



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

---

---

**POSGRADO EN CIENCIAS BIOLÓGICAS**

**FACULTAD DE CIENCIAS**

**Aspectos estructurales de los nucleosomas**

**T E S I S**

**QUE PARA OBTENER EL GRADO ACADÉMICO**

**DE MAESTRA EN CIENCIAS BIOLÓGICAS**

**(ORIENTACIÓN EXPERIMENTAL)**

**P R E S E N T A**

**D A N I E L A S O S A P E R E D O**

**DIRECTOR DE TESIS:**

**Dr. Pedro Eduardo Miramontes Vidal**

**MÉXICO, D.F.**

**MAYO, 2008**



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



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

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

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

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

## AGRADECIMIENTOS

Al Posgrado en Ciencias Biológicas. UNAM

Al apoyo recibido del CONACYT

A los miembros de mi comité tutorial por su tiempo y apoyo: Dr. Victor Valdés, Dr. Germinal Cocho y Dr. Pedro Miramontes Vidal.

## Agradecimientos

Un agradecimiento muy especial a Pedro Miramontes, por su gran apoyo, confiar en mí y hacer de este trabajo una experiencia muy enriquecedora para mi formación académica. Además de su invaluable amistad y las buenas pláticas a enormes distancias.

A Germinal y a Víctor Valdés, por el tiempo, la paciencia, las sugerencias y el apoyo que me dieron.

A Lorena Caballero y Arturo Becerra, por su gran apoyo y amistad.

A mis papás Luis Fernández y Cristina, a mi tía Rosalía y a mi abuela por su cariño y todos los ánimos que me han dado. Cri gracias por revisar este trabajo.

A mi Rodrigo por hacerme muy feliz, y hacerme recordar lo bonito que es ser niño y lo maravillosa que es la primaria.

A mi Yeyo, por todo el cariño, los buenos momentos y toda la alegría.

A Víctor Mireles por su apoyo en este trabajo y su valiosa amistad.

A mi familia y a mi tío Julio por todo el cariño y la confianza.

A mis queridísimas amigas Yssel, Gaby, Yetzi, Yara, Ade, Claudia, Tana, Ani, Naty y Lucía, por las miles de risas, sus consejos y todo su cariño.

# Índice general

<b>1. Introducción</b>	<b>2</b>
<b>2. Antecedentes</b>	<b>6</b>
2.1. Características estructurales del DNA . . . . .	8
2.1.1. Parámetros estructurales . . . . .	10
2.2. Nucleosomas . . . . .	13
2.3. Nucleosomas en arqueas . . . . .	18
<b>3. Secuencias del DNA asociadas a nucleosomas</b>	<b>23</b>
3.1. Posición de nucleosomas en eucariontes . . . . .	24
3.2. Posición de nucleosomas en arqueas . . . . .	29
<b>4. Metodología</b>	<b>33</b>
<b>5. Resultados</b>	<b>39</b>
<b>6. Discusión y Conclusiones</b>	<b>46</b>
<b>7. Apéndice</b>	<b>50</b>
Bibliografía	60

# Capítulo 1

## Introducción

En la naturaleza podemos observar la existencia de eventos periódicos, un ejemplo muy conocido es el de los relojes biológicos. La frecuencia de estos sucesos puede medirse para determinar cuál es el periodo con el que aparecen dichos ciclos. Sin embargo en muchas ocasiones las señales de frecuencia son difíciles de determinar cuando estas se mezclan.

Uno de los métodos más usados para distinguir diferentes períodos de señales mezcladas o con ruido, es el análisis de Fourier, el cual descompone una función dada en varias funciones periódicas (senos o cosenos) que permiten determinar la existencia de diferentes señales y definir los distintos períodos.

En la molécula del DNA existen formas muy diversas de secuencias que se repiten periódicamente y considerando que es posible cuantificar la distancia o el número de nucleótidos que hay entre un nucleótido determinado o un conjunto de nucleótidos y otro igual, se han podido utilizar métodos como por ejemplo el análisis de Fourier, para establecer diferentes patrones de periodicidad y su distribución en el genoma, por ejemplo el uso de codones en exones, los Alu repeats, y diversos patrones en las regiones de heterocromatina.

Otro evento con una periodicidad constante en el DNA es la presencia de proteínas nucleosómicas, en las que se enrolla el DNA, cuya función es de gran importancia para la organización del genoma y la regulación de la expresión de genes. Por esta razón los trabajos que se han realizado para la detección de secuencias asociadas al DNA son acercamientos interesantes para entender el comportamiento de estos mecanismos reguladores.

Este trabajo tuvo por objetivo probar una herramienta que permitiera detectar períodos en la secuencia de DNA que respondieran a un patrón de posibles secuencias asociadas a nucleosomas, esperando realizar en trabajos futuros la caracterización de estas secuencias y definir los niveles de restricción estructural del DNA que influyan en la funcionalidad del genoma y el sesgo implícito en la variabilidad genética.

Se ha visto que la asociación de proteínas al DNA se debe a reconocimientos estructurales en los que interactúan tanto la secuencia primaria del DNA como la estructura tridimensional de las proteínas. Sabemos que las proteínas nucleosómicas son proteínas estructuralmente muy conservadas por lo que las regiones de DNA en las que se asocian, deberían de mantener un patrón, el cual se ha buscado con relativo éxito en la sucesión de nucleótidos en muchos trabajos, algunos comentados más adelante.

En este trabajo se utilizó la transformada rápida de Fourier para encontrar periodos que correspondieran a la presencia de nucleosomas utilizando como señal, en vez de la sucesión de nucleótidos, las características fisico-químicas de los nucleótidos, así como la presencia de enlaces fuertes o débiles que se forman con la cadena complementaria de la doble hélice y la variación de la energía libre entre nucleótidos contiguos, es decir, la energía

liberada que se transforma en trabajo y calor a partir del rompimiento de los enlaces químicos en un sistema cerrado, estas son cualidades que comparten y diferencian a los nucleótidos y que puede conferir a un conjunto de nucleótidos las características físicas necesarias que le permitan enrollarse en los complejos nucleosómicos, por esto, consideramos estas cualidades como posibles señales para encontrar periodos relacionados con la presencia de nucleosomas y encontramos que con estas señales se pueden detectar periodos que presumiblemente corresponden a regiones de DNA espaciador o internucleosómico, por lo menos en las muestras analizadas y sin los sesgos experimentales comentados a lo largo de este trabajo y en la discusión.

En el capítulo de antecedentes se describen las características estructurales del DNA y los parámetros que se utilizaron para realizar este trabajo; se explica la composición de los nucleosomas, tanto en eucariontes como en arqueas y su función en el genoma. En el tercer capítulo se mencionan algunos de los trabajos y métodos que se han utilizado para la identificación de secuencias asociadas a nucleosomas en ambos dominios. En el capítulo cuatro se describe la metodología aplicada en el desarrollo del trabajo experimental; se explica cómo se manejaron las secuencias, tomadas al azar, de los genomas seleccionados y el análisis de Fourier, con el que se trabajaron los datos. En el capítulo cinco se exponen los resultados del análisis de secuencias con la metodología expuesta anteriormente y se presentan algunas de las gráficas en las que se observan los periodos encontrados en cada muestra de genoma. En el capítulo sexto se plantea una discusión sobre estos resultados obtenidos y sobre el método utilizado, comparándolo con los trabajos descritos en el capítulo de antecedentes; además, se plantean las posibles im-



plicaciones que tiene la presencia de patrones en la secuencia de nucleótidos como sesgos en la variabilidad genómica. Por último, en el apéndice se presentan el resto de las gráficas resultantes del análisis de periodicidad con los parámetros estructurales obtenidos de las secuencias genómicas.

# Capítulo 2

## Antecedentes

La organización del genoma eucarionte le ha conferido cualidades particulares a estos organismos, dentro de los niveles de organización del genoma se encuentra la unidad estructural más pequeña que es la de los nucleosomas (ver fig. 2.1), en estas estructuras se enrolla el DNA conduciendo al empaquetamiento del genoma para formar los cromosomas.

Los nucleosomas, son complejos de proteínas llamadas histonas, que involucran los niveles de organización primarios del genoma y son elementos básicos para entender una parte importante de las funciones estructurales del genoma, estabilizando, protegiendo y restringiendo físicamente la expresión génica y fomentando, muy probablemente, la especialización celular [23].

La conservación estructural en los eucariontes y la interacción de este complejo de proteínas con la secuencia de DNA, ha permitido observar una relación estrechamente dependiente de la conformación estructural tanto de las proteínas como de la misma secuencia de DNA a la que se unen.

Sin embargo, esta relación es dinámica debido a que hay factores externos que modifican estructuralmente al DNA y a dichas proteínas, permitiendo con ello el enrollamiento y desenro-

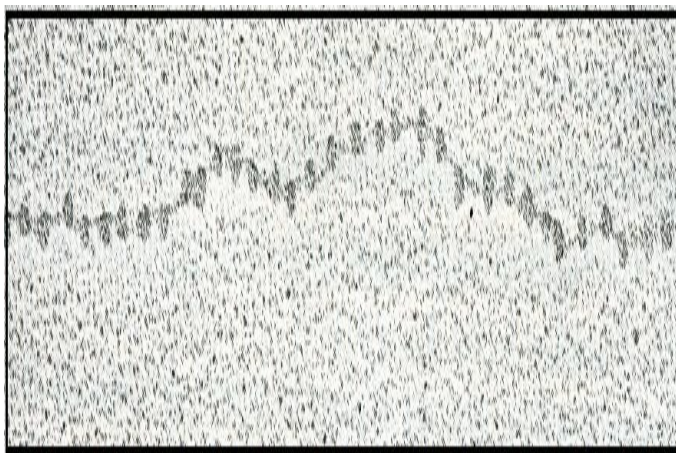


Figura 2.1: Nucleosomas observados en el microscopio electrónico (Alberts,1996)

llamamiento del DNA en el nucleosoma, exponiendo así secuencias génicas y sitios de activación transcripcional a los que se unen factores transcripcionales para que sean fácilmente expresadas. La presencia de nucleosomas en el DNA y todas estas implicaciones abren un campo de investigación muy interesante dentro del área conocida como *epigenética*, a la que se le atribuye la modulación de la expresión de diversos genes y el silenciamiento de otros [3], sin necesidad de cambiar la secuencia del DNA. Estos mecanismos de modulación en muchos casos son información heredable de la cual prescinde el DNA para su expresión correcta y con la falla o la herencia de señales epigenéticas se pueden ver involucradas enfermedades, como algunos casos de retraso mental y ciertos tipos de cáncer en humanos[10].

## 2.1. Características estructurales del DNA

La molécula del DNA está formada por bases nitrogenadas débiles (purinas y pirimidinas), moléculas de azúcar (desoxiribosa) y al menos un grupo fosforilo. A la molécula que se forma por la unión de una base nitrogenada con un azúcar se le llama *nucleósido*. El enlace glicosídico se da entre el carbono anomérico y el N(9) de una purina o el N(1) de una pirimidina.

Los nucleótidos se forman cuando una molécula de ácido fosfórico se esterifica al azúcar de un nucleósido. Esta esterificación ocurre en la posición 3' o 5' de un deoxiribonucleósido y en la posición 2',3',5' de un ribonucleósido.

Una cadena sencilla de nucleótidos se forma por la unión fosfodiéster en la región 3'-5' de los azúcares a los que se encuentran unidos los grupos fosfatos, la unión por puentes de hidrógeno de los nucleótidos con los nucleótidos de otra cadena complementaria y de forma antiparalela formarán la doble hélice del DNA [1]. Sin embargo la doble hélice es demasiado irregular pues puede adoptar tres tipos de estructuras, que dependes de la secuencia de nucleótidos, de la humedad y de la presencia de iones. Estas estructuras se conocen como B, Z y C (ver fig.2.2).

La hélice del DNA-B, considerada como la conformación nativa ideal y la que se observa funcionalmente en los seres vivos, tiene aproximadamente 10 pares de bases por vuelta [36]. Esto quiere decir que tiene un giro helicoidal de 36 grados por cada par de bases. Las bases aromáticas tienen espesores que le dan los enlaces de Van der Waals de  $3.4\text{Å}$  y están parcialmente apiladas una sobre otra, por lo que la hélice presenta una elevación por vuelta de  $34\text{Å}$ . Esta conformación origina que en el enrollamiento de la hélice se formen dos surcos uno mayor y otro menor (ver fig.2.3).

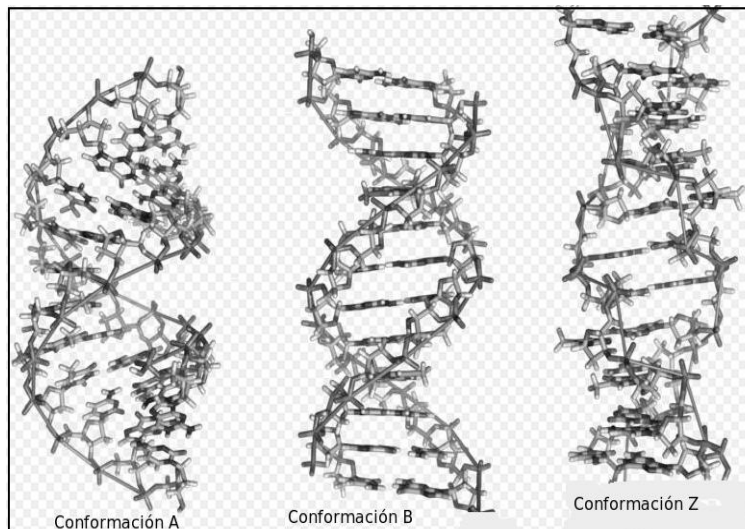


Figura 2.2: Conformaciones del DNA

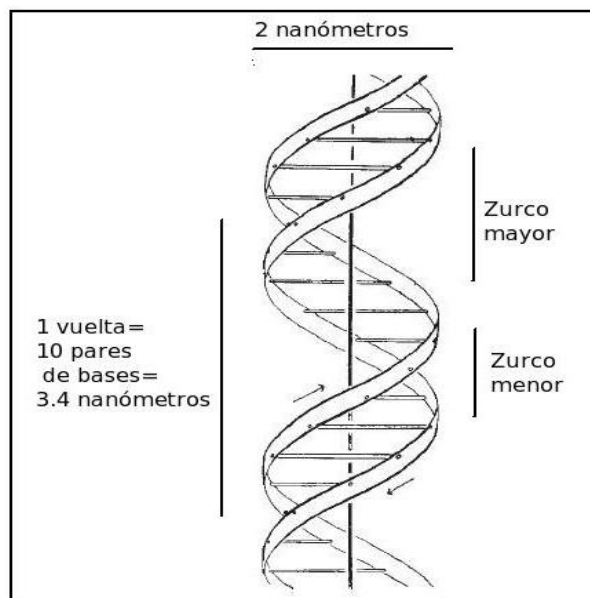


Figura 2.3: Estructura del DNA

En un primer orden, la conformación estructural del DNA, está determinada principalmente por la secuencia de nucleótidos y por las interacciones que derivan de los mismos, (tanto entre nucleótidos contiguos como con los nucleótidos que interaccionan para formar la doble hélice). En un segundo orden, está determinada por proteínas asociadas al DNA y algunos grupos químicos. Sin embargo, es determinante la configuración del primer orden asociado a la secuencia de nucleótidos, para la configuración resultante del DNA. Las características químicas y físicas de los nucleótidos confieren conformaciones singulares que no solamente caracterizará estructuralmente al genoma de una especie con una firma estructural (IDH) (ver fig. 2.4)[27], sino también afectarán la afinidad a proteínas que interaccionan con el DNA, como los factores de transcripción, replicación, histonas, etc.

### 2.1.1. Parámetros estructurales

Por su naturaleza química los nucleótidos se clasifican por su tamaño en dos diferentes grupos, purinas (Guanina y Adenina) y pirimidinas (Timina y Citocina). Las pirimidinas son compuestos heterocíclicos que contienen cuatro átomos de carbono y dos de nitrógeno, mientras que las purinas forman una estructura bicíclica que tiene una pirimidina fusionada a un anillo imidazol[18].

La secuencia de nucleótidos a lo largo de una hebra del DNA adquiere una conformación específica que le da las características típicas del DNA -B, y que está restringida por una posibilidad limitada de ordenamiento de purinas y pirimidinas. En todas las sucesiones de purinas-pirimidina (CA, TG y TA) la distancia tomada entre el protón del nucleótido y el protón anomérico es

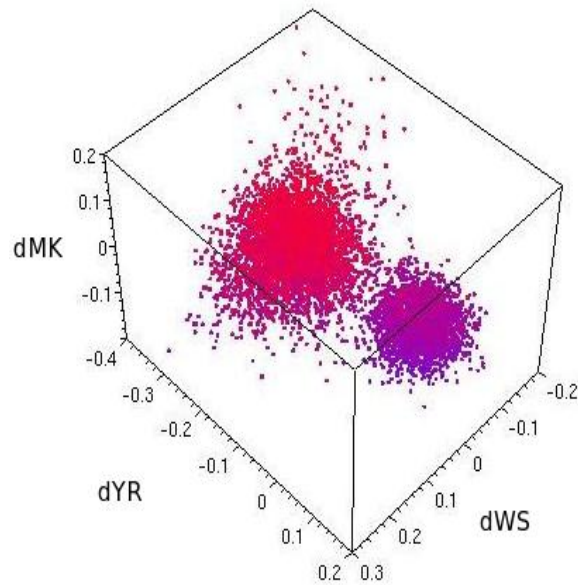


Figura 2.4: Firmas estructurales de los genomas de *R.norvegicus* y *Acrobacter sp.* Estas firmas representan la homogeneidad del DNA determinada por tres dicotomías que considera los siguientes parámetros estructurales: i) nucleótidos formadores de enlaces fuertes y débiles (dWS), ii) nucleótidos que por su naturaleza química se clasifican en purinas o pirimidinas (dYR) y iii) nucleótidos que tienen un grupo cetona o un grupo amino orientado al surco mayor de la doble hélice (dMK). La sucesión de nucleótidos tomando en cuenta estos parámetros es analizada en un algoritmo de distribución binaria llamado Índice de Homogeneidad del DNA (IDH), desarrollado por Miramontes *et al.* 1995

mayor de  $45 \text{ \AA}$ . En GC la distancia varía en un rango de  $3.8\text{-}4.5 \text{ \AA}$  mientras que entre AA el límite más bajo es de  $3.7 \text{ \AA}$  y el más alto es de  $4.2 \text{ \AA}$ . En las sucesiones de purina-purina, la distancia es afectada al menos por tres factores, i) por las secuencias adyacentes de purina-pirimidina, que causan un incremento en la distancia, ii) una serie de AT adyacente causa un decremento en la distancia y iii) una sustitución de bases en la posición

del vecino próximo más cercano permite cambios en la distancia entre GA y AA. Estas diferencias entre las separación de los nucleótidos en la cadena son una sonda de la anchura del surco menor característico del DNA-B, por lo que cualquier cambio en estos parámetros se reflejará en cambios en la anchura típica del surco menor de la conformación B del DNA[7].

Por otro lado, los dos tipos de enlace que se forman entre pares de bases de cadenas complementarias, es decir los enlaces fuertes (Guanina-Citocina) y los enlaces débiles (Adenina-Timina), le dan a la molécula del DNA una flexibilidad y estabilidad que no son uniformes y contribuye a la varianza entre los ángulos estructurales, lo que da a la molécula una estructura rugosa o lisa, dependiendo de esta varianza.

Las fuerzas termodinámicas en el DNA, juegan un papel importantísimo en el comportamiento de la molécula. En este sentido, la inmensa mayoría de los estudios en sistemas de ácidos nucleicos se han centrado en la caracterización termodinámica de la fusión o interacción de ácidos nucleicos, tanto monomolecularmente como bimolecularmente. Las fuerzas de interacción o fusión entre los nucleótidos, se refieren al punto de equilibrio termodinámico que las mantiene unidas y en donde la temperatura juega un papel central para evitar la desnaturalización del DNA, por lo que fragmentos de diferentes tamaños del DNA pueden ser caracterizados por la temperatura en la cual estas uniones se separan. Estos estudios tienen muchas aplicaciones, entre ellas, las utilizadas para hacer PCRs[25], el análisis de curvas de desnaturalización también permite saber la longitud de la cadena del DNA, el contenido de AT y localizar ciertos dominios del DNA; si se fraccionan las secuencias del DNA este tipo de análisis permite saber el orden de estos fragmentos, o si se tiene



secuenciada la mayor parte del genoma y sólo se necesitan saber algunas partes, este método puede servir para saber cuáles son los nucleótidos que falten y así complementar la secuenciación exacta [48].

Sin embargo, la secuencia de nucleótidos presenta un orden estructural mayor puesto que la torcedura de la doble hélice es variable, por lo que es necesario extender los análisis a eventos multimoleculares, es decir, que contemplen la fusión o interacción de más de dos moléculas[4].

Caracterizando el comportamiento del DNA con estos parámetros, Breslauer obtuvo una tabla de datos con la energía libre entre enlaces de nucleótidos contiguos, con lo cual se caracterizó una forma de señal físico-química que podría facilitar, como en este trabajo se propone, la detección de patrones en la secuencia del DNA.

## 2.2. Nucleosomas

Los nucleosomas consisten en 165 pb aproximadamente de superhélices de DNA acomodadas en dos giros alrededor de un núcleo de histonas, el cual está formado por un octámero de estas proteínas (H2A, H2B, H3 y H4), en cada nucleosoma hay dos histonas de cada una para formar este octámero (ver fig. 2.5). Cada nucleosoma está separado del siguiente por una longitud variable de 0 a 80 pb.

El DNA alrededor de estas proteínas está parcialmente accesible a proteínas regulatorias pero las histonas tienen tallos (residuos de aminoácidos) que pueden ser modificados químicamente en procesos reversibles (ver fig. 2.6) lo cual permite el movimiento y relajación del DNA con profundos efectos locales

en el complejo de la cromatina, permitiendo el acceso a proteínas regulatorias [1]. Por esto, los tallos de las histonas son imprescindibles para la remodelación del DNA que permite la expresión o inhibición de genes y secuencias reguladoras[2], mostrando también un proceso dinámico de rápida transferencia reversible de energía entre los dos estados que el DNA puede optar en relación con los nucleosomas: cerrado o abierto [26].

Dentro de las formas más conocidas y descritas de estas modificaciones están la metilación y la acetilación, en aminoácidos de los tallos de las histonas, pero también se ha encontrado que nucleótidos en el DNA tienen dichas modificaciones, como la metilación de la citosina, que además son señalizaciones heredables [17]. Se ha observado que estos procesos de modificación química ocurren regularmente en aminoácidos específicos y en combinación con la modificación química de otros aminoácidos en el mismo tallo [39]; también se ha observado que estos sitios son también muy conservados en diferentes organismos, por lo que se está trabajando en el desarrollo de un código de histonas y en el estudio de otros procesos de modificación química como fosforilización y ubiquitinación (ver fig. 2.8)[13].

Los tallos de las histonas tienen un papel importantísimo en la remodelación de la cromatina, en donde cualquier variación o mutación de estos afectará su actividad, implicándolas posiblemente en un papel evolutivo y participativo en la diversificación de los seres vivos. Muchos de los trabajos implicados con estructura y química de las histonas han sido trascendentes para entender el papel de las histonas en un contexto evolutivo, puesto que las histonas que forman el nucleosoma, H2A, H2B, H3 y H4, se cuentan entre las proteínas más conservadas. Este grado de conservación sugiere que en la función de las histonas parti-

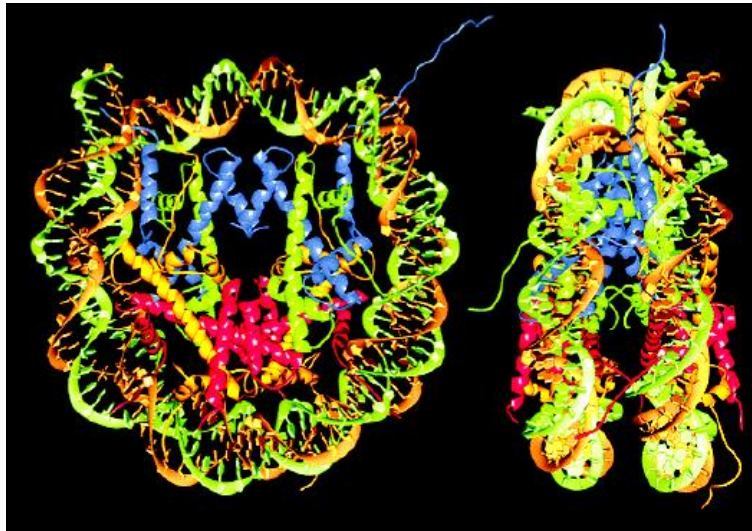


Figura 2.5: Nucleosoma (Fesenfeld, Groudine,2003)

cipan casi todos sus aminoácidos de modo que cualquier cambio en cualquier posición resulta deletéreo para la célula. Sin embargo, en estudios recientes se ha revelado que existen variantes de histonas que están implicadas en diversas funciones, como el caso de variantes de las histonas H2A, y H3 que presumiblemente tienen estas diversificaciones para asumir roles en el silenciamiento epigenético, en la expresión de genes y funciones centroméricas [14].

Un ejemplo es la histona H3 que en la región centromérica cambia por la CenH3 y cuya variación es altamente notable en los residuos de los tallos N-terminal. Tanto los tallos N-terminales como el plegamiento del dominio de la histona son sujetos a una adaptación evolutiva selectivamente positiva, como se ha visto en *Drosophila* al igual que en *Arabidopsis*, (organismos en los que se han estudiado estas variantes) ya que estas modificaciones estructurales permiten en la proteína una afinidad

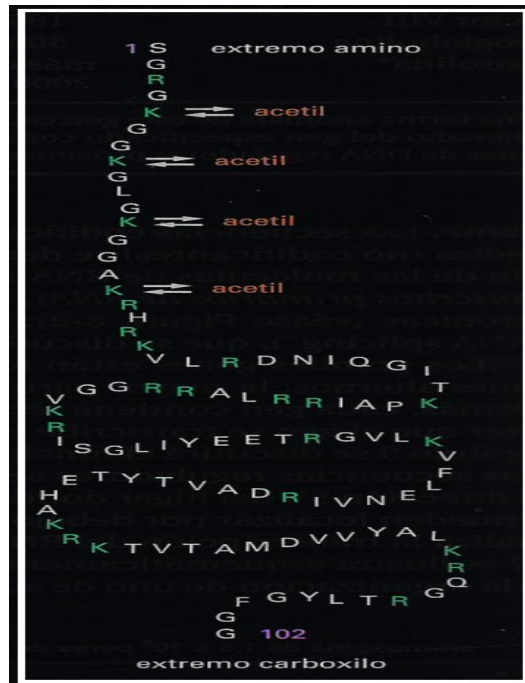


Figura 2.6: Tallo de la histona H4 (Alberts *et al*, 1996)

mayor al DNA, además de permitir la orientación centromérica requerida [8].

Existen otro tipo de histonas llamadas H1 son las histonas de mayor tamaño, están formadas por unos 220 aa y han sido menos conservadas evolutivamente que las histonas nucleosómicas [1]. La histona H1 es el mayor factor de estabilidad de la cromatina, el empaquetamiento con la H1 es operativamente importante en el silenciamiento de genes y formador de la heterocromatina (ver fig. 2.7) .

La H1 es dependiente también de la presencia de otras proteínas para su unión con el resto del nucleosoma como por ejemplo la ParaT, aunque en este ejemplo la función de la ParaT como inhibidor no es exclusiva para la H1 [15]. Sin embargo, la

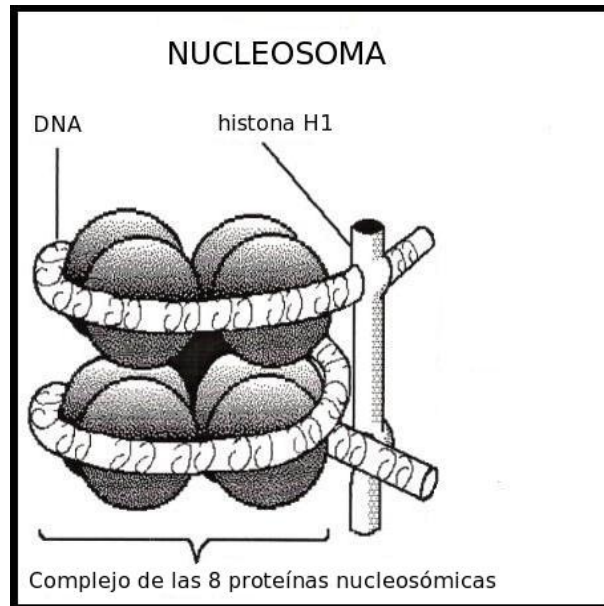


Figura 2.7: Función de la histona H1

formación de heterocromatina depende de la presencia de H1 y por tanto de cualquier otro factor que pueda inhibir esta unión [42].

Se ha visto también que la actividad conjunta de H1 y el complejo regulador Swi6 mantienen en un estado condensado a la heterocromatina debido a la formación estable de las interacciones entre los nucleosomas, lo cual sugiere que los eventos de unión de Swi6 son consistentes con un modelo estocástico de heterocromatina, que indica la conservación evolutiva de las propiedades de la heterocromatina para la unión a proteínas desde levadura hasta los mamíferos. También la metilación sobre la lisina 9 de la histona H3 es característica de la heterocromatina [43].

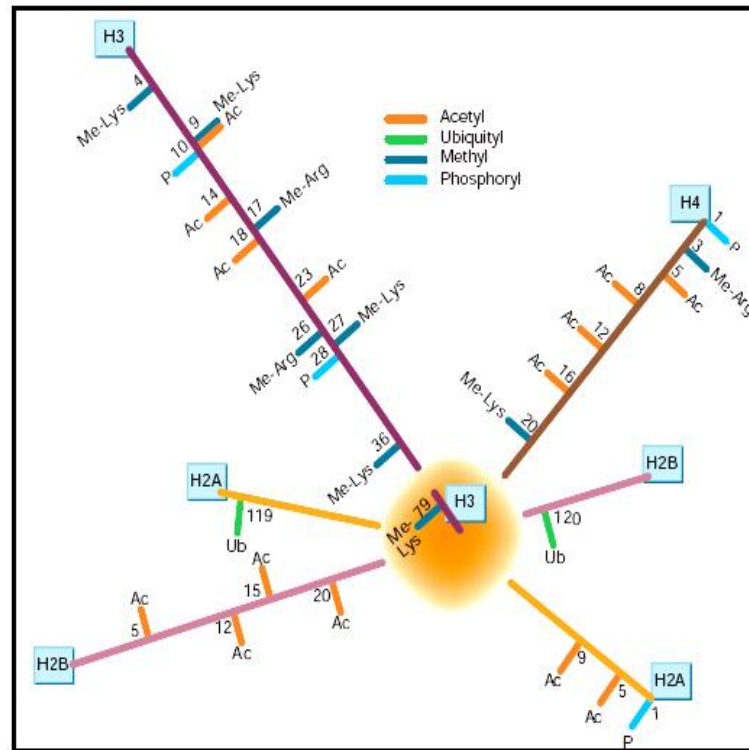


Figura 2.8: Código de Histonas(Felsenfeld, Groudine, 2003)

### 2.3. Nucleosomas en arqueas

El complejo de histonas ha sido de las proteínas más altamente conservadas entre los eucariontes, sugiriendo que la estructura fundamental de la cromatina involucra a un ancestro común en los eucariontes [13]. Sin embargo, los equivalentes a las histonas y estructuras simplificadas de la cromatina han sido también encontradas en organismos unicelulares del dominio *Archaeobacteria* [37].

Uno de los primeros trabajos en donde se habla de la presencia de nucleosomas en arqueas, es de Masaki Shioda y colabora-

dores [40], en el que mediante microscopía electrónica describen partículas protéicas asociadas al DNA, con 9.4 nm de diámetro y distribuidas repetidamente a lo largo de las fibras de DNA.

En 1991 Musgrave describe a la histona HMf, de la arquea hipertermofílica *Methanothermus fervidus*, unida a una doble cadena de DNA *in vitro*, formando estructuras compactas similares a nucleosomas. Estimaron que cada estructura de HMf-DNA contenía entre 90 y 150 pb. de DNA enrollado en un núcleo de cuatro moléculas de HMf [28].

Finalmente se logra la cristalografía de la histona HMk de arqueas reportada por Richard L. Fahrner (ver fig. 2.10) [12].

Algunos autores sugieren que además de que la presencia de nucleosomas en arqueas y eucariontes muestra un posible ancestro común, la estabilidad conferida por las histonas y la actividad de enzimas topoisomerasas reverso-girasas que las hace resistentes a la desnaturalización térmica, juegan una papel central en la hipótesis del origen hipertermofílico de la vida [29].

La función de las histonas en arqueas puede reflejar la función ancestral y por consiguiente el desarrollo que hubo originalmente de la cromatina como un importante mecanismo de regulación génica. El uso de las histonas en el empaquetamiento, ha sido tan benéfico para reclutar la estructura nucleosómica más compleja que se desarrolló subsecuentemente en los ancestros de los eucariontes modernos con un genoma tan expandido.

Las arqueas tienen histonas que contienen secuencias primarias comunes con el núcleo de histonas nucleosómicas en eucariontes y una estructura tridimensional que es esencialmente única del plegado de histonas. Las estructuras que se reportan en arqueas son similares al tetrámero de (H3 y H4)<sub>2</sub> en el centro del nucleosoma eucariótico, siendo esta una semejanza que compar-

ten las células arquea y eucaria, otro ejemplo son las polimerasas de RNA que tienen una secuencia primaria y complejidad de subunidades común con las polimerasas RNA eucarióticas, los promotores de arqueas contienen los elementos de la caja TATA y la iniciación de la transcripción en arqueas requiere la participación de proteínas, estructural y funcionalmente homólogas a las proteínas que se unen a la caja TATA y a los factores de transcripción (ver fig. 2.9) TFIIB [34]. Por ello el descubrimiento de histonas en arqueas es consistente con la relación ya establecida entre la maquinaria trascricional basal de arqueas y eucariontes. Los sistemas de reparación y replicación del DNA en eucariontes como en arqueas han mostrado muchas características en común que difieren de su contraparte en bacterias [35].

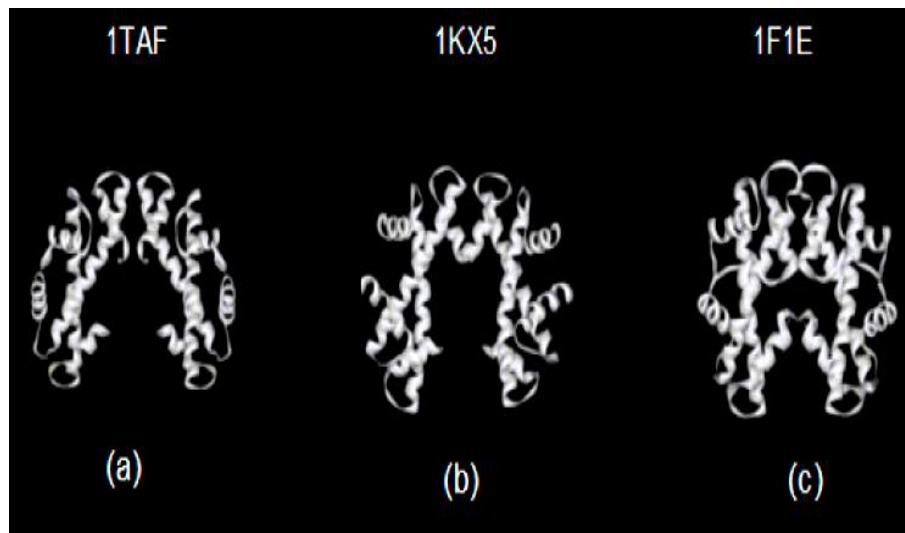


Figura 2.9: Estructuras de histonas y factores de transcripción

En contraste con los eucariontes, cuyos nucleosomas están formados por histonas que tienen entre 102 y 135 residuos de



aminoácidos, todas las histonas estudiadas en arqueas tienen una longitud similar, entre 66 y 69 residuos de aminoácidos, que al plegarse forman tres hélices  $\alpha$  (short $\alpha$ 1, long $\alpha$ 2, short $\alpha$ 3) separadas por dos cortas  $\beta$  plegadas (L1, L2). Son estables solamente plegadas en la configuración de dímero. En un dímero las dos  $\alpha$ 2, están alineadas antiparalelamente con las L1-L2a y las asas L1a-L2 están pareadas a cada final de las alineadas  $\alpha$ 2- $\alpha$ 2a. Las hélices  $\alpha$ 1 y  $\alpha$ 3 están plegadas hacia atrás en el mismo lado que  $\alpha$ 2 y todos los lados de las seis  $\alpha$ -hélices contribuyen a la formación del dímero del complejo hidrofóbico [35].



Figura 2.10: complejo de histonas en arqueas, (Fahrner, 2001)

Una de las diferencias entre la relación de histonas/DNA en arqueas y eucariontes es que las histonas de arqueas HMf y HTz envuelven al DNA en vueltas con sentido negativo, por lo que la topología que conforman los nucleosomas de arqueas, a dife-

rencia de los nucleosomas eucariontes que tienen preferentemente un enrollamiento positivo, las arqueas lo tienen en sentido opuesto[30].

El conjunto de histonas en eucariontes tiene una extensión adicional, que no se presenta en las arqueas, de secuencias en la región N y C terminal de las histonas, estas extensiones cargan residuos que les permiten tener modificaciones junto con los complejos remodeladores de la cromatina lo que facilita la unión de factores de transcripción y permiten la remodelación post-transcripcional, regulando la estructura de la cromatina y la expresión génica. Sin embargo, en arqueas no hay evidencia de que las histonas tengan tallos, por otro lado se ha demostrado que la transcripción en arqueas se lleva a cabo aún en presencia de nucleosomas, aunque de manera más lenta pero no es bloqueada [50]. Tres histonas de arqueas HMvA, MM0929 y MJ1647 de tres especies methanococales, *Methanocaldococcus janaschii*, *Methanococcus voltae*, y *Methanococcus maripaludis* tienen respectivamente 30 residuos aproximadamente de extensión en la región C-terminal. Estas tres secuencias están relacionadas pero no muestran una relación obvia con las extensiones de las histonas eucarióticas [35].

## Capítulo 3

# Secuencias del DNA asociadas a nucleosomas

Las herramientas bioinformáticas nos han permitido analizar secuencias de DNA para inferir o aproximarnos a los procesos evolutivos o por lo menos al diferente comportamiento que han tenido los genomas y genes de diferentes linajes, para modelar o predecir funciones biológicas. En el extenso campo que abarca la bioinformática, no se ha hecho a un lado la posibilidad de utilizar dichas herramientas así como la integración de otras disciplinas, con el propósito de detectar secuencias asociadas a nucleosomas debido a la importancia de la actividad nucleosómica en la regulación epigenética de la célula.

La obtención de datos experimentales ha sido esencial pero también lo ha sido el desarrollo de herramientas computacionales de análisis que predigan o ubiquen la posición de los nucleosomas en la secuencia de DNA, así como el desarrollo de algoritmos estadísticos y la implementación de diferentes consideraciones para inferir las mejores aproximaciones, como es diferenciar la información que contiene la secuencia, tomando en cuenta por ejemplo, los diferentes tipos de promotores de

transcripción genética, los diferentes tipos de genes y muchos otros aspectos sobre la funcionalidad de las secuencias génicas; en otros casos tomando en cuenta características fisicoquímicas de las secuencias.

Muchos de estos trabajos, comentados más adelante, se han aproximado de maneras muy distintas a predecir la posición de nucleosomas en el DNA, tomando como premisa y obteniendo como conclusiones la existencia de patrones de secuencias de DNA asociados a nucleosomas, los cuales en muchos casos son consistentes con las características generales que le permitan a la molécula de DNA la flexibilidad necesaria para envolverse en los nucleosomas, pero no son idénticos entre sí, pues en los resultados de estos trabajos los patrones varían, entre otras cosas, con la funcionalidad de la secuencia.

### **3.1. Posición de nucleosomas en eucariontes**

Entre los primeros trabajos desarrollados para predecir la posición de nucleosomas en el DNA de genomas eucariontes está el trabajo de Lowary y Windom, quienes desarrollando una ecuación para medir la energía libre en fracciones de ocupación de posiciones específicas del DNA y encontraron que al medir las diferencias de energía libre entre dos clones de secuencias de eritrocitos de pollo, las cuales separan de nucleosomas y luego reconstruyen; se pueden describir dos formas de afinidad del DNA a nucleosomas, por posición rotacional y por transicional; la transicional se refiere a la extensión que puede seleccionar un octámero de histonas en un espacio contiguo y particular de 146 pb en vez de cualquier otro espacio de igual longitud adelante o atrás de estos 146 pb. La posición rotacional es sólo una

degeneración de la transicional, en la cual un conjunto de posiciones discretas y de transición difieren por integrales múltiples de DNA repetido y todos estos sitios pueden ser ocupados indistintamente por los octámeros de histonas. Mediante esta descripción desarrollaron una ecuación que calcula la probabilidad de posición de nucleosomas en la secuencia de DNA, resultando que al menos 95 % de la masa total de DNA tienen afinidad a los octámeros de histonas en los nucleosomas [24].

Para detectar interrelaciones entre perfiles potenciales de nucleosomas y características funcionales de secuencias con cierto significado en cuanto a estructura y función, Levitsky desarrolló un método de algoritmos genéticos utilizando estadística bayesiana para la localización de dinucleótidos dentro de regiones con sitios de formación de nucleosomas. Estos métodos estiman la frecuencia local de dinucleótidos y detecta estructuras en bloque de sitios de formación de nucleosomas mediante una partición local en un contexto específico de dinucleótidos.

El programa desarrollado analiza diversas clases de secuencias genómicas de DNA, considerando como clases a los diferentes promotores y sus distintos patrones de expresión: exones, intrones, sitios de *splicing*, *loci* de regiones control, secuencias no flanqueadas, DNA satélite, y elementos móviles de inserción. Demostró que el potencial de nucleosomas en regiones promotoras de genes *housekeeping* y en genes expresados en muchos tejidos son considerablemente más bajos que los de regiones promotoras en genes expresados en tejidos específicos. La interfase de este programa se puede encontrar en internet [22]. El conjunto de sitios nucleosómicos usados para el análisis, está recopilado en la base de datos SiteSequences, extraída del banco de datos del EMBL de acuerdo con los códigos y posiciones indicados en

la base de datos NUCLEOSOMAL DNA hecha por Ioshikhes y Trinov, 1993[31].

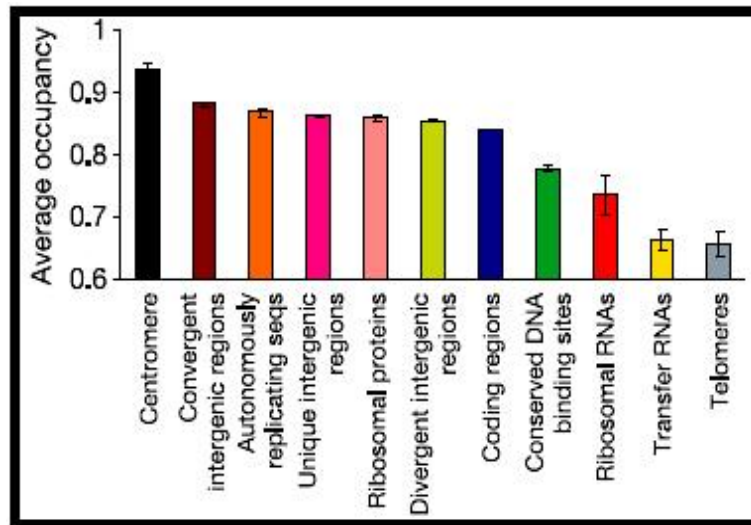


Figura 3.1: Porcentaje de ocupación de nucleosomas en el DNA (Segal *et al.*, 2006)

En mapeos experimentales de nucleosomas en el genoma de levadura, se ha encontrado que existe una diferencia en el porcentaje de nucleosomas para distintas regiones del genoma [38] (ver fig 3.1) y que la posición de nucleosomas es diferente para promotores que son regulados positivamente en contraste con los que son regulados negativamente por nucleosomas. También los patrones son distintos en las regiones intergénicas. Los genes regulados positivamente por nucleosomas tienden a no tener cajas TATA mientras que los genes inhibidos por nucleosomas tienden a tener cajas TATA, es decir, los genes que tienden a tener cajas TATA se encuentran relativamente compactos con arreglos de secuencias asociadas a nucleosomas sobre esta región promotora (ver fig. 3.2). Sugiriendo que estos dos modos

opuestos de regulación podrían venir de diferentes arquitecturas de la cromatina[19].

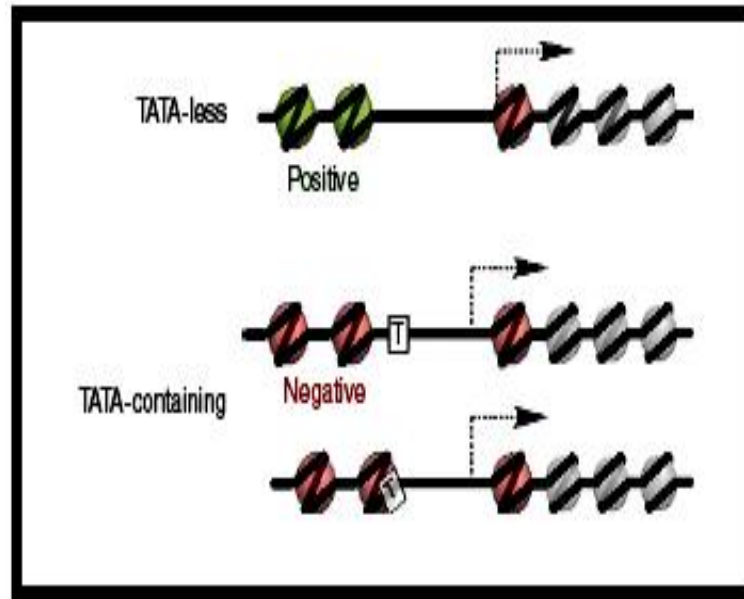


Figura 3.2: Arreglo de nucleosomas en dos clases de promotores (Ioshikhes *et al.*, 2006)

El trabajo experimental de Yuan [49], en donde obtienen todas las secuencias de DNA unidas a nucleosomas en el genoma completo de *Sacharomyces cerevisiae*, fue una gran oportunidad para desarrollar trabajos teóricos que predicen la posición de nucleosomas y compararlos con los datos de este trabajo, ya que una de las grandes limitantes en muchos trabajos se debe a los pocos datos experimentales que se tienen. Apoyados en este primer trabajo se estudió un grupo de regiones promotoras bien caracterizadas e identificadas con una fuerte periodicidad dentro de 199 alineaciones centrales de nucleosomas, concluyendo que existe una organización intrínseca en la posición de secuencias

de DNA con nucleosomas al menos en el 50% del genoma *in vivo* en experimentos observados en levaduras y se ha propuesto que los genomas eucariontes tienen un código para la posición de nucleosomas que está ligado a las funciones específicas del cromosoma. En este mismo trabajo se encontró que aproximadamente cada 10 pb. existe una periodicidad de dinucleótidos AA/TT/AT que oscilan en fase entre ellos y fuera de fase con dinucleótidos periódicos de GC que a su vez se encuentran aproximadamente cada 10 pb (ver fig. 3.3) [38]. Aunque en otros trabajos, también realizados en levaduras, se ha encontrado que esto resulta ser cierto sólo para un subconjunto de nucleosomas, en donde si bien se observan señales intrínsecas de la posición de nucleosomas, las señales son más inhibitorias como más variables en regiones promotoras que en regiones de marcos abiertos de lectura (open reading frames), siendo estos resultados inconsistentes con la propuesta de un código de posición de nucleosomas propuesto en el trabajo de Segal *et al.*[16].

Otra característica encontrada en secuencias de espacios intergénicos, intrones y exones en genes humanos, que son potenciales para la formación de nucleosomas, es que mantienen una correlación negativa con el contenido de GC independientemente de la longitud de la secuencias, esto probablemente es una explicación para el significado funcional de regiones con diferentes contenidos de GC [45]. Se ha encontrado también que los dinucleótidos CA son también importantes para la unión a nucleosomas y que el decámero TATAAACGCC muestra una gran afinidad a los mismos [47] [16], así como repeticiones de TGGA impiden la formación de nucleosomas y repeticiones múltiples de dinucleótidos A-T incrementan la accesibilidad para la unión de factores de transcripción en secuencias cercanas [6]. Se sabe



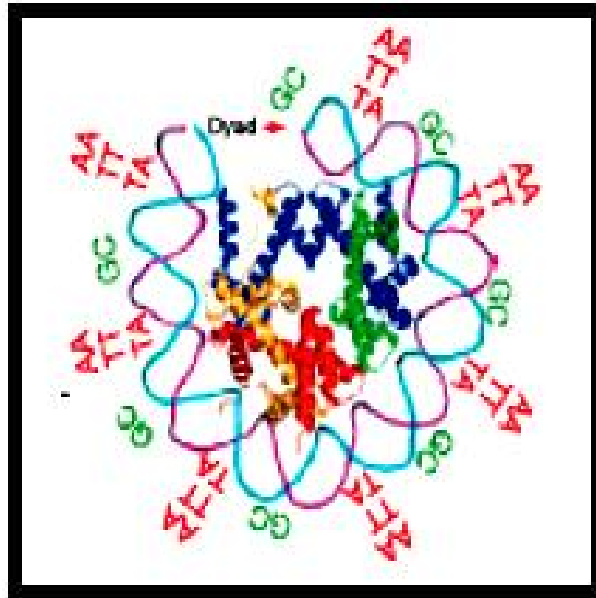


Figura 3.3: Secuencias de DNA unidas a nucleosomas (Segal *et al.*, 2006)

que pequeñas regiones del DNA ricas en secuencias de A-T y espaciadas por un número integral de nucleótidos, permiten que la molécula gire alrededor del nucleosoma (ver fig.3.4) [1]. Existe evidencia que en secuencias con repeticiones periódicas cada 10.4 pb de dinucleótidos AA y TT se encuentran nucleosomas asociados[5][16]. No obstante, parece ser que son un conjunto general de señales las reponsables de la composición del DNA y de la asociación con los nucleosomas[16].

### 3.2. Posición de nucleosomas en arqueas

Las arqueas mesófilas, termófilas e hipertermófilas miembros todos de las euriarqueas tienen homologías en las secuencias primarias, el plegamiento de histonas, las estructuras cuaternarias

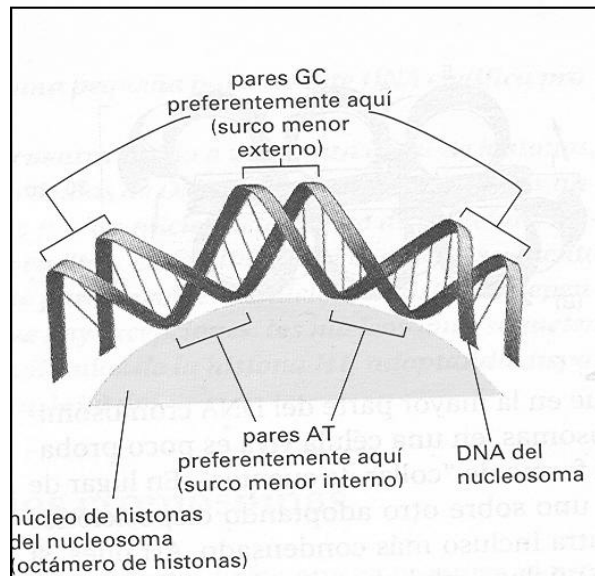


Figura 3.4: Surcos del DNA unidos al conjunto nucleosómico (Lewin, 1994)

y la formación de dímeros con el conjunto de histonas, H2A, H2B, H3 y H4, que forman los nucleosomas eucariontes [33]. Las histonas de arqueas tienen una unión similar y un enrollamiento en el DNA dentro de los complejos nucleosómicos los cuales comparten muchas características en común con los tetrámeros formados por histonas eucarióticas  $(H3+H2)_2$ , ambos tetrámeros también tienen dependencias similares con las secuencias de DNA a las que se unen preferentemente en localizaciones específicas tanto *in vivo* como *in vitro* [20].

En los trabajos desarrollados para caracterizar las regiones del DNA unidas a nucleosomas se encontró que las secuencias de DNA cercanas a las regiones nucleosómicas en arqueas tienen cuatro copias repetidas de secuencias TTTAAAGCCG que dan como resultado que la molécula de DNA incrementa su afini-

dad por las histonas de arqueas HMfB relativa a la dominancia cuantitativa de repeticiones helicoidales de los dinucleótidos TA que incrementan esta afinidad antes establecida. La alternancia de regiones ricas en A/T y G/C son localizadas en donde se encuentran los surcos menor y mayor respectivamente. Cuando las moléculas se incorporan en los nucleosomas, la alternancia de regiones ricas en A/T y en G/C se coloca en donde los surcos menores y el surco mayor sean comprimidos respectivamente y sean ubicados de cara al complejo de histonas. La conformación de estas copias de secuencias repetidas (A/T)<sub>3</sub> NN (G/C)<sub>3</sub> NN le proporciona flexibilidad a la molécula del DNA(ver fig.3.5). No obstante, se han encontrado diferencias de afinidades a las moléculas de DNA entre las dos histonas HMFA y la HMfB, las cuales se explican por distintos residuos en el carboxilo terminal de la región  $\alpha$ -hélice 3 de la histona. Los autores del trabajo antes mencionado sugieren que estas diferencias en la preferencia de secuencias puede anticipar un nuevo mecanismo estructural de la regulación de la expresión génica en arqueas[21].

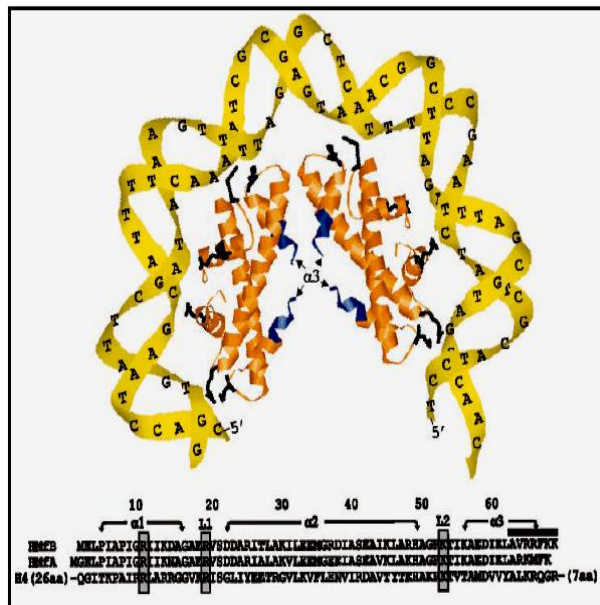


Figura 3.5: Secuencias conservadas de DNA asociado a nucleosomas en Arqueas (Bailey *et al*, 2002)

# Capítulo 4

## Metodología

Una función  $f(x)$  periódica es aquella donde existe un valor  $T > 0$ , tal que para cualquier  $x$  en el dominio de dicha función se tiene que  $f(x+T) = f(x)$  (ver fig.4.1). Las funciones periódicas más conocidas son  $sen(x)$ ,  $cos(x)$ ,  $tan(x)$ . etcétera. La suma, diferencia, producto o cociente de dos funciones de periodo  $T$ , es otra función de periodo  $T$  [44].

Muchas funciones están compuestas por varias señales, es decir por la suma de dos o más funciones periódicas (fig. 4.2), los genomas fungen como una función de este tipo, en la que se registran varias señales, y para analizar la composición de cualquier función se puede utilizar el análisis de Fourier

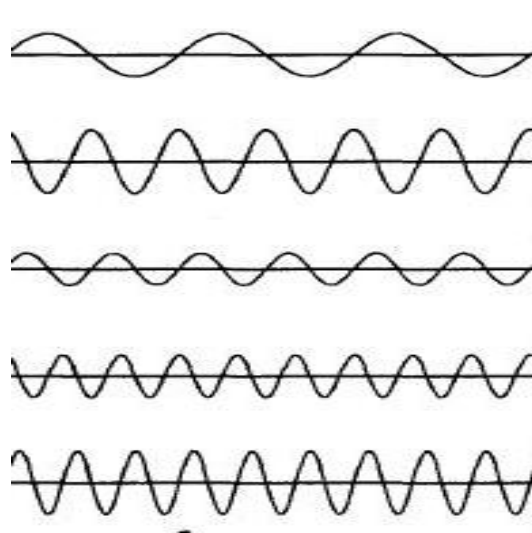


Figura 4.1: Períodos

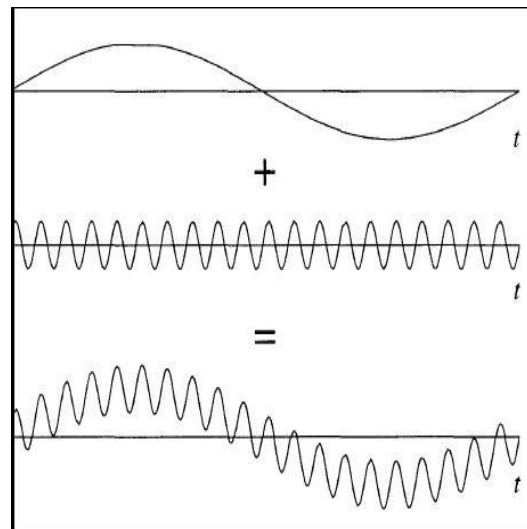


Figura 4.2: Suma de dos señales periódicas

El análisis de Fourier se basa en la descomposición de una función dada, en varias funciones en términos de cualquier fun-

ción periódica conocida ( $\text{sen}x$ ,  $\text{cos}x$ , etcétera) y de diferentes frecuencias que al recombinarse nos den la función original. Con este análisis se puede describir cualquier función periódica sin importar lo complicada que sea (fig.4.4).

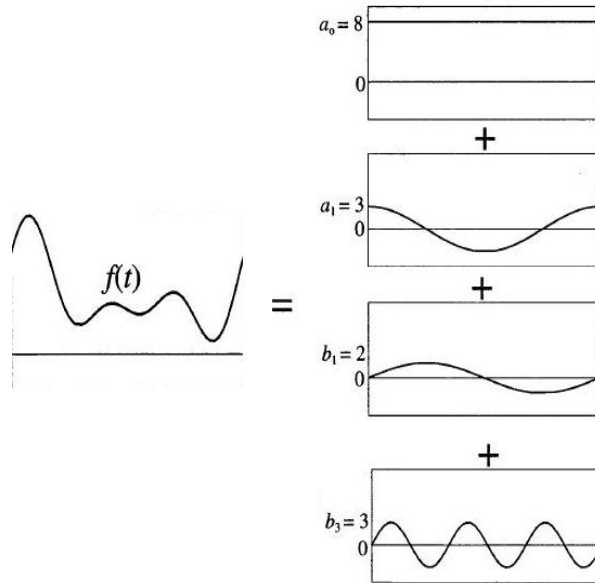


Figura 4.3: descomposición de una función como la suma de cuatro funciones periódicas

La transformada rápida de Fourier es un algoritmo que nos permite hacer el análisis de Fourier para representar así nuestra señal en el dominio de las frecuencias o el proceso que transforma una función a otra [41].

Los datos que estamos analizando con la información de nucleótidos representan un dominio finito, pero lo suficientemente grande para encontrar los períodos que buscamos, períodos en donde estén presentes señales de nucleosomas. La energía de interacción de dinucleótidos fue tomada de los experimentos realizados por Breslauer y Marky, en donde todos los valores se obtuvieron de la disrupción de la interacción existente entre las bases nucleotídicas a 1 M Na, 25C y con un pH7. Las unidades de entalpía están dadas en Kcal/mol.

Interacción NN	$\Delta H^\circ$	$\Delta S^\circ$
AA/TT	9 . 1	24.0
AT/TA	8. 6	23.9
TA/AT	6. 0	16.6
CA/GT	5. 8	12.9
GT/CA	6. 5	17.3
CT/GA	7. 8	20.8
GA/CT	5. 6	13.5
CG/GC	11. 9	27.8
GC/CG	11. 1	26.7
GG/CC	11. 0	26.6

De la base de datos del NCBI se bajaron algunos genomas representativos de eucariontes, arqueas y como control el genoma de *E. coli*. Los organismos utilizados fueron Homo sapiens, Pan troglodytes, Arabidopsis thaliana, Gallus gallus, Sacharomyces cerevisiae, Methanococcus jannashii, Thermoplasma volcanium



y *Escherichia coli*. De los cuales se obtuvieron pedazos de secuencias aproximadamente del mismo tamaño (2000pb).

Estas secuencias fueron transformadas mediante un programa en *Perl* a señales de energía libre correspondiente a los enlaces entre nucleótidos contiguos, utilizando la tabla de Breslauer. De la misma forma se transformó la secuencia para registrar señales de nucleótidos formadores de enlaces fuertes (GC) por un lado, como de enlaces débiles (AT) por el otro y se tomó un registro de las dos señales al mismo tiempo, obteniendo cuatro fuentes de datos con estas señales.

Para el registro de las señales de enlaces Fuertes se le asignaron a los nucleótidos G y C el valor 1 y al resto 0 y para el registro de señales de enlaces débiles se le asignó a los nucleótidos A y T el valor de 1 y al resto 0. Para el registro de ambos tipos de enlaces se asignó valor de 1 a los enlaces fuertes y -1 a los enlaces débiles.

Así se obtuvieron cuatro series de datos para cada genoma, tres correspondientes a los nucleótidos formadores de enlaces fuertes y débiles, y uno a la energía de entropía entre dinucleótidos contiguos.

Para cada serie de datos, se realizó un análisis de periodicidad utilizando como herramienta la transformada rápida de Fourier y se compararon los resultados.

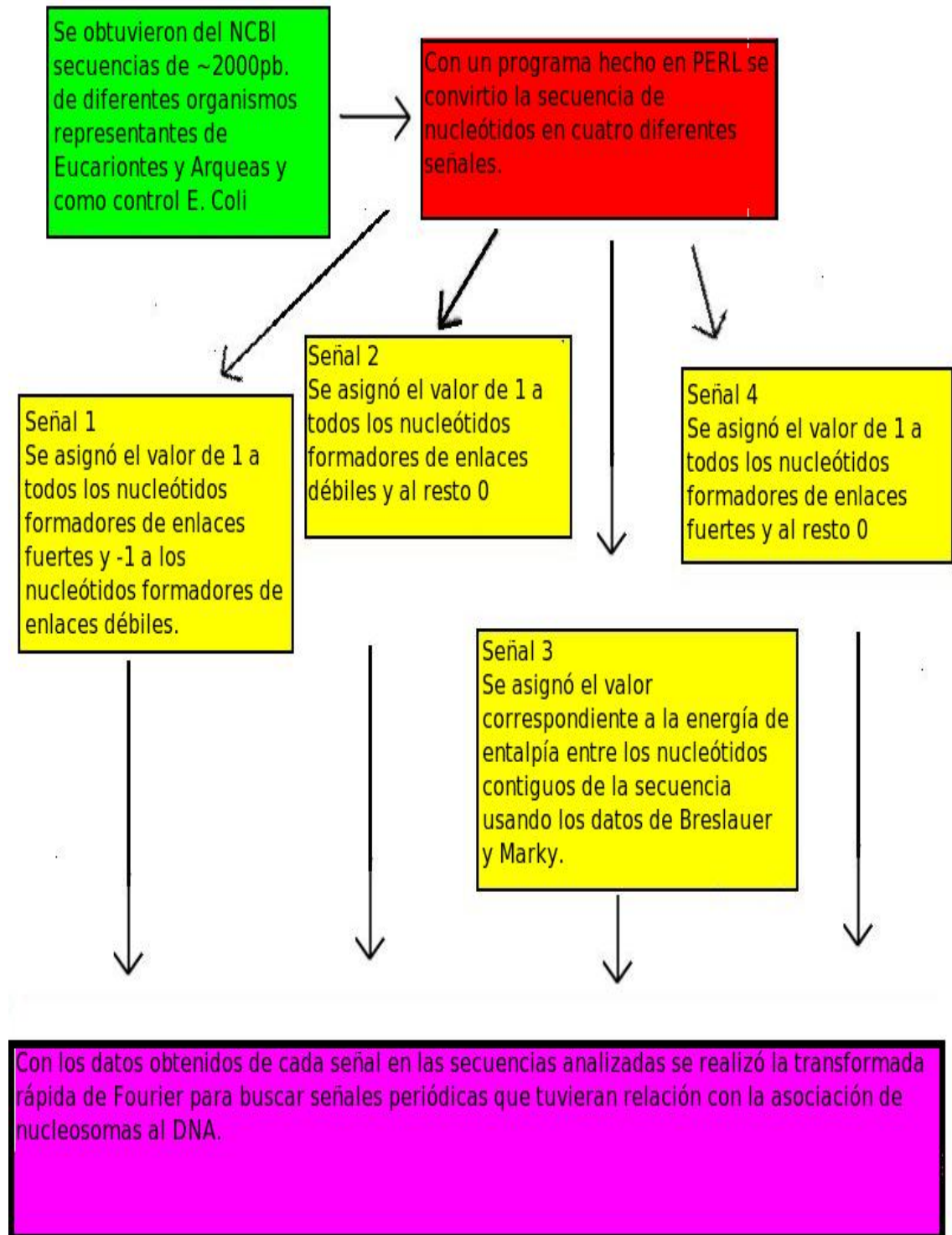


Figura 4.4: Diagrama de la Metodología

# Capítulo 5

## Resultados

En esta sección se muestran algunas de las gráficas obtenidas a partir de las secuencias de cada uno de los organismos, analizadas con la transformada rápida de Fourier, registrando señales de enlaces fuertes y débiles juntas y por separado, así como las de energía libre y entalpía entre dinucleótidos contiguos obtenidas con las medidas de Breslauer y Marky, en cada uno de los organismos.

Se tomaron secuencias con tamaños de 2000pb aproximadamente de cada muestra, contemplando que los períodos buscados para formar nucleosomas pudieran estar representados en las series de nucleótidos obtenidas.

Decimos que una señal se repite si después de cierta cantidad de pares de bases vuelve a aparecer el mismo patrón. La cantidad de pares de bases entre cada repetición le llamamos *el periodo*. En una misma secuencia pueden convivir varios periodos si hay segmentos que vuelven a aparecer con distintas cantidades de pares de bases entre repeticiones. Los picos obtenidos en cada gráfica [del espectro de Fourier] indican los periodos presentes en esa señal. Es decir, si hay un pico correspondiente a un periodo 80, quiere decir que hay algún segmento de la señal que se repi-

te cada 80 pares de bases. Es necesario indicar que cuando nos referimos a una repetición de un segmento de DNA, dicha repetición bien puede ser aproximada. Es decir, que lo que vuelve a aparecer es una región casi idéntica, pero no igual a la anterior. esto es importante porque si la señal que se trabaja es la de las energías libres en lugar de las bases, entonces habrá segmentos con energía de enlace casi idéntica pero cuya sucesión de bases sea diferente.

En *E.Coli* se observa claramente un pico de periodo 86 (ver apéndice);juntas aunque en otros trabajos no se han encontrado nucleosomas en bacterias, se ha reportado la presencia de proteínas *histone-like* que se unen al DNA, regulan la transcripción de genes y contribuyen a la organización del nucleoide [9], sin embargo no se encontraron reportes de la presencia periódica de la unión de estas proteínas al genoma bacteriano.

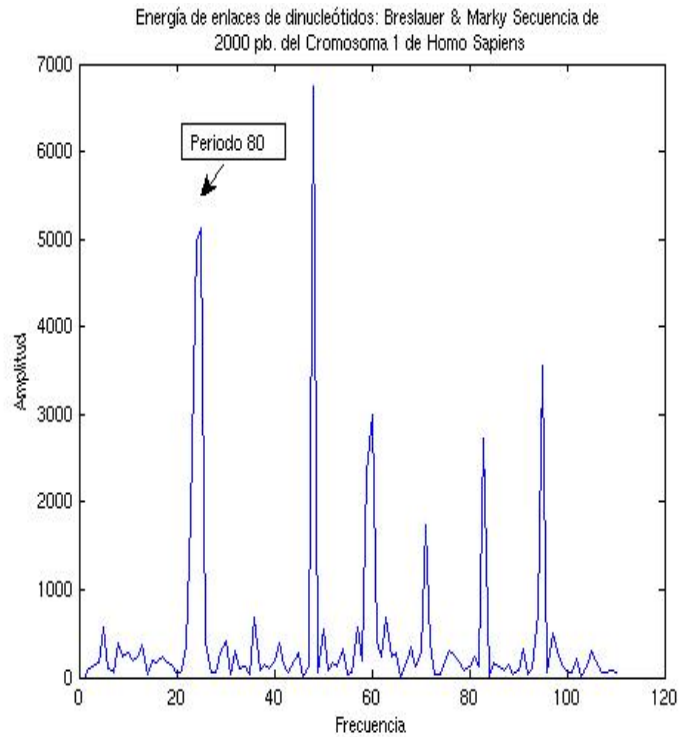


Figura 5.1: cromosoma 21 de *Homo sapiens*

Los datos que se obtuvieron de *Homo sapiens*, muestran gráficas con picos de periodo 80 con mejor claridad que el resto de las muestras. En todas las gráficas de los organismos analizados se muestran períodos que presumiblemente corresponden a espacios internucleósomáticos (cerca o menores de 80). También se observan picos que corresponden a múltiplos de los mismos períodos y aunque no se muestran en las gráficas, indican que la periodicidad en estos múltiplos se repite en la muestra lo que es congruente con el registro buscado.

Las gráficas que se obtuvieron de un mismo organismo, varían para cada registro pero se mantienen muy cerca o con el mismo periodo, en los registros para *Homo sapiens* las cuatro señales indican el mismo periodo; como puede observarse al comparar las gráficas mostradas para señales de enlaces de entalpía y señales de enlaces débiles en esta sección.

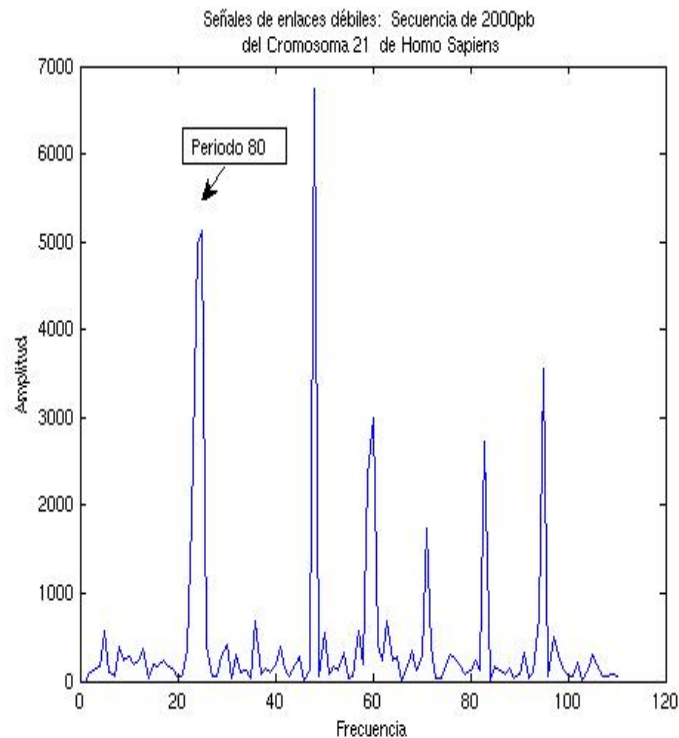


Figura 5.2: cromosoma 21 de *Homo sapiens*

Lo mismo sucede para los registros de *Pan troglodytes*, pero el periodo varía ligeramente con respecto del humano (ver apéndice) que indica que las señales utilizadas, en estos dos organismos son igualmente sensibles a estos períodos.

En el resto de las gráficas se observan períodos cercanos a 80 o de 80, para un mismo organismo, como por ejemplo *Arabidopsis thaliana*. Dependiendo del tipo de señal analizada (enlaces

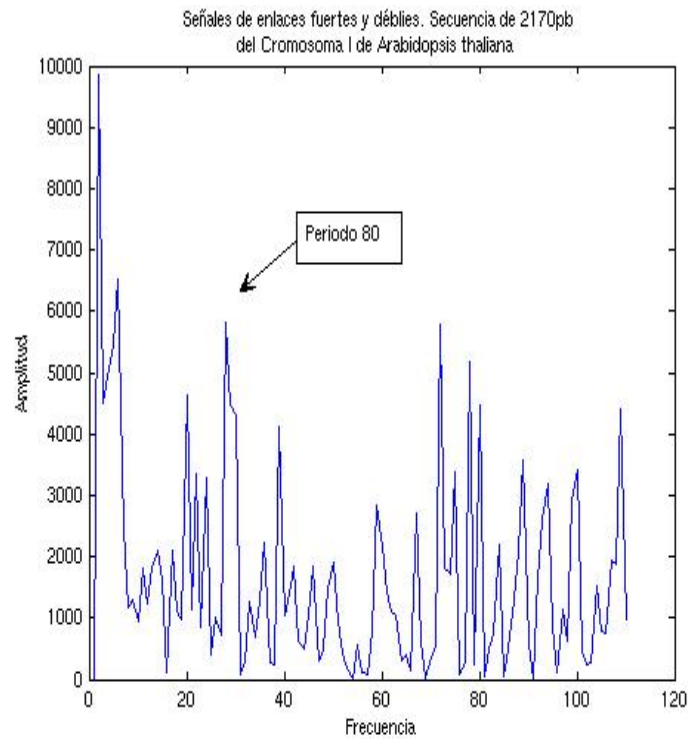


Figura 5.3: cromosoma I de *Arabidopsis thaliana*

fuertes, débiles, ambos o entalpía entre nucleótidos contiguos) la señal se puede matizar de diferente forma con otros períodos y variar muy poco como se muestra en las gráficas que se encuentran en el apéndice. En algunas de las gráficas se puede ver que los enlaces débiles o fuertes son los que tiene mejor claridad, sin embargo en ninguno se pierde la señal, lo que presumiblemente corresponde al DNA espaciador.

En todas las muestras de organismos eucariontes encontramos patrones similares en dichos períodos pero también los encontramos en los dos organismos pertenecientes a las arqueas. Aunque los registros para *Thermoplasma volcanium* son más cercanos al resto de los eucariontes que *Methanocaldococcus jannaschii* (ver apéndice) tampoco excede el rango del periodo buscado. Lo

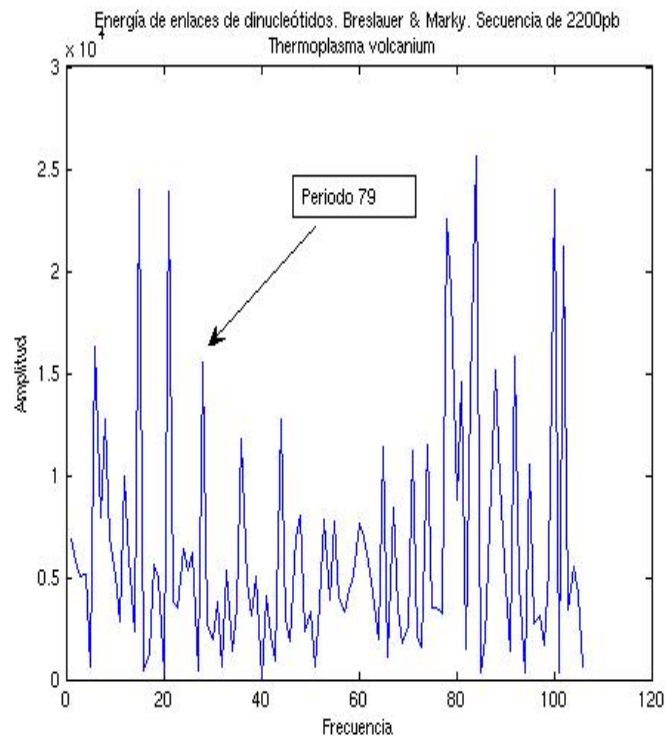


Figura 5.4: *Thermoplasma volcanium*

que esta indicando que para las mismas señales en eucariontes se registran patrones periódicos similares en la composición del genoma de arqueas en las muestras analizadas y que probablemente se debe al mismo patrón de asociación a proteínas. La amplitud de la señal en todas las gráficas para los períodos bus-



cados dependerá de la región genómica en la que se encuentren las muestras analizadas, debido a que el DNA es un acervo de diversas señales que también son periódicas.

# Capítulo 6

## Discusión y Conclusiones

La reciente obtención de genomas eucariontes, que en su mayoría no codifica para proteínas, ha dejado en claro la posibilidad de una organización tal que le permita tener regulaciones epigenéticas que le confieran con esto, la regulación transcripcional y como consecuencia, la especialización celular.

La presencia de nucleosomas en el DNA es fundamental en este sentido, ya que son la unidad básica de organización del genoma, lo cual hace muy interesantes los estudios de las relaciones entre DNA y nucleosomas.

El DNA se caracteriza por tener nucleosomas unidos cada 197 pb aproximadamente. Los nucleosomas como se ha mencionado anteriormente, son proteínas altamente conservadas tanto en la secuencia primaria como en el plegamiento y la afinidad de estas al DNA también es dependiente de la secuencia de nucleótidos. En muchos trabajos se han encontrado una serie de patrones, ligeramente variables en los nucleótidos que permiten hacer más flexible a la molécula para enrollarse alrededor de los complejos protéicos del DNA. También se tiene una base de datos de las secuencias de DNA asociado a nucleosomas obtenidas experimentalmente, en donde dichas secuencias asociadas

a nucleosomas se clasifican por la función que tiene el DNA que se asocia a estas proteínas. El programa de predicción de secuencias de DNA asociados a nucleosomas utiliza esta base de datos y mediante métodos heurísticos predicen la probabilidad de asociación de la secuencia al DNA. Sin embargo, en estos análisis la búsqueda está limitada a la comparación con los pocos datos de secuencias obtenidas experimentalmente en eucariontes.

El genoma está relegado a muchas restricciones tanto físicas como funcionales, dentro de las restricciones físicas se pueden asociar muchos comportamientos que podrían modelarse, como lo es la periodicidad. Con este tipo de análisis encontramos que las características fisicoquímicas de los nucleótidos producen señales que nos pueden dar información del periodo de un conjunto de nucleótidos que comparten las mismas cualidades. Con este estudio utilizamos las energías de entropía de nucleótidos contiguos medidas por Breslauer, así como la presencia y ausencia de nucleótidos generadores de enlaces fuertes y débiles cuya forma de distribución proporciona a la molécula las características que necesita para flexibilizarse, de manera que pueda enrollarse formando nucleosomas, demostrando ser señales muy sensibles para rastrear patrones periódicos en el DNA.

En las muestras analizadas, se logran ver períodos que varían en regiones cercanas a 80, lo que presumiblemente corresponde a espacios internucleosómicos o DNA espaciador, dándonos la idea de que existe un parámetro común para detectar un tipo de constitución de las secuencias que se unen a nucleosomas en cualquier organismo y sin las limitantes experimentales que se presentan en los trabajos antes citados.

El DNA no es estacionario, tiene muchos patrones periódicos, además de la presencia de nucleosomas, existen diversas señales

periódicas como los, *Alu repeats* uso de codones en exones, regiones de microsatélites, etcétera [11] por lo que en las muestras analizadas observamos otros picos y la amplitud de los períodos va a depender de la región que tomamos al azar, y que por el tamaño de las secuencias analizadas (2000pb) era seguro suponer que encontraríamos regiones asociadas a nucleosomas.

La presencia de histonas en arqueas y la similitud en los períodos encontrados, en las dos muestras analizadas, puede sugerir otro argumento respecto a la cercanía evolutiva con los eucariontes, además de darnos una idea de los principios de organización del genoma; aunque en la muestra de *E. coli* se encontró un pico de periodo muy claro de 86, no existen reportes de la presencia de nucleosomas en bacterias pero se sabe que existen unas proteínas similares llamada proteínas *histone-like*, en las que tampoco han reportado periodicidad, sólo se conoce que se unen al DNA, que permiten la organización del nucleoide y que se asocian a regiones específicas del DNA para inhibir o regular la transcripción génica[32].

La conservación de patrones observada en el DNA que se asocia a nucleosomas, es una cualidad interesante de los límites evolutivos a los que puede conllevar esta restricción estructural, y sobretodo si se observa que la organización del genoma es un factor imprescindible en la complejidad en los organismos dentro de todos los niveles de la escala evolutiva.

Las restricciones estructurales del DNA para asociarse a nucleosomas dejan abierta la pregunta sobre los límites que afectan la funcionalidad de estas secuencias. La relación que hay entre patrones de restricción estructural y funcionalidad de la secuencia no está clara, es decir, no se conoce a qué niveles de restricción pueda llegar esta relación y qué implicaciones tenga en

la funcionalidad de las secuencias, y sobretodo la consecuencia evolutiva que puede tener en los organismos.

Aunque en este trabajo no se puedan determinar los límites de restricción estructural y el efecto en la funcionalidad del DNA, sí permite mostrar que existe una manera más precisa y diferente para estudiar estas asociaciones que no responden a la sucesión exacta de nucleótidos sino a distribuciones energéticas en la molécula del DNA que le confieren las características necesarias para flexibilizarse y asociarse a proteínas nucleosómicas, y muestra que esta puede ser una herramienta para la búsqueda de patrones evolutivos en los genomas. Finalmente se propone en trabajos futuros encontrar con estas herramientas los niveles y límites de las restricciones estructurales que intervienen en la funcionalidad de la secuencia del DNA.

# Capítulo 7

## Apéndice

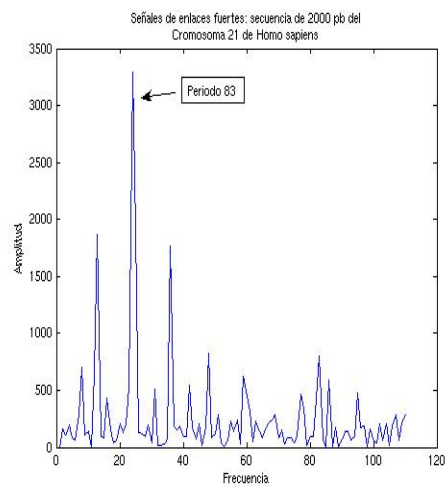


Figura 7.1: Gráfica que representa señales de enlaces fuertes del Cromosoma 21 de *Homo sapiens*, en la que se observa un pico muy bien definido que corresponde al periodo 83

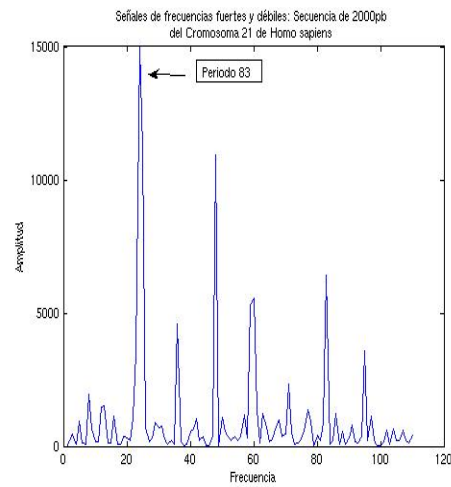


Figura 7.2: Gráfica del registro para señales fuertes y débiles del cromosoma 21 de *Homo sapiens*, en donde también se observa un pico de periodo 83

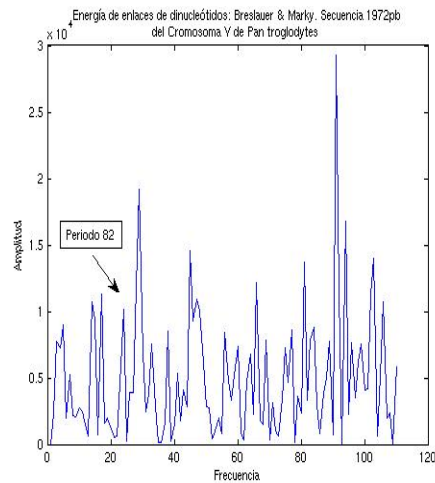


Figura 7.3: Cromosoma I de *Pan troglodytes* En esta gráfica se observa un periodo de 82, aunque la amplitud de los picos no es tan grande como en la anterior, esta muy definido.

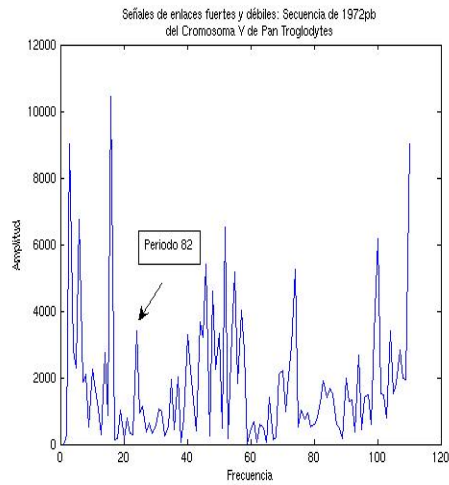


Figura 7.4: Esta gráfica corresponde a las señales de enlaces fuertes y débiles del Cromosoma I de *Pan troglodytes*, en donde también se encontró un pico de periodo de 82

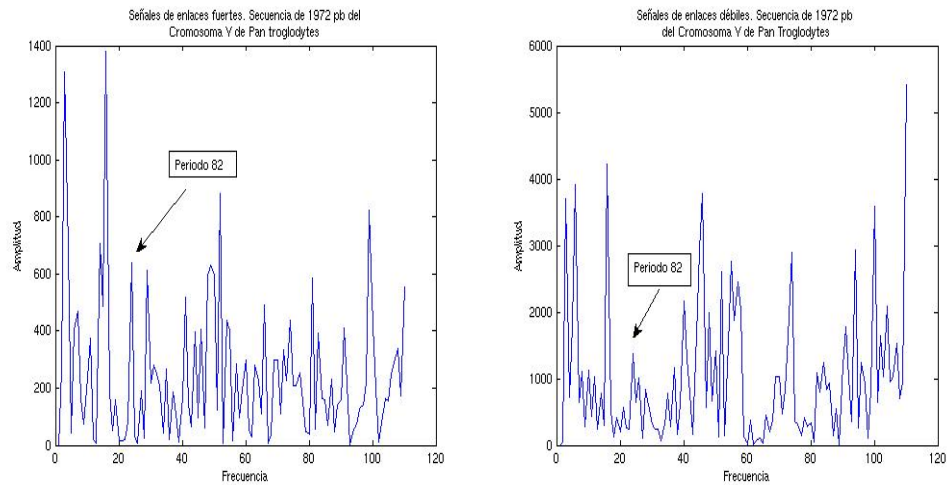


Figura 7.5: En estos registros de Cromosoma I de *Pan troglodytes* las señales parecen mostrar picos de menor amplitud, pero también están representados picos de períodos que presumiblemente corresponden al DNA espaciador



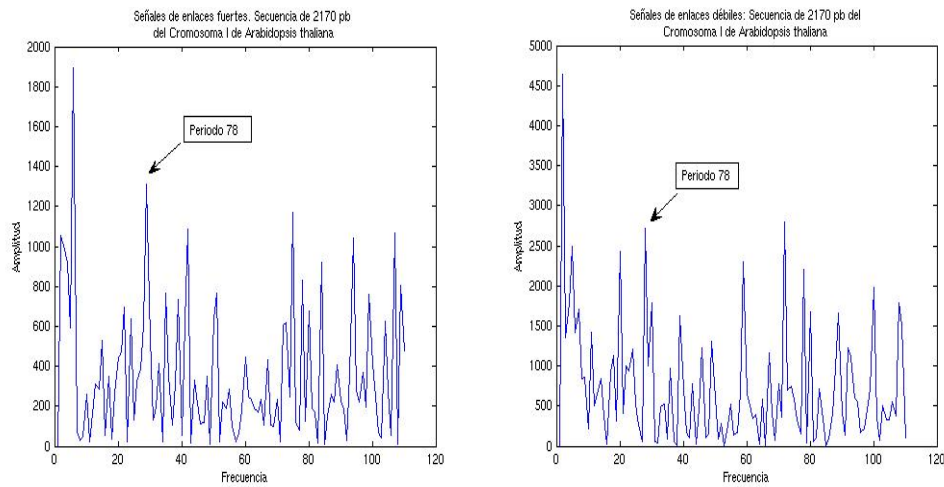


Figura 7.6: Estos son los registros de una secuencia del Cromosoma I *Arabidopsis thaliana*, que muestran un pico muy amplio y definido de periodo 78

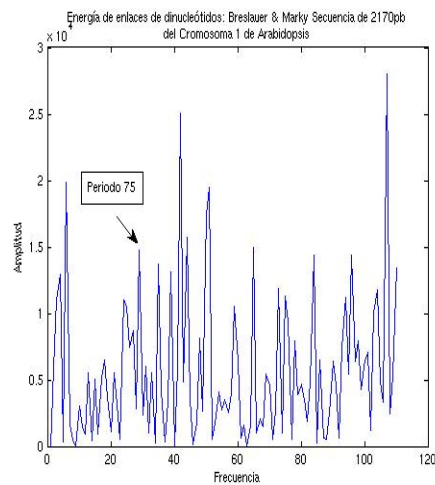


Figura 7.7: En esta gráfica se puede observar un pico bien definido de periodo 75 del Cromosoma I *Arabidopsis thaliana*

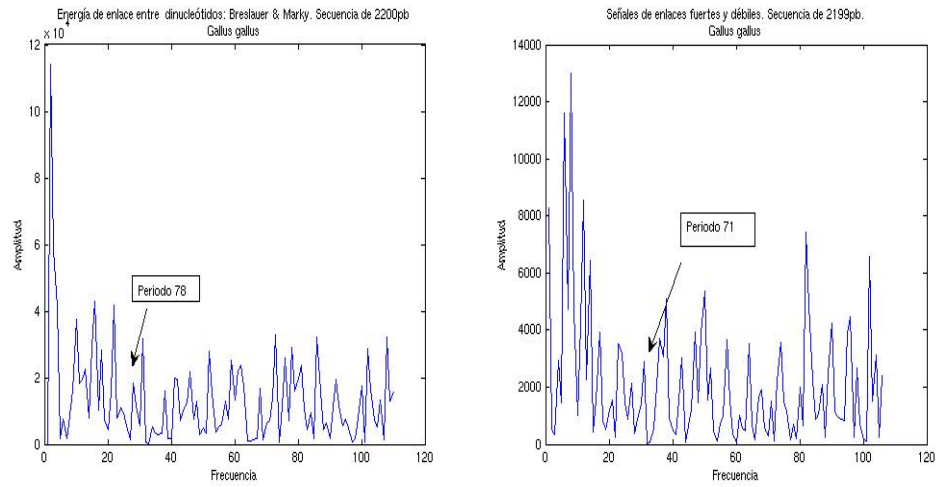


Figura 7.8: Estas gráficas son del Cromosoma I de *Gallus gallus*, con estas señales los picos son pequeños pero se observan muy bien períodos de 78

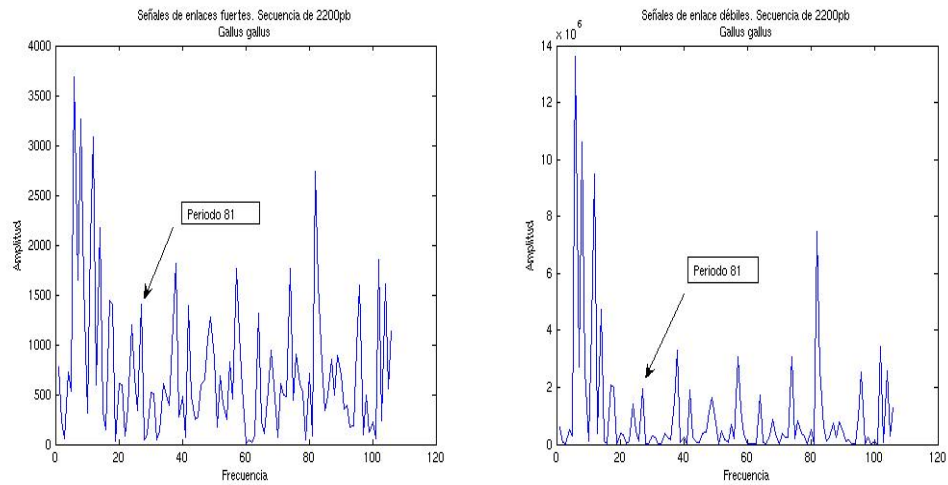


Figura 7.9: Cromosoma I de *Gallus gallus*. en la gráfica para enlaces fuertes la señal es de mayor amplitud, pero en ambas gráficas se registran también períodos de 81

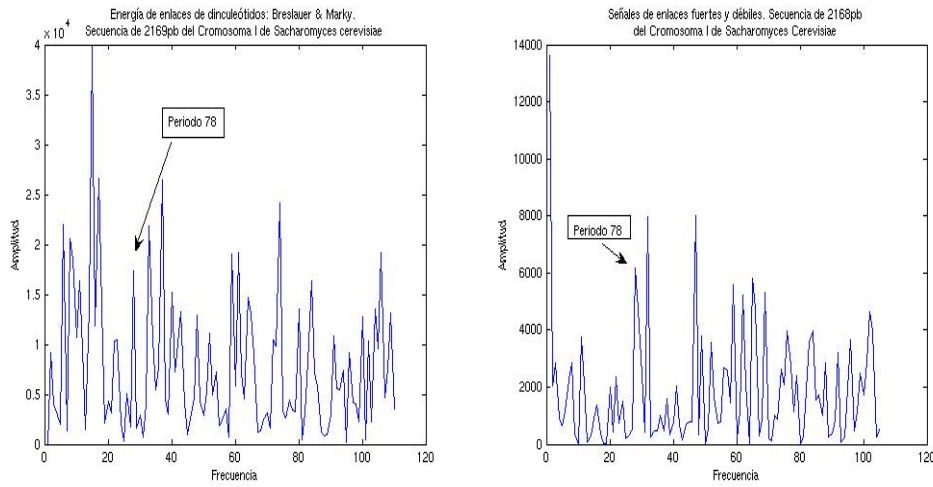


Figura 7.10: Cromosoma I de *Sacharomyces cerevisiae*, en ambos registros, tanto los de señales fuertes y débiles como los de energía de entalpía entre enlaces de nucleótidos se observan picos de periodo 78.

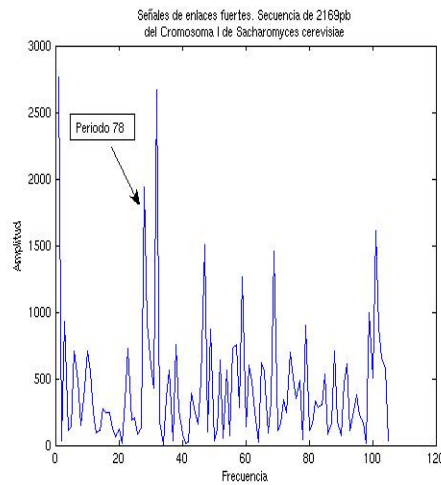


Figura 7.11: En esta gráfica de la secuencias obtenida Cromosoma I de *Sacharomyces cerevisiae* para señales fuertes se observa mucho mejor el pico que corresponde al periodo 78 como en los registros anteriores.

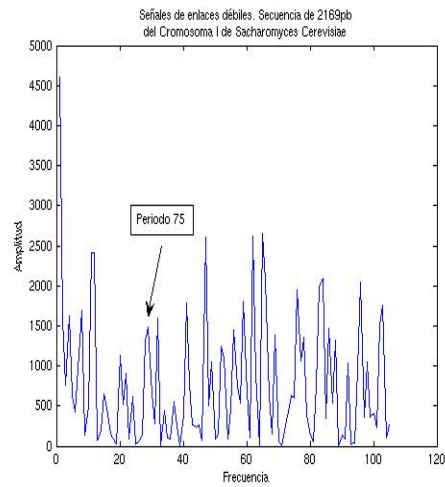


Figura 7.12: Esta es la gráfica que se obtuvo del Cromosoma I de *Saccharomyces cerevisiae* para señales débiles, aquí se logra ver un pico definido de periodo 75.

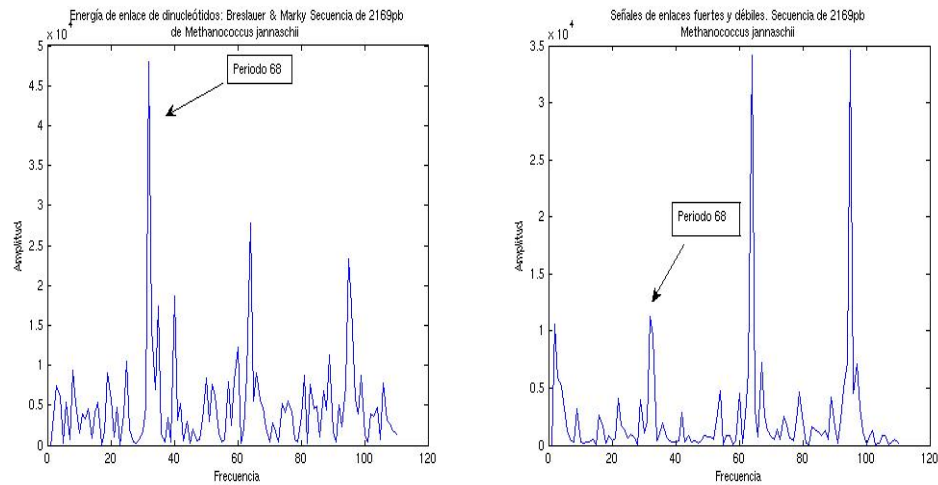


Figura 7.13: En estas gráficas para la secuencia de *Methanococcus jannaschii* se observan muy claros y bien definidos picos de periodo 68, aunque la señal parece ser de menor amplitud para enlaces fuertes y débiles.

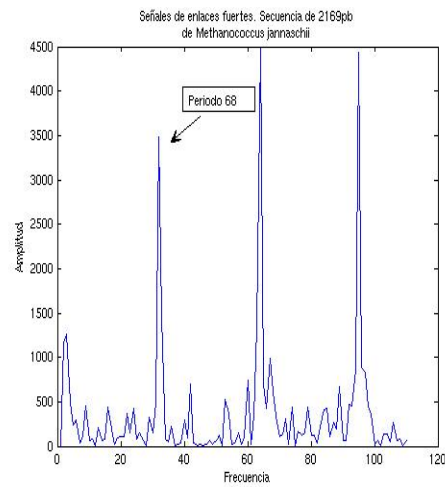


Figura 7.14: Este es el registro para señales fuertes de la secuencia de *Methanococcus jannaschii*, también se observa un pico muy grande de periodo 68

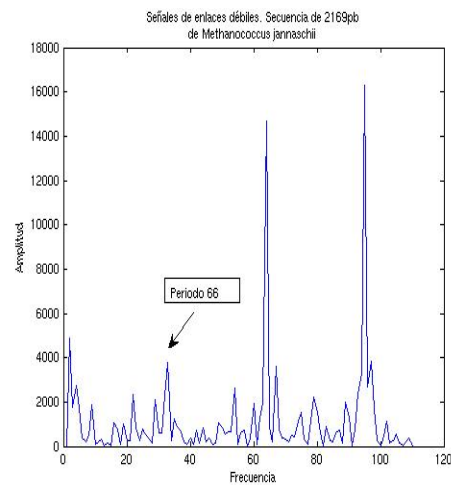


Figura 7.15: Con las señales de enlaces débiles de la secuencia de *Methanococcus jannaschii* se observa un pico definido y pequeño de periodo 66

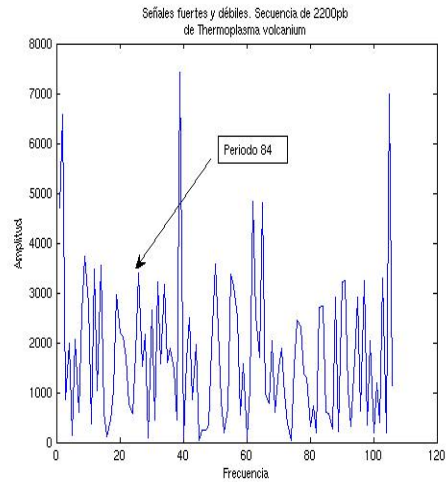


Figura 7.16: Esta es la gráfica de la secuencia de *Thermoplasma volcanium*, en donde se ven con estas señales un periodo muy claro de 84

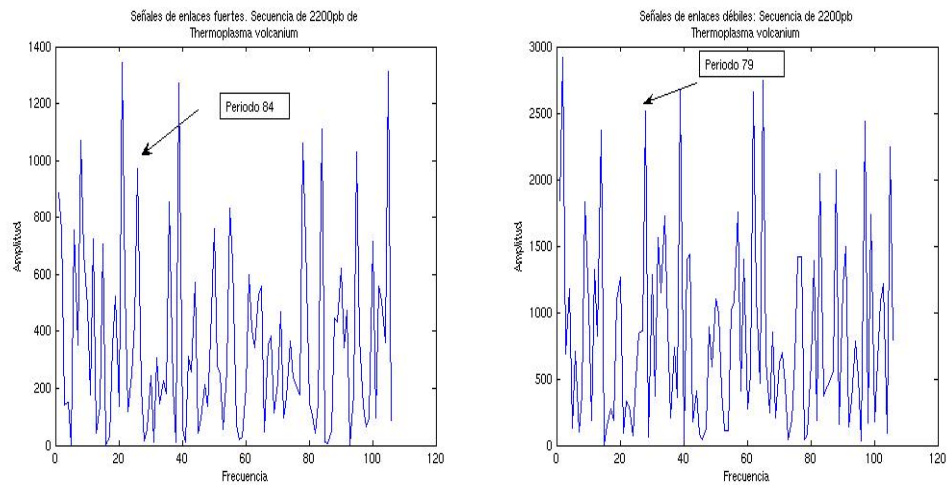


Figura 7.17: Estas son las gráficas de señales fuertes y débiles de la secuencia de *Thermoplasma volcanium*

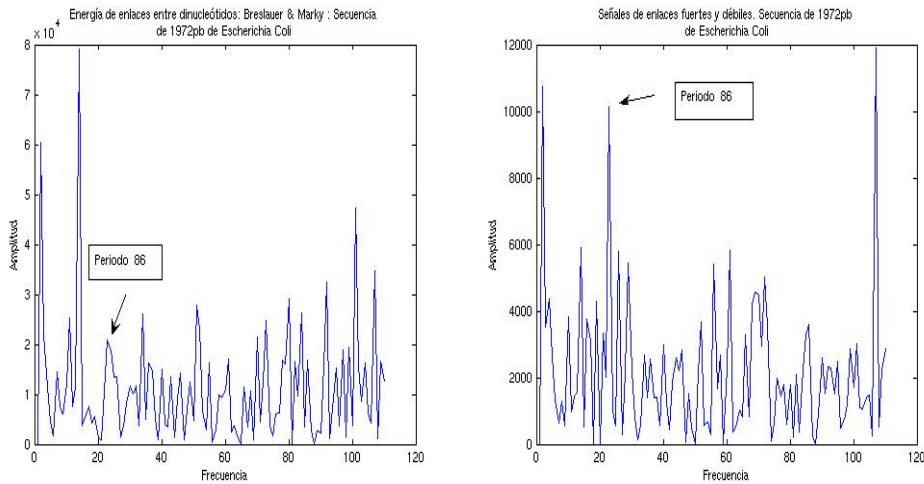


Figura 7.18: En estos registros para la secuencia de *Escherichia coli*, identificaron períodos de 86 aunque la señal se ve de mayor amplitud para los registros de enlaces fuertes y débiles

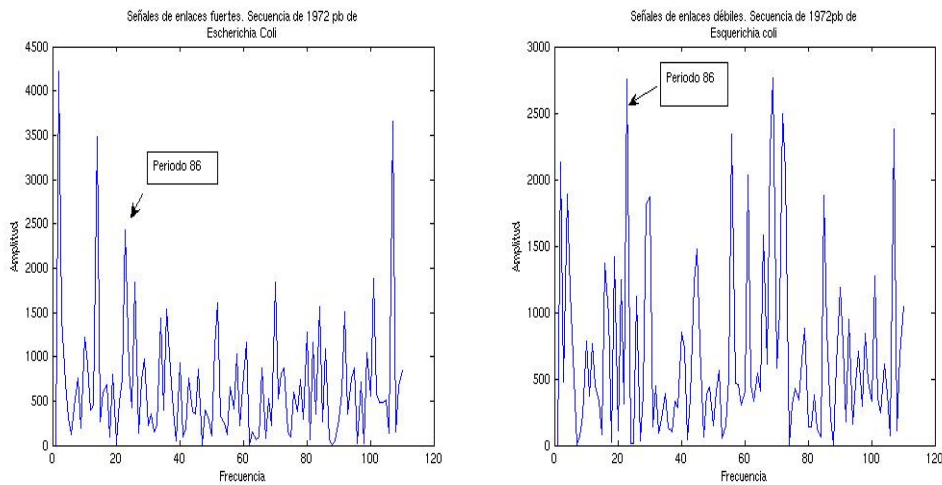


Figura 7.19: En estos registros para la secuencia de *Escherichia coli*, también se observan períodos de 86

# Bibliografía

- [1] Alberts *et al.*, 1996. Biología Molecular de la Célula, Ediciones Omega, S.A. Pags.366-371
- [2] Bertin A, Leforestier A, Durand D, Livolant F.,2004.Role of histone tails in the conformation and interactions of nucleosome core particles.Biochemistry. 27;43(16):4773-80.
- [3] Bonnie Burgess-Beusse, Catherine Farrell, Miklos Gaszner, Michael Litt, Vesco Mutskov, Felix Recillas-Targa, Melanie Simpson, Adam West, and Gary Felsenfeld,2002. The insulation of genes from external enhancers and silencing chromatin.PNAS vol. 99 Suppl. 4 16433-16437.
- [4] Breslauer K. J., Marky L.A.,1987. Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. Biopolymers. 26(9): 1601-1620.
- [5] Cohanin, A.B, kashi, Y. and Tifnov, E.N, 2005 Yeast nucleosome DNA pattern: deconvolution from genome sequences of *S. Cerevisiae*. J. Biomol. Struct. Dyn. 22: 687-694
- [6] Cao, H. Widlun, H.R. Simonsson,T. and Kubista,M. 1998 TGGA repeats impair nucleosome formation. J.Mol. Biol. 281:253-260



- [7] Chuprina VP, Lipanov AA, Fedoroff Oyu, Kim SG, Kihtar A, Reid BR, (1977) Sequence effects on local DNA topology, Proc Nat H Acad, Sci USA **8820** 9087-91
- [8] Cooper JL, Henikoff S., 2004 Adaptative evolution of the histon fold domain in centromeric histones. Mol Biol Evol Sep; 21(9):1712-8.
- [9] Dorman CJ, Deighan P.,2003 Regulation of gene expression by histone-like proteins in bacteria.Curr Opin Genet Dev. Apr,13(2):179-84
- [10] Gerda Egger, Gangning Liang, Ana Aparicio and Peter A. Jones,2004.Epigenetics in human disease and prospects for epigenetic therapy.Nature 429, 457-463
- [11] Stephen T Eskesen, Frank N Eskesen, Brian Kinghorn, and Anatoly Ruvinsky,2004.Periodicity of DNA in exons.BMC Mol Biol. 2004; 5: 12.
- [12] Fahrner RL, Cascio D, Lake JA, Slesarev A.,2001. An ancestral nuclear protein assembly: Cristal structure of the Methenopyrus kandleri histone. Protein Sci Oct 10 (10):2002-7
- [13] Felsenfeld, Groudine, 2003 Controlling the double helix Nature. 2003 Jan 23;421(6921):448-53.
- [14] Malik HS, Henikoff S.,2003 Phylogenimics of nucleosome. Nat Struct Biol. Nov;10 (11):882-91
- [15] Martic G, Karetsov Z, Kefala K, Politou AS, Clapier CR, Straub T, Papamarcaki T.,2005.Parathymosin affects the binding of linker histone H1 to nucleosomes and remodels chromatin structure. J Biol Chem.;280(16):16143-50

- [16] Heather E. Peackham, Robert E. Thurman, Yutao Fu, John A Stamatoyannopoulos, William Stafford Noble, Kevin Struhl, and Zhiping Weng Nucleosome positionin signals in genomic DNA. 2007. *Genome res.* 17: 1170-1177
- [17] R Holliday, The inheritance of epigenetic defects. *Science*, Vol 238, Issue 4824, 163-170
- [18] Horton, Moran, Ochs, Rowm, Scrimgeour, 1981, *Bioquímica*, Prentice-hall. Hispano América
- [19] Ioshikhes IP, Albert I, Zanton SJ, Pugh BF., 2006. Nucleosome positions predicted through comparative genomics. *Nat Genet.* Oct;38(10):1210-5.
- [20] Kathryn A. Bailey, Suzette L. Pereira, Jonathan Windom and John N. Reeve (2000) Arqueal Histone Selection of Nucleosome Positioning Sequences and the Procaryotic Origin of Histone-dependent Genome Evolution. *J.Mol.Biol.* **303**, 25-34.
- [21] Bailey KA, Marc F, Sandman K, Reeve JN. 2002. Both DNA and histone fold sequences contribute to archaeal nucleosome stability. *J Biol Chem.* 2002 Mar 15;277(11):9293-301. Epub 2001 Dec 21.
- [22] <http://www.mgs.bionet.nsc.ru/mgs/programs/recon/>
- [23] Lewin, R. (2005). *Genes VIII*. Prince Hall
- [24] Lowary PT, Widom J., 1997. Nucleosome packaging and nucleosome positioning of genomic DNA. *Proc Natl Acad Sci U S A.* Feb 18;94(4):1183-8.

- [25] Jean-Louis Mergny, Laurent Lacroix, 2003 Analysis of Thermal and Melting Curves. *Oligonucleotides*. December 1, 2003, 13(6): 515-537
- [26] Miroslav Tomshk, Haocheng Zheng, Ken van Holde, Jordanka Zlatanova, and Sanford H. Leuba, 2005. Fast, long-range, reversible conformational fluctuations in nucleosomes revealed by single-pair fluorescence resonance energy transfer. *PNAS* Vol.102 No.9 3278-3283.
- [27] Miramontes P, Medrano L, Cerpa C, Cedergen R, Ferbyre G, Cocho G, (1995) Structural and Thermodynamic properties of DNA uncover different evolutionary histories. *J Mol. Evol*; **40**(6): 698-704
- [28] Musgrave D, Sandman KM, Reeve JN. DNA binding by the archaeal histone HMf results in positive supercoiling. *Prec Natl Acad. Sci USA* Dec 1; 88(23): 10397-401
- [29] Musgrave D, Zhang X, Dinger M., 2002. Archaeal genome organization and stress responses: implications for the origin and evolution of cellular life. *Astrobiology*; 2(3): 241-53.
- [30] Musgrave D, Forterre P, Slesarev A., 2000. Negative constrained DNA supercoiling in archaeal nucleosomes. *Mol Microbiol*. Jan; 35(2): 341-9.
- [31] <http://www.mgs.bionet.nsc.ru/mgs/programs/recon/>
- [32] Taku Oshima, Shu Ishikawa, Ken Kurikawa, Hirofumi AIBA and Naotake Ogasawara, 2006. *Escherichia coli* Histone-Like Protein H-NS Preferentially Bind to Horizontally Acquired DNA in Association with RNA Polymerase. *DNA Research* 13 141-153.

- [33] Pereira SL, Reeve JN, (1998) Histones and nucleosomes in Archaea and Eukarya: a comparative analysis. *Extremophiles* Aug;**2**(3):141-8
- [34] Pereira SL, Grayling RA, Lurz R, Reeve JN.,1997.Archaeal nucleosomes.*Proc Natl Acad Sci U S A*.Nov 11;94(23):12633-7.
- [35] J.N Reeve, K.A.Bailey,W-t.Li,F. Marc,K.Sandman and D.J.Soares (2004) Archaeal histones:structures, stability and DNA binding. *Biochem Soc Trans* **32** (Pt 2):227-30.
- [36] D. Rhodes, A. Klug,1981.Sequence-dependent helical periodicity of DNA. *Nature* 292, 378 - 380
- [37] Sandman K, Pereira SL, Reeve JN.,1998,Diversity of prokaryotic chromosomal proteins and the origin of the nucleosome.*Cell Mol Life Sci.*; 54(12):1350-64
- [38] Segal E, Fondufe-Mittendorf Y, Chen L, Thstrm A, Field Y, Moore IK, Wang JP, Widom J.,2006.A genomic code for nucleosome positioning.*Nature*.Aug 17;442(7104):772-8.
- [39] Shukla A, Stanojevic N, Duan Z, Shadle T, Bhaumik SR.2006. Functional analysis of H2B-Lys-123 ubiquitination in regulation of H3-Lys-4 methylation and recruitment of RNA polymerase II at the coding sequences of several active genes in vivo *J Biol Chem*. Jul 14;281(28):19045-54.
- [40] M Shioda, K Sugimori, T Shiroya, and S Takayangi, 1989. Nucleosome like Structures Associated with Chromosomes of the Archaeobacterium *Halobacterium salinarium*. *J Bacteriol*. August; 171(8): 4514-4517

- [41] E.M. Stein, G Weiss, 1971. Introduction to Fourier Analysis on Euclidean Spaces. Princeton University Press.
- [42] Thierry Cheutin, Adrian J. McNairn, Thomas Jenuwein, David M. Gilbert, Prim B. Singh, Tom Misteli, 2003. Maintenance of Stable Heterochromatin Domains by Dynamic HP1 Binding. *Science* Vol. 299. no. 5607, pp. 721 - 725
- [43] Thierry Cheutin, Stanislaw A. Gorski, Karen M. May, Prim B. Singh, and Tom Misteli. 2004 In vivo Dynamics of Swi6 in Yeas: Evidence for Stochastic Model of Heterochromatin. *Molecular and Cellular Biology*, Vol. 24 No.8 3157-3167.
- [44] Tolstov Georgi P. Fourier Series. 1962. Dover. USA.
- [45] Vinogradov AE, 2004. Noncoding DNA, isochores and gene expression: nucleosome formation potential. *Nucleic Acids Research*. Vol.33 No.2 559-563
- [46] Vinogradov AE. 2004, Evolution of genome size: multilevel selection, mutation bias or dynamical chaos? *Curr Opin Genet Dev*. 2004 Dec;14(6):620-6
- [47] Widlund H, R. Kudvalli, P.N. Bengtsson, M., Cao, H., Tullius, T.D., and Ubitsa, M. 1999. Nucleosome structural features and intrinsic properties of the TATAAACGCC repeat sequences. *J. Biol. Chem.* 274: 31847-31852
- [48] M. Ya Azbel, DNA Sequencing and Melting Curve, 1979, PNAS, vol. 76 no.1 101-105

- [49] Yuan, GC. Liu Y.J., Dion, M.F, Slack, M.D. Wu L.F., Altschuler, S.J., and Tando, O.J. 2005. Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* 309:626-630.
- [50] Xie Y, Reeve JN.,2003 In vitro transcription assays using components from *Methanothermobacter thermoautotrophicus*. *Methods in Enzymology*, Volume 370,Pages 66-72