasanizadeh, S. Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science* **295**, 2080-2083 (2002).
171. Nielsen, P.R. et al. Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. *Nature* **416**, 103-107 (2002).
172. Min, J., Zhang, Y. & Xu, R.-M. Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes & Development* **17**, 1823-1828 (2003).
173. Tajul-Arifin, K., Teasdale, R., Ravasi, T., Hume, D.A. & Mattick, J.S. Identification and analysis of chromodomain-containing proteins encoded in the mouse transcriptome. *Genome Res* **13**, 1416-1429 (2003).
174. Ahringer, J. NuRD and SIN3 histone deacetylase complexes in development. *Trends Genet* **16**, 351-356 (2000).
175. Flanagan, J.F. et al. Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature* **438**, 1181-1185 (2005).
176. Lusser, A., Urwin, D.L. & Kadonaga, J.T. Distinct activities of CHD1 and ACF in ATP-dependent chromatin assembly. *Nat. Struct. Mol. Biol* **12**, 160-166 (2005).
177. Pray-Grant, M.G., Daniel, J.A., Schieltz, D., Yates, J.R. & Grant, P.A. Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. *Nature* **433**, 434-438 (2005).
178. Zhang, P. et al. Structure of human MRG15 chromo domain and its binding to Lys36-methylated histone H3. *Nucleic Acids Res* **34**, 6621-6628 (2006).
179. Sun, B. et al. Molecular basis of the interaction of *Saccharomyces cerevisiae* Eaf3 chromo domain with methylated H3K36. *J. Biol. Chem* **283**, 36504-36512 (2008).
180. Brasher, S.V. et al. The structure of mouse HP1 suggests a unique mode of single peptide recognition by the shadow chromo domain dimer. *EMBO J* **19**, 1587-1597 (2000).
181. Boswell, R.E. & Mahowald, A.P. tudor, a gene required for assembly of the germ plasm in *Drosophila melanogaster*. *Cell* **43**, 97-104 (1985).
182. Selenko, P. et al. SMN Tudor domain structure and its interaction with the Sm proteins. *Nat Struct Mol Biol* **8**, 27-31 (2001).
183. Brahms, H. et al. The C-terminal RG dipeptide repeats of the spliceosomal Sm proteins D1 and D3 contain symmetrical dimethylarginines, which form a major B-cell epitope for anti-Sm autoantibodies. *J. Biol. Chem* **275**, 17122-17129 (2000).
184. Yap, K.L. & Zhou, M.-M. Keeping it in the family: diverse histone recognition by conserved structural folds. *Crit. Rev. Biochem. Mol. Biol* **45**, 488-505 (2010).
185. Sanders, S.L. et al. Methylation of histone H4 lysine 20 controls recruitment of Crb2 to sites of DNA damage. *Cell* **119**, 603-614 (2004).
186. Huyen, Y. et al. Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. *Nature* **432**, 406-411 (2004).

187. Adams-Cioaba, M.A. & Min, J. Structure and function of histone methylation binding proteins. *Biochem. Cell Biol* **87**, 93-105 (2009).
188. Botuyan, M.V. et al. Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. *Cell* **127**, 1361-1373 (2006).
189. Klose, R.J. et al. The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36. *Nature* **442**, 312-316 (2006).
190. Klose, R.J. et al. The retinoblastoma binding protein RBP2 is an H3K4 demethylase. *Cell* **128**, 889-900 (2007).
191. Whetstine, J.R. et al. Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. *Cell* **125**, 467-481 (2006).
192. Huang, Y., Fang, J., Bedford, M.T., Zhang, Y. & Xu, R.-M. Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. *Science* **312**, 748-751 (2006).
193. Lee, J., Thompson, J.R., Botuyan, M.V. & Mer, G. Distinct binding modes specify the recognition of methylated histones H3K4 and H4K20 by JMJD2A-tudor. *Nat. Struct. Mol. Biol* **15**, 109-111 (2008).
194. Taverna, S.D. et al. Yng1 PHD finger binding to H3 trimethylated at K4 promotes NuA3 HAT activity at K14 of H3 and transcription at a subset of targeted ORFs. *Mol. Cell* **24**, 785-796 (2006).
195. Vermeulen, M. et al. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* **131**, 58-69 (2007).
196. Bienz, M. The PHD finger, a nuclear protein-interaction domain. *Trends Biochem. Sci* **31**, 35-40 (2006).
197. Mizuguchi, G., Tsukiyama, T., Wisniewski, J. & Wu, C. Role of nucleosome remodeling factor NURF in transcriptional activation of chromatin. *Mol. Cell* **1**, 141-150 (1997).
198. Wysocka, J. et al. A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442**, 86-90 (2006).
199. Wismar, J. et al. The *Drosophila melanogaster* tumor suppressor gene lethal(3)malignant brain tumor encodes a proline-rich protein with a novel zinc finger. *Mech. Dev* **53**, 141-154 (1995).
200. Usui, H., Ichikawa, T., Kobayashi, K. & Kumanishi, T. Cloning of a novel murine gene Sfmtb, Scm-related gene containing four mbt domains, structurally belonging to the Polycomb group of genes. *Gene* **248**, 127-135 (2000).
201. Trojer, P. et al. L3MBTL1, a histone-methylation-dependent chromatin lock. *Cell* **129**, 915-928 (2007).
202. Bonasio, R., Lecona, E. & Reinberg, D. MBT domain proteins in development and disease. *Seminars in Cell & Developmental Biology* **21**, 221-230 (2010).
203. Stec, I., Nagl, S.B., van Ommen, G.J. & den Dunnen, J.T. The PWWP domain: a potential protein-protein interaction domain in nuclear proteins influencing differentiation? *FEBS Lett* **473**, 1-5 (2000).
204. Ge, Y.-Z. et al. Chromatin targeting of de novo DNA methyltransferases by the PWWP domain. *J. Biol. Chem* **279**, 25447-25454 (2004).

205. Wang, Y. et al. Regulation of Set9-mediated H4K20 methylation by a PWWP domain protein. *Mol. Cell* **33**, 428-437 (2009).
206. Wysocka, J. et al. WDR5 Associates with Histone H3 Methylated at K4 and Is Essential for H3 K4 Methylation and Vertebrate Development. *Cell* **121**, 859-872 (2005).
207. Schuetz, A. et al. Structural basis for molecular recognition and presentation of histone H3 By WDR5. *EMBO J* **25**, 4245-4252 (2006).
208. Han, Z. et al. Structural Basis for the Specific Recognition of Methylated Histone H3 Lysine 4 by the WD-40 Protein WDR5. *Molecular Cell* **22**, 137-144 (2006).
209. Couture, J.-F., Collazo, E. & Trievel, R.C. Molecular recognition of histone H3 by the WD40 protein WDR5. *Nat Struct Mol Biol* **13**, 698-703 (2006).
210. Li, H. et al. Structural basis for lower lysine methylation state-specific readout by MBT repeats of L3MBTL1 and an engineered PHD finger. *Mol. Cell* **28**, 677-691 (2007).
211. Hughes, R.M., Wiggins, K.R., Khorasanizadeh, S. & Waters, M.L. Recognition of trimethyllysine by a chromodomain is not driven by the hydrophobic effect. *Proc. Natl. Acad. Sci. U.S.A* **104**, 11184-11188 (2007).
212. Ma, J.C. & Dougherty, D.A. The Cation- π Interaction. *Chemical Reviews* **97**, 1303-1324 (1997).
213. Ruthenburg, A.J. et al. Histone H3 recognition and presentation by the WDR5 module of the MLL1 complex. *Nat. Struct. Mol. Biol* **13**, 704-712 (2006).
214. Sprangers, R., Groves, M.R., Sinning, I. & Sattler, M. High-resolution X-ray and NMR structures of the SMN Tudor domain: conformational variation in the binding site for symmetrically dimethylated arginine residues. *J. Mol. Biol* **327**, 507-520 (2003).
215. Boyer, L.A., Latek, R.R. & Peterson, C.L. The SANT domain: a unique histone-tail-binding module? *Nat Rev Mol Cell Biol* **5**, 158-163 (2004).
216. Agaloti, T., Chen, G. & Thanos, D. Deciphering the transcriptional histone acetylation code for a human gene. *Cell* **111**, 381-392 (2002).
217. Ruthenburg, A.J., Li, H., Patel, D.J. & Allis, C.D. Multivalent engagement of chromatin modifications by linked binding modules. *Nat. Rev. Mol. Cell Biol* **8**, 983-994 (2007).
218. Cosma, M.P., Tanaka, T. & Nasmyth, K. Ordered recruitment of transcription and chromatin remodeling factors to a cell cycle- and developmentally regulated promoter. *Cell* **97**, 299-311 (1999).
219. Wang, G. et al. Conservation of heterochromatin protein 1 function. *Mol. Cell. Biol* **20**, 6970-6983 (2000).
220. Bernstein, E. et al. Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Mol. Cell. Biol* **26**, 2560-2569 (2006).
221. Senthilkumar, R. & Mishra, R.K. Novel motifs distinguish multiple homologues of Polycomb in vertebrates: expansion and diversification of the epigenetic toolkit. *BMC Genomics* **10**, 549 (2009).
222. López, A.J. Developmental role of transcription factor isoforms generated by alternative splicing. *Dev. Biol* **172**, 396-411 (1995).

223. Talavera, D., Orozco, M. & de la Cruz, X. Alternative splicing of transcription factors' genes: beyond the increase of proteome diversity. *Comp. Funct. Genomics* 905894 (2009).doi:10.1155/2009/905894
224. Kampa, D. et al. Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res* **14**, 331-342 (2004).
225. Modrek, B., Resch, A., Grasso, C. & Lee, C. Genome-wide detection of alternative splicing in expressed sequences of human genes. *Nucleic Acids Res* **29**, 2850-2859 (2001).
226. Blencowe, B.J. Alternative splicing: new insights from global analyses. *Cell* **126**, 37-47 (2006).
227. Smith, C.W. & Valcárcel, J. Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem. Sci* **25**, 381-388 (2000).
228. Modrek, B. & Lee, C.J. Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. *Nat. Genet* **34**, 177-180 (2003).
229. Neu-Yilik, G., Gehring, N.H., Hentze, M.W. & Kulozik, A.E. Nonsense-mediated mRNA decay: from vacuum cleaner to Swiss army knife. *Genome Biol* **5**, 218 (2004).
230. Convery, E. et al. Inhibition of homologous recombination by variants of the catalytic subunit of the DNA-dependent protein kinase (DNA-PKcs). *Proc. Natl. Acad. Sci. U.S.A* **102**, 1345-1350 (2005).
231. Hassan, A.H. et al. Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. *Cell* **111**, 369-379 (2002).
232. Syntichaki, P., Topalidou, I. & Thireos, G. The Gcn5 bromodomain co-ordinates nucleosome remodelling. *Nature* **404**, 414-417 (2000).
233. van Leeuwen, F. & van Steensel, B. Histone modifications: from genome-wide maps to functional insights. *Genome Biol* **6**, 113 (2005).
234. Lincoln, R.J., Boxshall, G.A. & Clark, P.F. *A dictionary of ecology, evolution, and systematics / R.J. Lincoln, G.A. Boxshall, and P.F. Clark.* (Cambridge University Press: Cambridge ; New York :, 1982).
235. Russo, V., Martienssen, R. & Riggs, A. *Epigenetic mechanisms of gene regulation.* (Cold Spring Harbor Laboratory Press: NY, 1996).
236. Holliday, R. Mechanisms for the control of gene activity during development. *Biol Rev Camb Philos Soc* **65**, 431-471 (1990).



PUBLICACIONES

PUBLICACIONES

*“Y si la encuentras pobre, Ítaca no te ha engañado
sabio como te has vuelto con tantas experiencias,
habrás comprendido lo que significan las Ítacas”.*

El trabajo presentado en esta tesis ha dado lugar a tres publicaciones que reflejan diferentes aspectos de los dominios funcionales de los reguladores epigenéticos (se enumeran a continuación en el orden cronológico de su publicación):

1. Papel de los dominios de interacción en la interpretación del código de histonas, poniendo de manifiesto que el proceso de reclutamiento de los reguladores epigenéticos, mediante las modificaciones post-traduccionales de las histonas, es fundamental para la regulación génica a corto y largo plazo. (*Publicación 1*)
2. Papel del *splicing* alternativo como un mecanismo de modulación funcional de los reguladores epigenéticos a través de cambios en su estructura de dominios. (*Publicación 2*)
3. Caracterización de la variabilidad estructural de los componentes principales de la interacción entre histonas y dominios efectores (la cola de la histona, por una parte, y el lugar de unión del dominio efector, por otro) (*Publicación 3*)

INFORME DEL DIRECTOR DEL FACTOR DE IMPACTO DE LOS ARTICULOS PUBLICADOS.

Publicación 1. Do protein motifs read the histone code? Xavier de la Cruz, **Sergio Lois**, Sara Sánchez-Molina y Marian A. Martínez-Balbás. *BioEssays* 27:164-175, 2005.

Este trabajo es una revisión de la literatura sobre la relación entre los diferentes dominios funcionales presentes en los reguladores epigenéticos, y el código de histonas. Se publicó en la revista *BioEssays* que recoge revisiones de gran impacto en el área de la biología molecular, evolución, etc. Su factor de impacto 4.479 confirma que el trabajo presentado tiene acceso a una audiencia destacada.

Publicación 2. The functional modulation of epigenetic regulators by alternative splicing. **Sergio Lois**, Noemí Blanco, Marian Martínez-Balbás y Xavier de la Cruz. *BMC Genomics*, 8:252, 2007. Este trabajo destaca por el amplio volumen de datos recogidos sobre los reguladores epigenéticos y su AS, y por el análisis de dichos datos en el contexto de la regulación de la expresión génica. *BMC Genomics* es una revista consagrada a publicar investigación original en diferentes áreas de la genómica, con un índice de impacto de 4.21, que garantiza su difusión entre la comunidad científica.

Publicación 3. Characterization of structural variability sheds light on the specificity determinants of the interaction between effector domains and histone tails. **Sergio Lois**, Naiara Akizu, Gemma Mas de Xaxars, Iago Vázquez, Marian Martínez-Balbás y Xavier de la Cruz *Epigenetic* 5:2, 137-148, 2010. Este trabajo aporta un destacado análisis estructural de la interacción histona-dominio efector y descubre dos nuevos determinantes de la especificidad de dicha interacción. Se ha publicado en la revista *Epigenetics*, de destacado impacto en el campo de la epigenética y con un índice de impacto de 4.58 que así lo confirma.

PUBLICACIÓN 1

Do protein motifs read the histone code?

Xavier de la Cruz, **Sergio Lois**, Sara Sánchez-Molina y Marian A. Martínez-Balbás.

BioEssays 27:164-175, 2005

TÍTULO:

¿Los motivos proteicos leen el código de histonas?

RESUMEN:

Según la hipótesis del código de histonas, las modificaciones químicas presentes en las colas de las histonas constituirían lugares de unión para diferentes familias de proteínas con la capacidad de alterar la estructura de la cromatina. En este artículo se describe de forma exhaustiva aquellos dominios funcionales presentes en dichas proteínas que tienen la capacidad de interactuar con las modificaciones covalentes de las histonas: Bromodominio, Cromodominio y dominio SANT. En particular, se revisa la distribución de estos dominios entre las diferentes familias de enzimas modificadoras de la cromatina conocidas (HAT, HDAC, HMT y enzimas remodelantes de la cromatina) y se intenta caracterizar la contribución de estos dominios en la interpretación del código de histonas.

Entre los resultados de dicha publicación cabe destacar, que los Bromodominios, Cromodominios y dominios SANT pueden encontrarse en las tres familias de enzimas estudiadas, o sus complejos, sin embargo, su distribución es desigual. Esta distribución diferencial entre los dominios de interacción y las actividades enzimáticas, confirma la idea de que estos dominios confieren propiedades específicas de unión a las diferentes familias de enzimas. Además, se discute cómo otros factores tales como la variabilidad de secuencia en el lugar de unión y residuos colindantes, el número de copias del dominio y cambios alostéricos inducidos por la interacción proteína-proteína tras la unión con la cromatina, pueden influir en la especificidad de unión de dichos dominios.

Do protein motifs read the histone code?

Xavier de la Cruz,^{2,3} Sergio Lois,³
Sara Sánchez-Molina,^{1,3} and Marian A. Martínez-Balbás^{1,3*}

Summary

The existence of different patterns of chemical modifications (acetylation, methylation, phosphorylation, ubiquitination and ADP-ribosylation) of the histone tails led, some years ago, to the histone code hypothesis. According to this hypothesis, these modifications would provide binding sites for proteins that can change the chromatin state to either active or repressed. Interestingly, some protein domains present in histone-modifying enzymes are known to interact with these covalent marks in the histone tails. This was first shown for the bromodomain, which was found to interact selectively with acetylated lysines at the histone tails. More recently, it has been described that the chromodomain can be targeted to methylation marks in histone N-terminal domains. Finally,

the interaction between the SANT domain and histones is also well documented. Overall, experimental evidence suggests that these domains could be involved in the recruitment of histone-modifying enzymes to discrete chromosomal locations, and/or in the regulation their enzymatic activity. Within this context, we review the distribution of bromodomains, chromodomains and SANT domains among chromatin-modifying enzymes and discuss how they can contribute to the translation of the histone code. *BioEssays* 27:164–175, 2005.

© 2005 Wiley Periodicals, Inc.

The histone code hypothesis

The packing of the eukaryotic genome into chromatin provides the means for compaction of the entire genome inside the nucleus. However, this packing restricts the access to DNA of the many regulatory proteins essential for biological processes like replication, transcription, DNA repair and recombination.⁽¹⁾ There are two mechanisms that can counterbalance the repressive nature of chromatin, allowing access to nucleosomal DNA: (i) covalent modification of histone tails like acetylation, methylation, phosphorylation and ubiquitination;^(2–5) and (ii) altering of the nucleosomal structure by enzymes utilising energy from ATP hydrolysis.⁽⁶⁾

In the early nineties, it was proposed that histone covalent modifications can work as recognition signals, directing the binding to chromatin of non-histone proteins that determine chromatin function.^(7,8) More recently, it has been hypothesized that specific tail modifications and/or their combinations constitute a code, the histone code, that determines the transcriptional state of the genes.^(9–11) According to this hypothesis, “multiple histone modifications, acting in a combinatorial or sequential fashion on one or multiple tails, specify unique downstream functions”.⁽⁹⁾

In the last years, an increasing amount of experimental data has provided clear support for the different aspects of the histone code hypothesis, contributing to refine and improve it.^(For review 12,13) One important point that has been addressed by different authors is the idea that the histone code must use combinations of modifications.⁽⁹⁾ For example, H3 methylated at K9 could initiate chromatin condensation and silencing^(14,15) but, in the context of methylated H3K4 and H4K20, methyl-K9 H3 helps to maintain active marks by allowing the binding of BRAHMA, the enzyme of the remodelling dSWI/SNF complex.⁽¹⁶⁾

¹Instituto de Biología Molecular de Barcelona. CID. Consejo Superior de Investigaciones Científicas (CSIC).

²Institut Català per la Recerca i Estudis Avançats (ICREA). Barcelona, Spain.

³Institut de Recerca Biomèdica de Barcelona-Parc Científic de Barcelona (IRBB-PCB) Barcelona. Spain.

Funding agency: This work was supported by grants from the Ministerio de Ciencia y Tecnología SAF2002-00741 and PB98-0468 to MMB and BIO2003-09327 to XdlC. SL is recipient of a studentship from the Ministerio de Ciencia y Tecnología. SSM is recipient of a studentship from the Generalitat de Catalunya.

*Correspondence to: Marian A. Martínez-Balbás, Institut de Recerca Biomèdica de Barcelona-Parc Científic de Barcelona (IRBB-PCB) Josep Samitier 1-5. 08028 Barcelona. Spain.

E-mail: mmbbmc@ibmb.csic.es.

DOI 10.1002/bies.20176

Published online in Wiley InterScience (www.interscience.wiley.com).

Abbreviations: HAT, histone acetyltransferase; HDAC, histone deacetylase; HMT, histone methyltransferase; Chromodomain, chromatin organization modifier; SANT, “SWI3, ADA2, N-CoR and TFIIB B”; NURD, nucleosome-remodelling and deacetylase complex; SAGA, Spt-Ada_Gcn5 acetyltransferase; SUV39H1, suppressor of variegation 3–9 homologue 1; SWI/SNF, switching-defective/sucrose non-fermenting; PCAF, p300/CBP associated factor; NURF, nucleosome remodelling factor, CBP, CREB-binding protein; RSC, remodel the structure of chromatin.; NuA4, nucleosome acetyltransferase of histone H4; SMRT/N-CoR, silencing mediator of retinoid and thyroid receptor/nuclear receptor co-repressor. MLL, mixed lineage leukaemia; E(Z), enhancer of Zeste.

Another aspect of the histone code hypothesis that has received much attention, is the idea that modifications on the same or different histone tails may be interdependent. That is, the fact that modification in one residue can determine that of another either in *cis* or, more surprisingly, in *trans*. In the first case, it has been shown that methylation of H3K4 has two important effects: it blocks both the binding of the remodelling deacetylation complex NURD and the methylation of H3K9, thereby preventing the placement of silencing marks.⁽¹⁷⁾ As an example of *trans* effect, we can mention that ubiquitination of H2B K123 is required for an efficient methylation of H3K4.⁽¹⁸⁾

Recently, two new concepts have been introduced to understand the basis of the signalling by combinations of histone modifications, the concepts of “binary switches” and “modification cassettes”.⁽¹⁹⁾ In the former, neighbouring modifications act together, while for the latter residues in linear strings of densely modifiable sites can have different biological readouts, depending on their modification state.

Combinations of modifications appear to be important for both short- and long-term transcriptional regulation, and have been described in different systems. In the first case, there are clear experimental results showing that regulation of rapid transcriptional processes usually requires a cascade of modification events. For example, activation of the IFN- β gene requires acetylation of several lysines in histones H3 and H4 that mediate the recruitment of the SWI/SNF complex and TFIID, respectively.⁽²⁰⁾ In the case of long-term regulation, there is less experimental evidence, although it is now clear that some specific histone modifications have the potential to exert long-term effects. For example, H3 methylated at K9 has the potential to initiate chromatin condensation and silencing.^(14,15) Also, Czermin, Müller and colleagues have shown that H3 methylated by the E(Z) complex, binds specifically to polycomb protein, suggesting a direct relationship between H3 methylation and silencing by PcG complex.^(21,22)

Finally, it has to be mentioned that recently the histone code hypothesis has been extended to the nucleosome code hypothesis, by proposing that high-order chromatin is largely dependent on the local concentration and combination of differentially modified nucleosomes.⁽¹⁰⁾

Chromatin-binding domains and the translation of the histone code

An important issue when considering the histone code is how it is translated. More precisely, how the combinations of modifications that constitute the code are recognised and then translated into a given functional effect. According to the histone code hypothesis, the histone modification marks would provide the binding sites for a series of effector proteins that would affect chromatin function.⁽⁹⁾ Interestingly, histone-modifying enzymes are unable to access their substrates unless targeted there. In other words, the same enzyme will not modify all histones in all genes, at the same time; only that

subset of genes that have recruited the modifying enzyme to the promoter will be regulated by it. This highlights the relevance of the targeting process, at the origin of the selectivity of the enzyme action, as an important feature of the regulation by histone modification. Within this context, it is clear that protein domains able to interact with chromatin and/or its modified components—like bromodomains, chromodomains or the SANT domains—can play a crucial role in the targeting process. These domains (or protein modules, as they have been named⁽¹⁹⁾) could contribute to both the recognition of specific patterns of modifications, as well as to their setting at given locations. Here, we explore this idea by analysing the distribution of chromatin-binding domains (bromodomain, chromodomain and SANT domain) among chromatin-modifying enzymes (HAT, HDAC, HMT and ATP-dependent remodelling enzymes). To this end, we first review the main structural and functional characteristics of the chromatin-binding domains, together with their distribution among chromatin-modifying enzymes. Then, on that basis we consider their possible role in the translation of the histone code, as targeting elements of the histone-modifying enzymes in the context of short-term regulation. Finally, we discuss their role in the long-term regulation processes.

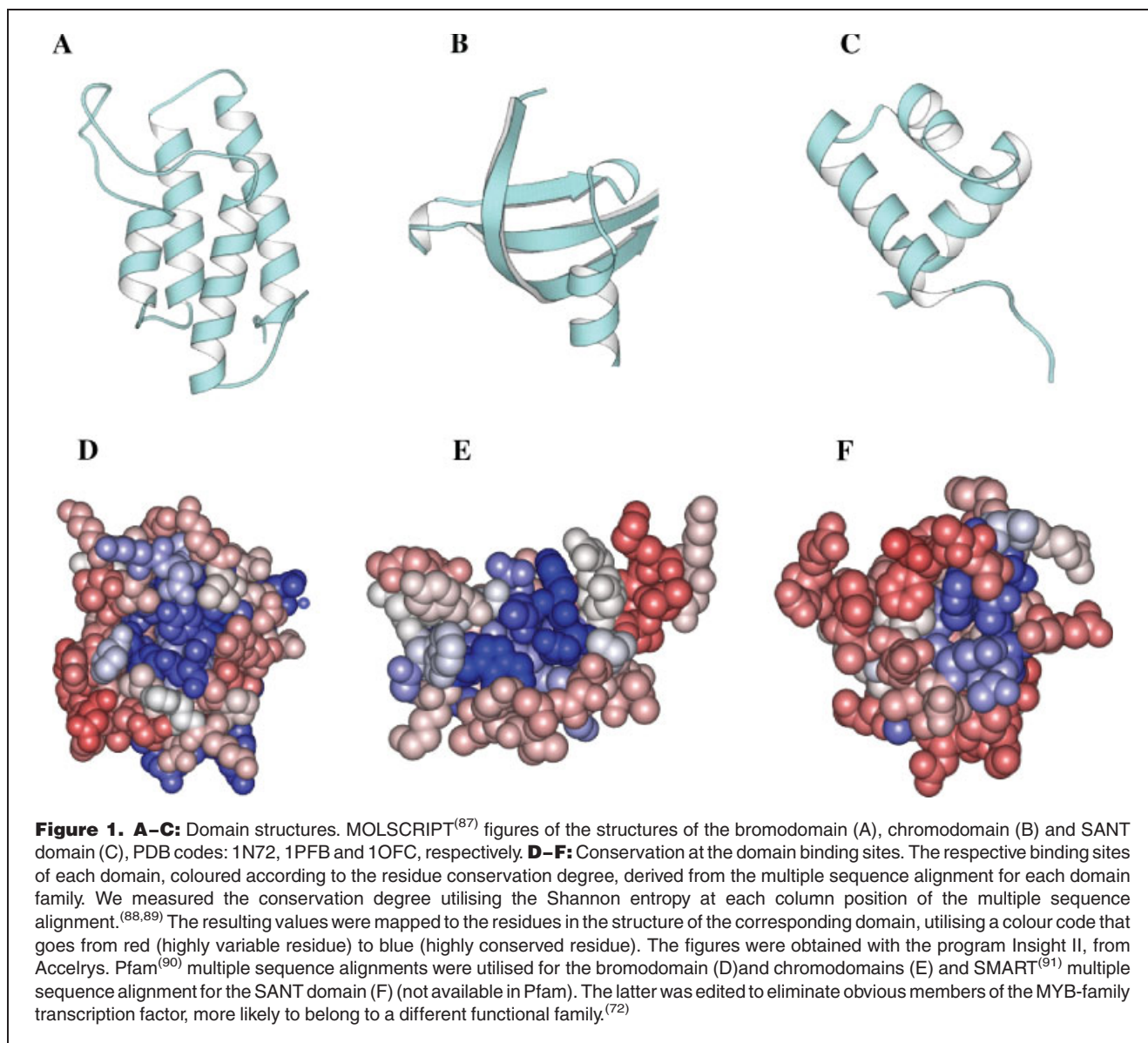
The bromodomain

Bromodomains are small protein domains that form an extensive family.⁽²³⁾ The first reported bromodomain was found in the *Drosophila* Brahma protein.⁽²⁴⁾ Bromodomains were later found in many chromatin-associated proteins and most histone acetyltransferases.^(25,26)

Structure and function

The three-dimensional structure of a prototypical bromodomain from the histone acetyltransferase PCAF shows an unusual left-handed four-helix bundle (Fig. 1A).^(27,28) A long loop between helices αZ and αA is packed against the loop connecting helices B and C to form a surface-accessible hydrophobic pocket, located at one end of the four-helix bundle. This unique feature is conserved in the bromodomain family and can be seen in the bromodomain structure of human GCN5, *S. cerevisiae* GCN5p and human TAF_{II}250.^(29–31)

The bromodomain role in chromatin remodelling was suggested some time ago, on the basis of yeast genetic studies.⁽²⁴⁾ However, its biological function was confirmed after the more recent discovery that bromodomains function as acetyl-lysine binding domains.^(28–31) Initially, *in vitro* studies showed that bromodomains preferentially bind acetylated peptides, leading to speculation that acetylated histone tails could become targets for the binding of bromodomain-containing proteins.^(28,30–32) This has been recently confirmed by Hassan and colleagues, who have shown that the SWI/SNF complex is retained to the chromatin by previous histone acetylation by SAGA or NuA4, and after removal of the



transcription factor (Gal4-VP16).⁽³³⁾ The retention requires the bromodomain of Swi2/Snf2 subunit of the SWI/SNF complex.⁽³³⁾ Further, the SAGA complex itself is anchored to acetylated arrays, following removal of the activator, and can coordinate nucleosomal remodelling; however, this will only happen if the bromodomain of the Gcn5 subunit is intact, providing a self-perpetuating mark tethered to a small chromatin domain.⁽³⁴⁾

If bromodomains play a role in enzyme targeting to the chromatin, then one would expect a high conservation degree at their binding sites independently of the chromatin-modifying enzyme carrying them. In Fig. 1D, we plot a view of the binding site, with all residues coloured according to their conservation degree, as derived from the multiple sequence alignment for

the domain family. In accordance with the proposed role, we observe that highly conserved residues tend to define, or cluster around, the domain binding site.

Overall, these data confirm that the bromodomain has the ability to bind acetylated histone tails *in vivo*, with an apparent independence of the protein to which it belongs, and this ability can be utilised by different chromatin-remodelling enzymes to find and/or act on their targets.

Distribution among chromatin-modifying enzymes

In accordance with the above mentioned functional data, we find that the bromodomain is widely distributed among the different enzymes that acetylate, methylate or remodel chromatin (Table 1).

Table 1. Summary table of the HAT-, HMT-, and ATP-dependent chromatin-remodelling enzymes carrying bromodomains, chromodomains or SANT domains as part of their sequence, in different organisms

Protein activity	Domain	Organism	Gene	# Domains	Residue specificity	Molecular function
Chromatin Remodeling	BROMO	<i>D. melanogaster</i>	BRM	1	—	Transc. coactiv./corepres.
		<i>H. sapiens</i>	BRM	1	—	Transc. coactiv./corepres.
		<i>S. cerevisiae</i>	BRG1	1	—	Transc. coactiv./corepres.
			SWI2/SNF2	1	—	Transc. coactiv./corepres.
			STH1	1	—	Transc. coactiv./corepres.
Histone Acetyltransferase	BROMO	<i>A. thaliana</i>	GCN5	1	—	Transc. coactiv.
		<i>C. elegans</i>	1E91	1	—	Transc. coactiv.
			CBP-1	1	—	Transc. coactiv.
		<i>H. sapiens</i>	P300	1	H3: K14, K18	H2A, H2B
			PCAF	1	H3: K14	H4: K5, K8
			CBP	1	H3: K14, K18	H4: K8
			TAF250	2	H3: K14, K18	H2A, H2B
			GCN5	1	H3: K14	H4
			GCN5	1	H3: K14	H4: K8, K16
		<i>M. musculus</i>	GCN5	1	H3: K14	H4: K8, K16
			PCAF	1	H3: K14	H4: K8, K16
			CBP	1	H3: K14, K18	H4: K5, K8
			LOC330129	1	—	H2A, H2B
Histone Methyltransferase	BROMO	<i>H. sapiens</i>	MLL	1	H3: K4	Transc. coactiv.
				2	—	Transc. coactiv.
Chromatin Remodeling	CHROMO	<i>A. thaliana</i>	CHD-3	2	—	Transc. corepres.
		<i>C. elegans</i>	LET-418 (MI-2 like)	2	—	Transc. corepres.
			CHD-3	2	—	Transc. corepres.
		<i>D. melanogaster</i>	CHD-3	2	—	Transc. corepres.
		<i>H. sapiens</i>	CHD-3	2	—	Transc. corepres.
			CHD-4	2	—	Transc. corepres.
			CHD-5	2	—	Transc. corepres.
		<i>O. sativa</i>	P0018C10.33	2	—	Unknown
				2	—	Unknown
				2	—	Unknown
Histone Acetyltransferase	CHROMO	<i>H. sapiens</i>	CDY-1	1	—	Transc. corepres.
			CDY-2	1	—	Transc. corepres.
			TIP60	1	H2A: K5	HIV tat interaction
			MORF4L1	1	H4: K5, K8, K12, K16	H4: K5, K8, K12, K16
		<i>D. melanogaster</i>	MORF4L1	1	—	H3
		<i>M. musculus</i>	CG6121	1	—	Transc. coactiv.
		<i>R. norvegicus</i>	MORF4L1	1	—	Unknown
			TIP60	1	—	Unknown
		<i>S. pombe</i>	SPAC637.12C	1	H2A: K5	HIV tat interaction
			(Myst family)	1	—	Unknown
		<i>S. cerevisiae</i>	ESA1	1	H4: K5, K8, K12, K16	H4: K5, K8, K12, K16
			SU(VAR)3-9	1	H3: K9	H2A
Histone Methyltransferase	CHROMO	<i>D. melanogaster</i>	SU(VAR)3-9	1	H3: K9	Heterochr. Formation/Transc. corepres.
		<i>H. sapiens</i>	SUV39H1	1	H3: K9	Heterochr. Formation/Transc. corepres.
			SUV39H2	1	H3: K9	Heterochr. Formation/Transc. corepres.
		<i>M. musculus</i>	SUV39H1	1	H3: K9	Heterochr. Formation/Transc. corepres.
				1	H3: K9	Heterochr. Formation/Transc. corepres.

								H3: K9	1	SUV39H2			Heterochr. Formation/Transc. corepres.
								H3: K9	1	SUVAR39-1			Heterochr. Formation/Transc. corepres.
								H3: K9	1	SUVAR39-2			Heterochr. Formation/Transc. corepres.
								H3: K9	1	CLR4			Heterochr. Formation/Transc. corepres.
									2	ISW1			Transc. coactiv./corepres.
									2	ISW2			Transc. coactiv./corepres.
									2	ISWI	SANT	Chromatin Remodeling	Transc. coactiv./corepres.
									2	SNF2L			Transc. coactiv./corepres.
									2	SNF2H			Transc. coactiv./corepres.
									1	MEDEA			Gene Silencing
									1	EZA1			Gene Silencing
								H3: K27	2	E(Z)			Gene Silencing
								H3: K27	2	EZH1			Gene Silencing
									2	EZH2			Gene Silencing
								H3: K27	2	EZH1	SANT	Histone Methyltransferase	Gene Silencing
									2	EZH2			Gene Silencing
									1	EZ1			Gene Silencing
									1	EZ2			Gene Silencing
									1	EZ3			Gene Silencing

Bromodomain and HAT enzymes

The bromodomain is present in the members of *Arabidopsis*, human and mouse GCN5/PCAF family, human and mouse CBP/p300, human TAF_{II}250, TAF1L, acetyltransferases and two putative HAT enzymes: Q9N3N7 (involved in membrane transport in *C. elegans*) and LOC330129 in mouse, which encodes for a protein similar to PCAF. The presence of bromodomains in many HAT enzymes suggests that self-perpetuation of the HAT at acetylated locus through interactions between their bromodomains and acetylated histones could be a common feature for these enzymes (Fig. 2B).

Bromodomain and HMT enzymes

Bromodomains are also part of some HMT enzymes: Ash1, RIZ (member of the RIZ family), and MLL (members of the TRX proteins). MLL also has an MBD domain that may recognise methylated DNA.⁽³⁵⁾

Bromodomain and ATP-dependent remodelling enzymes

Remodelling enzymes that utilise ATP to alter chromatin structure also have bromodomains: SNF2 α and SNF2 β from human, as well as their homologous SNF2 in *Saccharomyces cerevisiae* and Brahma in *Drosophila*, all of them members of the SWI/SNF complexes, and STH1, subunit of the yeast RSC complex.

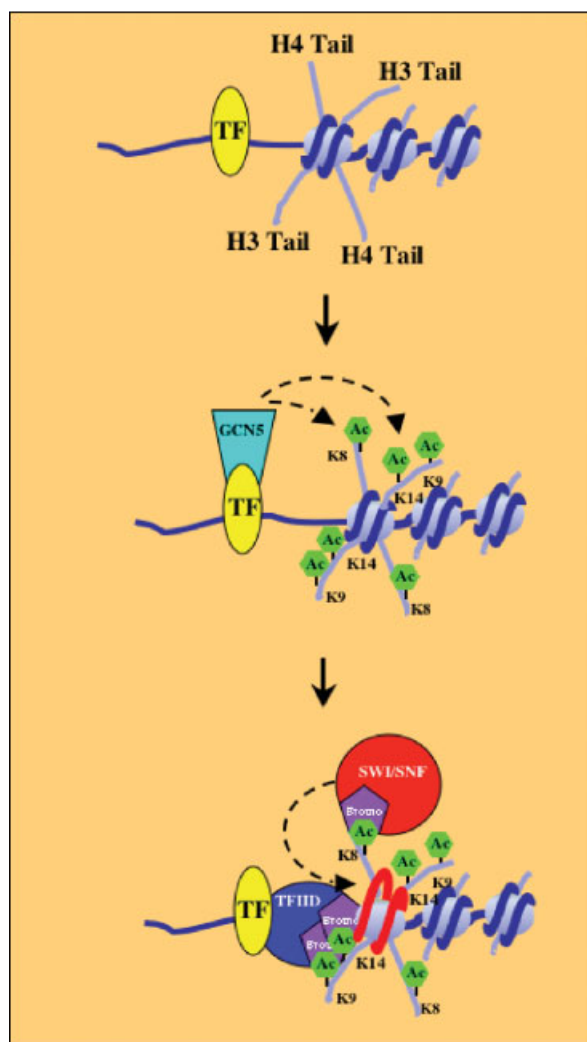
Bromodomains can also be found in subunits, with no catalytic activity, of remodelling complexes, where they could help the latter in the recognition of previously modified chromatin or to stabilize the interaction of the complex. For example, bromodomains can be found in (1) *C. elegans*, human, mouse and rat MTA-1 protein, subunit of the MTA-1 complex, (2) in human and mouse ACF1, subunit of the CHRAC complex, (3) in yeast spt7, component of the SAGA complex, and (4) in yeast RSC1 and 2, components of the RSC complex. In the latter, the bromodomain is essential for the RSC remodelling function, although it is not required for complex assembly.⁽³⁶⁾

Taken together, these data suggest that specific interactions between some ATP remodelling enzymes and chromatin could be stabilised by the bromodomain Ac-lys interaction, helping to establish the final remodelled chromatin structure, in accordance with work by Agaliote and colleagues.⁽²⁰⁾

The chromodomain

The chromodomain was first identified as a common domain between two distinct regulators of chromatin structure in *Drosophila*: HP1 and Polycomb.⁽³⁷⁾ Later, chromodomains have been found in many other chromatin regulators: (i) remodelling factors involved in causing conformational changes by ATP-dependent movement of nucleosomes and (ii) histone acetyltransferases and methyltransferases.^(38,39)

A



B

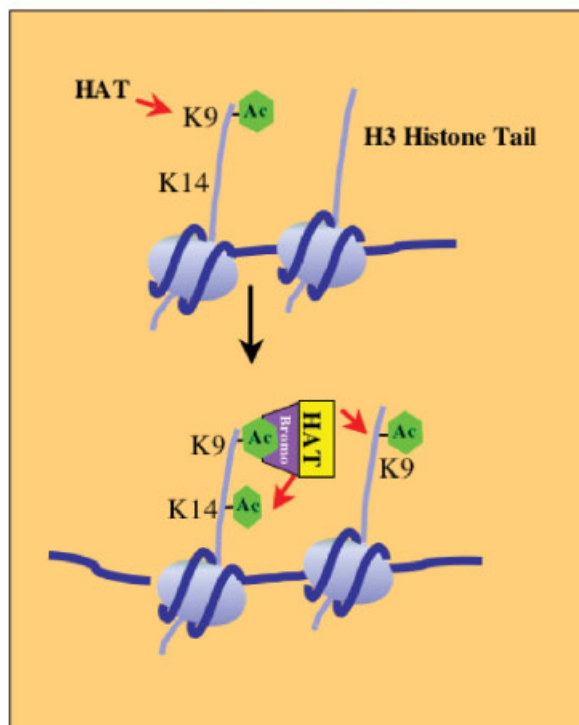


Figure 2. Model for cooperation of the chromatin modifying enzymes and the chromatin-binding domains (illustrated for the case of bromodomains). **A:** Possible functional interactions between acetylated histone tails and bromodomains containing enzymes that lead to a cascade of events to activate transcription. GCN5 acetyltransferase is recruited to the gene promoter by interaction with a transcription factor. GCN5 acetylates H3K9, H3K14 and H4K8. Finally, the bromodomain-containing transcription complexes SWI/SNF and TFIID are recruited to the promoter by specific interactions between their bromodomain and specifically acetylated histone tails. **B:** Possible positive feedback in chromatin signalling mediated by specific interactions between acetylated histone tails and HAT enzymes containing bromodomains that leads to self-perpetuation of activating marks on chromatin.

Structure and function

The structure of the HP1 chromodomain, in complex with a peptide from histone 3 with Lys 9 methylated, consists of a three-stranded antiparallel β -sheet supported by an α -helix that runs across the sheet (Fig. 1B).⁽⁴⁰⁾ The binding pocket for the N-methyl group is provided by three aromatic side chains that become ordered on binding of the peptide.^(28,41,42)

The finding of the chromodomain in two completely different epigenetic repressors, like HP1 and Polycomb,

immediately suggested that chromodomains can have chromatin-related functions.^(37,43) Although the role of the chromodomain within these proteins is not yet fully understood, experimental evidence points to an involvement in protein–protein interactions.^(40,44) In particular, recent work from different laboratories has shown that the HP1 chromodomain can recognise methylation of Lys 9 in histone H3, thus directing the binding of other proteins to control chromatin structure and gene expression.^(14,15,45) The structure of the chromodomain

complex with a histone H3 peptide that includes methylated Lys 9 explains how the binding can take place, with the lysine side chain almost fully extended and surrounded by residues conserved in many chromodomains.⁽⁴⁰⁾ The latter is particularly relevant, as it would support the chromatin-binding ability of chromodomains independently of the protein to which they belong.

It has to be mentioned that, apart from recognising methyl-lysines, chromodomains can also serve to DNA and/or RNA recognition.^(41,42,46–48) For example MOF and MSL-3 use their chromodomains to bind the non-coding roX RNA, crucial for the integrity and targeting of the *Drosophila* dosage compensation complex to the X chromosome;^(49–51) this situation is similar to the role of the HP1 chromodomain in heterochromatin formation, but recognizing RNA instead methyl-Lys. It is interesting to notice that it has been suggested that non-coding RNAs could be involved in some epigenetic regulations and some enzymes, as the chromodomain-containing enzyme SUV39 that trimethylates H3K9, requires an unidentified RNA.⁽⁵²⁾

As mentioned before, the chromodomain is also able to bind DNA. For example, the chromodomain of Mi-2 binds the nucleosome but, surprisingly, deletion of all histone tails does not eliminate such an interaction.⁽⁴⁷⁾ The latter is maintained thanks to a sequence unspecific binding to the nucleosomal DNA.⁽⁴⁷⁾ In contrast to the methyl-lysine binding (see above), structural determinants indicative of RNA- or DNA-binding chromodomain have not yet been identified, although two point mutations on the MOF chromodomain eliminate binding to RNA in vitro.⁽⁴⁸⁾

As for bromodomains, we have looked at the binding site conservation degree of chromodomains (Fig. 1E). We find a substantial conservation degree, supporting a similar role in enzyme targeting for these domains.

The previous data support the idea that chromodomains, like bromodomains, are able to identify and bind specific tags in histones, in particular methylated lysines, and this binding would be vital to recruit different chromatin-modifying enzymes to their targets.

It has also been reported that the chromodomain of HP1 can interact with histone H3 in the absence of the N-terminal tail; however, while this interaction may contribute to chromatin binding in general, it does not explain the specific targeting of HP1.⁽⁵³⁾

Distribution among chromatin-modifying enzymes

The distribution of the chromodomain among these enzymes is more restricted than that of the bromodomain. However, we have found it in HAT, HMT and ATP-dependent chromatin-remodelling enzymes (Table 1).

Chromodomain and HAT enzymes

Chromodomains can be found in human and mouse MORF4L1, identified on the basis of the ability of these

proteins to induce replicative senescence in immortal human cell lines.⁽⁵⁴⁾ The members of this family have a clear similarity to Msl-3 and Eaf3p, both known components of multisubunit histone acetyltransferase complexes.⁽⁵⁵⁾ Msl-3 is a component of the dosage compensation complex that acetylates histone H4 on the male X chromosome at multiple sites.⁽⁵⁶⁾ Eaf3p is a component of the yeast NuA4 HAT complex that carries the Esa1p HAT protein.⁽⁵⁷⁾ This complex also functions by specifically acetylating histone H4 in vivo and has been linked to transcriptional activation and nucleosome remodelling in yeast and flies.^(58,59)

The chromodomain is also present (1) in rat and human Tip60, and *Saccharomyces pombe* SPAC637.12C proteins, members of the Myst family of HAT, and (2) in human and mouse CDY1 and 2 proteins, which contain a putative HAT domain and are components of the heterochromatin-like complexes that act as gene repressors during spermatogenesis.^(60,61)

Chromodomain and HMT enzymes

The chromodomain can be found in different HMT enzymes, in the members of the suvar-3-9 family in *Drosophila* [su(var)3-9], yeast (clr4), insect [su(var)3-9] and mammalian [su(var)3-9H1 and H2].⁽⁶²⁾

Chromodomain and ATP-dependent remodelling enzymes

Chromodomains are found in (1) members of the SNF2/RAD54 helicase family (CHD3 in *Arabidopsis thaliana*, *Drosophila* and *C. elegans*),^(63–65) (2) Mi-2 and human Mi-2 β or CHD4,^(66–68) (3) Mi-2 α or CHD3,⁽⁶⁸⁾ and (4) CHD5.⁽⁶⁹⁾ Mi-2 enzyme remodels chromatin thanks to its ATP hydrolysis ability, and is part of a complex called NURD⁽⁵⁵⁾ that can also deacetylate histones. This suggests that NURD could be specifically targeted to previously methylated chromatin through the chromodomain.

Finally, a chromodomain can also be found in a putative remodelling enzyme, Q8LJJ7, from *Oryza sativa*.

The SANT domain

The SANT domain was identified as a small motif, approximately 50 amino acids, present in nuclear receptor corepressors. Sequence and structure analysis show a clear similarity between the SANT domain and the DNA-binding domain (DBD) of c-Myb related proteins.⁽⁷⁰⁾

Structure and function

The SANT domain consists of three α -helices, each with a bulky aromatic residue, arranged in a helix-turn-helix motif (Fig. 1C).⁽⁷¹⁾ The overall structure is similar to that of Myb DBD, although the SANT domain is functionally divergent from the canonical Myb DBD.⁽⁷²⁾ The SANT domain is present in some ATP-dependent remodelling enzymes complexes: yeast Swi3p, Rsc8p, BAF155/170 and *Drosophila* ISWI. It has been

shown that the SANT domain is essential for the *in vivo* functions of yeast Swi3p, Ada2p and Rsc8p, subunits of three chromatin-remodelling complexes. The general role of the SANT domain is to stabilise, through direct binding, histone N-terminal tails in a conformation favouring their binding to the modifying enzymes, and the subsequent catalytic process.^(73,74) Although the SANT domain interacts primarily with unmodified histone tails, we decided to include it here because (1) it has a central role in chromatin remodelling, being the unique histone-interaction module that couples histone binding to enzyme catalysis, and (2) it is present, like bromodomains and chromodomains, in the two enzyme classes responsible of chromatin modifications (enzymes that catalyse the histones covalent modifications and complexes using ATP hydrolysis). The preference of the SANT domain for unmodified histone tails suggests that histone deacetylation could increase its affinity for histone tails. Interaction with unacetylated histone tails could block the binding of HATs, thus maintaining the deacetylated state, as proposed by Yu and colleagues.⁽⁷⁵⁾

When looking at the conservation degree of the SANT domain residues (Fig. 1F) we find a less-clear trend than for bromo or chromodomains. While significant, the conservation degree of binding site residues is smaller in this case; this can be attributed to the high functional degeneracy of the underlying family (see above), and to the less specific nature of its binding.

Distribution among chromatin-modifying enzymes

The SANT domain is broadly present among ATP-dependent remodelling enzymes and their complexes, but it can also be found in HMTs (Table 1), and in proteins forming part of complexes with HAT and HDAC activities, thus suggesting an important role in regulating chromatin accessibility.

SANT domain and HAT enzymes

We have not found any HAT or HDAC enzyme with a SANT domain as part of its sequence, although some components of HAT and HDAC complexes can have it: (1) SPR1 from *C. elegans*, part of the Co-REST corepressor complex, essential for HDAC1 activation,⁽⁷⁶⁾ (2) human and mouse N-CoR, that interact with HDAC7 and together with Sin3 and HDAC,⁽⁷⁷⁾ and (3) ADA2 proteins, ADA2 α and ADA2 β in yeast, mouse, rat and human. The latter deserve further mention, as ADA2 proteins form part of the SAGA, ADA/GCN5 and PCAF histone acetyl-transferase complexes. Interestingly, it has been observed that ADA and SAGA complexes containing a deletion of the ADA2 SANT domain show a reduced ability to bind non-acetylated histone tails, being inactive in nucleosomal HAT assays.^(73,78)

SANT domains are also present in several subunits of other co-repressor complexes that possess HDAC activity, such as MLL and SMRT. In the latter, the SANT domain functions as a

histone-tail interaction domain that binds to non-acetylated histone H4 peptides.⁽⁷⁵⁾ In addition, the presence of the SANT domain enhances the HDAC activity of the SMRT-HDAC3 complex, by increasing the affinity of the latter for histone tails.⁽⁷⁵⁾

SANT domain and HMT enzymes

SANT domains can be found in several members of the polycomb group of proteins, involved in the repression of homeotic genes and with HMT activity: MEDEA and EZA1 in *Arabidopsis*, EZ1-3 in maize, EZ in *Drosophila*, and EZH1-2 in both mouse and human.

SANT domain and ATP-dependent remodelling enzymes

Some ATP-remodelling enzymes have a SANT domain: *Drosophila* ISWI, (the catalytic subunit of the remodelling complex NURF, CHRAC and ACF), yeast ISWI1 and ISWI2 (catalytic subunits of ISWI1 and ISWI2 complexes, respectively) and human SNF2L and SNF2H (member of the RSF complex). The SANT domain is also present in: yeast Swi3, a component of the SWI/SNF complex; human and mouse MTA1 and MTA2, which are part of the remodelling, deacetylating, complexes NURD, RSC8p and BAF155/170. The presence of the SANT domain in these enzymes suggests that targeting, or stabilization, of the chromatin-enzyme interaction could happen frequently by direct interaction between the SANT domain and the chromatin component. This would help coupling the histone-tail binding and enzymatic activity, as has been previously suggested by Boyer and colleagues.^(73,74)

All of these data suggest that the SANT domain can mediate interactions between remodelling enzymes and their chromatin substrates. More precisely, the SANT domain could contribute: (1) to the recruitment of chromatin modifying enzymes, or (2) to help the interaction between histones and the enzymes. As mentioned before, the latter would follow from the SANT-histone interaction that would improve the histone binding and subsequent catalysis by the modifying enzymes.⁽⁷³⁾

What is the role of chromatin-binding domains?

The data previously discussed (Table 1) show that bromodomains, chromodomains, and SANT domains can be found in the three chromatin-modifying enzymes, or their complexes, although with an unequal distribution among them. While in some cases the enzyme may carry more than one copy of the same domain (Table 1), no combination of different domains has been found (Table 1). Overall, the differential distribution of the chromatin-binding domains among the chromatin-modifying enzymes (Table 1) is in accordance with the idea that these domains can confer specific chromatin-binding properties to the different enzyme families. For example, acetylation at H4K8 could help the recruitment of the

remodelling complex, through interaction with the BRG1 bromodomain, thus contributing to prepare chromatin to be transcribed (Fig. 2A);⁽²⁰⁾ however, this would not be the case for chromatin-remodelling enzymes lacking bromodomains. Or, in the same way, acetylation at H3K14 could help recruitment of some bromodomain-carrying HMTs, that could set a specific combination of activation marks at a given locus, correlated with transcriptional activation, as would be the case for MLL that methylates at H3K4.⁽⁷⁹⁾ However, bromodomain-lacking HMTs, which methylate other positions (as Suvar39 at H3K9) and are involved in silencing, will not be targeted to that specific locus.

Within this context, the histone-binding specificity of domains becomes an important issue that deserves further comment. In particular, a critical question is why domains recognize specifically some modified lysines and not others. We discuss below three main contributions to domain binding specificity: (1) sequence variability at the domain binding site, and neighbouring residues, (2) domain copy number, and (3) allosteric changes induced by protein–protein interactions after chromatin binding.

As we have seen before, there is a substantial degree of sequence conservation at the domains binding site (Fig. 1D–F), supporting the overall conservation of function. However, sequence conservation is not complete and some variability is observed for the different domains that could modulate the domain-binding specificity. This can be illustrated by considering the case of bromodomains and chromodomains. For these two domain types, it seems that not all of them, or their acetylated or methylated targets behave similarly. In the case of chromodomains, swapping experiments have shown a non-uniform functional conservation of this domain in silencing assays.^(80,81) For example, chromodomain of HP1 recognizes methylated H3K9, while the chromodomains from polycomb (M33) and Mi2 do not bind tightly to methylated lysine residues,⁽¹⁴⁾ probably they are able to recognize other chromatin targets, as DNA or RNA. A similar situation is found for bromodomains, for example, the bromodomain in BRG1 binds the H4 tail acetylated at K8 and bromodomain of Brd2 interacts with acetylated H4K12,^(20,82) whereas the double bromodomain in TAF_{II}250 binds the H3 tail acetylated at K9 and K14 (Fig. 2A).^(20,31) In the case of the budding yeast SAGA HAT complex, Gcn5 and Spt7 subunits contain bromodomains capable of binding acetyl-lysines. However, while the Gcn5 bromodomain is essential for tethering SAGA to acetylated nucleosomes arrays *in vitro*, the bromodomain of Spt7 is dispensable. However, if swapped into Gcn5 subunit, the Spt7 bromodomain is capable of anchoring SAGA.⁽³³⁾ The latter suggests that specificity of chromatin-binding domains could depend, at least in part, on the protein context. In this particular case, it seems likely that amino acids flanking acetyl-lysines, as well as non-conserved amino acids in and around the bromodomain, could modulate the binding specificity of the latter.

The presence of more than one chromatin-binding domain can also be critical to determine domain-binding specificity. The data in Table 1 show that chromatin-modifying enzymes contain only one type of chromatin recognition motifs (Table 1). However, the number of domain copies may change, and some histone-modifying enzymes contain two tandem copies of the bromodomain, chromodomain or SANT domain. This domain duplication could contribute to the binding specificity, by increasing the stability of the enzyme-dimodified histone, when the modifications are appropriately spaced in the histone tail. Actually, this is the case for the TAF_{II}250 double bromodomain that binds to diacetylated H3 at K9 and K14.⁽³¹⁾

Finally, allosteric changes induced upon association with transcription factor complexes, and after interaction with the modified chromatin, can also determine domain binding specificity. For example SWI/SNF is recruited to the promoter through the association of the BRG1 bromodomain with the CBP-acetylated H4K8 tail.⁽²⁰⁾ The interaction with other acetylated residues in H3 or H4 may be possible *in vitro*; however, these interactions will not have sufficient strength to ensure stable binding of SWI/SNF to CBP and the promoter.⁽²⁰⁾

Histone-binding domains and long-term regulation

The ideas discussed above can explain how combinations of histone modifications and the chromatin-binding domains that recognize them could regulate short-term transcription. However, they do not completely address the critical issue of what is their contribution to the establishment and maintenance, and even the heritability, of long-term transcriptional states. At present, it is well established that histone modifications have the potential to exert long-term effects, for example H3 methylated at K9 could initiate chromatin condensation and silencing, in part due to its ability to bind proteins such as HP1 through their chromodomains.^(14,15) However, the potential of a single modification to have such long-term effects depends on the chromatin context in which the modification is present. For example, in the context of H3K4 and H4K20 methylated nucleosome, methyl-K9 H3 allows the binding of BRAHMA—the enzyme of the remodelling dSWI/SNF complex, a mark to maintain a long-term transcriptionally active locus.⁽¹⁶⁾

It has been proposed that chromatin-binding domains could play a central role in helping to establish and maintain long-term transcriptional states.⁽⁸³⁾ This would be due to the ability of some enzymes to form self-sustaining marks in chromatin, by stably binding their enzymatic products through the chromatin-binding domains (Fig. 2B).⁽⁸³⁾ This stable binding would in turn allow silencing and/or activating complexes to self-perpetuate their potential. For example, many HAT enzymes bind preferentially to acetylated peptides, *in vitro* binding assays, using their bromodomains.⁽⁸⁴⁾ More

particularly, it has been found that the SAGA complex remains anchored to acetylated arrays of nucleosomes through a GCN5 bromodomain, even after removal of the transcription factor VP16, thus providing a self-perpetuating activating mark on chromatin domain.⁽³³⁾ In contrast the NuA4 complex, which lacks a bromodomain, is not retained following removal of the activators.⁽³³⁾

In the same way, Czermin, Müller and colleagues have shown that Enhancer Zeste, (E(Z)), has methyltransferase activity^(21,22) and H3 methylated by the E(Z) complex binds specifically to polycomb protein, suggesting a direct relationship between H3 methylation by E(Z) and assembly of the PcG silencing complex.⁽²¹⁾

While these data suggest a role for chromatin-binding domains in the long-term transcriptional regulation, it is also quite clear that long-term transcriptional regulation is a complex process, requiring a subtle coordination between histone modifications, DNA methylation and binding of silencing RNA.^(85,86)

Conclusions

The histone code hypothesis provides a useful conceptual framework for understanding how gene expression is modulated through covalent marks in chromatin. Particularly, experimental data published during these last years reinforce the idea that the functional effect (activation or repression) of the translation of the histone code will depend (1) on the combination of histone marks laid down by the enzymes recruited to the gene (by a transcription factor or bromo/chromo/SANT interactions), and (2) also on the chromatin architecture of each gene.

Within this context, an issue that remains open is how chromatin-modifying enzymes are targeted to their histone templates. Recent studies show that some domains, able to specifically recognise histones—chromo, bromo and SANT domains—are also present in different chromatin-modifying enzymes—HAT, HMT and ATP-dependent remodelling enzymes—leading to the proposal that they could contribute to the targeting of histone-modifying enzymes to chromatin targets. Here we have reviewed the distribution of chromatin-binding domains among chromatin-modifying enzymes, finding that it is unequal and supporting the idea that these domains can confer specific chromatin-binding properties to the different enzyme families. In addition, we discuss how factors such as sequence variability, domain copy number and allosteric changes can contribute to modulate the domain-targeting properties.

References

- Konberg RD, Lorch Y. 1999. Chromatin-modifying and remodelling complexes. *Curr Opin Genet Dev* 9:148–151.
- Hebbes TR, Thorne AW, Crane-Robinson C. 1988. A direct link between core histone acetylation and transcriptionally active chromatin. *EMBO J* 7:1395–1402.
- Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun ZW, et al. 2000. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* 406:593–599.
- Wu J, Grunstein M. 2000. 25 years after the nucleosome model: Chromatin modifications. *Trends Biochem Sci* 25:619–623.
- Berger S. 2002. Histone modifications in transcriptional regulation. *Curr Opin Genet Dev* 12:142–148.
- Kingston RE, Narlikar GJ. 1999. ATP-dependent remodelling and acetylation as regulators of chromatin fluidity. *Genes Dev* 13:2339–2352.
- Turner BM, Birley AJ, Lavender J. 1992. Histone H4 isoforms acetylated at specific lysine residues define individual chromosomes and chromatin domains in *Drosophila* polytene nuclei. *Cell* 69:375–384.
- Turner BM. 1993. Decoding the nucleosome. *Cell* 75:5–8.
- Strahl BD, Allis CD. 2000. The language of covalent histone modifications. *Nature* 403:41–45.
- Jenuwein T, Allis CD. 2001. Translating the histone code. *Science* 293:1074–1080.
- Spotswood HT, Turner BM. 2002. An increasingly complex code. *J Clin Invest* 110:577–582.
- Turner BM. 2002. Cellular memory and the histone code. *Cell* 111:285–291.
- Marmorstein R. 2001. Protein modules that manipulate histone tails for chromatin regulation. *Nature Rev* 2:422–432.
- Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, et al. 2001. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* 410:120–124.
- Lachner M, O'Carroll D, Rea S, Mechtler K, Jenuwein T. 2001. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* 410:116–120.
- Beisel C, Imhof A, Greene J, Kremmer E, Sauer F. 2002. Histone methylation by *Drosophila* epigenetic regulator Ash1. *Nature* 419:857–862.
- Zegerman P, Canas B, Pappin D, Kouzarides T. 2002. Histone H3 lysine 4 methylation disrupts binding of nucleosome remodelling and deacetylase (NuRD) repressor complex. *J Biol Chem* 277:11621–11624.
- Sun Z-W, Allis CD. 2002. Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. *Nature* 418:104–108.
- Fischle W, Wang Y, Allis CD. 2003. Binary switches and modification cassettes in histone biology and beyond. *Nature* 425:475–479.
- Agalioti T, Chen G, Thanos D. 2002. Deciphering the transcriptional histone acetylation code for a human gene. *Cell* 111:381–392.
- Czermin B, Melfi R, McCabe D, Seitz V, Imhof A, et al. 2002. *Drosophila* enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* 111:185–196.
- Müller J, Hart CM, Francis NJ, Vargas ML, Sengupta A, et al. 2002. Histone methyltransferase activity of a *Drosophila* polycomb group repressor complex. *Cell* 111:197–208.
- Jeanmougin F, Wurtz J-M, Douarin B, Chambon P, Losson R. 1997. The bromodomain revisited. *TIBS* 22:151–153.
- Tankum JW, Deuring R, Scott MP, Kissinger M, Pattatucci AM, et al. 1992. *brhma*: a regulator of *Drosophila* homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2. *Cell* 68:561–572.
- Horn PJ, Peterson CL. 2001. The bromodomain: a regulator of ATP-dependent chromatin remodelling? *Front Biosci* 6:D1019–D1023.
- Zheng L, Zhou M. 2002. Bromodomain: an acetyl-lysine binding domain. *FEBS Lett* 513:124–128.
- Dhalluin C, Carlson JE, Zeng L, He C, Aggarwal AK, et al. 1999. Structure and ligand of a histone acetyltransferase bromodomain. *Nature* 399:491–496.
- Bottomley M. 2004. Structures of proteins that create or recognize histone modifications. *EMBO Rep* 5:464–469.
- Hudson BP, Martinez-Yamout MA, Dyson HJ, Wright PE. 2000. Solution structure and acetyl-lysine binding activity of the GCN5 bromodomain. *J Mol Biol* 304:355–370.
- Owen DJ, Ormaghi P, Yang J-C, Lowe N, Evans PR, et al. 2000. The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase Gcn5p. *EMBO J* 19:6141–6149.
- Jacobson RH, Landurner AG, King DS, Tjian R. 2000. Structure and function of a human TAF_{II}250 double bromodomain module. *Science* 288:1422–1425.

32. Winston F, Allis CD. 1999. The bromodomain: a chromatin-targeting module? *Nat Struct Biol* 6:601–604.
33. Hassan AH, Prochasson P, Neely KE, Galasinski SC, Chandy M, et al. 2002. Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. *Cell* 111:369–379.
34. Syntichaki P, Topalidou I, Thireos G. 2000. The Gcn5 bromodomain coordinates nucleosome remodelling. *Nature* 404:414–417.
35. Caldas C, Myeong-Hee K, MacGregor A, Cain D, Aparicio S, et al. 1998. Isolation and characterisation of a Pufferfish MLL (mixed lineage leukaemia)-like gene (fMll) reveals evolutionary conservation in vertebrate genes related *Drosophila tritorax*. *Oncogene* 16:3233–3241.
36. Kasten M, Szerlong H, Erdjument-Bromage H, Tempst P, Werner M, et al. 2004. Tandem bromodomains in the chromatin remodeler RSC recognize acetylated histone H3 Lys14. *EMBO J* 23:1348–1359.
37. Paro R, Hogness DS. 1991. Polycomb protein shares a homologous domain with a heterochromatin-associated protein of *Drosophila*. *Proc Natl Acad Sci USA* 88:263–267.
38. Jones DO, Cowell IG, Singh PB. 2000. Mammalian chromodomain proteins: Their role in genome organisation and expression. *Bioessays* 22:124–137.
39. Eisenberg JC. 2001. Molecular biology of the chromodomain: an ancient chromatin module comes of age. *Gene* 275:19–29.
40. Ball LJ, Murzina NV, Broadhurst RW, Raine AR, Archer S, et al. 1997. Structure of the chromatin binding (chromo) domain from mouse modifier protein 1. *EMBO J* 16:2473–2481.
41. Nielsen PR, Nietlispach D, Mott HR, Callaghan J, Bannister A, et al. 2002. Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. *Nature* 416:103–107.
42. Jacobs SA, Khorasanizadeh S. 2002. Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science* 295:2080–2083.
43. Reuter G, Spierer P. 1992. Position effect variegation and chromatin proteins. *Bioessays* 14:605–612.
44. Cowell IG, Austin CA. 1997. Self-association of chromo domain peptides. *Biochim Biophys Acta* 2:198–206.
45. Nakayama J, Rice JC, Strahl BD, Allis CD, Grewal SI. 2001. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* 292:110–113.
46. Jacobs SA, Taverna SD, Zhang Y, Briggs SD, Li J, et al. 2001. Specificity of the HP1 chromo domain for the methylated N-terminus of histone H3. *EMBO J* 20:5232–5241.
47. Bouazzone K, Mitterweiger A, Langst G, Imhof A, Becker PB, et al. 2002. The dMi-2 chromodomains are DNA binding modules important for ATP-dependent nucleosome binding and mobilization properties. *EMBO J* 21:2430–2440.
48. Akhtar A, Zink D, Becker PB. 2000. Chromodomains are protein-RNA interaction modules. *Nature* 407:405–409.
49. Gu W, Szauter P, Lecchesi JC. 1998. Targeting of MOF, a putative histone acetyl transferase, to the X chromosome of *Drosophila melanogaster*. *Dev Genet* 22:56–64.
50. Kelley RL, Meller VH, Gordadze PR, Roman G, Davis RL, et al. 1999. Epigenetic spreading of the *Drosophila* dosage compensation complex from roX RNA genes into flanking chromatin. *Cell* 98:513–522.
51. Meller VH, Gordadze PR, Park Y, Chu X, Stuckenholz C, et al. 2000. Ordered assembly of the roX RNAs into MSL complexes on the dosage-compensated X chromosome in *Drosophila*. *Curr Biol* 10:136–143.
52. Maison C, Bailly D, Peters AH, Quivy JP, Roche D, et al. 2002. Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. *Nat Genet* 30:329–334.
53. Nielsen AL, Oulad-Abdelghani M, Ortiz JA, Remboutsika E, Chambon P, et al. 2001. Heterochromatin formation in mammalian cells: interaction between histones and HP1 proteins. *Mol Cell* 7:729–739.
54. Bertram MJ, Berube NG, Hang-Swanson X, Ran Q, Leung JK, et al. 1999. Identification of a gene that reverses the immortal phenotype of a subset of cells and is a member of a novel family of transcription factors-like genes. *Mol Cell Biol* 19:1479–1485.
55. Bertam MJ, Pereira-Smith OM. 2001. Conservation of the MORF4 related gene family: identification of a new chromo domain subfamily and a novel protein motif. *Gene* 266:111–121.
56. Smith ER, Pannuti A, Gu W, Steurnagel A, Cook RG, et al. 2000. The *drosophilae* MSL complex acetylates histone H4 at lysine 16, a chromatin modification linked to dosage compensation. *Mol Cell Biol* 20:312–318.
57. Eisen A, Utley RT, Nourani A, Allard S, Schmidt P, et al. 2001. The yeast NuA4 and *Drosophila* MSL complexes contain homologous subunits important for transcriptional regulation. *J Biol Chem* 276:3484–3491.
58. Allard S, Utley RT, Savard J, Clarke A, Grant P, et al. 1999. NuA4, an essential transcription adaptor/histone H4 acetyltransferase complex containing Esa1p and the ATM-related cofactor Tra1p. *EMBO J* 18:5108–5119.
59. Roth SY, Denu JM, Allis CD. 2001. Histone acetyltransferases. *Annu Rev Biochem* 70:81–120.
60. Lahn BT, Tang ZL, Zhou J, Berndt RJ, Parvinen M, et al. 2002. Previously uncharacterized histone acetyltransferases implicated in mammalian spermatogenesis. *Proc Natl Acad Sci USA* 99:8707–8712.
61. Kleiman SE, Yogev L, Hauser R, Botchan A, Bar-Shira Maymon B, et al. 2003. Members of the CDY family have different expression patterns: CDY1 transcripts have the best correlation with complete spermatogenesis. *Hum Genet* 113:486–492.
62. Ivanova AV, Bonaduce MJ, Ivanov SV, Klar SJS. 1998. The chromo and SET domains of the Ctr4 protein are essential for silencing in fission yeast. *Nat Genet* 19:192–195.
63. Ogas J, Kaufmann S, Henderson J, Somerville C. 1999. PICKLE is a CHD3 chromatin-remodeling factor that regulates the transition from embryonic to vegetative development in *Arabidopsis*. *Proc Natl Acad Sci USA* 96:13839–13844.
64. Kehle J, Beuchle D, treuheit S, Christen B, Kennison JA, et al. 1998. dMi-2, a hunchback-interacting protein that functions in Polycomb repression. *Science* 282:1897–1900.
65. Zelewsky TV, Palladino F, Brunschwig K, Tobler H, Hajnal A, et al. 2000. The *C. elegans* Mi-2 chromatin-remodelling proteins function in vulval cell fate determination. *Development* 127:5277–5284.
66. Seelig HP, Moosbrugger I, Ehrfed H, Fink T, Renz M, et al. 1995. The major dermatomyositis specific Mi-2 autoantigen is a presumed helicase involved in transcriptional activation. *Arthritis Rheum* 38:389–1399.
67. Ahringer J. 2000. NuRD and SIN3 histone deacetylase complexes in development. *Trends Genet* 16:351–356.
68. Woodage T, Basrai MA, Baxevanis AD, Hieter P, Collins FS. 1997. Characterization of the CHD family of proteins. *Proc Natl Acad Sci USA* 94:11472–11477.
69. Schuster EF, Stoeger RJ. 2002. CHD5 defines a new subfamily of chromodomain-SWI2/SNF2-like helicases. *Mamm Genome* 13:117–119.
70. Aasland R, Stewart AF, Gibson T. 1996. The SANT domain: a putative DNA-binding domain in the SWI_SNF and ADA complexes, the transcriptional co-repressor N-CoR and TFIIIB. *Trends Biochem Sci* 21:87–88.
71. Grüne T, Brzeski J, Eberharter A, Clapier CR, Corona DF, et al. 2003. Crystal structure and functional analysis of a nucleosome recognition module of the remodeling factor ISWI. *Mol Cell* 12:449–460.
72. Ogata K, Morikawa S, Nakamura H, Sekikawa A, Inoue T, et al. 1994. Crystal structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices. *Cell* 79:639–648.
73. Boyer LA, Langer MR, Crowley KA, Tan S, Denu JM, et al. 2002. Essential role for the SANT domain in the functioning of multiple chromatin remodeling enzymes. *Mol Cell* 4:935–942.
74. Boyer LA, Latek RR, Peterson CL. 2004. The SANT domain: a unique histone-tail-binding module? *Nat Rev Mol Cell Biol* 5:158–163.
75. Yu J, Li Y, Ishizuka T, Guenther MG, Lazar MA. 2003. A SANT motif in the SMRT corepressor interprets the histone code and promotes histone deacetylation. *EMBO J* 13:3403–3410.
76. You A, Tong JK, Grozinger CM, Schreiber SL. 2001. CoRest is an integral component of the CoREST human histone deacetylase complex. *Proc Natl Acad Sci USA* 98:1454–1458.
77. Fischle W, Dequiedt F, Fillion M, Hendzel MJ, Voelter W, et al. 2001. Human HDAC7 histone deacetylase activity is associated with HDAC3 in vivo. *J Biol Chem* 276:35826–35835.
78. Sterner DE, Wang X, Bloom MH, Simon GM, Berger SL. 2002. The SANT domain of Ada2 is required for normal acetylation of histones by the yeast SAGA complex. *J Biol Chem* 277:8178–8186.
79. Milne TA, Briggs SD, Brock HW, Martin ME, Gibbs D, et al. 2002. MLL targets SET domain methyltransferase activity to Hox gene promoters. *Mol Cell* 5:1107–1117.

80. Platero JS, Harnett T, Eissenberg JC. 1995. Functional analysis of the chromo domain of HP1. *EMBO J* 14:3977–3986
81. Wang G, Ma A, Chow CM, Horsley D, Brown NR, et al. 2000. Conservation of heterochromatin protein 1 function. *Mol Cell Biol* 20:6970–6983.
82. Kanno T, Kanno Y, Siegel RM, Jang MK, Lenardo MJ, et al. 2004. Selective recognition of acetylated histones by bromodomain proteins visualized in living cells. *Mol Cell* 1:33–43.
83. Cosma MP, Tanaka T, Nasmyth K. 1999. Ordered recruitment of transcription and chromatin remodeling factors to a cell cycle- and developmentally regulated promoter. *Cell* 97:299–311.
84. Ornaghi P, Ballario P, Lena AM, Gonzalez A, Filetici P. 1999. Bromodomain of Gcn5p interacts in vitro with specific residues in the N terminus of histone H4. *J Mol Biol* 1:1–7.
85. Volpe TA, Kidner C, Hall IM, Teng G, Grewall SIS, et al. 2002. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* 297:1833–1837.
86. Richards EJ, Elgin SC. 2002. Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. *Cell* 108:489–500.
87. Kraulis PJ. 1991. MOLSCRIPT: A Program to Produce Both Detailed and Schematic Plots of Protein Structures. *J Appl Cryst* 24:946–950.
88. Shannon CE. 1948. A mathematical theory of communication. *Bell System Tech J* 27:379–423.
89. Valdar WS. 2002. Scoring residue conservation. *Proteins* 48:227–241.
90. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, et al. 2004. The Pfam Protein Families Database. *Nucleic Acids Res Database Issue* 32:D138–D141.
91. Letunic I, Goodstadt L, Dickens NJ, Doerks T, Schultz J, et al. 2002. Recent improvements to the SMART domain-based sequence annotation resource. *Nucleic Acids Res* 30:242–244.

PUBLICACIÓN 2

The functional modulation of epigenetic regulators by alternative splicing

Sergio Lois, Noemí Blanco, Marian Martínez-Balbás y Xavier de la Cruz

BMC Genomics, 8:252, 2007

TÍTULO:

La modulación funcional de los reguladores epigenéticos mediante *splicing* alternativo.

RESUMEN:

Los reguladores epigenéticos juegan un papel fundamental en el control de la expresión génica al modificar el estado local de la cromatina. Sin embargo, debido a su reciente descubrimiento, poco se sabe sobre su regulación. De entre los posibles niveles de regulación (transcripción, traducción, *splicing* alternativo del ARNm, etc.) esta publicación se centra en el estudio del *splicing* alternativo, y su capacidad para generar isoformas con un variado espectro funcional.

En lo que se refiere a los datos primarios, es decir genes y tipos de dominios analizados, etc..., este trabajo presenta una diferencia clara respecto al artículo anterior: nos beneficiamos de la mayor disponibilidad en las bases de datos públicas de nuevas familias de reguladores epigenéticos y de anotaciones de dominios involucrados en la adición, eliminación y reconocimiento de modificaciones de las histonas. Ello permitió que esta publicación fuera más exhaustiva e incluyese información sobre 160 genes que codifican diferentes reguladores epigenéticos (HATs, HDACs, HMTs, HDMTs, enzimas remodelantes de la cromatina dependientes de ATP, ubiquitininas, quinasas, etc...).

Los resultados de esta publicación muestran que aproximadamente el 49 % (70 % en humanos) de los genes analizados, expresan más de un transcrito, y que el 60 % (64 % en humanos) de estos genes presentan isoformas con al menos un dominio perdido o significativamente afectado. Los cambios introducidos en la estructura de dominios de las diferentes isoformas analizadas se asocian a diferentes mecanismos de regulación (represión de la función nativa de la enzima y creación de isoformas con nuevas funciones).

Finalmente, se discute el papel del *splicing* alternativo como mecanismo de modulación funcional de las enzimas modificadoras de la cromatina; considerando que estas enzimas controlan el estado transcripcional de grandes conjuntos de genes, se propone que la regulación epigenética de la expresión génica podría estar fuertemente regulada mediante *splicing* alternativo.

Research article

Open Access

The functional modulation of epigenetic regulators by alternative splicing

Sergio Lois^{1,2}, Noemí Blanco^{1,2}, Marian Martínez-Balbás*^{1,2} and Xavier de la Cruz*^{2,3}

Address: ¹Instituto de Biología Molecular de Barcelona. CID. Consejo Superior de Investigaciones Científicas (CSIC); 08028 Barcelona, Spain, ²Institut de Recerca Biomèdica-PCB; 08028 Barcelona, Spain and ³Institució Catalana de Recerca i Estudis Avançats (ICREA); Barcelona, Spain

Email: Sergio Lois - sergi@mmb.pcb.ub.es; Noemí Blanco - nbgbmc@ibmb.csic.es; Marian Martínez-Balbás* - mmbbmc@ibmb.csic.es; Xavier de la Cruz* - xavier@mmb.pcb.ub.es

* Corresponding authors

Published: 25 July 2007

Received: 24 April 2007

BMC Genomics 2007, 8:252 doi:10.1186/1471-2164-8-252

Accepted: 25 July 2007

This article is available from: <http://www.biomedcentral.com/1471-2164/8/252>

© 2007 Lois et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Epigenetic regulators (histone acetyltransferases, methyltransferases, chromatin-remodelling enzymes, etc) play a fundamental role in the control of gene expression by modifying the local state of chromatin. However, due to their recent discovery, little is yet known about their own regulation. This paper addresses this point, focusing on alternative splicing regulation, a mechanism already known to play an important role in other protein families, e.g. transcription factors, membrane receptors, etc.

Results: To this end, we compiled the data available on the presence/absence of alternative splicing for a set of 160 different epigenetic regulators, taking advantage of the relatively large amount of unexplored data on alternative splicing available in public databases. We found that 49 % (70 % in human) of these genes express more than one transcript. We then studied their alternative splicing patterns, focusing on those changes affecting the enzyme's domain composition. In general, we found that these sequence changes correspond to different mechanisms, either repressing the enzyme's function (e.g. by creating dominant-negative inhibitors of the functional isoform) or creating isoforms with new functions.

Conclusion: We conclude that alternative splicing of epigenetic regulators can be an important tool for the function modulation of these enzymes. Considering that the latter control the transcriptional state of large sets of genes, we propose that epigenetic regulation of gene expression is itself strongly regulated by alternative splicing.

Background

Epigenetic regulation of gene expression constitutes a fundamental mechanism by which a series of chromatin modifications allow the normal functioning of the cell under different conditions [1-3]. In particular, these modifications control the repressive effect of chromatin, which

limits the access of regulatory proteins to DNA, thus posing serious restraints to biological processes like replication, transcription, etc [4]. In agreement with this, an increasingly large amount of experimental data shows the relevance of chromatin modifications in development [5], disease [6], etc. For example, recent studies indicate that

histone modifications are involved in paternal X chromosome inactivation [7,8]. Work from Roopra and colleagues [9] shows that histone methylation regulates the tissue-dependent silencing of neuronal genes. Also, expression of Hox transcription factors is directly related to the presence of histone marks [10].

Chromatin modifications are produced by a series of chromatin-modifying enzymes (epigenetic regulators) that act on chromatin by either introducing histone modifications or by inducing ATP-dependent nucleosome remodelling. Histone modifications usually take place at histone tails and can introduce a wide variety of covalent marks including acetylation, methylation, phosphorylation, etc [2]. These marks provide a simple way to access nucleosomal DNA and normally have different functional consequences [2,11-14]. A synthetic view of the biological role of histone modifications is provided by the histone code hypothesis [1]. According to this hypothesis, the regulatory state of a gene is a function of these modifications and their combinations. Apart from histone-modifying enzymes, enzymes that utilise ATP to modify the nucleosomal structure, altering histone-DNA interactions [15], also give access to nucleosomal DNA. Interestingly, both mechanisms are coordinated and cooperate to finally give access to nucleosomal DNA. For example, it has been recently shown that the SWI/SNF complex is retained to the chromatin only if SAGA or NuA4 acetylate it [16].

As with transcription factors [17,18], the functional activity of chromatin-modifying enzymes must be regulated in order to produce gene expression patterns that are coherent with high-level biological processes, like development or tissue differentiation. However, little is yet known about how this regulation occurs, due to the recent discovery of these enzymes [2,3,19]. Among the possible regulation levels [18], like transcription, translation or mRNA splicing, in this work we have focused on the study of the latter. We have chosen alternative splicing for four different reasons. First, because recent data [20-23] strongly suggest that alternative splicing can introduce functionally relevant changes in chromatin-modifying enzymes. Second, because alternative splicing is already known to play an important role in gene expression regulation by modulating the functional properties of transcription factors [17,18], for example, alternative splicing can change the DNA-binding properties of transcription factors [24]; introduce or eliminate activating domains [25], increase the *in vivo* stability of a given isoform [26], etc. Third, because of the availability, in public databases, of a large amount of unexplored information on alternative splicing patterns of chromatin-modifying enzymes is available in public databases. And fourth, because the functional and regulatory impact of the most frequent alternative splicing events -in particular long sequence insertions/deletions- is

relatively easier to infer, particularly if it affects known protein domains [17].

In our work we have studied (i) whether, and to which extent, epigenetic regulators (ATP-dependent remodelling enzymes, histone acetyltransferases, deacetylases, methyltransferases, etc) have alternative splicing, and (ii) the impact of alternative splicing on the domain structure of these enzymes, with special focus on catalytic and interaction domains, which are known to play a key role [2,3,27,28]. We obtained the alternative splicing data from databases with very different curation protocols, going from literature surveys, like SwissProt [29], to that of highly automated methods based on sequence processing and EST data, like ENSEMBL [30]. Our results show that a substantial percentage of epigenetic regulators, 49 % (70 % for human genes), have alternative splicing. In addition, in more than 59 % of these cases alternative splicing changes affect either the catalytic or the interaction domain (Figure 1), suggesting the existence of functional regulatory effects comparable to those found in transcription factors [17].

Results and discussion

A set of 160 genes, from different species, of chromatin-modifying enzymes was considered in this work. These enzymes cover the following activities: ATP-dependent chromatin remodelling, histone acetylation, deacetylation, methylation, demethylation, phosphorylation, ubiquitination, and sumoylation. We find (Table 1) that 49 % of the genes show alternative splicing, with an average number of 2.8 isoforms per gene. In humans, this number goes up to 70 % (with 2.8 isoforms per gene), a value close to one of the largest estimates obtained for human, *e.g.* 74 % [31]. This result points to a significant



Figure 1
Alternative splicing pattern of human histone methyltransferase SUV39H2. Representation of the domain structure of three isoforms of SUV39H2, together with their sizes. Shown in red are the two domains, PRE-SET and SET that constitute the catalytic unit of the enzyme. The interaction domain, chromodomain, is shown in green. This domain is seriously damaged in the second isoform, and is unlikely to play any targeting role. The catalytic unit, on the contrary, remains intact in this second isoform, but is clearly damaged in the third isoform, with 28 % of the SET domain and the whole PRE-SET domain missing.

role of alternative splicing in the modulation of the functional properties of chromatin-modifying enzymes.

To characterise the functional variability introduced by alternative splicing in chromatin-modifying enzymes, we compared the different isoforms of the same gene at the protein sequence level, using the longest isoform as a reference. We focused our study on the changes affecting protein domains of known function, because they can be reliably interpreted in terms of biochemical/biological function [17]. For example, it has been experimentally shown that domain changes between isoforms can be associated to isoforms with [17,32]: a dominant-negative role, different binding affinities or new interaction partners, modified enzymatic activity, etc.

In our case, we observe that 60 % (64 % for human) of the genes with alternative splicing have isoforms with at least one missing, or significantly affected, domain (Table 1). These cases can be grouped according to the functional role of the domain: (i) changes in the catalytic domains; (ii) changes in the protein interaction domains; and (iii) drastic sequence reductions. There are only four exceptions to this broad classification, corresponding to the small, single-domain, human proteins: ubiquitin-conjugating enzyme E2A (UBE2A, 154 aas), casein kinase 2, alpha 1 polypeptide (CKII, 391 aas), NAD-dependent deacetylase sirtuin-2 (SirT2, 389 aas) and aurora kinase B (AURKB, 344 aas) for which interaction and catalytic domains coincide. In these cases, alternative splicing modifications will affect both functions.

We discuss below the three above-mentioned scenarios.

(i) Changes in the catalytic domains

In the human, we find several genes with isoforms that have the catalytic domain either missing or affected (Table 2). In a short isoform of the histone methyltransferase SUV39H2 (Figure 1), the catalytic unit is seriously damaged by the loss of the whole PRESET domain, and about 30 % of the SET domain. The situation seems different for chromatin remodelling SMARCA1's and kinase PRKDC's short isoforms, which only lack 11 % and 8 % of their respective catalytic domains (Table 2). However, visual inspection of the catalytic domains' structures shows that

the changes are far from being structurally neutral. The deletion affecting the helicase domain DEXHC of the chromatin-remodelling enzyme SMARCA1 involves an alpha helix linking two of the most extreme strands of the central beta sheet (Figure 2A). The deletion affecting the catalytic PI3_PI4_KINASE domain of the kinase PRKDC affects a beta sheet, eliminating one strand and altering the inter-strand connectivity (Figure 2B). In both cases, the changes will produce either structural strain, or significant rearrangements, likely to result in function loss/modification. Indeed, recent experimental data for kinase PRKDC [23] show that the protein kinase activity of the short isoform of this enzyme is lost.

Inactivation of the enzyme's catalytic function by alternative splicing is also found in one of maize methyltransferase mezz2's isoforms that has completely lost its SET domain (Table 2).

Two cases deserve additional comment. CARM1 (coactivator-associated arginine methyltransferase 1) has an alternative splice isoform, the catalytic domain of which, SKB1, is clearly damaged (48 % of the domain is lost). We have classed CARM1 within this section, even though an interaction domain has not yet been identified, because the full-length isoform is big enough (608 aas) to have both an interaction domain and a catalytic domain. The second case is that of RPS6KA5 (ribosomal protein S6 kinase, 90 kDa, polypeptide 5) which has two catalytic domains, but no interaction domain. In this case, lack of one of the catalytic domains may result in either an inactive or a less active protein. This situation would be equivalent to an amount regulatory mechanism similar to that described for other enzymes.

In general, alternative splicing isoforms with a missing catalytic domain may behave as dominant-negative regulators of the fully functional isoform, a well-known situation in the case of transcription factors [17,33]. This may be the case in chromatin-modifying enzymes. Indeed, a recently described PRKDC isoform with no protein kinase domain has no catalytic activity and shows slight inhibitory activity of the full-length isoform [23]. However, the situation may be more complex, as for example the short PRKDC isoform described here is able to participate in some DNA repair processes, despite having no kinase activity [23]. Thus we cannot rule out the possibility that, in some cases, isoforms lacking the catalytic unit may have functional roles other than being dominant-negative regulators.

(ii) Changes in the protein interaction domains

As for the previous case, the effect of alternative splicing can range from partial deletion to complete domain loss (Table 3). In the human, we find the latter in several

Table 1: Summary of the data utilised in this work

	Number of genes	Number of genes with AS	Number of genes with AS involving protein domains
All species	160	78	46
<i>Homo sapiens</i>	71	50	32
<i>Mus musculus</i>	31	21	10

Table 2: Cases for which alternative splicing sequence changes mainly affect catalytic domains

Gene name	Species	Reference Isoform Size	Alternative Isoform Size	Domains affected
SUV39H2	H.s.	410	230	PRESET, SET
SMARCA1 (SNF2L)	H.s.	1054	1033	DEXHC
PRKDC (DNA-PK)	H.s.	4127	4097	PI3_PI4 KINASE
RPS6KA5 (MSK1)	H.s.	802	549	PKINASE
EZH2	H.s.	751	376	SET
EHMT2 (G9a)	H.s.	1210	202	ANK, PRESET, SET*
CARM1 (PRMT4)	H.s.	608	412	SKBI
SETDB1	H.s.	1290	397	MBD, PRESET, SET, TUDOR
EHMT1	H.s.	1267	1153	SET
FBXL11 (JHDM1A)	H.s.	1162	856	JMJC
AOF2 (LSD1)	H.s.	876	852	AMINO_OXIDASE
GSG2 (HASPIN)	H.s.	798	314	PKINASE
PRDM2 (RIZ1)	H.s.	1718	1481	SET
Setdb1	M.m.	1308	488	MBD, PRESET, SET
Htatip	M.m.	546	492	MOZ_SAS
Fbxl10 (Jhdm1b)	M.m.	1309	776	JMJC
Fbxl10 (Jhdm1b)	M.m.	1309	656	JMJC
Jmjd1b (Jhdm2b)	M.m.	1562	1124	JMJC
fbxl10 (Jhdm1b)	X.l.	1259	738	JMJC
mez2	Z.m.	894	624	SET

genes, for example GCN5L2, MYST1 and MORF4L1. The first of them expresses two isoforms lacking the PCAF_N domain, which is involved in the interaction between the histone acetyltransferase GCN5L2 and CBP. For histone acetyltransferase MYST1, the chromodomain is lost together with a substantial part of the protein, but the catalytic domain is left intact. The case of the histone acetyltransferase MORF4L1 is somewhat surprising, as it is the short isoform that shows the chromodomain, after deletion of a sequence stretch that is in the middle of the domain's sequence in the long isoform [20].

In other cases the impact caused by alternative splicing changes is such that, from a functional point of view, it is essentially equivalent to a domain loss. In general, a simple measure, like size, is usually enough to understand the damaging nature of the change. This is the case of human histone methyltransferase SUV39H2 that has an isoform with only 68 % of its chromodomain (Figure 1). The deleterious effect of this deletion on protein function is supported by visual inspection of the corresponding domain structure that points to a disruption of important secondary structure elements (Figure 3A). Interestingly, even small changes are likely to inactivate the domain's function. For example, chromatin remodelling SMARCA2's bromodomain only loses 14 % of its residues, but analysis of the three-dimensional structure shows that a relevant alpha helix from the helix bundle structure is lost, pointing to a disruption of such a small structure (Figure 3B).

Lack of a whole interaction domain is also found in other species, for example in the short isoform of the mouse histone acetyltransferase Htatip (Tip60), which has a missing chromodomain (Table 3). It has to be noted that in this case a significant part of the protein is also missing (the short isoform is about half the size of the long isoform). Thus, while the catalytic domain, MOZ_SAS, is preserved, it may happen that some unknown domains are also lost. Interestingly, the case of the human histone acetyltransferase MORF4L1 also appears in mouse.

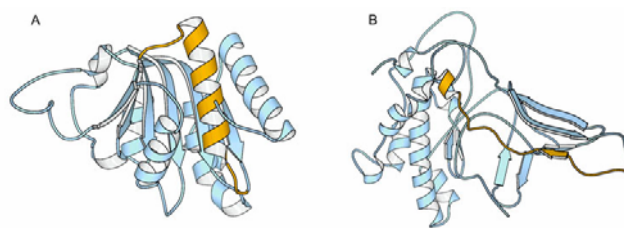


Figure 2
Impact of alternative splicing in catalytic domains. In all cases the part of the protein affected by alternative splicing is shown in yellow, while the remaining of the protein is shown in blue. (A) Domain DEXHC of human chromatin remodelling SMARCA1. Alternative splicing results in the loss of a α -helix. (B) Domain PI3_PI4_KINASE of kinase PRKDC. Alternative splicing results in the loss of a sequence stretch that has very distant ends. The figures were obtained using the MOLSCRIPT software [65].

Table 3: Cases for which alternative splicing sequence changes mainly affect interaction domains

Gene name	Species	Reference Isoform Size	Alternative Isoform Size	Domains affected
SUV39H2	H.s.	410	350	CHROMO
GCN5L2	H.s.	837	476	PCAF_N
GCN5L2	H.s.	837	427	PCAF_N
MYST-1	H.s.	467	300	CHROMO
SMARCA2 (BRM)	H.s.	1590	1572	BROMO
MLL	H.s.	3969	3931	PHD
MORF4LI	H.s.	362	333	CHROMO
MORF4LI	H.s.	362	323	CHROMO
FBXL10 (JHDM1B)	H.s.	1336	1326	LRR_RI
FBXL10 (JHDM1B)	H.s.	1336	1306	LRR_RI
JMJ2B (JHDM3B)	H.s.	1096	448	PHD, TUDOR
MLL2	H.s.	5265	4957	RING, PHD
MLL3	H.s.	4911	4029	PHD
NSD1	H.s.	2696	2593	PWWP
RNF40	H.s.	1001	838	RING, ZF_C3HC4
Morf4l1	M.m.	362	323	CHROMO
Htatip	M.m.	546	302	CHROMO
Fbxl11 (Jhdm1a)	M.m.	1161	494	ZF_CXXC
Fbxl11 (Jhdm1a)	M.m.	1161	338	ZF_CXXC
Jmjd2a (Jhdm3a)	M.m.	1064	1033	PHD, TUDOR
Jmjd2b (Jhdm3b)	M.m.	1086	1021	TUDOR
cbp-1	C.e.	2056	2045	ZNF_TAZ

In all these cases the *a priori* functional meaning of the loss of protein interaction domains is similar and would correspond to a down-regulation of the enzyme's activity. The underlying molecular mechanisms will vary depending on the nature of the interaction lost with the missing domain. If this interaction is required for the formation of a complex between the enzyme and its partners, necessary for the catalysis, down-regulation will result from the formation of inactive complexes. This is probably the case of the short isoform of histone acetyltransferase GCN5L2.

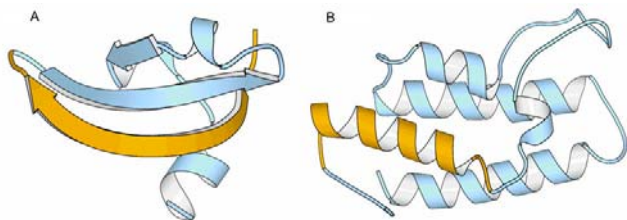


Figure 3
Impact of alternative splicing in interaction domains.

In all cases the part of the protein affected by alternative splicing is shown in yellow, while the remaining of the protein is shown in blue. (A) Chromodomain of human histone methyltransferase SUV39H2. One of the main strands of the β -sheet is missing in one of the alternative splice isoforms. (B) Bromodomain of human chromatin remodelling SMARCA2. One of the four helices of the helix bundle is lost in the alternative splice isoforms. The figures were obtained using the MOLSCRIPT software [65].

If the missing domain is responsible for substrate targeting, e.g. a chromodomain or a bromodomain, down-regulation will be a consequence of the enzyme being unable to reach its substrate. However, in this case another option is also possible, as the enzyme could be recruited to its reaction site after binding one of its complex's partners. The resulting effect on the regulation of gene expression may be substantially different in this case, as modification of the histone tail will take place. However, lack of the chromatin-binding domain will eliminate the positive feedback in chromatin signalling. The latter is mediated by specific interactions between the modified histone tails and the corresponding enzymes and leads to self-perpetuation of activating marks on chromatin. This effect has been recently proposed for enzymes carrying the bromodomain [16,34].

Lastly, we also find instances where alternative splicing is likely to result in small modulatory changes. For example, in histone methyltransferase MLL only one of the three PHD domains is affected by alternative splicing. The small size of the change, 11 % of the domain, and the fact that the other two PHD domains remain intact, points to a modulation of the enzyme's binding properties rather than to a complete inactivation. For *C.elegans*'s histone acetyltransferase cbp-1, the situation is similar as only one of the two copies of the protein interaction domain ZNF_TAZ is affected, by a small change that happens at a relatively neutral location (Figure 4).

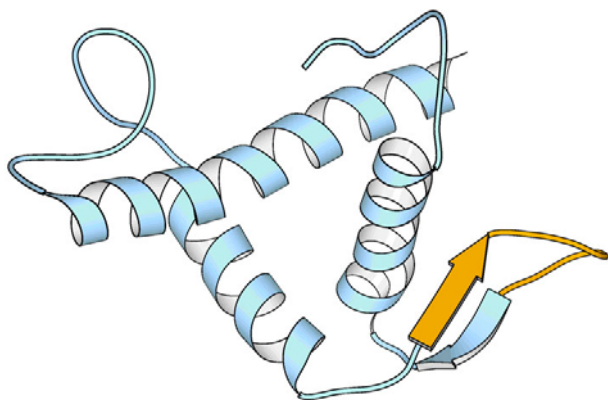


Figure 4
Impact of alternative splicing in the ZNFTAZ domain of *C.elegans*'s histone acetyltransferase *cbp-1*. A small strand (yellow) is lost in one of the alternative splice isoform. Only small changes can be expected from this deletion. The figure was obtained using the MOLSCRIPT software [65].

(iii) Drastic sequence changes

Generation of inactive isoforms constitutes a simple and powerful mechanism to regulate the amount of functional protein present in the cell [35-37]. Usually, inactive isoforms are short versions of the fully active protein in which most functional domains are missing [36]. For several genes we find isoforms that fit this description and

thus could be inactive isoforms (Table 4). In all of them the size reduction relative to the active protein is dramatic, between 35 % and 95 %, and most of the functional domains are lost or seriously damaged. For example, in the case of the human kinase ATM, the functional protein is 3056 residues long, whilst there is a short isoform associated to this gene with only 138 residues (Table 4). Catalysis-associated domains like FAT, FATC and PI3_PI4_KINASE, are missing from the short isoform, together with most of the non-annotated parts of the sequence. It is improbable that such isoform may have any functional role itself and is thus likely to be the result of the above-mentioned regulatory process. We observe a similar situation for ubiquitin-conjugating enzyme E2 A (UBE2A), which has two isoforms lacking 47 % and 22 % of the UBCC domain. The damaging effect of the missing sequence is supported by visual inspection of the corresponding domain structures (Figure 5).

It has to be noted, however, that short isoforms may not always be the consequence of a regulatory process aiming at reducing the amount of functional protein. In some genes, for example in the case ankyrin-3 [38], they have a specific functional role. This could also be the case for some of the transcripts mentioned in this section.

Furthermore, we cannot completely discard the possibility that some of these cases correspond to database annotation errors.

Table 4: Cases for which alternative splicing sequence changes result in drastically affected isoforms

Gene name	Species	Reference Isoform Size	Alternative Isoform Size	Domains affected
SETDB1	H.s.	1290	249	MBD, PRESET, SET, TUDOR
SETDB1	H.s.	1290	151	MBD, PRESET, SET, TUDOR
SMARCA2 (BRM)	H.s.	1590	278	HSA, BRK, DEXHC, HELICASE_C, BROMO
SMARCA2 (BRM)	H.s.	1590	254	HSA, BRK, DEXHC, HELICASE_C, BROMO
SMARCA2 (BRM)	H.s.	1590	236	HSA, BRK, DEXHC, HELICASE_C, BROMO
SMARCA2 (BRM)	H.s.	1590	119	HSA, BRK, DEXHC, HELICASE_C, BROMO
SMARCA4 (BRG1)	H.s.	1679	628	BRK, BROMO, DEXHC, HSA
SUV39H1	H.s.	412	409	CHROMO, PRESET, SET
MLL	H.s.	3969	511	BROMO, FYRC, FYRN, PHD, SET, ZF-CXXC
ATM	H.s.	3056	138	FAT, FATC, PI3_PI4 KINASE
MORF4L1	H.s.	362	235	MRG
EHMT1	H.s.	1267	825	ANK, PRESET, SET
WBP7 (MLL4)	H.s.	2715	582	ZF_CXXC, PHD, FYRC, FYRN, SET
Setdb1	M.m.	1308	500	MBD, PRESET, TUDOR
Stk4	M.m.	487	126	PKINASE
Htatip	M.m.	546	302	CHROMO
Suv39h2	M.m.	477	257	CHROMO, PRESET, SET
Fbxl10 (Jhdm1b)	M.m.	1309	114	JMJC, ZF_CXXC
Su(var)3-9	D.m.	635	475	CHROMO, PRESET, SET
mez2	Z.m.	894	341	SET

In the "Gene name" column we list the standard names of the proteins, although in some cases we also provide alternative names that are frequently used in the literature. In the "Species" column H.s., M.m., D.m., C.e., O.s., X.l. and Z.m. mean *Homo sapiens*, *Mus musculus*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Oryza sativa*, *Xenopus laevis* and *Zea mays*, respectively. The sizes of the different isoforms are given in amino acid number. *In this case, although the ANK protein interaction domain is lost, the NFSP transcription factor binding domain is retained.

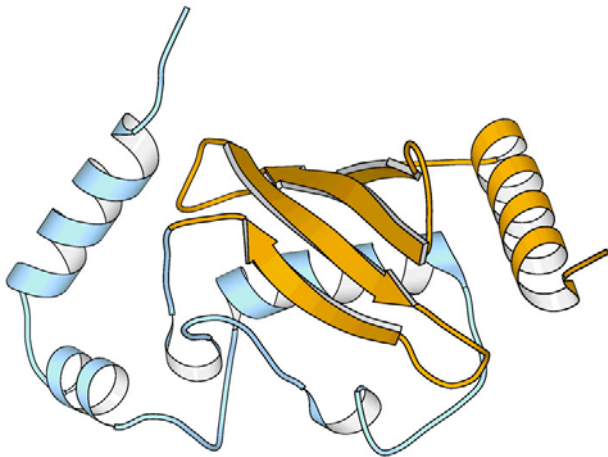


Figure 5
Alternative splicing of human ubiquitin-conjugating enzyme E2 A UBE2A. The part of the protein affected by alternative splicing is shown in yellow, and the remaining in blue. One can see that a α -helix and a whole β -sheet are lost in one of the isoforms, with a potentially very disruptive effect. The figure was obtained using the MOLSCRIPT software [65].

Conclusion

A common effect of alternative splicing is to produce isoforms lacking a given functional domain, pointing to an inhibitory role of the fully functional isoforms [17,36,39]. This correspondence between alternative splicing and protein function changes is a consequence of the modular structure of protein function, having been experimentally demonstrated in different instances [17]. Here we show that epigenetic regulators are no exception and that their alternative splicing patterns usually involve loss of the catalytic or the binding domain, resulting in short isoforms that could easily play the above-mentioned inhibitory role. They can also be the consequence of alternative splicing-based mechanisms for the regulation of product amount.

Thus, our results show how alternative splicing may regulate the functional role of chromatin-modifying enzymes. This is a first step towards the goal of understanding the biological impact of alternative splicing on epigenetic gene expression regulation. This goal, which in general is very difficult to attain [17], becomes particularly hard in our case, as epigenetic regulators act both at gene-specific and whole-genome levels [2,40]. They are involved in relevant biological processes like development [5] or disease [6] and, in addition, they may also act on proteins other than histones. Nonetheless, our results clearly support the idea that alternative splicing is likely to have a substantial

impact on the epigenetic regulation of large sets of genes, by regulating the activity of chromatin-modifying enzymes. One of the simplest mechanisms would be the co-expression of two alternative splice isoforms of one of these enzymes, a fully functional isoform and a dominant-negative inhibitor of the former, which may result in a reduced repression or activation of the set of genes controlled by this enzyme. To illustrate how this could happen, we can mention the case of G9a (EHMT2), a histone dimethyltransferase likely to play an important role in the repression of a large set of neuronal genes [9]. This repression, which can affect between 30 and 800 genes, is based on a chromatin-level mechanism [9] (Figure 6): (i) NFSP transcription factor would recruit histone dimethyltransferase G9a to the target genes; (ii) the latter would be silenced by G9a's dimethylation of histone tails at that location. It has been observed, that dominant-negative inhibition of G9a results in abrogation of this gene silencing [9]. In our case, we find that one of the G9a's isoforms has all the characteristics of a dominant-negative regulator (Table 2), as it has lost all its domains but the binding domain to NFSP transcription factor. We can speculate that this isoform could modulate the repression of this set of neuronal genes, in a similar way as G9a dominant-negative designed constructs [9] (Figure 6).

Methods

Dataset of epigenetic regulators

The list of chromatin-modifying enzymes was taken from five recent reviews on chromatin-modifying enzymes [2,3,19,41,42]. Note that DNA methyltransferases have not been considered. Subsequently we checked for the existence of alternative splicing for the corresponding genes in different databases: SwissProt [29], NCBI-Gene [43], Ensembl [44] and ASAP [45]. These databases have different annotation protocols, from manual annotation in SwissProt [29] to highly automatic procedures in Ensembl [44]. This allows increasing the coverage of our study. A discussion on possible error sources can be found at the end of the Materials and Methods section.

As shown in Table 5, the final dataset was constituted by 78 genes with alternative splicing, together with additional information on the species, protein name and function. Due to the different procedures followed in the different databases to obtain alternative splicing information we expect a complementary coverage of the alternative splicing patterns.

In general, the gene names used follow the international standards set for each species. Standard gene names were obtained: for human from the Human Gene Nomenclature Database [46]; for mouse from the Mouse Genome Database (MGD) [47]; for *D.melanogaster* from the Fly-Base [48], version FB2006_01; for *C.elegans* from the

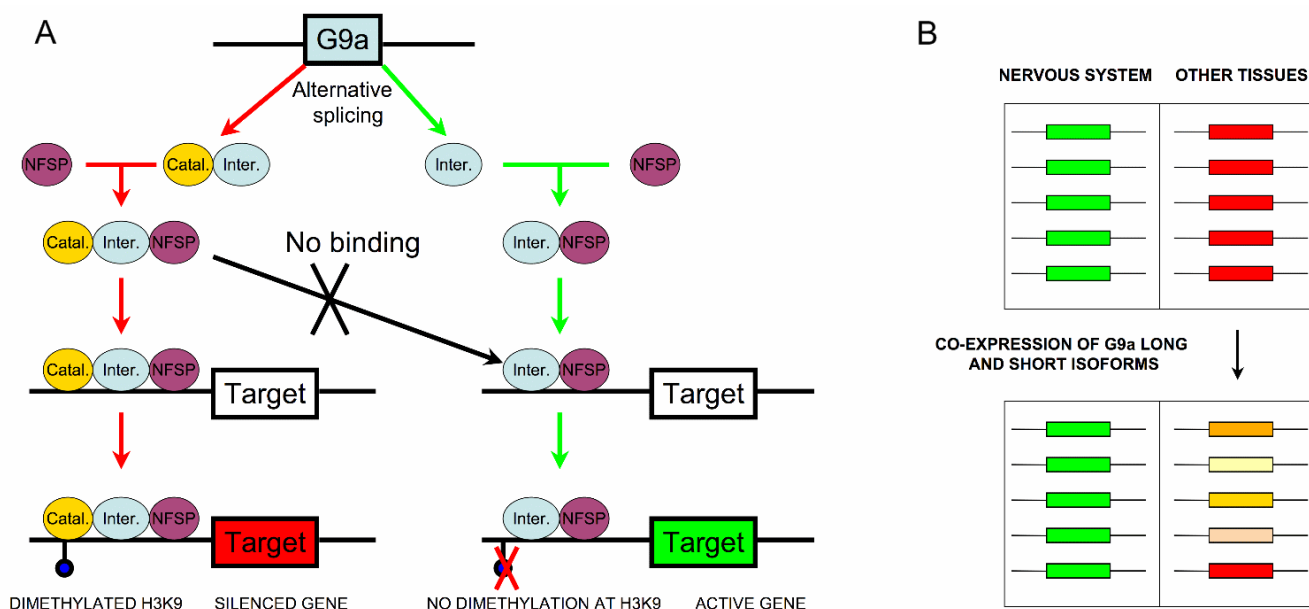


Figure 6
Hypothetical mechanism of regulation by alternative splicing of histone dimethyltransferase G9a function. (A) Experimental evidence indicates that histone dimethyltransferase G9a plays an important role in the silencing of neuronal genes in non-neuronal tissues [9]. In the proposed mechanism [9], shown here with red arrows, in non-neuronal tissues the transcription factor NFSP (shown in magenta) recruits the fully functional isoform of G9a (shown here with two domains: a binding domain in blue, and a catalytic domain in yellow) to a series of target genes that are subsequently silenced by G9a dimethylation of lysine-9 from histone H3. This mechanism may be inhibited/modulated by expression of the G9a short isoform (which only retains the NFSP transcription factor binding domain, Table 2), as shown here with green arrows. This isoform may behave as a dominant-negative inhibitor, as shown by the green arrows, blocking the access of the catalytically active isoform to the chromatin of the target gene. Absence of methylation marks in histone H3's lysine-9 would then result in an active gene. (B) The expression state of the target genes in both the nervous system (active, green colour) and in other tissues (silenced, red colour), as a result of the silencing, combined, action of NFSP and G9a. Co-expression of both the long and the short isoforms may result in the modification of the expression state of the target genes in non-neuronal tissues. These target genes may now show varying degrees of activity, as a result of the dominant-negative inhibitor role played by the short isoform (described in (A)).

WormBase [49], release WS166; for *Z.mays* from MaizeGDB [50].

The detailed exon structure of the isoforms studied in this work is provided in an additional file [see Additional file ExonStructure.xls].

Possible error sources

As explained in the previous section, alternative splicing data are obtained from different databases and come from different sources -e.g. literature, processing of ESTs- therefore they will have a different error attached to them. Unfortunately, it is not possible to provide a reliability measure for each observation, but we can discuss the reliability of the general trends observed and how the possi-

ble sources of error affect the main conclusions of our work.

First, we observe that the overall trends we find in our dataset coincide with those previously observed by other authors that have studied alternative splicing in more general sets of genes. In particular, the fact that insertions/deletions of domain size prevail in our dataset is in agreement with previous observations [39]. Also the corresponding mechanisms for function modulation - dominant-negative inhibition, amount regulation- have been proposed and observed for other genes [17], although the biological context and expected impact are obviously different. Some of the very short isoforms we have obtained can be artifactual but they may also consti-

Table 5: List of genes showing alternative splicing

Gene Symbol	Species	Function	Protein name
CDY-1	H.s.	A	chromodomain protein, Y-linked, 1
GCN5L2	H.s.	A	GCN5 general control of amino-acid synthesis 5-like 2 (yeast)
HAT1	H.s.	A	histone acetyltransferase 1
HTATIP (TIP60)	H.s.	A	HIV-1 Tat interacting protein
MORF4L1	H.s.	A	mortality factor 4 like 1
MYST1	H.s.	A	MYST histone acetyltransferase 1
NCOA-1	H.s.	A	Nuclear receptor coactivator 1
TAF1 (TAF250)	H.s.	A	TATA box binding protein (TBP)-associated factor, 250 kDa
CARM1 (PRMT4)	H.s.	M	coactivator-associated arginine methyltransferase 1
DOT1L	H.s.	M	DOT1-like, histone H3 methyltransferase (<i>S. cerevisiae</i>)
EHMT2 (G9a)	H.s.	M	euchromatic histone-lysine N-methyltransferase 2
EZH2	H.s.	M	enhancer of zeste homolog 2 (<i>Drosophila</i>)
MLL	H.s.	M	Myeloid/lymphoid or mixed-lineage leukemia
PRMT1	H.s.	M	protein arginine methyltransferase 1
SETD8 (PR-SET7, SET8)	H.s.	M	SET domain containing (lysine methyltransferase) 8
SETDB1	H.s.	M	SET domain, bifurcated 1
SUV39H1	H.s.	M	suppressor of variegation 3-9 homolog 1 (<i>Drosophila</i>)
SUV39H2	H.s.	M	suppressor of variegation 3-9 homolog 2 (<i>Drosophila</i>)
ATM	H.s.	P	ataxia telangiectasia mutated
ATR	H.s.	P	ataxia telangiectasia and Rad3 related
AURKB	H.s.	P	aurora kinase B
MAP3K12 (DLK/ZIP)	H.s.	P	Mitogen-activated protein kinase 12
PRKDC (DNA-PK)	H.s.	P	protein kinase, DNA-activated, catalytic polypeptide
RPS6KA5 (MSK1)	H.s.	P	ribosomal protein S6 kinase, 90kDa, polypeptide 5
RPS6KA4 (MSK2)	H.s.	P	ribosomal protein S6 kinase, 90kDa, polypeptide 4
CHD-3	H.s.	R	chromodomain helicase DNA binding protein 3
CHD-4	H.s.	R	chromodomain helicase DNA binding protein 4
SMARCA1 (SNF2L)	H.s.	R	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 1
SMARCA2 (BRM)	H.s.	R	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 2
SMARCA4 (BRG1)	H.s.	R	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4
UBE2A	H.s.	U	ubiquitin-conjugating enzyme E2A (RAD6 homolog)
CKII	H.s.	P	casein kinase 2, alpha 1 polypeptide
EHMT1	H.s.	M	Histone-lysine N-methyltransferase, H3 lysine-9 specific 5
GSG2 (HASPIN)	H.s.	P	Serine/threonine-protein kinase Haspin
FBXL11 (JHDM1A)	H.s.	DM	JmjC domain-containing histone demethylation protein 1A
FBXL10 (JHDM1B)	H.s.	DM	JmjC domain-containing histone demethylation protein 1B
JMJ1B (JHDM2B)	H.s.	DM	JmjC domain-containing histone demethylation protein 2B
JMJ2B (JHDM3B)	H.s.	DM	JmjC domain-containing histone demethylation protein 3B
JMJ2C (JHDM3C)	H.s.	DM	JmjC domain-containing histone demethylation protein 3C
AOF2 (LSD1)	H.s.	DM	Lysine-specific histone demethylase 1
MLL2	H.s.	M	Myeloid/lymphoid or mixed-lineage leukemia protein 2 (ALL1-related protein)
MLL3	H.s.	M	Myeloid/lymphoid or mixed-lineage leukemia protein 3 homolog
WBP7 (MLL4)	H.s.	M	WW domain-binding protein 7 (Myeloid/lymphoid or mixed-lineage leukemia protein 4) (<i>Trithorax</i> homolog 2)
MLL5	H.s.	M	myeloid/lymphoid or mixed-lineage leukemia 5 (<i>trithorax</i> homolog, <i>Drosophila</i>)
NSD1	H.s.	M	H3-K36-HMTase and H4-K20-HMTase
PRMT5	H.s.	M	Protein arginine N-methyltransferase 5
PRDM2 (RIZ1)	H.s.	M	PRDM2 (PR domain containing 2, with ZNF domain)
RNF40	H.s.	U	E3 ubiquitin-protein ligase BRE1B (RING finger protein 40)
SETDB2	H.s.	M	Histone-lysine N-methyltransferase SETDB2
SIRT2	H.s.	DA	NAD-dependent deacetylase sirtuin-2
Gtf3c4	M.m.	A	General transcription factor IIIC, polypeptide 4
Htatip	M.m.	A	HIV-1 tat interactive protein, homolog (human)
Morf4l1	M.m.	A	mortality factor 4 like 1
Ncoa-1	M.m.	A	Nuclear receptor coactivator 1
Ehmt2	M.m.	M	euchromatic histone lysine N-methyltransferase 2
Ezh2	M.m.	M	enhancer of zeste homolog 2 (<i>Drosophila</i>)
Prmt1	M.m.	M	protein arginine N-methyltransferase 1
Carm1 (Prmt4)	M.m.	M	protein arginine N-methyltransferase 4
Setdb1	M.m.	M	SET domain, bifurcated 1

Table 5: List of genes showing alternative splicing (Continued)

Suv39h1	M.m.	M	suppressor of variegation 3-9 homolog 1 (<i>Drosophila</i>)
Suv39h2	M.m.	M	suppressor of variegation 3-9 homolog 2 (<i>Drosophila</i>)
Stk4	M.m.	P	serine/threonine kinase 4
Myst2 (Hbo1)	M.m.	A	Histone acetyltransferase MYST2
Fbxl11 (Jhdm1a)	M.m.	DM	JmjC domain-containing histone demethylation protein 1A
Fbxl10 (Jhdm1b)	M.m.	DM	JmjC domain-containing histone demethylation protein 1B
Jmjd1a (Jhdm2a)	M.m.	DM	JmjC domain-containing histone demethylation protein 2A
Jmjd1b (Jhdm2b)	M.m.	DM	JmjC domain-containing histone demethylation protein 2B
Jmjd2a (Jhdm3a)	M.m.	DM	JmjC domain-containing histone demethylation protein 3A
Jmjd2b (Jhdm3b)	M.m.	DM	JmjC domain-containing histone demethylation protein 3B
Ring1A	M.m.	U	E3 ubiquitin-protein ligase RING1
Rnf20	M.m.	U	E3 ubiquitin-protein ligase BRE1A
Su(var)3-9	D.m.	M	Suppressor of variegation 3-9
trx	D.m.	M	trithorax
Taf1	D.m.	P	TBP-associated factor 1
brm	D.m.	R	Brahma
cbp-1	C.e.	A	Bromodomain
fbxl10 (jhdmlb)	X.l.	DM	JmjC domain-containing histone demethylation protein 1B
mez2	Z.m.	M	Polycomb protein EZ2

In the "Species" column H.s., M.m., C.e., O.s., X.l. and Z.m. mean *Homo sapiens*, *Mus musculus*, *Caenorhabditis elegans*, *Oryza sativa*, *Xenopus laevis* and *Zea mays*, respectively. In the column "Function" A, DA, DM, M, P, U and R mean Acetylation, deacetylation, demethylation, methylation, phosphorylation, ubiquitination and chromatin remodelling, respectively.

tute a possible regulatory mechanism [51]. In fact very short isoforms have been described for the genes in our study, e.g. for MLL [52].

At a more detailed level, in the case of data from ASAP [45], the authors provide an error estimate of less than 2 % [53]. To decrease it more, we discarded all the ASAP isoforms for a given gene, when none of them coincided with the longest isoform provided by another database. For the remaining databases the error estimates will vary, even within the database. For example, in the case of SwissProt [29], protein records are manually annotated, but the evidence supporting a given isoform may vary from one gene to another. Nonetheless, SwissProt [29] has been utilised in many bioinformatics studies on alternative splicing due to the high quality of the data [39,54-59]. In the case of Ensembl [44], the predictive nature of the annotations suggests that there may be a certain amount of false positives. The latter may be more frequent in the case of very short isoforms, although it has to be mentioned that these isoforms are usually supported by a substantial amount of evidence from EST data and other databases.

For all these reasons, we believe that the overall conclusions of this work will not be substantially affected by possible errors in the data.

Domain annotation

The domain structure of the different isoforms was obtained utilising CD-Search [60]. This program identifies the functional domains present in a protein sequence. We focused our analysis on the Pfam [61] and Smart [62] domain definitions. COG (Tatusov et al., 2001) definitions were not available for all the species and for this rea-

son they were not utilised (no significant differences were observed when utilised in this analysis). Because in some cases domain boundaries for the same domain would change slightly from one database to another, we combined the two definitions in a consensus domain definition, as follows: the location of the N-terminal domain was taken to be the minimum of the Pfam [61] and Smart [62] values; for the C-terminal end, instead of the minimum, we took the maximum of the Pfam [61] and Smart [62] values. For example, if a given domain occupies positions 3-75 and 8-82 according to the Pfam and Smart definitions, respectively, in our consensus definition it will go from position 3 to position 82.

We eliminated from the domain mapping all the domains with functional annotations of no, or unclear, meaning within the context of this work, that is: microbial domains, like viral capsid domains, and Pfam B domains [61]. In Table 6 we provide a list of the domains affected by alternative splicing mentioned in this work.

Classification of the alternative splicing events

Our study focused on those alternative splicing events that affect any of the known domains, as it is easier to infer their functional impact [17]. In general, epigenetic regulators are multidomain proteins that have both catalytic and interaction domains. Because the functional role of a given isoform will depend on which of these domains has been affected by alternative splicing, we grouped the observed isoforms according to the biochemical nature of the affected domain(s): (i) alternative splicing affects the catalytic domains; (ii) alternative splicing affects the protein interaction domains; and (iii) alternative splicing affects results in drastic sequence reductions. An alterna-

Table 6: List of domains affected by alternative splicing in chromatin-modifying enzymes

Domain name	Function	Enzyme name
AMINO_OXIDASE	Catalytic	AOF2
ANK	Protein-Protein Interaction	EHMT1, EHMT2
BRK	Unknown	SMARCA2, SMARCA4
BROMO	Interaction (Acetylated Lysines)	SMARCA2, SMARCA4, MLL
CHROMO	Interaction (Methylated Lysines)	SUV39H1, SUV39H2, Suv39h2, Su(var)3-9, MYST-1, MORF4L1, Morf4l1, Htatip
DEXHC	Catalytic	SMARCA1, SMARCA2, SMARCA4
FAT	Interaction/Modulate catalysis	ATM
FATC	Interaction/Modulate catalysis	ATM
FYRC	Probably not-catalytic	MLL, WBP7
FYRN	Probably not-catalytic	MLL, WBP7
HELICASE_C		SMARCA2
HSA	Probably DNA binding	SMARCA2, SMARCA4
JMJC	Catalytic	FBXL11, fbxl10 (from <i>Mus musculus</i> and <i>Xenopus laevis</i>), jmjd1b
LRR_RI	Interaction	FBXL10
MBD	DNA binding	SETDB1, Setdb1
MOZ_SAS	Catalytic	Htatip
MRG	Interaction	MORF4L1
PCAF_N	Interaction with CBP	GCN5L2
PHD	Intra- and Intermolecular interactions	MLL, MLL2, MLL3, JMJD2B, Jmjd2a, WBP7
PI3_PI4_KINASE	Catalytic	PRKDC, ATM
PKINASE	Catalytic	AURKB, GSG2, RPS6KA5, stk4
PRESET	Interaction-Catalysis	SUV39H1, SUV39H2, Suv39h2, Su(var)3-9, SETDB1, Setdb1, EHMT1, EHMT2
PWWP	Unknown	NSD1
RING	Interaction	MLL2, RNF40
SET	Catalytic	PRDM2, SUV39H1, SUV39H2, suv39h2, Su(var)3-9, SETDB1, Setdb1, mez2, MLL, WBP7, EHMT1, EHMT2, EZH2
SKBI	Catalytic	CARM1
UBCC	Whole protein	UBE2A
TUDOR	Interaction	Jmjd2a, Jmjd2b, JMJD2B, SETDB1, Setdb1
ZF_C3HC4	Interaction	RNF40
ZF_CXXC	Interaction	Fbxl10, Fbxl11, MLL, WBP7,
ZNF_TAZ	Interaction	cbp-1

Table 7: Templates utilised for comparative modelling

Protein name	Size Ref.	Species	Domain name	PDB code	% Seq. Id.
cbp-1	2056	C.e.	ZNF_TAZ	1L8C	75
SMARCA1 (SNF2L)	1054	H.s.	DEXHC	1Z6A	38
SMARCA2 (BRM)	1590	H.s.	BROMO	1N72	26
SUV39H2	410	H.s.	CHROMO	1KNA	47
SUV39H2	410	H.s.	SET	1MVH	39
UBE2A	154	H.s.	UBCC	1JAS	95
PRKDC (DNA-PK)	4127	H.s.	PI3-PI4 KINASE	1E8Y	29

In the "Species" column H.s. and C.e., mean *Homo sapiens* and *Caenorhabditis elegans*, respectively. The size of the whole protein is given in amino acid number. % Seq.Id. is the percentage of sequence identity between the target and the template sequences. The PDB code is the code of the template structure utilised for the comparative modelling in the PDB database [64].

tive splicing event belongs to the first class when the corresponding sequence change mainly affects the catalytic domains, but the resulting isoform retains at least one of its binding domains (i.e. keeps its binding ability). Alternative splicing events are classified in the second group when the sequence change mainly affects the interaction domains, but not the catalytic unit. Finally, alternative splicing events belong to the third class when both the catalytic and the binding domains are affected by the sequence change. Four proteins were not included in this classification, ubiquitin-conjugating enzyme E2A (UBE2A, 154 aas), casein kinase 2, alpha 1 polypeptide (CKII, 391 aas), NAD-dependent deacetylase sirtuin-2 (SirT2, 389 aas) and aurora kinase B (AURKB, 344 aas) because they only have a single domain which plays both a catalytic and a binding role and therefore large alternative splicing sequence changes are very likely to affect both functions simultaneously.

Structure analysis

Direct structural information was not available for none of the proteins considered in this work. However, in some cases the changes produced by alternative splicing embraced a part of the sequence for which structural information was available from a homolog. In these cases, this part was modelled utilising the well known, standard, modelling package MODELLER [63], and using the structure of the homolog as a template. The latter was obtained from the PDB database [64]. A list of cases, together with the domains involved, the homologs utilised, and the sequence identities between the latter and our proteins, is shown in Table 7.

Structural models are utilised throughout the article to illustrate the location of alternative splicing changes and to help understand/infer their functional impact. The conclusions that can be drawn from the use of these models are limited by the following facts: (i) in general, epigenetic regulators are multidomain proteins, while the structures correspond to only one of these domains; (ii) the structural changes resulting from certain sequence changes may be difficult to predict. It is clear that the structural analysis would benefit from taking into account the structure of the whole protein, but this information is not yet available for the proteins in our dataset or for their homologs, neither close nor remote. This would be a serious problem if our aim were to predict with high accuracy the structural/functional changes resulting from alternative splicing. However, our goal is more coarse-grained, as what we want to see is whether alternative splicing changes result in the presence or absence of the biochemical function associated to a given domain. When the sequence change affects the whole domain, by far the most frequent situation, it is reasonable to assume that the resulting protein has lost this activity and that it may

function as a regulator (e.g. a dominant-negative inhibitor) of the full-length isoform, something that has been experimentally confirmed in the case of transcription factors [17], among others.

If the sequence change does not reach the domain size the situation is more complex, because it is more difficult to decide whether it will result in complete function loss, modulation of an original function or creation of a new function. Without further structural data we cannot provide a definite answer for none of our cases. However, in some instances the nature of the sequence change is not compatible with preservation, or smooth modulation, of the domain's function. This happens when the domain is small and the sequence change is large, or it affects the protein core or any important secondary structure element. In these cases we have proposed that the most likely effect of alternative splicing is that of a regulator of the fully functional isoforms, something that has been already observed in the case of the epigenetic regulator SMARCA1 [22].

Finally, we cannot reject the possibility that some of the regions affected by alternative splicing may be intrinsically disordered, as has been recently proposed [59]. However, if the sequence stretch affected by alternative splicing encompasses a whole protein domain the functional interpretation will remain the same, as it is independent of whether the domain in question is structured or disordered. If the affected stretch is of sub-domain size, the situation could be different if we knew that the domain involved is disordered. However, this is unlikely as the domains affected by alternative splicing discussed here are homologues, sometimes very close, of domains with known three-dimensional structure (Table 7).

Abbreviations

aas: amino acids.

Authors' contributions

SL obtained the set of manually curated data, annotated them with the alternative splicing and protein domain information. NB contributed to design the study and to its testing. MM-B and XdC conceived the study, designed most of the testing and wrote the article. All authors read and approved the final manuscript.

Additional material

Additional file 1

Exon structure of the isoforms studied. The file provides a description of the exon structure of the isoforms analysed in the present article (Table 2). Most of the data were obtained after querying the ENSEMBL [44] and NCBI Gene databases [43]. Part of the data were also obtained after aligning the target isoform with the genome of the corresponding species or using the SEDB package [66]. Finally, in four cases (*GSG2*, *Imjdb1*, *fbxl10* and *mez2*, from human, mouse, frog and maize, respectively) no information could be retrieved. The structure of the file is the following: the first column corresponds to the name of the genes; the second column corresponds to the isoform size; the third column corresponds to the organism; and the following columns correspond to the exons constituting the isoform. Each gene is preceded by a line with these fields and the order of each exon within the gene (exons with no order number correspond to parts of the isoform sequence for which the exon could not be identified). For each gene the data given in the first line correspond to the longest, full-length, isoform; data in the following lines correspond to the remaining isoforms. The numbers within each exon cell correspond to its size in amino acids. A colour code was used to distinguish constitutive exons (red), alternative initiation sites (yellow), intron retentions (green), and sequence stretches with no exon(s) assigned (lilac).

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-8-252-S1.xls>]

Acknowledgements

The authors are grateful to the SwissProt team for their support. XdC acknowledges funding from the Spanish government (grants BIO2003-09327, BIO2006-15557). MM-B and NB acknowledge funding from the Spanish government (grants SAF2002-00741, SAF2005-01285, Gen2003-20642, CSD2006-00049, and BFU2006-01493/BMC). NB acknowledges financial support from the Parc Científic de Barcelona. SL acknowledges financial support from the Consejo Superior de Investigaciones Científicas.

References

- Strahl BD, Allis CD: **The language of covalent histone modifications.** *Nature* 2000, **403(6765)**:41-45.
- Peterson CL, Laniel MA: **Histones and histone modifications.** *Curr Biol* 2004, **14(14)**:R546-551.
- de la Cruz X, Lois S, Sanchez-Molina S, Martinez-Balbas MA: **Do protein motifs read the histone code?** *Bioessays* 2005, **27(2)**:164-175.
- Kornberg RD, Lorch Y: **Chromatin-modifying and -remodeling complexes.** *Curr Opin Genet Dev* 1999, **9(2)**:148-151.
- Margueron R, Trojer P, Reinberg D: **The key to development: interpreting the histone code?** *Curr Opin Genet Dev* 2005, **15(2)**:163-176.
- Egger G, Liang G, Aparicio A, Jones PA: **Epigenetics in human disease and prospects for epigenetic therapy.** *Nature* 2004, **429(6990)**:457-463.
- Okamoto I, Otte AP, Allis CD, Reinberg D, Heard E: **Epigenetic dynamics of imprinted X inactivation during early mouse development.** *Science* 2004, **303(5658)**:644-649.
- Mak W, Nesterova TB, de Napoles M, Appanah R, Yamanaka S, Otte AP, Brockdorff N: **Reactivation of the paternal X chromosome in early mouse embryos.** *Science* 2004, **303(5658)**:666-669.
- Roopra A, Qazi R, Schoenike B, Daley TJ, Morrison JF: **Localized domains of G9a-mediated histone methylation are required for silencing of neuronal genes.** *Mol Cell* 2004, **14(6)**:727-738.
- Chambeyron S, Bickmore WA: **Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription.** *Genes Dev* 2004, **18(10)**:1119-1130.
- Hebbes TR, Thorne AW, Crane-Robinson C: **A direct link between core histone acetylation and transcriptionally active chromatin.** *Embo J* 1988, **7(5)**:1395-1402.
- Wu J, Grunstein M: **25 years after the nucleosome model: chromatin modifications.** *Trends Biochem Sci* 2000, **25(12)**:619-623.
- Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun ZW, Schmid M, Opravil S, Mechtler K, Ponting CP, Allis CD, et al.: **Regulation of chromatin structure by site-specific histone H3 methyltransferases.** *Nature* 2000, **406(6796)**:593-599.
- Berger SL: **Histone modifications in transcriptional regulation.** *Curr Opin Genet Dev* 2002, **12(2)**:142-148.
- Kingston RE, Narlikar GJ: **ATP-dependent remodeling and acetylation as regulators of chromatin fluidity.** *Genes Dev* 1999, **13(18)**:2339-2352.
- Hassan AH, Prochasson P, Neely KE, Galasinski SC, Chandy M, Carrozza MJ, Workman JL: **Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes.** *Cell* 2002, **111(3)**:369-379.
- Lopez AJ: **Developmental role of transcription factor isoforms generated by alternative splicing.** *Dev Biol* 1995, **172(2)**:396-411.
- Latchman DS: **Eukaryotic Transcription Factors.** Third edition. London: Academic Press; 1998.
- Kouzarides T: **Chromatin modifications and their function.** *Cell* 2007, **128(4)**:693-705.
- Tominaga K, Pereira-Smith OM: **The genomic organization, promoter position and expression profile of the mouse MRG15 gene.** *Gene* 2002, **294(1-2)**:215-224.
- Tajul-Arifin K, Teasdale R, Ravasi T, Hume DA, Mattick JS: **Identification and analysis of chromodomain-containing proteins encoded in the mouse transcriptome.** *Genome Res* 2003, **13(6B)**:1416-1429.
- Barak O, Lazzaro MA, Cooch NS, Picketts DJ, Shiekhhattar R: **A tissue-specific, naturally occurring human SNF2L variant inactivates chromatin remodeling.** *J Biol Chem* 2004, **279(43)**:45130-45138.
- Convery E, Shin EK, Ding Q, Wang W, Douglas P, Davis LS, Nickoloff JA, Lees-Miller SP, Meek K: **Inhibition of homologous recombination by variants of the catalytic subunit of the DNA-dependent protein kinase (DNA-PKcs).** *Proc Natl Acad Sci U S A* 2005, **102(5)**:1345-1350.
- Kozmik Z, Czerny T, Busslinger M: **Alternatively spliced insertions in the paired domain restrict the DNA sequence specificity of Pax6 and Pax8.** *Embo J* 1997, **16(22)**:6793-6803.
- Foulkes NS, Mellstrom B, Benusiglio E, Sassone-Corsi P: **Developmental switch of CREM function during spermatogenesis: from antagonist to activator.** *Nature* 1992, **355(6355)**:80-84.
- Cox JS, Walter P: **A novel mechanism for regulating activity of a transcription factor that controls the unfolded protein response.** *Cell* 1996, **87(3)**:391-404.
- Sterner DE, Berger SL: **Acetylation of histones and transcription-related factors.** *Microbiol Mol Biol Rev* 2000, **64(2)**:435-459.
- Cheng X, Collins RE, Zhang X: **Structural and sequence motifs of protein (histone) methylation enzymes.** *Annu Rev Biophys Biomol Struct* 2005, **34**:267-294.
- Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, et al.: **The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003.** *Nucleic Acids Res* 2003, **31(1)**:365-370.
- Curwen V, Eyraas E, Andrews TD, Clarke L, Mongin E, Searle SM, Clamp M: **The Ensembl automatic gene annotation system.** *Genome Res* 2004, **14(5)**:942-950.
- Johnson JM, Castle J, Garrett-Engle P, Kan Z, Loerch PM, Armour CD, Santos R, Schadt EE, Stoughton R, Shoemaker DD: **Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays.** *Science* 2003, **302(5653)**:2141-2144.
- Graveley BR: **Alternative splicing: increasing diversity in the proteomic world.** *Trends Genet* 2001, **17(2)**:100-107.
- Stamm S, Ben-Ari S, Rafalska I, Tang Y, Zhang Z, Toiber D, Thanaraj TA, Soreq H: **Function of alternative splicing.** *Gene* 2005, **344**:1-20.
- Syntichaki P, Topalidou I, Thireos G: **The Gcn5 bromodomain coordinates nucleosome remodelling.** *Nature* 2000, **404(6776)**:414-417.

35. Smith CW, Valcarcel J: **Alternative pre-mRNA splicing: the logic of combinatorial control.** *Trends Biochem Sci* 2000, **25(8)**:381-388.
36. Modrek B, Lee CJ: **Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss.** *Nat Genet* 2003, **34(2)**:177-180.
37. Neu-Yilik G, Gehring NH, Hentze MW, Kulozik AE: **Nonsense-mediated mRNA decay: from vacuum cleaner to Swiss army knife.** *Genome Biol* 2004, **5(4)**:218.
38. Hopitzan AA, Baines AJ, Ludosky MA, Recouvreur M, Kordeli E: **Ankyrin-G in skeletal muscle: tissue-specific alternative splicing contributes to the complexity of the sarcolemmal cytoskeleton.** *Exp Cell Res* 2005, **309(1)**:86-98.
39. Kriventseva EV, Koch I, Apweiler R, Vingron M, Bork P, Gelfand MS, Sunyaev S: **Increase of functional diversity by alternative splicing.** *Trends Genet* 2003, **19(3)**:124-128.
40. van Leeuwen F, van Steensel B: **Histone modifications: from genome-wide maps to functional insights.** *Genome Biol* 2005, **6(6)**:113.
41. Li B, Carey M, Workman JL: **The role of chromatin during transcription.** *Cell* 2007, **128(4)**:707-719.
42. Shi Y, Whetstone JR: **Dynamic regulation of histone lysine methylation by demethylases.** *Mol Cell* 2007, **25(1)**:1-14.
43. Maglott D, Ostell J, Pruitt KD, Tatusova T: **Entrez Gene: gene-centered information at NCBI.** *Nucleic Acids Res* 2005, **33(Database):**D54-58.
44. Birney E, Andrews D, Caccamo M, Chen Y, Clarke L, Coates G, Cox T, Cunningham F, Curwen V, Cutts T, et al.: **Ensembl 2006.** *Nucleic Acids Res* 2006:D556-561.
45. Lee C, Atanelov L, Modrek B, Xing Y: **ASAP: the Alternative Splicing Annotation Project.** *Nucleic Acids Res* 2003, **31(1)**:101-105.
46. Wain HM, Lush MJ, Ducluzeau F, Khodiyar VK, Povey S: **Genew: the Human Gene Nomenclature Database, 2004 updates.** *Nucleic Acids Res* 2004:D255-257.
47. **Mouse Genome Database** [<http://www.informatics.jax.org>]
48. **FlyBase** [<http://www.flybase.org>]
49. **WormBase** [<http://www.wormbase.org>]
50. Lawrence CJ, Seigfried TE, Brendel V: **The maize genetics and genomics database. The community resource for access to diverse maize data.** *Plant Physiol* 2005, **138(1)**:55-58.
51. Modrek B, Lee C: **A genomic view of alternative splicing.** *Nat Genet* 2002, **30(1)**:13-19.
52. Nam DK, Honoki K, Yu M, Yunis JJ: **Alternative RNA splicing of the MLL gene in normal and malignant cells.** *Gene* 1996, **178(1-2)**:169-175.
53. Resch A, Xing Y, Modrek B, Gorlick M, Riley R, Lee C: **Assessing the impact of alternative splicing on domain interactions in the human proteome.** *J Proteome Res* 2004, **3(1)**:76-83.
54. Kondrashov FA, Koonin EV: **Origin of alternative splicing by tandem exon duplication.** *Hum Mol Genet* 2001, **10(23)**:2661-2669.
55. Boue S, Vingron M, Kriventseva E, Koch I: **Theoretical analysis of alternative splice forms using computational methods.** *Bioinformatics* 2002, **18(Suppl 2)**:S65-73.
56. Furnham N, Ruffe S, Southan C: **Splice variants: a homology modeling approach.** *Proteins* 2004, **54(3)**:596-608.
57. Valenzuela A, Talavera D, Orozco M, de la Cruz X: **Alternative splicing mechanisms for the modulation of protein function: conservation between human and other species.** *J Mol Biol* 2004, **335(2)**:495-502.
58. Wang P, Yan B, Guo JT, Hicks C, Xu Y: **Structural genomics analysis of alternative splicing and application to isoform structure modeling.** *Proc Natl Acad Sci U S A* 2005, **102(52)**:18920-18925.
59. Romero PR, Zaidi S, Fang YY, Uversky VN, Radivojac P, Oldfield CJ, Cortese MS, Sickmeier M, LeGall T, Obradovic Z, et al.: **Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms.** *Proc Natl Acad Sci U S A* 2006, **103(22)**:8390-8395.
60. Marchler-Bauer A, Bryant SH: **CD-Search: protein domain annotations on the fly.** *Nucleic Acids Res* 2004:W327-331.
61. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, et al.: **The Pfam protein families database.** *Nucleic Acids Res* 2004:DI38-141.
62. Letunic I, Copley RR, Schmidt S, Ciccarelli FD, Doerks T, Schultz J, Ponting CP, Bork P: **SMART 4.0: towards genomic data integration.** *Nucleic Acids Res* 2004:DI42-144.
63. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A: **Comparative protein structure modeling of genes and genomes.** *Annu Rev Biophys Biomol Struct* 2000, **29**:291-325.
64. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE: **The Protein Data Bank.** *Nucleic Acids Res* 2000, **28(1)**:235-242.
65. Kraulis P: **MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures.** *Journal of Applied Crystallography* 1991, **24(5)**:946-950.
66. Leslin CM, Abyzov A, Ilyin VA: **Structural exon database, SEDB, mapping exon boundaries on multiple protein structures.** *Bioinformatics* 2004, **20(11)**:1801-1803.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp



PUBLICACIÓN 3

Characterization of structural variability sheds light on the specificity determinants of the interaction between effector domains and histone tails.

Sergio Lois, Naiara Akizu, Gemma Mas de Xaxars, Iago Vázquez, Marian Martínez-Balbás y Xavier de la Cruz

Epigenetic 5:2, 137-148, 2010

TITULO:

La caracterización de la variabilidad estructural vierte luz sobre los factores determinantes de la especificidad de la interacción entre los dominios efectores y las colas de las histonas.

RESUMEN:

La caracterización estructural de la interacción entre el extremo N-terminal de las histonas y los dominios efectores (Bromodominios, Cromodominios, PHD *fingers*, etc....) es fundamental para la comprensión de los aspectos mecánicos de la regulación epigenética de la expresión génica. En los últimos años muchos investigadores han aplicado esta aproximación a sistemas específicos, proporcionando una visión cada vez más rica, pero todavía fragmentaria, de la interacción histona-dominio. En este artículo, se ha utilizado esta información para caracterizar estructuralmente los dos componentes de esta interacción: los péptidos de histona y los lugares de unión de los dominios de interacción (centrándonos en aquellos que reconocen lisinas metiladas), y enriquecer nuestro conocimiento sobre los determinantes de su especificidad.

Los resultados obtenidos muestran que los lugares de unión de los dominios efectores son estructuralmente variables, pero presentan ciertas características comunes que permiten su clasificación en tres grupos principales: *flat-groove*, *narrow-groove* y *cavity-insertion*. Adicionalmente, el resultado de nuestro análisis, en contexto con los trabajos de otros investigadores, contribuye a clarificar los orígenes de la especificidad de la interacción entre la cola de la histona y el dominio efector: (a) existencia de diferentes regiones en el lugar de unión y, (b) diferencias en la transición desorden-orden experimentadas por diferentes péptidos de histona tras la unión con la proteína efectora.

Characterization of structural variability sheds light on the specificity determinants of the interaction between effector domains and histone tails

Sergi Lois,¹ Naiara Akizu,³ Gemma Mas de Xaxars,² Iago Vázquez,² Marian Martínez-Balbás³ and Xavier de la Cruz^{2,4,*}

¹Institut de Medicina Preventiva i Personalitzada del Càncer; Badalona, Spain; ²Computational Biology and Bioinformatics laboratory; and ³Molecular Signaling and Chromatin Function laboratory; Instituto de Biología Molecular de Barcelona; Consejo Superior de Investigaciones Científicas; Parc Científic de Barcelona; Barcelona, Spain; ⁴Institució Catalana de Recerca i Estudis Avançats; Barcelona, Spain

Key words: histone-effector interaction, modification of histone tails, histone binding mode, epigenetic regulator specificity, histone disorder, lysine post-translational modifications

Abbreviations: BS, whole binding site; MBS, binding site of the modified residue; AUTODOM, comparisons between BS from different versions of the same structure (experimental replicas, or different binding states, etc.); INTRADOM, INTERDOM, comparisons between BS from effectors with the same fold or different folds, respectively; ASA, accessible surface area

Structural characterization of the interaction between histone tails and effector modules (bromodomains, chromodomains, PHD fingers, etc.) is fundamental to understand the mechanistic aspects of epigenetic regulation of gene expression. In recent years many researchers have applied this approach to specific systems, thus providing a valuable but fragmentary view of the histone-effector interaction. In our work we use this information to characterize the structural features of the two main components of this interaction, histone peptides and the binding site of effector domains (focusing on those which target modified lysines), and increase our knowledge on its specificity determinants.

Our results show that the binding sites of effectors are structurally variable, but some clear trends allow their classification in three main groups: flat-groove, narrow-groove and cavity-insertion. In addition, we found that even within these classes binding site variability is substantial. These results in context with the work from other researchers indicate that there are at least two determinants of binding specificity in the binding site of effector modules. Finally, our analysis of the histone peptides sheds light on the structural transition experienced by histone tails upon effector binding, showing that it may vary depending on the local properties of the sequence stretch considered, thus allowing us to identify an additional specificity determinant for this interaction.

Overall, the results of our analysis contribute to clarify the origins of specificity: different regions of the binding site and, in particular, differences in the disorder-order transitions experienced by different histone sequence stretches upon binding.

Introduction

Post-translational modifications—acetylation, methylation, ubiquitination, etc.—of the histone tails play a fundamental role in the regulation of gene expression by altering the chromatin state to either active or repressed.¹ These marks are recognized by a set of effector modules able to bind the modified amino acids and part of their neighboring residues.^{1,2} Consequently, the structural characterization of the histone-effector interaction constitutes an important step towards understanding the molecular mechanisms of epigenetic regulation of gene expression, and how their failure leads to disease.

In recent years many studies have addressed this goal² and there is now available a significant number of effector structures in either their apo- and/or holo-forms (where the bound compounds are usually histone peptides carrying different marks). The results of these studies have shed light on different aspects of the histone-effector interaction.²⁻⁴ Structural analyses have been used to comprehend the binding specificities of trimethylated lysines towards different chromodomains;³ they have also enhanced our mechanistic understanding of the recognition of specific histone marks, particularly of the lysine methylation states (mono-, di- and trimethylation). In particular, it has been shown that modified lysines bind to an aromatic cage constituted by two to four residues and an acidic side chain, with cation- π

*Correspondence to: Xavier de la Cruz; Email: xdmcri@ibmb.csic.es

Submitted: 11/13/09; Accepted: 12/29/09

Previously published online: www.landesbioscience.com/journals/epigenetics/article/11079

interactions and steric restraints playing a fundamental role.² Considering together the known structures of histone-effector complexes, Taverna and colleagues² have identified two main binding modes for marked lysines related to binding specificity: cavity-insertion and surface-groove. In the cavity-insertion mode, the modified lysine side chain is buried in a deep, narrow pocket with the ability to filter out ligands by size; on the contrary, in the surface-groove mode the lysine side chain is not subjected to such strict restraints. Histone peptides are intrinsically disordered;⁴ however, when bound to effectors they adopt an extended conformation.² On this basis it has been proposed that the conformational transition experienced by histone tails between unbound and bound states corresponds to a disorder-order transition,⁴ rather than to an induced-fit mechanism.² This is important because binding energetics depends on whether structural transitions are involved, and their kind.

The above findings clarify our view of the histone-effector interaction and point to some important open issues about the structural properties of the histone-effector interaction, and the corresponding biophysical/biochemical consequences. For example, our knowledge of specificity determinants is still incomplete. Data from different authors^{2,3} suggest that apart from the binding site of the modified residue (to which we will refer as MBS from now on) the remainder of the binding site may also play a role. Therefore, classifying whole binding sites (to which we will refer as BS from now on), and relating the resulting classes with the known MBS classes would allow a better understanding of how binding specificity is distributed over the different BS components. In addition, although it will not be considered here, good classifications of BS (1) can be used to improve comparative modeling and docking studies of histone-effector complexes, facilitating the identification of BS and binding modes in domains for which no structural information is available or is restricted to their apo forms, (2) allow a better understanding of the evolution of function mechanisms within protein families,⁵⁻⁷ and (3) allow the identification of possible sources of cross-reactions in designed ligands.⁸ In addition, to improve our knowledge of specificity determinants it is also important to understand how binding affects the histone peptide structure. For example, the degree to which the extended conformations in the different complexes are similar (it is known that extended conformations are structurally heterogeneous⁹), if there are conserved structural motifs indicating the existence of structural propensities that could favor/disfavor the unbound-bound transition, whether this transition is the same for all modified lysines (e.g., H3K4, H3K9, H3K27, etc.). Identification of any structural trend in histone peptides would help to improve our quantitative understanding of the disorder-order transition associated to histone binding and to see if histone modifications owe part of their functional effect to shifts in this transition.

In this article we present the results of our work on two of these issues: the structural characterization of the BS of effectors, and that of the bound histone peptides. The BS of a series of effectors were compared using both visual and automatic structure comparison methods; peptide structures were also compared following a similar approach. Our results showed that BS could

be classified in three main structural classes: cavity-insertion, flat-groove and narrow-groove. These classes did not completely coincide with those identified for the MBS² pointing to a partition of the BS in specificity determinants. Regarding histone peptides, our results indicated that apart from a certain amount of structural heterogeneity, almost all of them shared a small, hook-like motif involving the side chain atoms of the modified lysine and some of the nearby main chain atoms. The different sequence propensities for the ARK and RTK sequences corresponding to this motif show that these tripeptides undergo a different disorder-order transition upon binding that constitutes a specificity determinant for the histone-effector interaction.

Results

Visual analysis of the BS. Visual analysis was done using a molecular surface¹⁰ representation of the BS, as it averages out atomic detail thus providing a global-shape view which allowed the classification of BS. Inspection of the available complexes (Table 1) showed that BS could be classified in three classes: flat-groove, narrow-groove and cavity-insertion. We found that there was not an exact mapping between these classes and Taverna and colleagues'² classes. Our cavity-insertion class was essentially the same as theirs. However, their MBS surface-groove class mapped to our two remaining BS classes: the flat-groove (Fig. 1) and narrow-groove (Fig. 2). The former included BS from the PHD, the Tudor and the double chromodomain effectors. The narrow-groove class was constituted only by BS from single chromodomains. It was the more homogeneous class and for this reason we used it to illustrate BS structural variability arising from structural homology by choosing three representatives. It should be noted that two of these (chromodomains from mouse HP1 beta, PDB code: 1GUW; and from Polycomb, PDB code: 1PDQ) had to be slightly trimmed before identifying them as members of this class.

The cavity insertion class (Fig. 3) included BS from the bromodomain, the tandem Tudor and the WDR5 effectors. For these cases, apart from the tunnel-like cavity, it was hard to identify other common features. Probably the most different case was that of the bromodomain complex as it also displayed features from the narrow-groove class.

Use of the molecular surface representation highlighted a prominent feature in the members of the flat-groove class (Fig. 1): a protruding part from the effector (corresponding to an aromatic ring) anchored the histone peptide, which adopted a hook-like structure at this locus. This motif was also found in the representatives of the narrow-groove class (Fig. 2), although the role of the tryptophan aromatic ring was not so relevant. A more different variant of the motif was also found in the bromodomain member of the cavity-insertion mode (Fig. 3). Interestingly, in all cases the peptide local structure was very similar.

Automatic structure comparison of the BS. Use of a hard-sphere representation, which preserves atomic detail, served to evidence a clear structural heterogeneity among BS (Figs. 1–3), between and within the classes previously defined. Both for the flat-groove and narrow-groove representatives we found that the

aromatic cage was relatively well preserved, but the remainder of the BS showed clear intra-class differences (Figs. 1 and 2). When looking at the conserved motif, we found that while the structure of the histone atoms involved looked well preserved, this was not the case for the effector's contacting atoms (Figs. 1–3). We decided to use automatic structure comparison methods to assess this structural variability and see whether it was relevant or, on the contrary, it was comparable to experimental noise and therefore negligible.

We did an all-against-all comparison of the BS structures using the program MAMMOTH¹¹ (see Materials and Methods). For each comparison we obtained an alignment and an associated rmsd (root-mean-square deviation) value. The rmsd was used as a measure of the structural variability between BS. As mentioned before BS from effectors with the same fold, which corresponded to instances of structure divergence, looked relatively similar on visual inspection (Fig. 2). On the contrary, BS from effectors with different folds (e.g., Tudor and PHD effectors from the flat-groove class), which corresponded to examples of structure convergence, looked more different (Fig. 1). Consequently, in these cases similarities could be harder to find by MAMMOTH. To take into account this heterogeneity in the comparisons we grouped them into two separate classes: INTRADOM and INTERDOM, which corresponded to comparisons between BS from effectors with the same and different folds, respectively. The rmsd distributions for these two classes were compared with (1) the coordinate uncertainty of experimental origin, which varies between 0.1 Å and 1 Å,¹² and (2) the rmsd values resulting from the comparison between different versions of the same effector (apo- and holo-forms, complexes with different ligands), what we called the AUTODOM class.

We found that all INTRADOM values were above 1 Å, therefore the corresponding structural diversity could not be attributed to noise of experimental origin. In accordance with this, we found that INTRADOM and AUTODOM distributions were significantly different (Kolmogorov-Smirnov test: p value ~ 0), although they showed a certain degree of overlap (Fig. 4A). This small overlap was mostly due to the fact that values of AUTODOM above 3 Å corresponded to comparisons involving BS defined from different ligands, more prone to give incorrect alignments. The fact that INTRADOM was different from AUTODOM indicated that structural differences between homologs were larger than those arising from a mere local side-chain rearrangement such as that occurring after substrate binding. To complete the rmsd analysis we computed the fraction of degenerate positions (those positions where the aligned atoms were different, e.g., C and O) present in the alignments (Fig. 4B). We found that AUTODOM and INTRADOM distributions were different (Kolmogorov-Smirnov test: p value ~ 0) although their overlap was larger than in the case of rmsd.

AUTODOM and INTERDOM rmsd distributions were different (Kolmogorov-Smirnov test: p value ~ 0), and only had a small overlap, but INTERDOM and INTRADOM distributions showed a substantial overlap (Fig. 4A) suggesting that the structural variability for convergent and divergent cases was similar. However, inspection of the INTERDOM alignments

Table 1. Histone-effector complexes used in this work

Ligand	Effector domain	Modified residues	PDB codes
Histone peptides	Chromodomain	H3K4me2K9me2 ^a	1GUW ^b
		H3K9me1	1Q3L
		H3K9me2	1KNA
		H3K9me3	1KNE
		H3K27me3	1PDQ, 1PFB
Double Chromodomain		H3K4me3T3ph	2B2T
		H3K4me3R2me2	2B2U
		H3K4me1	2B2V
		H3K4me2	2B2W
		H4K16ac	1E6I
PHD finger		H3K4me2	2FSA
		H3K4me3	2F6J, 2FUU, 2G6Q
Tandem Tudor		H4K20me2	2IG0
Double Tudor		H3K4me3	2GFA
WDR5		H3K4	2CO0
		H3K4me	2H6K
		H3K4me2	2H6N, 2CNX, 2G99, 2G9A, 2H13
		H3K4me3	2H6Q
Non-histone ligand	Chromo Shadow/Chromodomain ^c	Chr. Ass. Fac.	1S4Z
		EMPSY	2FMM
		Bromodomain	
		TAT peptide	1JM4B
		p53 C-t	1JSPB
		NP1	1WUG
		NP2	1WUM
MIB	1Z55A		
TTR	2D82A		

^aLigand. For histones only the modified, or the normal target, residues are shown. ^bPDB code. ^cChromo Shadow domains were treated as chromodomains, as their structural comparison with chromodomains gave excellent alignments.

showed that this was not true. For a set of the most promising cases from this distribution we found that all MAMMOTH alignments were meaningless (when analyzed visually) and their low rmsd values (near 4 Å) were essentially due to the small number of aligned atoms. There was only one exception: the comparison between the human CHD1 tandem chromodomains (PDB code: 2B2W) and the PHD finger from human BPTF (PDB code: 2FSA) (Fig. 5). In this case the alignment superimposed the motif found in the previous section and present in both BS. These results confirmed our visual analysis of the hard-sphere representations showing that structural variability among convergent cases was larger than what was suggested by overall shape similarity. Indeed, while MAMMOTH was able to produce

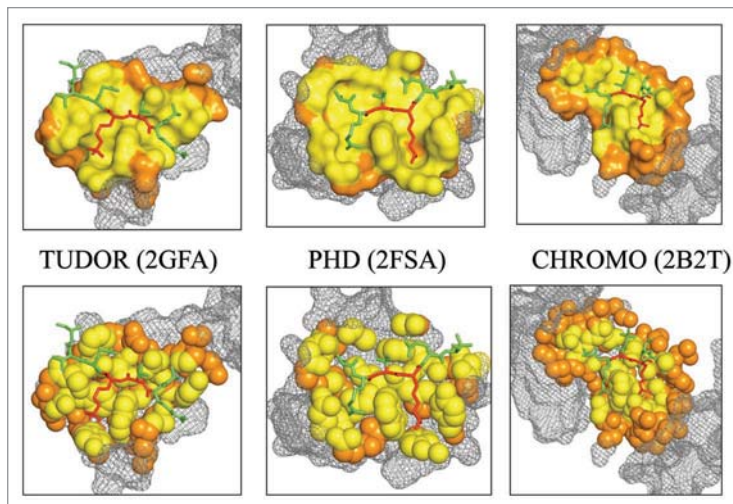


Figure 1. Classification of the effector BS. The flat-groove class. In this figure and in the next two figures (Figs. 2 and 3) we show three representative members of each class with their names and PDB codes, plotted using two representations: molecular surface (upper row) and hard sphere (lower row). Shown in yellow are the BS atoms in contact with the histone peptide, and in orange those included in the BS to improve shape representation.²⁵ The remainder of the effector is represented with a grey mesh. The peptide is represented in green, except for the part corresponding to the conserved hook-like motif, shown in red.

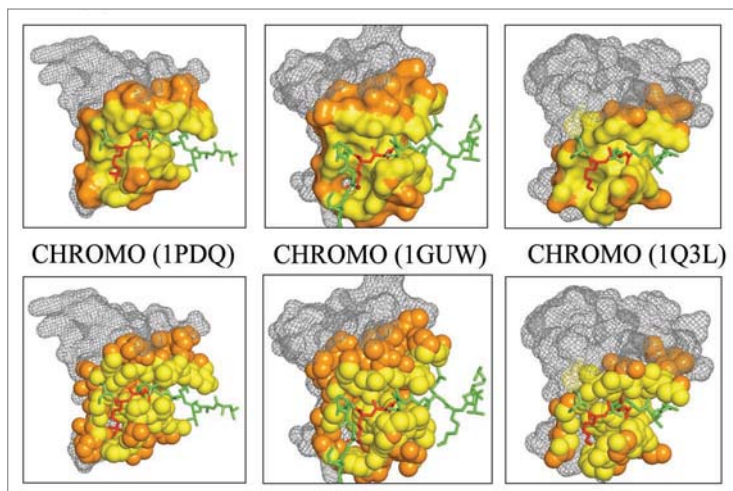


Figure 2. Classification of the effector BS. The narrow-groove class. For an explanation on the molecular representations and color codes, see legend to Figure 1.

reasonable alignments between divergent BS, this was not the case for convergent BS.

Common ligand features. Histone peptides tend to adopt an extended conformation when bound to effectors.² Here we explore the degree of conservation of this extended conformation. To this end we superimposed a series of representative histone peptides (PDB codes: 1GUW, 2B2T, 1PDQ, 2FUU, 2G6Q, 2GFA, 1E6I, 2IG0) against the structure of the histone peptide from the histone-chromodomain complex from *Drosophila melanogaster*

(PDB code: 1Q3L), arbitrarily chosen as reference. The resulting multiple structure alignment allowed us to identify two regions in the bound peptides (Fig. 6A): a highly conserved hook-like motif (shown in red) and a non-conserved region, corresponding to the remainder of the peptide (the parts upstream and downstream the conserved motif). Within the non-conserved regions we could distinguish a clear cluster corresponding to peptides binding narrow-groove class members (Fig. 6B). As expected, peptides bound to flat-groove class members showed a larger variability in the non-conserved regions (Fig. 6C). Most of this variability could be attributed to the peptide bound to the tandem chromodomains (PDB code: 2B2T); the similarity between the remaining peptides was closer to that found between narrow-groove members.

The hook-like motif was highly conserved (Table 2), irrespective of the effector's BS class. It involved the following atoms: the side chain (except the terminal N and its bound methyl or acetyl groups) and the N and C_α main chain atoms from the modified lysine; the main chain atoms of the residue at position -1 relative to the modified lysine; and the C and O main chain atoms of the residue at position -2. An incomplete version of the hook-like motif was also found in the peptides bound to the bromodomain (PDB code: 1E6I) and to the tandem Tudor effector (PDB code: 2IG0), which lacked the residue at position -2 and therefore the corresponding C and O main chain atoms. The motif was not found in the WDR5 complexes, where the peptide adopted a different structure upstream the modified lysine.

Given the high conservation degree of the hook-like motif, we decided to characterize its contribution to the histone-effector interaction. To this end we computed the percentage of accessible surface area (ASA) buried by the motif upon complex formation, relative to the ASA buried by the peptide and by the whole histone. The first percentage was directly obtained from the structures of the complexes. The average ASA buried by the motif was ~113 Å² (Table 2). Relative to the ASA buried by the whole peptide, the values of the hook-like motif are high, between 10% and 25% of the total, indicating that the motif identified also played an important role in the binding affinity of the peptide-effector interaction.

The size of the histone peptides in our dataset varies substantially (Table 1) and may not properly reflect what happens in complexes in which whole histones are involved. For this reason we decided to obtain an estimate of the ASA buried by the motif relative to that buried by the whole histone. The latter could not be directly computed from the structures available and was estimated applying the following assumptions: (i) that the ASA buried by histones had to have an upper threshold, i.e., it had to tend asymptotically to a given value; and (ii) that this value should resemble for all effectors, given the similar sizes of their BS. The first assumption was based on the fact that in our dataset the largest peptides tended to have unbound terminal ends. The second assumption, which implies

an averaging over all binding modes, was motivated by visual inspection of the peptide-effector complexes, which showed that peptide-effector interfaces had similar sizes. A plot of the peptide buried ASA vs. peptide length for the complexes in our dataset showed an asymptotic behavior, with buried ASA approaching 1,160 Å² (Fig. 7) as peptide length increased. We took this value as an approximation of the ASA buried by whole histones upon effector binding. Using this value we found that the contribution of the hook-like motif to an average histone-effector interaction was near 10% (Table 2), confirming its relevance when an approximate but more realistic scenario was considered.

To refine our analysis we broke down the motif's buried ASA into atomic contributions. The result (Fig. 8) showed a clear general trend defined by the presence of two peaks: one for the C and O main chain atoms of the residue at position -2 relative to the modified lysine, and the other for the lysine side chain itself. However, underlying this trend there was a clear variability that reflected both the BS and the histone sequence variabilities. This was established by grouping the results according to the residue sequence of the hook-like motif (two tripeptides only: ARK and RTK, Fig. 8B). The variability within groups reflected the underlying BS variability: ARK motifs, which were bound to homolog BS (belonging to the narrow-groove class), had smaller variability than RTK motifs, which were bound to BS from the flat-groove class (more structurally heterogeneous). The differences between groups reflected the sequence differences between hook-like motifs.

Discussion

The availability of structural information for different histone-effector complexes provides a good opportunity to advance our understanding of the molecular basis of epigenetic regulation, by shedding light on the biophysical/functional properties of this interaction and of its components. In this article we have studied the structural variability of histone-effector complexes focusing on the effector BS, and on the bound histone peptides. We found that there was an incomplete equivalence between our BS-level classes and Taverna and colleagues'² MBS-level classes: their surface-groove class could be found in both our flat-groove (Fig. 1) and narrow-groove (Fig. 2) representatives. This discrepancy is not contradictory, as these classes arise from considering BS at two different levels (whole and part, respectively). It indicates that the BS of effectors can be broken down in two complementary parts, MBS and remainder of the BS, and that for effectors with the same MBS (like chromodomain and PHD) differences in binding specificity/affinity will be determined by the rest of the BS. This idea is supported by both mutagenesis and sequence data. For example, for the

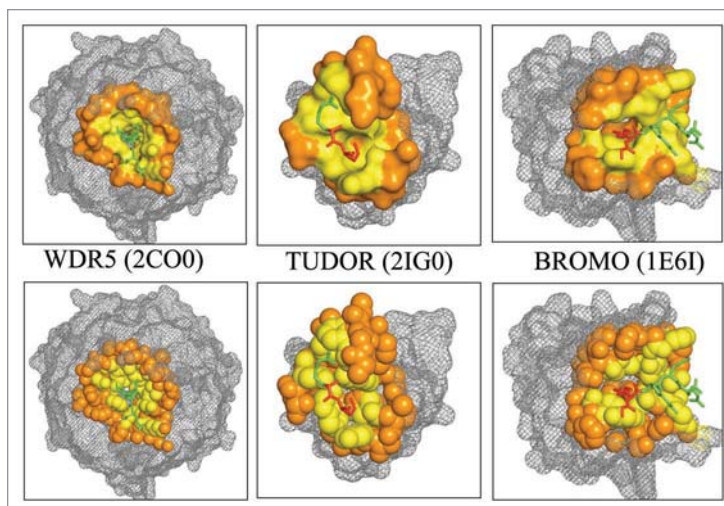


Figure 3. Classification of the effector BS. The cavity-insertion class. For an explanation on the molecular representations and color codes, see legend to Figure 1.

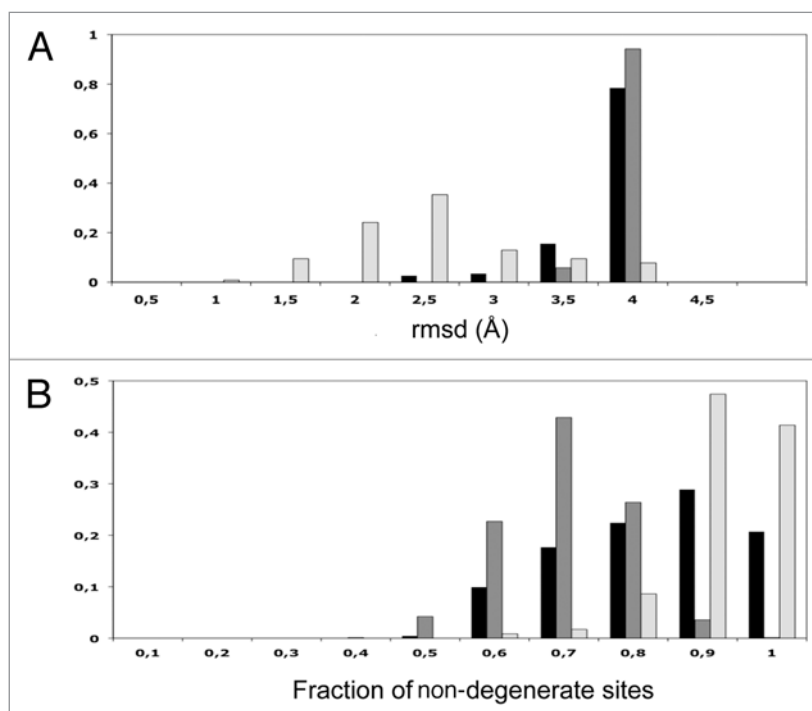


Figure 4. Frequency histograms of the rmsd (A) and fraction of non-degenerate sites (B) distributions. Three distributions are shown for each parameter: AUTODOM (light grey), INTRADOM (black) and INTERDOM (dark grey).

second Tudor domain (PDB code: 2GFA) of Jumonji domain-containing protein 2A, mutations Asp945Ala and Asp945Arg reduce and eliminate H3K4me3 binding, respectively.¹³ Because this residue is in contact with the histone peptide-but not with the modified lysine, this result is in accordance with the role of specificity-responsible residues proposed for BS residues outside the MBS.¹³ This is also the case for the Tyr1500Ala mutant of Tumor

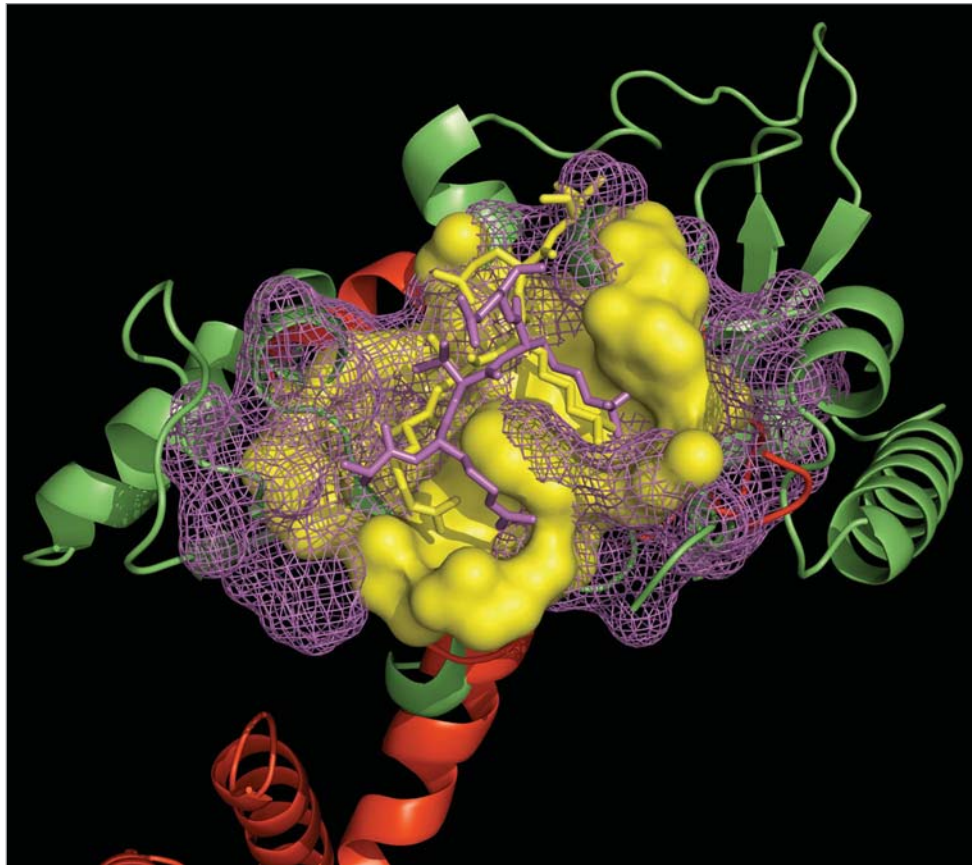


Figure 5. Structure alignment of the BS from the tandem chromodomains of human CHD1 (PDB code: 2B2W) and the PHD finger from human BPTF (PDB code: 2FSA). The alignment was obtained with MAMMOTH,¹¹ without using the structure of the bound peptides. The BS of the CHD1 chromodomains (represented with a mesh) and the bound peptide are shown in magenta, the remainder of the structure was represented with a green ribbon. The BS of the PHD finger (represented with a continuous molecular surface) and the bound peptide are shown in yellow, the remainder of the structure was represented with a red ribbon. The good coincidence of the peptide structures shows the location of the common structural motif, formed by a protruding tryptophan side chain and a hook-like peptide substructure.

suppressor p53-binding protein 1 tandem Tudor domain (PDB code: 2IG0) and the Val26Met mutant of *Drosophila* HP1 chromodomain (PDB code: 1Q3L) which reduce binding affinity for H4K20me₂,¹⁴ and abolish H3 binding,¹⁵ respectively. In both cases the mutated residues belong to the BS but are outside the MBS region, i.e., they are not in contact with the modified histone residues, thus confirming their contribution to histone binding. From the sequence point of view, comparison of the chromodomain sequences from HP1 and Polycomb, shows that some BS residues outside the MBS region are very different in nature, in spite of occupying equivalent positions in the structure alignment of both effectors. For example, residues from the *Drosophila* HP1 chromodomain (PDB code: 1Q3L) Glu23, Val26, Asp62 and Cys63 are respectively paired with residues Val25, Ala28, Leu64 and Asp65 from the *Drosophila* Polycomb chromodomain (PDB code: 1PDQ). While sequence evidence is not as strong as mutagenesis studies, the fact that residues in contact with the bound histone peptide are so different in their physico-chemical nature supports the idea that they contribute to the substrate specificity differences between these two families.³

Results of the automatic comparison of BS from homolog effectors (BS of the chromodomains from the narrow-groove class) point in the same direction. It would seem that because of the high structural similarity of these BS, which have the same MBS, they would have the same substrate specificity. However, this is not always the case, as shown for the Polycomb and HP1 chromodomains. In spite of both belonging to the same BS class, narrow-groove and MBS class, surface-groove, these chromodomains bind two different substrates,³ as mentioned before: H3 trimethylated lysines K27 and K9, respectively. Fischle and colleagues³ have shown that distinct BS features outside the MBS can play an important role determining the different substrate specificities of Polycomb and HP1 chromodomains. Our results generalize this observation, first by showing that the variability between homolog BS, described by the INTRADOM distribution (Fig. 4A), is larger than experimental noise. Therefore, there are non-trivial structural differences between BS that may result in their having different interaction profiles likely to introduce subtle specificity differences, such as those described in the case of the Polycomb and HP1 chromodomains.³ In addition, the comparison between the INTRADOM and AUTODOM rmsd

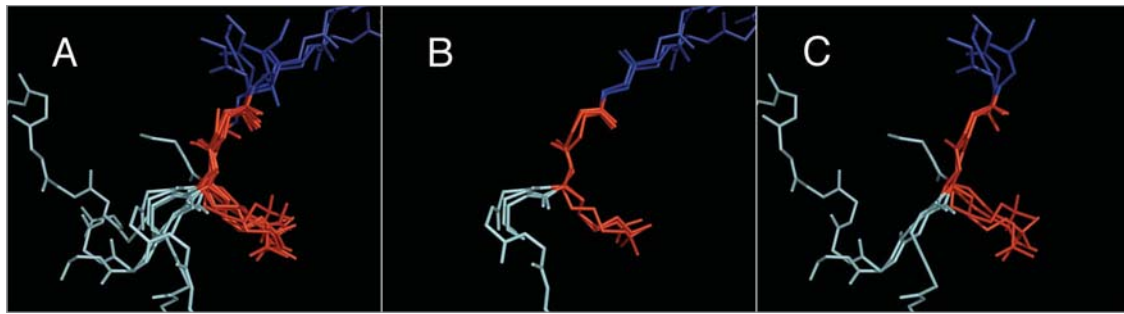


Figure 6. The hook-like motif in bound histone peptides. (A) Structural alignment of the histone peptides bound a set of representative effectors (PDB codes: 1Q3L, 1GUW, 2B2T, 1PDQ, 2FUU, 2G6Q, 2GFA, 1E6I and 2IG0). Shown in red is the common motif, the upstream (N-terminal end) and the downstream (C-terminal end) variable regions are shown in dark and light blue, respectively. (B) Same as in (A) but only for the peptides bound to the BS of the narrow-groove class only (PDB codes: 1Q3L, 1GUW and 1PDQ). (C) Same as in (A) but only for the peptides bound to the BS of the flat-groove class only (PDB codes: 2B2T, 2FUU, 2G6Q and 2GFA).

Table 2. Structural characterization of the hook-like motif in a set of representative effectors

Chromodomain ^a	1Q3L ^b	- ^c	150.0 ^d	25.0 ^e	13.0 ^f	ARK ^g	(g, g ⁺ , g) ^h
Chromodomain	1GUW	0.7	125.0	15.0	11.0	ARK	(g, g ⁺ , g)
Chromodomain	1PDQ	0.4	138.0	16.0	12.0	ARK	(g, g ⁺ , g)
Chromodomain	2B2T	1.0	126.0	25.0	11.0	RTK	(g, g ⁺ , g)
PHD	2FUU	0.7	70.0	10.0	6.0	RTK	(g, g ⁺ , g)
PHD	2G6Q	0.7	107.0	18.0	9.0	RTK	(g, g ⁺ , g)
Tudor	2GFA	0.3	107.0	19.0	9.0	RTK	(g, g ⁺ , g)
Tudor	2IG0	1.0	71.0	22.0	6.0	RK	-

^aEffector module. ^bPDB code. ^crmsd relative to 1Q3L. ^dASA buried by the hook-like motif upon binding. ^ePercentage of the hook-like buried ASA relative to the peptide's total. ^fPercentage of the hook-like buried ASA relative to the whole histone estimate. ^gResidue sequence of the hook-like motif.

^hState of the rotational angles determining the hook-like main chain structure (see text).

distributions (Fig. 4A) can be used to shed some light on the mechanism underlying the structural differences between homolog BS. It has been shown that distributions equal in nature to the AUTODOM rmsd distribution reflect the native-state dynamics of the protein.¹⁶ The fact that INTRADOM and AUTODOM rmsd distributions are different indicates that structural differences between homolog BS are unlikely to arise from the stabilization, through substrate binding, of common conformational states from their native dynamics. Rather, they are more likely to correspond to states specific to each effector that, in turn, determine the substrate binding mode. In other words, specificity differences between homolog effectors would have a structural component, complemented by the chemical differences resulting from sequence divergence, of which the percentage of non-degenerate sites (Fig. 4B) is an approximate measure. Confirmation of this idea constitutes a challenging problem that would require the use of simulation techniques beyond the structure analysis tools used in this article.

Our analyses of the peptide structures gave a picture consistent with the BS classification: we found a substantial degree of variability arising from differences between the corresponding effectors (Fig. 6B and C). However, we could also identify a structural motif, the hook-like motif, present in most peptides regardless of the effector class (Fig. 6). Analysis of the interaction pattern of this motif allows refining our view on the partition of the BS

according to specificity determinants. When the residue sequence of the hook-like motif was RTK we could see a high variability in its interaction pattern (Fig. 8B). As explained before, this variability was related to differences in the binding effectors. This indicates that the structure of atoms nearby, but outside, the MBS are involved in specificity-determining interactions. On the contrary, when the sequence of the hook-like motif was ARK its interaction pattern was well conserved. As all the ARK motif-carrying peptides were bound to homolog effectors (the chromodomains from the narrow-groove class) this confirms that the substrate specificity determinants towards different histone peptides involve effector atoms in contact with histone atoms outside the conserved sequence residue. This is in accordance with the results obtained by Fischle and colleagues³ in the case of Polycomb and HPI chromodomains.

Our view on specificity determinants was completed by the analysis of the histone peptide structures and of their local structure propensities. It is known that histone N-terminal tails are intrinsically disordered⁴ and undergo a disorder-order transition when binding effectors. Because for interactions involving disordered proteins specificity depends on this transition,¹⁷ we decided to see whether our data could shed some light upon its nature, by focusing on the composition of the disordered state. We restricted our analysis to the disordered state of ARK and RTK tripeptides, as they correspond to the residue sequence of

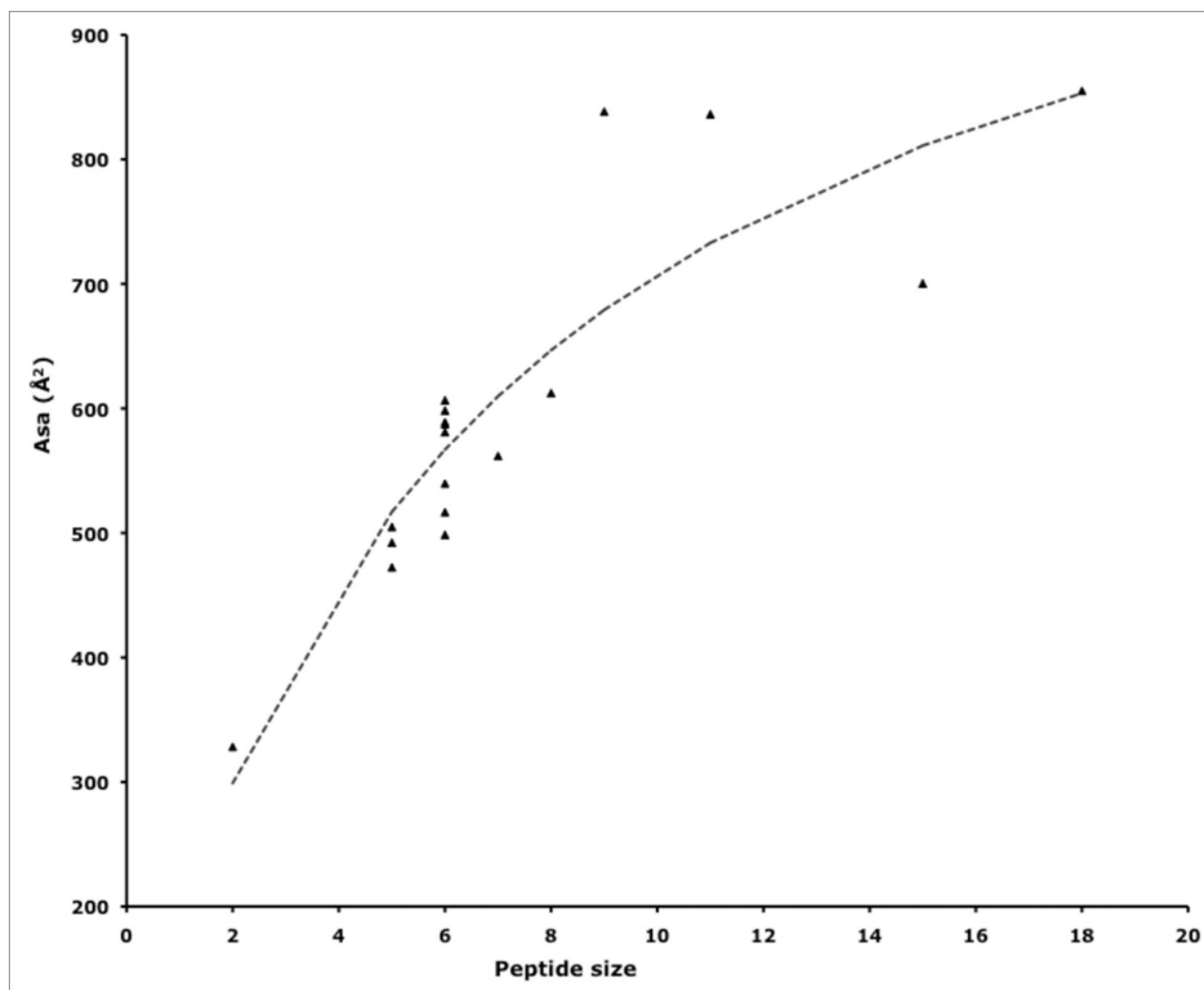


Figure 7. Histone peptide ASA buried upon effector binding vs. peptide length. The observations corresponding to the different peptides are shown with a triangle. The dashed line represents the curve of equation $1159.5 - [7611.6 / (6.0 + \text{peptide length})]$, obtained after a non-linear fit of the data.

the hook-like motifs (Table 2) and contribute an average of 44% of the ASA buried by histone peptides upon binding. The high conservation degree of the hook-like motif (Fig. 6 and Table 2) suggested that ARK and RTK tripeptides could have a strong propensity towards this structure. To test whether this was the case we analyzed all the tripeptides with these sequences present in a non-redundant set of the PDB.¹⁸ We found that ARK and RTK adopted a variety of structures (Table 3), many of them different from those found in bound histone peptides. In addition, the structural propensities of these tripeptides varied between them: for example, RTK was most commonly found in the state it adopts in histone-effector complexes than ARK. We also checked the secondary structure of both tripeptides, finding again clear differences: ARK tripeptides were more frequently found as part of α -helices (59%) than β -strands (3%), while for RTK tripeptides the differences between both states were smaller (18 and 11%, respectively). These results are in accordance with recent molecular dynamics simulations of an 18-residue peptide encompassing the first fifteen residues of histone H3.¹⁹ These simulations show that the sequence stretch between

K4 and K9 (which includes the ARK motif associated to K9) populates more frequently the helical state, while the first four N-terminal residues (which include the RTK motif associated to K4) populate more frequently the extended structure (Fig. 2 in Liu and Duan¹⁹). The consistency between (1) Liu and Duan's simulations¹⁹ and (2) our database statistics strongly supports the idea that the disordered state, and consequently the disorder-order transition, of ARK and RTK are different. The balance¹⁷ between the contribution of this transition and the atomic contacts made upon binding will determine the strength of the histone-effector interaction and the binding specificity of effectors towards given modifications. Experimental results²⁰ together with results from the aforementioned molecular dynamics simulation¹⁹ suggest that histone modifications could modulate the formation of the histone-effector complex by shifting the histone tail equilibrium population from the helical to the extended state, or vice versa. The resulting effect will combine with the contribution of the interaction between the histone peptides and the binding sites specificity determinants to produce the final binding specificity for the histone-effector interaction.

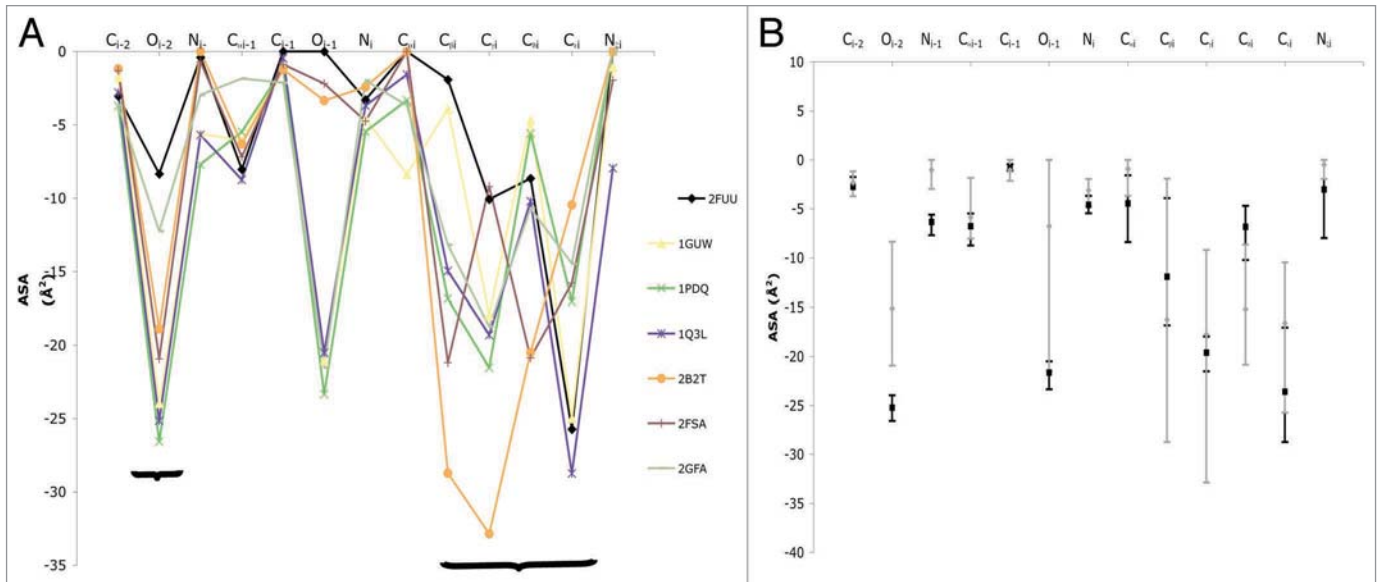


Figure 8. Interaction pattern of the histone peptide atoms from the hook-like motif. (A) ASA buried by each atom of the hook-like motif for a series of histone-effector complexes (see the figure legend). The location of the two main peaks is shown with brackets. (B) Same as (A) but here the results were averaged according to the residue sequence of the hook-like motif: ARK (black) and RTK (grey). The upper and lower bars indicate the maximal and minimal values for each atom.

The specificity of whole multidomain epigenetic regulators for given chromatin loci could then be the result of several specific histone-effector interactions, as postulated by Ruthenburg and colleagues²¹ in their multivalence model.

We are still far from a quantitative model embracing all these effects and allowing the raising of very general, experimentally testable predictions on the histone-effector interaction. However, our results can already be used to make concrete predictions on how specificity of the histone-effector interaction is modulated. For example, it has been shown that the chromodomains of Polycomb family members display differential binding to H3K27me3.²² In particular, some of them are able to bind H3K9me3, like the chromodomains from HP1 proteins. For example, Cbx7 chromodomain is able to bind both H3K9me3 and H3K27me3, and Cbx4 chromodomain prefers H3K9me3. Interestingly, in the list of Polycomb BS residues outside the MBS region and their equivalent HP1 residues (first paragraph of the Discussion) we find that Ala28 from Polycomb is paired with Val26 from HP1, a residue shown to play an important role in histone binding.¹⁵ A look at the multiple sequence alignment for the Polycomb family²³ shows that the Cbx4 and Cbx7 residues equivalent to *Drosophila* Polycomb Ala28 are valines. This strongly suggests that this mutation in one of Polycomb's specificity determinants may play an important role in the specificity shift towards H3K9me3 observed for Cbx4 and Cbx7. While there may be other associated residue changes, this example illustrates how our results can be used to shed light on the regulation of the histone-effector interaction. Also, low-resolution knowledge of the disordered state populated by histone peptides nearby modified residues, such as that provided by our database study (Table 3), may be used to obtain clues on the nature of the effector BS. For example, peptides with disordered states populating

Table 3. Distribution of states of the torsional dihedrals ϕ_{i-1} , ψ_{i-1} and ϕ_i from the ARK and RTK tripeptides found in a non-redundant subset of the PDB

ϕ_{i-1} ^a	ψ_{i-1}	ϕ_i	ARK ^b	RTK ^c
g^- ^d	t	g^-	50, 6	28, 3
g^-	t	t	15, 1	8, 2
t	t	g^-	12, 2	3, 3
g^+ ^e	g^+	g^-	11, 6	38, 6
g^-	g^+	t	5, 1	11, 4
t	t	t	2, 8	0, 5
g^+	t	g^-	1, 1	0
g^-	g^-	g^-	0, 6	4, 3
g^-	t	g^+	0, 3	1, 1
t	g^+	t	0, 3	0
t	g^+	g^+	0, 3	0
g^-	g^-	t	0	2, 7
t	g^+	g^-	0	1, 6

^a i indicates the position of the last residue, the lysine. ^bTotal number of observations: 352. ^cTotal number of observations: 184. ^dThe three states are defined as follows: $60 < g^+ < 180$; $-60 < t < 60$; $-180 < g^- < -60$. ^eShown in bold is the state found in the histone-effector complexes.

many conformations may require more interactions with the effector BS to form a stable complex than those with less heterogeneous disordered states. This in turn will require larger or deeper BS, able to form more interactions with the histone than flatter BS.

The work presented here focuses on lysine modifications, which represent a subset of all histone tail modifications.² A complete reproduction of our analyses for other target residues—Arg,

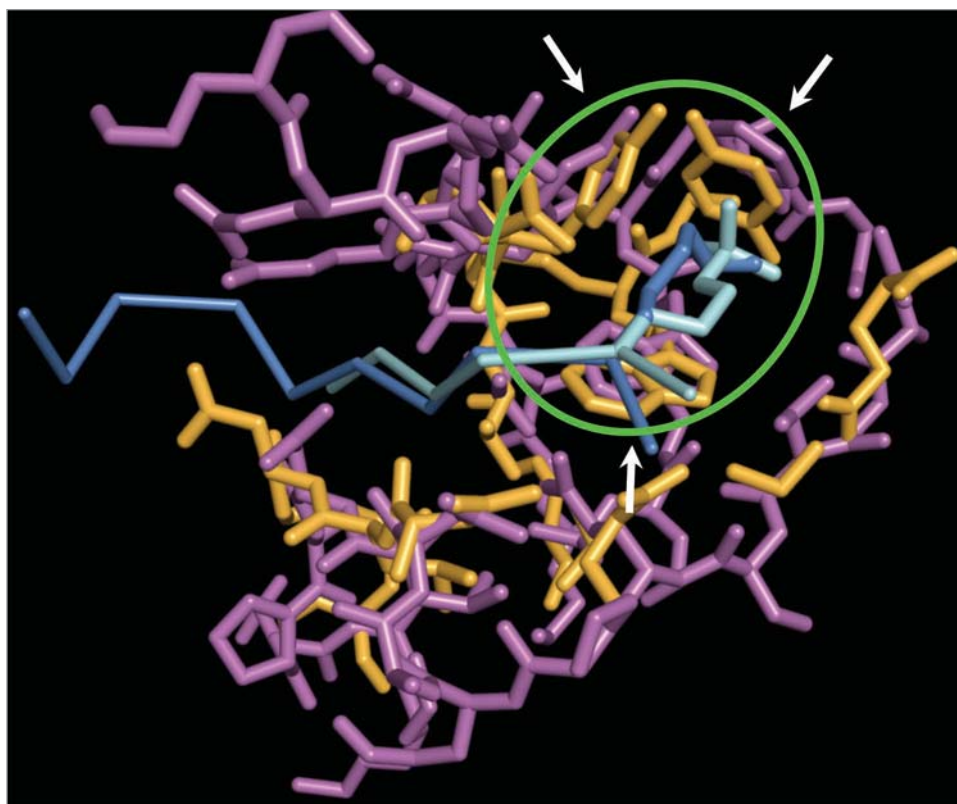


Figure 9. Performance of the structure comparison program MAMMOTH¹¹ for the alignment of binding sites. Automatic structure alignment of the BS of two HP1 beta chromodomains, from mouse (pink, PDB code: 1GUW) and fruit fly (orange, PDB code: 1KNE), respectively. The alignment was obtained with MAMMOTH¹¹ without using the structure of the bound peptides. These are shown here, in dark blue (bound to the mouse chromodomain) and light blue (bound to the fruit fly chromodomain), to highlight the quality of the BS alignment. The oval green box shows the good coincidence in the aromatic residues (signaled with white arrows) defining the aromatic cage.

Ser, Tyr, etc. and modifications—phosphorylation, sumoylation, etc., is not yet feasible for lack of structural data on effectors and their complexes with histone peptides. However, one can follow the protocol described here and use PDB¹⁸ data to obtain a first characterization of the disordered state in the neighborhood of the different histone target residues. Indeed, one can obtain approximately 477 and 24 observations for each possible tri- (20^3 amino acid combinations) and quadriptide (20^4 amino acid combinations), respectively (values obtained using the size distribution in the structure list available at Dunbrack's server,²⁴ non-redundant at 90% sequence identity and 2.5 Å resolution). While not all peptides are equally sampled, these sample sizes indicate that one could obtain a first idea of the disordered state of each peptide, and compare them as done in this work (Table 3) to explore how their disorder-order transitions vary.

Materials and Methods

Structure of the histone-effector complexes. The structures of the complexes used are listed in Table 1, with their PDB¹⁸ codes. The effectors involved were: bromodomain, chromodomain, Tudor, PHD and WDR5. Most of the histone tails were from histone H3 and variants (from different species), although in two cases they were from histone H4. They had residues, mostly

lysines, with different modifications. In a very few cases the complexes involved non-histone molecules (proteins and organic molecules) instead of histone peptides. These complexes were included to increase the sampling of the BS variability, although in some cases the automatic alignment procedure could not identify the common parts between them (see below).

The BS. BS atoms are: all effector atoms contacting any ligand atom (atom-atom distance lower than 5 Å), and some neighboring atoms, included to improve the shape representation of the BS.²⁵ The latter are obtained from the cavity pattern of the effector. First, we computed the effector cavities using SURFNET.²⁶ Second, we found all cavities with more than 50% of their atoms in contact with the ligand; all the atoms of these cavities were considered as BS atoms. Finally, atoms from other cavities were added if they were in contact with all the previously included effector atoms (atom-atom distance lower than 5 Å).

The automatic alignment of BS. BS were aligned using the program MAMMOTH.¹¹ This program was conceived for the alignment of C_α -traces and is very fast. Because its alignment algorithm is based on geometrical principles it can be applied to our problem.

To confirm that MAMMOTH alignments were meaningful, we explored manually a large number of them with the package PyMol.²⁷ We found that for the vast majority of cases

MAMMOTH would give reasonable results, aligning sets of atoms that defined a similar shape. As an example, in Figure 9 we show the alignment between two chromodomains binding two slightly different histone peptides.

It has to be mentioned that the contouring atoms that define a BS may remain constant or vary among different versions of a given structure. That is, at some locations we may find different atoms, e.g., an aromatic carbon or a polar nitrogen, depending on the BS version considered. This degeneracy has three different origins: sequence divergence, natural dynamics of the protein and technical indeterminacies. When comparing BS, degenerate positions may be excluded or included in the alignment. We decided to include them (i.e., MAMMOTH was allowed to align any pair of atoms, regardless of their nature) and complement our results with a measure of their abundance, i.e., the percentage of non-degenerate positions in the alignment. This may help to understand the degree of structural and chemical divergence at the BS.

Peptide alignment. Eight representative peptides were aligned following a semi-automatic procedure. We aligned eight peptides (taken from the complexes with PDB codes: 1GUW, 2B2T, 1PDQ, 2FUU, 2G6Q, 2GFA, 1E6I, 2IG0) against the histone peptide in the histone-chromodomain complex from *Drosophila melanogaster* (PDB code: 1Q3L), which was arbitrarily chosen. For each comparison the protocol followed was: (1) visually identify an initial set of equivalent atoms in both peptides; (2) superimpose this atom set using the Kabsch algorithm²⁸ and if the rmsd is above 1.1 Å the alignment is discarded; (3) explore visually the resulting alignment and identify any possible additional atom pairs; (4) if there are new possible pairs, add them to the original atom set and go to step (2), otherwise the protocol is finished.

Atomic surface area computations. Buried atomic accessible surface area (ASA) is an atomic contact descriptor related to the contact's free energy contribution²⁹ and for this reason constitutes a valuable tool in the study of molecular interactions. The percentage of accessible surface area buried by the hook-like motif upon complex formation was computed as follows: $ASA_{hook}(\text{peptide in isolation}) - ASA_{hook}(\text{complexed peptide})/BASA_{histone}$, where $BASA_{histone}$ was the estimate of the ASA buried by the whole histone upon complex formation (it was obtained as explained in the RESULTS section). The two peptide ASA values, $ASA_{hook}(\text{peptide in isolation})$ and

$ASA_{hook}(\text{complexed peptide})$, were obtained with the program NACCESS.³⁰

Structural propensity of the ARK and RTK tripeptides. In the hook-like motif we can distinguish two sources of structural variety: the lysine side chain torsional angles, and the main chain angles ϕ_{i-1} , ψ_{i-1} and ϕ_i . We focused our analysis on the latter, which are related to the main structural features of the disorder-order transition experienced by histone tails upon binding. To see if their highly conserved states on bound peptides were due to intrinsic sequence propensities, we looked in a non-redundant subset of the PDB¹⁸ structural database for instances of ARK and RTK tripeptides (corresponding to the hook-like motifs in our dataset) and obtained the values of their ϕ_{i-1} , ψ_{i-1} and ϕ_i torsionals ($60 < g^+ < 180$; $-60 < t < 60$; $-180 < g^- < -60$) from the output of the DSSP program.³¹

Conclusions

Using structural analyses of a series of histone-effector complexes we have characterized the structural variability of the two main components of the histone-effector interaction: histone tails and the binding site of effector modules. The results of this analysis have crystallized in a coherent classification for the latter that allows us to propose that, in general, the BS of effector domains is partitioned in two different specificity determinants (the MBS and the remainder of the BS). In addition, structural analysis of the bound histone peptides and their sequences led to the identification of an additional specificity determinant: the disorder-order transition. This transition, which takes place upon histone binding, would be a specific property of the sequence nearby the post-translationally modified histone residue (e.g., it would be different for H3K4 and H3K9). Overall, our results contribute to clarify the specificity origins for the histone-effector interaction.

Acknowledgements

We thank R. Jackson and the members of the Molecular Modeling and Bioinformatics group for helpful comments. We also thank both referees for constructive comments on our work. This work was supported by the Spanish Ministerio de Educación y Ciencia (grant numbers BIO2006-15557, BFU2006-01493, CSD2006-00049); and the Spanish Ministerio de Ciencia e Innovación (grant numbers BFU2009-11527, BFU2009-11144).

References

- de la Cruz X, Lois S, Sanchez-Molina S, Martinez-Balbas MA. Do protein motifs read the histone code? *Bioessays* 2005; 27:2-4.
- Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol* 2007; 14:1025-40.
- Fischle W, Wang Y, Jacobs SA, Kim Y, Allis CD, Khorasanizadeh S. Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev* 2003; 17:1870-81.
- Hansen JC, Lu X, Ross ED, Woody RW. Intrinsic protein disorder, amino acid composition and histone terminal domains. *J Biol Chem* 2006; 281:1853-6.
- Anantharaman V, Aravind L, Koonin EV. Emergence of diverse biochemical activities in evolutionarily conserved structural scaffolds of proteins. *Curr Opin Chem Biol* 2003; 7:12-20.
- Gherardini PF, Wass MN, Helmer-Citterich M, Sternberg MJ. Convergent evolution of enzyme active sites is not a rare phenomenon. *J Mol Biol* 2007; 372:817-45.
- Todd AE, Orengo CA, Thornton JM. Evolution of protein function, from a structural perspective. *Curr Opin Chem Biol* 1999; 3:548-56.
- Weskamp N, Hullermeier E, Klebe G. Merging chemical and biological space: Structural mapping of enzyme binding pocket space. *Proteins* 2009; 76:317-30.
- de la Cruz XF, Mahoney MW, Lee B. Discrete representations of the protein C α chain. *Fold Des* 1997; 2:223-34.
- Connolly ML. Analytical molecular surface calculation. *J Appl Cryst* 1983; 16:548-58.
- Lupyan D, Leo-Macias A, Ortiz AR. A new progressive-iterative algorithm for multiple structure alignment. *Bioinformatics* 2005; 21:3255-63.
- DePristo MA, de Bakker PI, Blundell TL. Heterogeneity and inaccuracy in protein structures solved by X-ray crystallography. *Structure* 2004; 12:831-8.
- Huang Y, Fang J, Bedford MT, Zhang Y, Xu RM. Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. *Science* 2006; 312:748-51.
- Botuyan MV, Lee J, Ward IM, Kim JE, Thompson JR, Chen J, et al. Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. *Cell* 2006; 127:1361-73.

15. Jacobs SA, Taverna SD, Zhang Y, Briggs SD, Li J, Eissenberg JC, et al. Specificity of the HP1 chromo domain for the methylated N-terminus of histone H3. *EMBO J* 2001; 20:5232-41.
16. Best RB, Lindorff-Larsen K, DePristo MA, Vendruscolo M. Relation between native ensembles and experimental structures of proteins. *Proc Natl Acad Sci USA* 2006; 103:10901-6.
17. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 2005; 6:197-208.
18. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, et al. The Protein Data Bank. *Acta Crystallogr D Biol Crystallogr* 2002; 58:899-907.
19. Liu H, Duan Y. Effects of posttranslational modifications on the structure and dynamics of histone H3 N-terminal Peptide. *Biophys J* 2008; 94:4579-85.
20. Wang X, Moore SC, Laszczak M, Ausio J. Acetylation increases the alpha-helical content of the histone tails of the nucleosome. *J Biol Chem* 2000; 275:35013-20.
21. Ruthenburg AJ, Li H, Patel DJ, Allis CD. Multivalent engagement of chromatin modifications by linked binding modules. *Nat Rev Mol Cell Biol* 2007; 8:983-94.
22. Bernstein E, Duncan EM, Masui O, Gil J, Heard E, Allis CD. Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Mol Cell Biol* 2006; 26:2560-9.
23. Senthilkumar R, Mishra RK. Novel motifs distinguish multiple homologues of Polycomb in vertebrates: expansion and diversification of the epigenetic toolkit. *BMC Genomics* 2009; 10:549.
24. Wang G, Dunbrack RL Jr. PISCES: a protein sequence culling server. *Bioinformatics* 2003; 19:1589-91.
25. Yeturu K, Chandra N. PocketMatch: a new algorithm to compare binding sites in protein structures. *BMC Bioinformatics* 2008; 9:543.
26. Laskowski RA. SURFNET: a program for visualizing molecular surfaces, cavities and intermolecular interactions. *J Mol Graph* 1995; 13:323-30.
27. DeLano WL. The PyMOL Molecular Graphics System. Palo Alto, California: DeLano Scientific LLC 2009.
28. Kabsch W. A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Cryst A* 1978; 34:827-8.
29. de La Cruz X, Calvo M. Use of surface area computations to describe atom-atom interactions. *J Comput Aided Mol Des* 2001; 15:521-32.
30. Hubbard SJ, Thornton JM. NACCESS. Department of Biochemistry and Molecular Biology, University College London 1993.
31. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 1983; 22:2577-637.