

UNIVERSIDAD DE BARCELONA

DEPARTAMENTO DE BIOQUÍMICA Y BIOLOGÍA MOLECULAR

**Estudio teórico sobre la influencia del solvente en la
estructura y dinámica del ADN**

MANUEL RUEDA BORREGO

2005

Índice

| | | |
|------------|--|-----------|
| I | Introducción..... | 1 |
| 1.1 | Estudio teórico sobre la influencia del solvente en la estructura y dinámica del ADN | 4 |
| 1.2 | Referencias bibliográficas | 5 |
| II | Métodos: Modelado Molecular de Ácidos Nucleicos..... | 7 |
| 2.1 | Métodos Macroscópicos..... | 10 |
| 2.2 | Métodos Mesoscópicos..... | 10 |
| 2.3 | Métodos Microscópicos..... | 11 |
| 2.3.1 | El campo de fuerzas o <i>forcefield</i> | 11 |
| | • Términos enlazantes..... | 13 |
| | • Términos no enlazantes..... | 15 |
| | • Eficacia de los campos de fuerzas para Ácidos Nucleicos..... | 18 |
| 2.3.2 | Métodos basados en campo de fuerzas..... | 19 |
| | • Mecánica Molecular..... | 19 |
| | • Dinámica Molecular..... | 21 |
| | • Condiciones de la Simulación..... | 24 |
| 2.4 | Recogida de datos y Análisis..... | 27 |
| | • Dinámica Esencial del ADN..... | 30 |
| 2.5 | Referencias bibliográficas..... | 38 |
| III | Estructura y Dinámica de los Ácidos Nucleicos..... | 47 |
| 3.1 | Estructura Primaria..... | 49 |
| 3.1.1 | Propiedades físico-químicas..... | 50 |
| 3.1.2 | Conformación..... | 51 |
| 3.2 | Estructura Secundaria..... | 56 |
| 3.2.1 | Fuerzas estabilizantes: Puentes de hidrógeno y apilamiento..... | 57 |
| 3.2.2 | Estructuras canónicas de los Ácidos Nucleicos..... | 60 |
| | • Formas Dextrógiras..... | 61 |
| | • Forma Levógira..... | 63 |
| 3.2.3 | Estructuras no canónicas de Ácidos Nucleicos..... | 65 |
| | • G-cuadrúplex..... | 65 |
| 3.3 | Efectos del entorno en la estructura y dinámica del ADN..... | 67 |
| 3.3.1 | El agua..... | 67 |
| 3.3.2 | Los iones..... | 68 |
| 3.3.3 | Interacción del ADN con ligandos..... | 70 |
| 3.4 | Métodos experimentales para el estudio de los Ácidos Nucleicos en fase gas..... | 74 |
| 3.5 | Los Ácidos Nucleicos en fase gas..... | 80 |
| 3.6 | Referencias bibliográficas..... | 83 |

| | | |
|------------|---|------------|
| IV | Resultados..... | 93 |
| 4.1 | Exploración de la atmósfera de iones alrededor del ADN..... | 95 |
| 4.1.1 | Exploring the Counterion Atmosphere around DNA. What can be learned from Molecular Dynamics Simulations..... | 97 |
| 4.2 | La estructura y dinámica del ADN en vacío..... | 111 |
| 4.2.1 | The Structure and Dynamics of DNA in the Gas Phase..... | 113 |
| 4.2.2 | Nature of minor-groove DNA interactions in the Gas Phase..... | 123 |
| 4.2.3 | G-quadruplexes can maintain their structure in the Gas Phase..... | 143 |
| 4.3 | Referencias bibliográficas..... | 184 |
| V | Discusión de resultados..... | 191 |
| 5.1 | Exploración de la atmósfera de iones alrededor del ADN..... | 193 |
| 5.2 | La estructura y dinámica del ADN en fase gas..... | 195 |
| 5.3 | Referencias bibliográficas..... | 203 |
| VI | Conclusiones..... | 207 |
| VII | Otras publicaciones..... | 211 |
| | Agradecimientos..... | 215 |

Capítulo I
INTRODUCCIÓN

La molécula de ADN, pese a su aparente simplicidad estructural comparada con las proteínas, es extremadamente rica en información biológica (Watson & Crick, 1953). Su estructura molecular la hace idónea para el almacenamiento y regeneración de la información y esto la convierte en la portadora ideal del código genético. En ella, se encuentra información tanto a nivel de apareamientos canónicos A·T y G·C, como en su estructura tridimensional global, especialmente en la configuración de los surcos, los cuales modulan la interacción con proteínas y otras moléculas que modifican su funcionalidad. Adicionalmente, la información contenida en los surcos puede ser aprovechada para modificar el comportamiento de los ácidos nucleicos mediante, por ejemplo, la unión de pequeños ligandos específicos (Blackburn & Gait, 1990; Bloomfield et al., 2000; Rawn, 1989; Saenger, 1984).

Aunque los primeros trabajos experimentales sobre la estructura de los ácidos nucleicos (basados en difracción de fibras) apuntaban ya hacia cierta diversidad conformacional, no fue hasta que se obtuvieron las primeras estructuras cristalográficas que se comenzó a contemplar la posibilidad de que pequeños cambios locales en la estructura pudieran tener implicaciones biológicas (MacKerell & Nilsson, 2000). En la actualidad, los estudios sobre estructura basados en Rayos-X o RMN (Resonancia Magnética Nuclear) son tremendamente útiles para explorar la diversidad estructural de los ácidos nucleicos. No obstante, presentan limitaciones para generar una visión dinámica global de los mismos. El estudio de los ácidos nucleicos mediante métodos teóricos ofrece la posibilidad de potenciar y extender la información estructural accesible, mediante la obtención de información dinámica a nivel atómico. Esto brinda la posibilidad de profundizar en el estudio de la sutil relación entre estructura, energética y dinámica, propiedades directamente relacionadas con la función biológica (Cheatham & Kollman, 2000; Cheatham & Young, 2001; Karplus & McCammon, 2002; Norberg & Nilsson, 2002; Orozco et al., 2003; Orozco et al., 2004).

Dentro de los métodos teóricos para el estudio de biomoléculas, la técnica de Dinámica Molecular con solvente explícito se ha convertido en una herramienta esencial debido a su capacidad para generar una descripción atomística de los

movimientos estructurales que ocurren en solución (Auffinger & Westhof, 1998; Cheatham, 2004; Orozco et al., 2004). Esta técnica ha hecho posible el estudio con éxito de numerosas estructuras basadas en ácidos nucleicos, tal y como muestran recientes revisiones sobre el campo (Auffinger & Westhof, 1998; Beveridge & McConnell, 2000; Cheatham, 2004; Cheatham & Kollman, 2000; Cheatham & Young, 2001; Giudice & Lavery, 2002; MacKerell & Nilsson, 2000; Norberg & Nilsson, 2002; Orozco et al., 2003; Orozco et al., 2004; Swarnalatha et al., 2002). La mejora de los campos de fuerza y de los algoritmos de simulación, así como la mayor disponibilidad de recursos computacionales, augura una gran extensión en el estudio por Dinámica Molecular de estructuras de ácidos nucleicos.

1.1 Objetivos

El objetivo fundamental de esta tesis ha sido profundizar en el conocimiento de la influencia del solvente en la estructura y dinámica del ADN mediante las técnicas computacionales disponibles en la actualidad. Se han realizado dos bloques diferenciados, cuyos objetivos concretos serían:

1- El estudio de una secuencia de ADN en solución con diferentes fuerzas iónicas con el objetivo de racionalizar cómo los contraiones modulan las propiedades estructurales y dinámicas de la molécula.

2- El estudio de diversos ácidos nucleicos canónicos y no canónicos en ausencia de solvente, bajo condiciones similares a las que tienen lugar en los experimentos de *espectrometría de masas por ionización con electrospray* (ESI-MS). Este bloque, a su vez, está dividido en tres trabajos:

i/-Estudio de la doble hélice de ADN en condiciones de vacío propias de los experimentos de ESI-MS.

ii/-Estudio de la naturaleza de complejos no covalentes entre *minor groove binders* (mG-binders) y ADN de doble cadena en condiciones de vacío.

iii/-Análisis del comportamiento de estructuras de ADN formadas por cuatro cadenas (G-cuadrúplex) en condiciones de vacío.

1.2 Referencias bibliográficas

Auffinger, P. & Westhof, E. (1998). Simulations of the molecular dynamics of nucleic acids. *Curr Opin Struct Biol* 8, 227-236.

Beveridge, D. L. & McConnell, K. J. (2000). Nucleic acids: theory and computer simulation, Y2K. *Curr Opin Struct Biol* 10, 182-196.

Blackburn, G. M. & Gait, M. J. (1990). *Nucleic Acids in Chemistry and Biology*, IRL Press., New York.

Bloomfield, V. A., Crothers, D. M. & Tinoco, I. (2000). *Nucleic Acids. Structures, Properties and Functions*, University Science Books, Sausalito.

Cheatham, T. E., 3rd. (2004). Simulation and modeling of nucleic acid structure, dynamics and interactions. *Curr Opin Struct Biol* 14, 360-367.

Cheatham, T. E., 3rd & Kollman, P. A. (2000). Molecular dynamics simulation of nucleic acids. *Annu Rev Phys Chem* 51, 435-471.

Cheatham, T. E. & Young, M. A. (2001). Molecular Dynamics Simulation of Nucleic Acids: Successes, Limitations and Promise. *Biopolymers* 56, 232-256.

Giudice, E. & Lavery, R. (2002). Simulations of nucleic acids and their complexes. *Acc Chem Res* 35, 350-357.

Karplus, M. & McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nat Struct Biol* 9, 646-652.

MacKerell, A. D., Jr. & Nilsson, L. (2000). *Nucleic Acid Simulations*. University of Nottingham.

- Norberg, J. & Nilsson, L. (2002). Molecular dynamics applied to nucleic acids. *Acc Chem Res* 35, 465-472.
- Orozco, M., Perez, A., Noy, A. & Luque, F. J. (2003). Theoretical methods for the simulation of nucleic acids. *Chem Soc Rev* 32, 350-364.
- Orozco, M., Rueda, M., Blas, J. R., Cubero, E., Luque, F. J. & Laughton, C. A. (2004). Molecular dynamics simulations of nucleic acids. *Encyclopedia of Computational Chemistry*.
- Rawson, J. D. (1989). *Bioquímica*, Interamericana-McGrawHill, Madrid.
- Saenger, W. (1984). *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
- Swarnalatha, R., Leclerc, F. & Karplus, M. (2002). DNA Polymorphism: A comparison of Force Fields for Nucleic Acids. *Biophys J* 84, 1421-1449.
- Watson, J. D. & Crick, F. H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 171, 737-738.

Capítulo II

MÉTODOS: MODELADO MOLECULAR DE ÁCIDOS NUCLEICOS

El término Modelado Molecular designa el proceso general de describir sistemas químicos complejos en términos de modelos físicos realistas con el propósito último de entender y predecir propiedades *macroscópicas* de los mismos (van der Spoel et al., 2004).

La modelización molecular de ácidos nucleicos presenta unas características únicas debidas fundamentalmente a la dificultad de representar simultáneamente tanto interacciones atómicas a nivel de par de bases, como fenómenos globales a escala cromosómica. Esta diversidad de tamaños condiciona la elección de una aproximación teórica adecuada para su estudio.

El primer paso en la simulación de un ácido nucleico es la elección de un funcional que conecte las coordenadas tridimensionales del sistema con su energía (Leach, 2001). Idealmente, se puede recurrir a las técnicas mecano-cuánticas e intentar resolver la ecuación de Schrödinger del sistema. Desafortunadamente, dichas técnicas requieren un coste computacional excesivo y por ello su utilidad está limitada a estudios de reactividad o a casos donde se requiere una descripción detallada de la estructura de algunas bases nucleotídicas (Cheatham & Kollman, 2000).

Los métodos clásicos son una alternativa a las metodologías cuánticas para abordar el estudio de sistemas químicos de gran tamaño en los que no se producen alteraciones en enlaces covalentes, ni cambios drásticos en la distribución electrónica. En ellos, el Hamiltoniano cuántico riguroso es reemplazado por uno parametrizado a partir de datos empíricos, lo que se conoce como *campo de fuerzas* (Frenkel & Smit, 2002; Leach, 2001; McCammon & Harvey, 1987).

En el campo de la modelización molecular de ácidos nucleicos se puede clasificar los métodos clásicos (Orozco et al., 2003) en:

2.1-macroscópicos (ideal-elástico)

2.2-mesoscópicos (intermedios)

2.3-microscópicos (atómicos)

2.1 Modelos macroscópicos

En estos modelos se asume que el ácido nucleico es una *cuerda* ideal flexible, cuyo movimiento está regido por una serie de ecuaciones diferenciales relacionadas con su comportamiento polimérico macroscópico. Estos métodos son de baja resolución, pero muy eficientes computacionalmente, lo que permite el estudio de propiedades supra-moleculares tales como el empaquetamiento cromosómico o la superhelicidad de las hebras de ADN. Normalmente, son utilizados para estudiar propiedades de fragmentos de gran longitud, donde las propiedades a extraer son mayoritariamente independientes de los detalles de la secuencia de estudio.

2.2 Modelos mesoscópicos

En el caso de los ácidos nucleicos, el uso de modelos intermedios entre la representación macroscópica y microscópica implica la segmentación de la molécula a fragmentos. Estos bloques son considerados como entidades rígidas conectadas entre sí por partes de mayor flexibilidad. Estos modelos permiten tratar fragmentos largos de ADN introduciendo de modo aproximado características dependientes de secuencia.

El tamaño de los bloques dependerá del tipo de problema a analizar y puede ser desde un par de bases hasta cientos de ellas. Incrementar el tamaño de los bloques introduce una mayor velocidad de cálculo, pero produce una reducción en la resolución del mismo. Para calcular la energía mesoscópica de deformación se suele recurrir a potenciales elásticos clásicos, cuya parametrización se puede realizar en base a observaciones de estructuras de alta resolución, propiedades macroscópicas o simulaciones de dinámica molecular.

2.3 Modelos microscópicos

Dentro de esta categoría existen dos grupos: los basados en *variables colectivas* y los *modelos atomísticos*. Los primeros se basan en el conocimiento previo de varios parámetros relacionados con la estructura de los ácidos nucleicos, los cuáles permiten reducir sus grados de libertad conformacional. Un ejemplo de ellos serían los modelos *helicoidales*, donde la estructura está representada por una serie de parámetros rotacionales y traslacionales referidos a nivel de par de bases o bases aisladas. Los modelos microscópicos atomísticos, no obstante, utilizan una descripción detallada de todos los grados de libertad del sistema, usualmente en el espacio cartesiano. Estos métodos han sido la columna vertebral de esta tesis, y por ello serán explicados en más detalle a continuación.

2.3.1 El campo de fuerzas

Un *campo de fuerzas* está compuesto por dos componentes distintos:

-*Conjunto de ecuaciones* (llamadas funciones de potencial) usadas para determinar energía potencial a partir de estructuras.

-Los *parámetros empíricos* usados para dichas ecuaciones.

Es posible utilizar diferentes parámetros con un mismo conjunto de ecuaciones, aunque es aconsejable comprobar la correspondencia entre los mismos. Existen diversos campos de fuerzas en la bibliografía (Orozco et al., 2004), siendo AMBER (Weiner et al., 1986) y CHARMM (Karplus, 1983) los más utilizados en el campo de los ácidos nucleicos.

Los diferentes campos de fuerzas hacen uso implícito de la aproximación de Born-Oppenheimer, la cual permite separar el movimiento nuclear del electrónico. Mediante ella se puede asumir que los núcleos se desplazan sujetos al potencial efectivo creado por la distribución electrónica. Esto permite trabajar exclusivamente con las

coordenadas nucleares, pudiendo así describir el efecto del potencial electrónico mediante una serie de ecuaciones simples y de parámetros empíricos (Leach, 2001). A pesar de que los funcionales posibles son muy diversos, la mayoría de los campos de fuerzas usados en el estudio de los ácidos nucleicos emplean términos muy similares para evaluar la energía del sistema. De ahora en adelante, nos centraremos en el *campo de fuerzas* de AMBER, pero los fundamentos del mismo son extrapolables a cualquier otro.

Dentro de un campo de fuerzas, la función $E_{sis}(r,p)$ que describe la energía potencial de un sistema depende de las coordenadas r del sistema y de los parámetros p del campo de fuerzas que se emplee. Dicha función se puede expresar de manera básica tal como muestra la ecuación (2.1),

$$E_{sis} = E_{\text{términos enlazantes}} + E_{\text{términos no enlazantes}} \quad (2.1)$$

donde el primer término hace referencia a la energía de interacción entre pares de átomos enlazados entre si y el segundo incluye la energía de interacción entre átomos no enlazados.

Los términos enlazantes incluyen contribuciones debidas a los enlaces covalentes, los ángulos de enlace y las torsiones propias e impropias. Los términos no enlazantes quedan definidos por un término de atracción-repulsión de tipo Lennard-Jones para las fuerzas de van der Waals y un término Coulómbico para las interacciones electrostáticas. Además de estos términos, la mayoría de los campos de fuerzas permiten la inclusión de otros parámetros especiales, como las restricciones de posición o términos asociados a la energía de polarización.

- **Términos enlazantes**

- i) Término de enlace covalente o *stretching*

Existen diferentes aproximaciones para representar el perfil de energía del enlace químico dependiendo del comportamiento armónico o enarmónico del mismo. Para casos donde el comportamiento del enlace sea del tipo enarmónico se suele recurrir a un potencial de Morse (Leach, 2001; Scott et al., 1999; van der Spoel et al., 2004; van Gunsteren et al., 1996) o aproximaciones polinomiales (Leach, 2001; van der Spoel et al., 2004; van Gunsteren et al., 1996). Normalmente, se estudian sistemas de comportamiento armónico donde las distancias de enlace están próximas al equilibrio y la distorsión de enlace se puede describir mediante el oscilador armónico de Hooke (ecuación 2.2).

$$E_{str} = \sum_{enl} K_{str} (l - l_{eq})^2 \quad (2.2)$$

donde l es la distancia de enlace, l_{eq} el valor de equilibrio y K_{str} la constante de fuerza.

- ii) Término de ángulo o *bending*

La vibración del ángulo entre tres átomos puede ser representado en su esquema más sencillo por un potencial armónico tipo Hooke (ecuación 2.3), aunque cuando la distorsión respecto al equilibrio es mayor se recurre al uso de funciones más complejas (Leach, 2001; Scott et al., 1999; van der Spoel et al., 2004; van Gunsteren et al., 1996).

$$E_{bnd} = \sum_{ang} K_{bnd} (\Theta - \Theta_{eq})^2 \quad (2.3)$$

donde K_{bnd} se refiere a la constante de enlace y Θ_{eq} es el ángulo de equilibrio.

iii) Términos de torsión

Los términos de torsión (E_{tor}), a diferencia de los anteriores, no tienen un comportamiento armónico basal, sino que presentan una periodicidad que hace necesario expresar sus perfiles energéticos como series de Fourier (ecuación 2.4). Cuanto más complejo es el perfil de energía de la torsión, más términos se necesitarán para describirlo (van der Spoel et al., 2004; van Gunsteren et al., 1996). En general, la mayoría de las energías torsionales se ajustan a una expansión de Fourier de orden tres.

$$E_{tor} = \sum_{tor} \sum_n \frac{V_n}{2} (1 + \cos(n\Phi - n)) \quad (2.4)$$

donde V_n representa barrera de torsión asociada a cada término de la función, n la periodicidad del término, Φ el ángulo diedro y n el ángulo de fase.

Este tipo de torsión recibe el nombre de *torsión propia* para diferenciarla de otro tipo especial de torsión, la *torsión impropia*, que se utiliza para forzar a determinados átomos a mantenerse en el plano generado por otros tres. Este tipo de torsiones se aplican especialmente para reproducir adecuadamente las vibraciones de fuera del plano de átomos unidos a sistemas aromáticos. En el campo de fuerzas de AMBER, se aplica una expresión similar a la torsión convencional (ecuación 2.5) (Case et al., 2004b).

$$E_{tor} = \sum_{tor} \sum_n \frac{V_n}{2} (1 + \cos(2\xi_n - 180)) \quad (2.5)$$

donde V_n representa la constante de fuerza de cada término de la función, n la periodicidad del término y ξ el ángulo diedro.

- **Términos no enlazantes**

- i) Término electrostático

La distribución de carga de una molécula puede ser representada utilizando cargas puntuales fraccionales normalmente centradas en los núcleos, las cuales han sido optimizadas para reproducir las propiedades electrostáticas de la misma (Leach, 2001). La interacción electrostática (E_{ele}) queda, pues, representada según la ley de Coulomb (ecuación 2.6).

$$E_{ele} = \sum_i \sum_{j>i} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (2.6)$$

donde q_i y q_j son las cargas puntuales de cada átomo, r_{ij} la distancia entre átomos y ϵ la constante dieléctrica del medio.

La mayoría de los campos de fuerzas actuales (en su versión no polarizable) usan cargas derivadas de funciones de onda obtenidas a nivel *Hartree-Fock/6-31G(d)*, las cuales sobreestiman la polarización. En simulaciones acuosas donde el solvente está tratado explícitamente ($\epsilon=1$), esta sobreestimación ayuda a corregir los fenómenos de polarización relacionados con la solvatación.

- ii) Término de *van der Waals*

Es usado tanto para evitar la fusión de átomos no unidos de carga opuesta, como para representar las interacciones atractivas a corta distancia fruto de las fuerzas dispersivas. La forma general de las expresiones que se utilizan para describir este tipo de potencial se reduce a la suma de dos fuerzas: una repulsiva para las distancias cortas entre los átomos, y otra atractiva para distancias mayores. Las dos tienden asintóticamente a cero a medida que aumenta la distancia.

La expresión más utilizada para calcular este término es la denominada 12-6 (o potencial de Lennard-Jones), donde el término de repulsión depende de la doceava potencia de la distancia entre átomos y el término de atracción depende de la sexta potencia de la distancia (van der Spoel et al., 2004) (ecuación 2.7).

$$E_{vdw} = \sum_i \sum_{j>i} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) \quad (2.7)$$

donde A_{ij} y B_{ij} son constantes para cada par de átomos relacionados con sus radios de van der Waals y durezas, y r_{ij} es la distancia entre ellos.

Para minimizar los efectos que se producen en los bordes de un sistema finito en fase condensada se suele recurrir a las *condiciones periódicas de contorno* (PBC). Mediante ellas, los átomos del sistema se incluyen dentro de una caja tridimensional, la cual estará rodeada de réplicas idénticas a si misma en todas las direcciones del espacio. Durante la simulación, cuando una molécula se mueve en la caja central, su imagen periódica en cada una de las otras cajas se moverá exactamente con la misma orientación y de la misma manera. Por tanto, si una molécula abandona la caja central, una de sus imágenes entrará por la cara opuesta de la misma.

Cuando se usen PBC o en sistemas muy grandes, el cálculo de las interacciones no enlazantes entre átomos unidos por tres o más enlaces químicos (véase figura 2.1) será muy costoso computacionalmente, ya que implicará el cálculo de todas las interacciones entre pares atómicos posibles $O(N^2)$, siendo N el número de átomos. Por este motivo, se suelen realizar algunas restricciones. Las más comunes serían:

-Truncación de las interacciones. Todas las interacciones no enlazantes decaen a medida que se incrementa la distancia entre los átomos implicados. Ello da soporte a la estrategia de calcular sólo las interacciones entre partículas incluidas dentro de un cierto radio. La generación de la lista de pares de átomos incluidos dentro de la distancia de corte es una operación costosa en términos computacionales, así que suele realizarse periódicamente. La distancia de corte está relacionada con las dimensiones máximas de

la caja, soliendo ser inferior a la mitad del eje mayor. Esta estrategia reduce en gran medida el número de interacciones a calcular (van der Spoel et al., 2004). No obstante, conlleva ciertos problemas en sistemas cargados donde el decaimiento de la energía con la distancia es muy lento. Se pueden llevar a cabo algunas medidas para suavizar esta discontinuidad, como el empleo de funciones de *switching* o la asignación de *grupos cargados* (Bhandarkar et al., 2005; van der Spoel et al., 2004; van Gunsteren et al., 1996). Estas medidas se suelen utilizar tanto para las interacciones de van der Waals como para las electrostáticas.

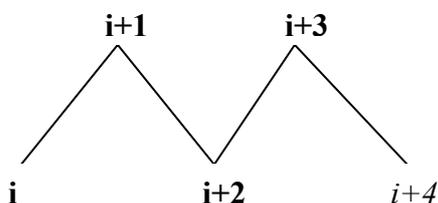


Figura 2.1. Representación esquemática de las interacciones del átomo i con sus vecinos. $i+1$ e $i+2$ son exclusiones del átomo i . Para la interacción entre i e $i+3$ se suele utilizar un factor de escalado.

-Interacciones de largo-alcance. El tratamiento inadecuado de las interacciones electrostáticas de larga distancia es capaz de generar inestabilidades en el sistema, por lo que es recomendable el uso de métodos como el *campo de reacción*, el *método de celda múltiple* o el *sumatorio de Ewald*, que permiten introducir su efecto dentro del cálculo. El método de Ewald, originariamente creado para calcular las interacciones de largo alcance en cristales replicados periódicamente (Ewald, 1921), es el más utilizado en simulaciones de ácidos nucleicos con condiciones periódicas de contorno. La idea del mismo es transformar el cálculo de sumatoria de todas las interacciones electrostáticas posibles en otros dos términos de convergencia mucho más rápida (ecuación 2.8).

$$E_{ele} = Ee_{dir} + Ee_{rec} + E_0 \quad (2.8)$$

donde Ee_{dir} se refiere al *espacio directo*, Ee_{rec} se refiere al *espacio recíproco* y E_0 es una constante.

Cuando se utiliza este método, las interacciones de rango-corto o de *espacio directo* son calculadas mediante un potencial Coulómbico modificado, mientras que el resto o de *espacio recíproco* son calculadas vía sumatoria de vectores. En el cálculo del espacio directo se utilizan unas distribuciones gaussianas para representar las cargas originales, que hacen converger más rápidamente el término que en su forma original. Dichas distribuciones son posteriormente contrarestadas mediante distribuciones gaussianas de carga inversa en el cálculo del espacio recíproco. Este último término converge más rápidamente que el potencial coulómbico original (Leach, 2001). Una versión modificada del método de Ewald es el *Particle Mesh Ewald* (PME) (Darden et al., 1993), donde se optimiza el cálculo del espacio recíproco a través de la transformación mediante interpolación *B-spline* de las cargas puntuales a mallas tridimensionales. Dichas mallas son posteriormente tratadas mediante transformadas de Fourier, las cuales reducen notablemente el costo computacional.

- **Eficacia de los campos de fuerzas para Ácidos Nucleicos**

De entre la diversidad de campos de fuerzas que existen para la representación de ácidos nucleicos, las últimas versiones de CHARMM (CHARMM-27) y AMBER (AMBER-99) son las más populares. Estos campos de fuerzas han sido capaces de reproducir estructuras de diversos ácidos nucleicos y de predecir diversas propiedades macroscópicas de los mismos (Auffinger & Westhof, 1998; Beveridge et al., 2004; Beveridge & McConnell, 2000; Cheatham, 2004; Cheatham & Kollman, 2000; Cheatham & Young, 2001; Giudice & Lavery, 2002; Lankas et al., 2003; MacKerell & Nilsson, 2000; Norberg & Nilsson, 2002; Orozco et al., 2003; Orozco et al., 2004; Swarnalatha et al., 2002).

En los últimos años, AMBER-99 se ha convertido en el campo de fuerzas de defecto, especialmente para el estudio de formas inusuales de ácidos nucleicos. Esto, y nuestra experiencia propia (Cubero et al., 2003; Cubero et al., 2002; Cubero et al., 2001a; Cubero et al., 2000; Cubero et al., 2001b; Harris et al., 2001; Soliva et al., 2000; Soliva et al., 1998) han hecho patente que este sea el campo de fuerzas de elección para simulaciones de ácidos nucleicos.

2.3.2 Métodos basados en campos de fuerzas

El resultado elemental de un cálculo clásico es la energía potencial asociada a una configuración de los átomos del sistema. Esta información puede ser procesada *a posteriori* usando diferentes algoritmos, lo que da lugar a un abanico de métodos computacionales. Se explicarán a continuación los métodos clásicos utilizados en esta tesis, la Mecánica Molecular (MM) y la Dinámica Molecular (MD).

- Mecánica molecular (MM): Minimización de energía

La función de energía potencial de una macromolécula tiene un perfil muy complejo. En una superficie de potencial se encuentra un mínimo global y varios mínimos locales, donde las derivadas de la función de energía potencial con respecto a las coordenadas son cero y las segundas derivadas son positivas (Frenkel & Smit, 2002; Leach, 2001).

Conociendo toda la superficie de energía potencial es posible derivar la función de partición (ecuación 2.9) y determinar todo el comportamiento del sistema. Desgraciadamente, para sistemas grandes como los ácidos nucleicos, la energía es función de $3N$ (coordenadas cartesianas) y por tanto el número de mínimos locales es tan grande que es imposible muestrearlo completamente y es de hecho muy difícil, sino imposible, localizar el mínimo absoluto. Sí que es posible, dada una configuración inicial, encontrar el mínimo energético más cercano (Leach, 2001). A esto proceso se le conoce como *minimización de energía* (figura 2.2).

$$Q = \sum_{\text{estados}} e^{-E_i / k_B T} \quad (2.9)$$

donde Q es la función de partición, E_i es la energía en un estado, k_B es la constante de Boltzmann y T es la temperatura.

Los métodos más utilizados hoy en día para este propósito son los basados en gradientes, los cuáles hacen uso de la primera derivada de la función de energía

potencial respecto a la posición de los átomos. Entre ellos, los más frecuentemente usados son el *steepest descent* y el *conjugate gradients*. El primero, que utiliza una dirección de búsqueda basada sólo en el gradiente, es idóneo cuando la posición de partida está alejada del mínimo, mientras que el segundo, que guarda una memoria de las direcciones previas, es más eficiente cuando la búsqueda está cercana al mínimo (Leach, 2001). En la práctica, para la mayoría de sistemas, lo más frecuente es realizar una combinación de ambos, donde *steepest descent* se utiliza primero y *conjugate gradients* después.

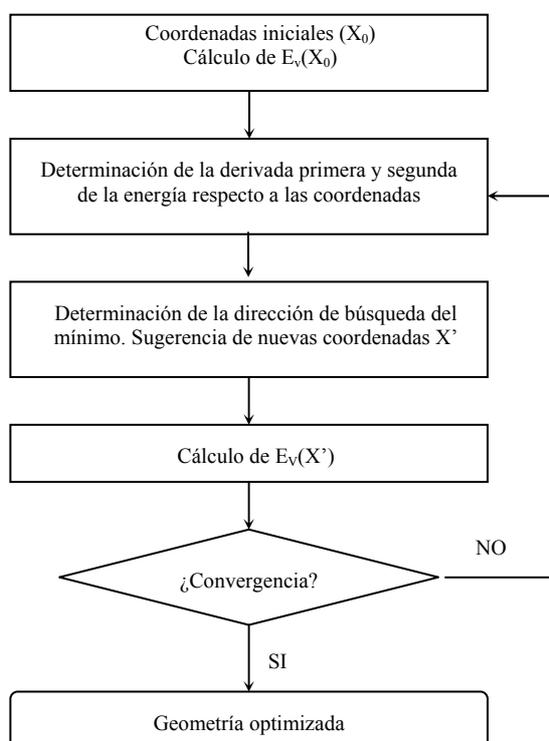


Figura 2.2 Algoritmo básico de los métodos de mecánica molecular.

La minimización de energía genera configuraciones individuales de mínima energía que en muchos casos, pueden ser suficiente para predecir ciertas propiedades de un sistema. Ahora bien, el conocimiento de una sola estructura, incluso si ésta pertenece al mínimo global, puede ser insuficiente a la hora de calcular propiedades *macroscópicas* de los mismos. De hecho, las propiedades macroscópicas de sistemas en *equilibrio* son el resultado de un *promedio temporal* de todos los microestados que aparecen durante el

tiempo de medida experimental. Desgraciadamente, no es posible simular todo el comportamiento dinámico para sistemas con un número de átomos *macroscópicos* (del orden de 10^{23}). Para solventar esta limitación Boltzmann y Gibbs postularon que era posible sustituir el promediado temporal de un sistema por un conjunto de *réplicas* del mismo, todas ellas consideradas de forma simultánea. Gracias a esto, dentro del marco de la mecánica estadística, asumiendo un comportamiento *ergódico*, es posible reemplazar el *promedio temporal* por lo que se conoce como *promedio sobre las réplicas* (Allen & Tildesley, 1989; Chandler, 1987; Leach, 2001). La función de partición pasará a ser entonces una *función de partición de las réplicas*, a partir de la cual se podrán obtener varias propiedades termodinámicas como la energía interna o la entalpía (Leach, 2001). Conviene señalar, por último, que dado que no se pueden enumerar todas las configuraciones posibles de un sistema, se ha de recurrir a las *técnicas de muestreo*, con las se obtienen promedios temporales aproximados, pero estadísticamente significativos.

Existen dos métodos para la generación de promedios aproximados representativos: Monte Carlo (MC) y Dinámica Molecular (MD). El primero de ellos es muy útil acoplado a métodos de variables colectivas, pero en simulaciones atomísticas de ácidos nucleicos la Dinámica Molecular es la técnica de elección. Esta potente técnica, además de representar el espacio configuracional del sistema de acuerdo con la ley de Boltzmann, permite representar su evolución a lo largo del tiempo, lo que la hace especialmente útil para seguir, por ejemplo, transiciones conformacionales (Karplus, 2003; Karplus & McCammon, 2002).

- Dinámica molecular (MD)

Se trata de una técnica *determinista* ya que permite predecir el estado de un sistema a partir del estado actual, definido en función de sus posiciones y momentos (Leach, 2001). Su fundamento está basado en las leyes de la mecánica clásica, mediante las cuales se integran las ecuaciones del movimiento con el fin de generar las sucesivas configuraciones, permitiendo obtener así una secuencia temporal de la evolución del sistema.

Las nuevas configuraciones se obtienen solucionando la ecuación diferencial contenida en la segunda ley de Newton (ecuación 2.10).

$$\frac{d^2 x_i}{dt^2} = \frac{F_i}{m_i} \quad (2.10)$$

que describe el movimiento de una partícula i de masa m_i a lo largo de la coordenada x_i siendo F_i la fuerza aplicada y donde t es el tiempo.

La complejidad de la superficie de potencial requiere que las ecuaciones del movimiento sean integradas numéricamente en pequeños pasos sucesivos. En cada paso se calculan las fuerzas en los átomos y posteriormente se combinan con las posiciones actuales para generar las nuevas posiciones. Los átomos son entonces desplazados a sus nuevas posiciones y se vuelve a comenzar el ciclo. En la figura 2.3 se muestra un esquema de un algoritmo básico de Dinámica Molecular.

Dentro de los algoritmos de integración numérica de las ecuaciones de Newton, el más usado es el de *leap-frog* o de salto de rana, donde las posiciones y las propiedades dinámicas se aproximan como una expansión de una serie de Taylor (ecuaciones 2.11, 2.12 y figura 2.4). Otros algoritmos con fundamentos similares utilizados en Dinámica Molecular serían el de *Verlet*, *Beeman* o modelos de *predictor-corrector* (Allen & Tildesley, 1989; Leach, 2001; Scott et al., 1999; van der Spoel et al., 2004; van Gunsteren et al., 1996).

La elección del *tiempo de integración* adecuado ha de ser un compromiso entre el tipo de muestreo que se necesite y la duración de los fenómenos a observar. En el caso de biomoléculas, los movimientos de frecuencia más rápida son los debidos a las vibraciones de los enlaces, los cuales se encuentran en el rango del femtosegundo (fs). Por tanto, el tiempo de integración que se debería utilizar para que no hubiese inestabilidades tendría que ser suficientemente pequeño como para reproducir estos movimientos, lo que supone una seria limitación en la escala temporal del proceso a simular. Una posibilidad de reducir el coste computacional consiste en aprovechar el hecho de que en los sistemas grandes los movimientos más rápidos contribuyen poco al

movimiento global. De esta manera es posible fijar la distancia de enlace atómica mediante, por ejemplo, el algoritmo de SHAKE (Ryckaert et al., 1977). Evitando estos rápidos movimientos es posible incrementar el tiempo de integración a 2 fs, sin introducir inestabilidades manifiestas en el sistema.

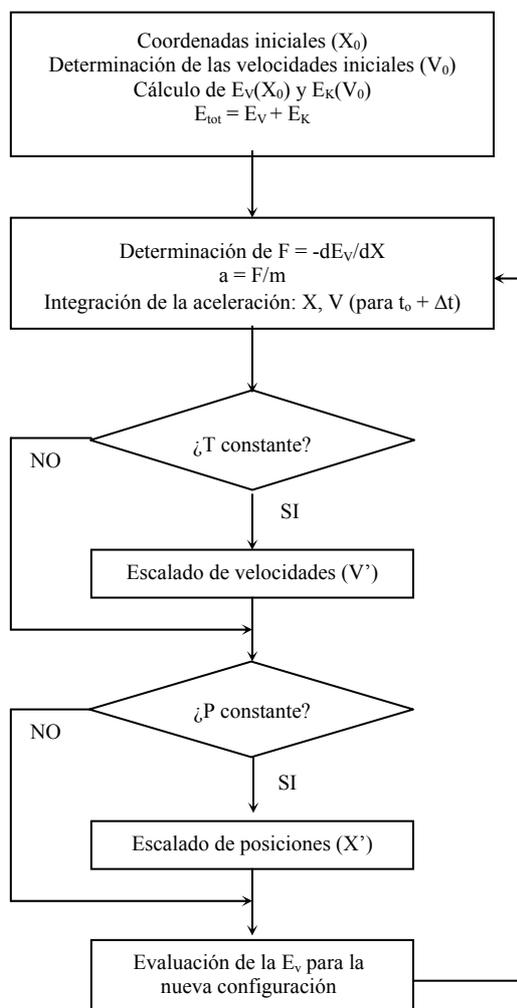


Figura 2.3 Algoritmo básico de los métodos de dinámica molecular.

$$v\left(t + \frac{\Delta t}{2}\right) = v\left(t - \frac{\Delta t}{2}\right) + \frac{F(t)}{m} \Delta t \quad (2.11)$$

$$r(t + \Delta t) = r(t) + v\left(t + \frac{\Delta t}{2}\right) \Delta t \quad (2.12)$$

donde r son las posiciones al tiempo t y velocidades v al tiempo $t - \Delta t/2$. Se recalculan las posiciones y las velocidades usando las fuerzas F determinadas por las posiciones al tiempo t .

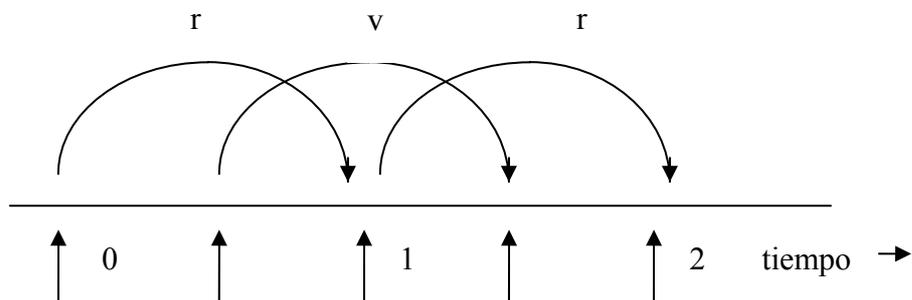


Figura 2.4 El método de integración de salto de rana (*leap-frog*). El algoritmo es llamado así debido a que r y v van saltando como ranas entre sus espaldas

Condiciones de la simulación de Dinámica Molecular

El punto inicial de cualquier simulación es siempre una **configuración inicial**. Dicha estructura podrá provenir de datos experimentales de Rayos-X, Resonancia Magnética Nuclear o de un modelo teórico. Asimismo, precisa un conjunto de velocidades iniciales, que se asignarán al azar siguiendo una distribución de tipo Maxwell-Boltzmann, tal que la energía cinética asociada se corresponda a la energía térmica diana para la temperatura de trabajo. Finalmente, requiere fijar las **condiciones de simulación**: el número de partículas (N), el volumen (V), la temperatura (T), la presión (P) o la energía total del sistema (E). En función de la combinación de los mismos se distingue entre varios colectivos:

- microcanónico (NVE)
- isotérmico-isobárico (NPT)
- canónico (NVT)

siendo NPT y NVT los más usados en dinámica molecular de ácidos nucleicos.

Como se ha descrito con anterioridad, la finalidad última de las simulaciones de ácidos nucleicos es reproducir propiedades *macroscópicas* a partir de modelos, atómicos en este caso. Así pues, para reproducir propiedades *fisiológicas* del ADN,

además de mantener la temperatura y presión dentro de unos rangos, se debe introducir explícitamente el efecto del solvente en el cálculo.

El uso de condiciones periódicas de entorno (PBC), permiten que una simulación se lleve a cabo utilizando un número reducido de moléculas de solvente y a la vez represente un sistema altamente hidratado. El tamaño y morfología de la caja periódica suele estar relacionado con la geometría de nuestra molécula y con la distancia de corte que se aplica a las interacciones de van der Waals. Esto es así para evitar fenómenos relacionados con la periodicidad, o al menos, minimizarlos cerca del ácido nucleico. En sistemas donde una dimensión predomina sobre el resto, como es el caso de las dobles hebras de ADN, normalmente se utilizan cajas ortorómbicas de tipo rectangular, minimizando así el número de aguas necesarias para rodear a la molécula. Desgraciadamente, el tiempo de correlación rotacional del ADN está dentro de la escala del nanosegundo y, dado que cada vez los tiempos de simulaciones son mayores, es de esperar la presencia de rotaciones en el ADN, que pueden provocar interacción directa con las cadenas de las celdas periódicas vecinas produciendo fenómenos de *crystalización* (Cheatham, 2004; Orozco et al., 2004). Para solucionar esto último, cada vez se suele recurrir con más frecuencia a la utilización de un tipo de cajas que poseen un perfil parecido al esférico, pero manteniendo las aristas, permitiendo así su replicación tridimensional en el espacio. Las más usadas son el *octaedro truncado* (14 caras, $x=y=z$, $\alpha=\beta=\gamma=109.47^\circ$) y el *dodecaedro rómbico* (12 caras, $x=y=z$, $\alpha=\gamma=60^\circ$, $\beta=90^\circ$). Dichas cajas permiten rotar libremente en su interior al polímero, pero presentan la desventaja de que su uso puede ser computacionalmente prohibitivo para secuencias largas de ADN debido a la gran cantidad de aguas que se han de incluir en ellas.

En cálculos periódicos, si se trabaja bajo condiciones isotérmicas-isobáricas se tendrá que controlar tanto la temperatura como la presión. En el caso de la temperatura existen diferentes opciones. La primera de ellas sería el re-escalado de las velocidades mediante la aplicación de un factor. Otra alternativa sería mantener la temperatura acoplada a un baño térmico externo fijado a la temperatura deseada. Dichos baños actúan como fuente de energía térmica, añadiendo o eliminando calor del sistema en

función de las necesidades. Las velocidades son también re-escaladas en cada paso, en función de la diferencia de temperatura entre el baño y el sistema. Los métodos más conocidos son el de Berendsen (Berendsen et al., 1984) y el de Nosé-Hoover (Hoover, 1985; Nosé, 1984). Por lo que respecta a la presión, los métodos para mantenerla constante son análogos a los utilizados para la temperatura.

Tras la primera integración de las ecuaciones de Newton se pasa a un proceso de **equilibrado**, el propósito de cuál es llevar a nuestro sistema a un equilibrio a partir de la configuración original de posiciones y momentos. Durante este periodo se suelen monitorizar varios parámetros tales como la energía potencial o la temperatura, hasta que éstos llegan a un valor estable. La duración de dicho periodo es variable y depende en gran medida del sistema de estudio. En el caso de los ácidos nucleicos se puede llegar a una situación de equilibrio apenas después de 500 picosegundos, aunque la tendencia en el futuro serán tiempos de equilibrado mayores.

Una vez superado el equilibrado se entra en la llamada **fase de producción**. La duración de la misma siempre será un compromiso entre recursos computacionales, espacio de disco y la capacidad de racionalizar dichos resultados. La evolución dinámica de nuestra estructura quedará recogida en forma de coordenadas cartesianas, almacenadas normalmente como una *trayectoria*.

Un punto clave en un cálculo de Dinámica Molecular es estimar cuál ha de ser la longitud óptima de una simulación. En general, se estima que la misma debería de ser al menos 10 veces mayor que la escala temporal del proceso a estudiar. En la práctica, esto significa que sólo procesos acontecidos en la escala del nanosegundo podrán ser simulados con los ordenadores actuales, aunque la evolución de los algoritmos y recursos computacionales hace vislumbrar la escala del microsegundo como algo no muy lejano en el tiempo. Que ello sea posible dependerá en gran medida de la mejora en los algoritmos de paralelización. Dos de los programas más utilizados para el cálculo extensivo de Dinámica Molecular como son AMBER (Case et al., 2004a) y NAMD (Kale et al., 1999) usan estas técnicas de paralelización, haciendo posible el estudio de sistemas mayores y a una escala temporal también mayor.

2.4 Recogida de datos y análisis

Las simulaciones de Dinámica Molecular generan largas *trayectorias*, el estudio de las cuáles es cada vez más complejo dado que la razón señal/ruido es más baja (Orozco et al., 2005). En la práctica, por motivos históricos, la tendencia en el caso de los ácidos nucleicos ha sido salvar las coordenadas cada picosegundo, obteniendo así una representación muy exhaustiva de la evolución temporal de los mismos. Sin embargo, al extenderse el tiempo de simulación, trayectorias guardadas cada picosegundo se hacen excesivamente grandes, lo que dificulta su análisis. Por ello resulta interesante emplear frecuencias de grabado de coordenadas más bajas (sin olvidar, no obstante, que la separación entre las mismas debería de ser menor que la escala temporal del fenómeno a estudiar). Para tener un muestreo macroscópico correcto, en teoría se deberían recopilar 10 estructuras o *snapshots* por cada grado de libertad, lo que implica que para un sistema de N átomos el número de estructuras debería ser $10 \times 3N$. En los ácidos nucleicos existen numerosas restricciones torsionales, lo que limita a sólo 7 grados de libertad por nucleótido de entre los aproximadamente 65 que le corresponderían al contar el número total de átomos ($3N$). De esta manera, para tener un muestreo *estadísticamente significativo* de una MD de ácidos nucleicos sólo se necesitan $10 \times 7 \times M$ *snapshots* (siendo M el número de nucleótidos) (Orozco et al., 2005). Así, para una simulación de 10 nanosegundos de una doble hebra de 12 pares de bases, se necesitan aproximadamente 1700 estructuras (70×24 *bp*) para representar la trayectoria en términos de varianza total. Así pues, siempre que no se este interesado en *movimientos de alta frecuencia* o en *reordenaciones del solvente*, existirá la opción de salvar las coordenadas cada 5 picosegundos con lo que el coeficiente de ahorro de espacio en disco sería de 1/5.

Una vez se tiene almacenada la trayectoria, se puede proceder a su estudio, ya sea mediante visualización o mediante análisis numérico. Existen en el mercado programas para visualizar y analizar trayectorias de dinámica molecular, algunos de los más usados serían VMD (*Visual Molecular Dynamics*) (Humphrey et al., 1996) y *ptraj* (Case et al., 2004a). Es frecuente también el uso de aplicaciones propias para análisis específicos

sobre trayectorias de MD. Se pueden distinguir tres grupos de análisis en función del tipo de información a obtener de las coordenadas:

- 1- Información estructural promediada
- 2- Información estructural dinámica
- 3- Cálculos de estabilidad

1- Información estructural promediada

En el caso de los ácidos nucleicos es posible calcular una serie de parámetros estructurales promedio. Entre ellos:

- *Parámetros helicoidales*, como por ejemplo, el *twist* o el *roll* (Lavery & Sklenar, 1988; Lu & Olson, 2003).

- Relacionados con pautas de *distancias* atómicas como podrían ser los puentes de hidrógeno, los tiempos de residencia de aguas o los mapas de contacto entre residuos cercanos (figura 2.5).

- *Descriptor*es genéricos de la estructura como la *superficie accesible al solvente* (SAS) (Hubbard, 1993) o el *factor de sección de colisión* (CCS) (Helden et al., 1993; Shvartsburg & Jarrold, 1996; von Helden et al., 1991).

- *Energéticos* como, por ejemplo, la energía de interacción entre bases o el potencial electrostático molecular.

- *Estructuras promedio* de fragmentos de trayectoria, que sirven para obtener una visión aproximada del sistema.

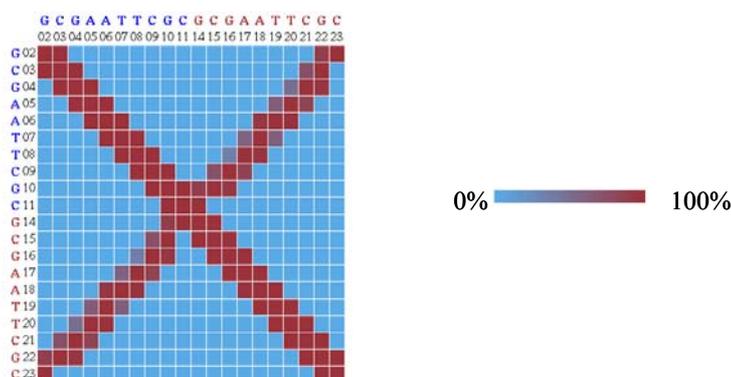


Figura 2.5 Mapa de contacto entre residuos proveniente del último nanosegundo de una MD del dodecámero de Dickerson (extremos no incluidos, distancia de corte 3.5 Å).

2-Información estructural dinámica

Una trayectoria contiene información detallada de las fluctuaciones temporales de una molécula. Esta información puede ser procesada de diferentes maneras, por ejemplo, se puede calcular el *RMSD* (desviación cuadrática media) respecto a una estructura de referencia, la cual suele ser la original o una estructura promediada de la parte final de la trayectoria (ecuación 2.13).

$$RMSd_k = \left[\frac{1}{N} \sum_{l=1}^{3N} (x_{kl} - x_l)^2 \right]^{1/2} \quad (2.13)$$

donde x_{kl} representa la l -coordenada en la estructura k , x_l es el valor de la l coordenada en la estructura de referencia y N es el número de átomos

Una alternativa al RMSD sería la desviación cuadrática media con múltiples estructuras de referencia, lo que puede dar lugar a mapas bidimensionales donde se comparan todos los *snapshots* de la trayectoria. Este tipo de análisis se suele complementar con análisis de subgrupos o *clusters*, con la finalidad de encontrar familias conformacionales en las trayectorias (figura 2.6)

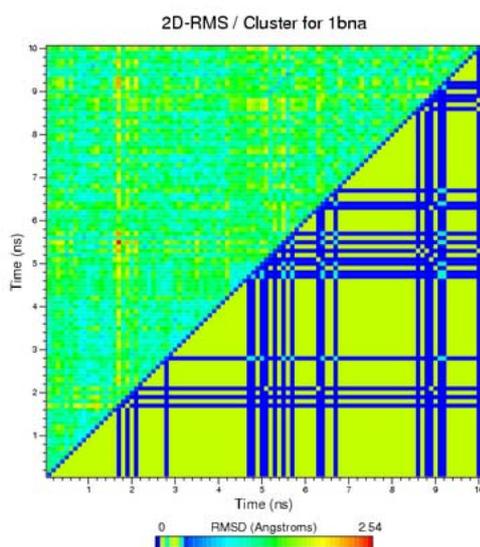


Figura 2.6 2D-RMS (arriba izquierda) y análisis de cluster (abajo derecha) de una simulación de MD de 10 ns de un dodecámero (1BNA).

Otro parámetro relacionado con las fluctuaciones atómicas sería el cálculo de los *B-factors* (ecuación 2.14), que tienen la ventaja de poder ser comparados con valores experimentales.

$$B=(8\pi^2/3) \langle \Delta r^2 \rangle \quad (2.14)$$

donde Δr es la fluctuación atómica con respecto a la posición promedio

Se puede tener un estimado de la compactación de una estructura mediante el cálculo del *radio de giro* (ecuación 2.15)

$$R_{gyr} = \left(\frac{\sum_i r_i^2 m_i}{\sum_i m_i} \right)^{1/2} \quad (2.15)$$

donde m_i es la masa del átomo i y r_i es la posición del átomo i respecto el centro de masas de la molécula.

Además de los parámetros mencionados anteriormente, una de las técnicas de mayor poder dentro del estudio del comportamiento dinámico del ADN es la técnica de *dinámica esencial*.

- **Dinámica esencial del ADN**

Los ácidos nucleicos podrían ser considerados como construcciones biológicas que han sido diseñadas para permitir ciertos tipos de movimientos internos que les permiten realizar su función en la célula. Estos movimientos internos normalmente implican correlaciones complejas de grupos de átomos, pero su naturaleza es inherente a la estructura y las interacciones de la molécula (Amadei et al., 1993). Es de gran interés por tanto derivar dichos movimientos y poder así reducir el complejo comportamiento de los mismos a unas pocas pautas de deformación esenciales.

Las correlaciones entre movimientos atómicos pueden ser derivadas de simulaciones de Dinámica Molecular a través de la diagonalización de la matriz de covarianza

siguiendo un esquema de análisis de componentes principales (PCA) (Amadei et al., 1993) (ecuación 2.16).

$$C = \text{cov}(x) = \langle (x - \langle x \rangle)(x - \langle x \rangle)^T \rangle \quad (2.16)$$

donde $\langle \rangle$ hace referencia a un promedio temporal y x son las coordenadas para cada átomo en la trayectoria; en la práctica, todas las estructuras se referencian a una estructura promedio común, lo que elimina movimientos globales de traslación y rotación en el cálculo de la matriz de covarianza.

En el caso del ADN dichas matrices pueden ser construidas en base a varios modelos de coordenadas diferentes: cartesianas, cartesianas ponderadas por masa, helicoidales o matrices en espacio de distancias (diedros). De ahora en adelante nos referiremos a matrices de covarianza generadas a partir de coordenadas cartesianas aunque los fundamentos del método son análogos para el resto de matrices.

Cuando se trabaja con un número suficiente de configuraciones independientes (al menos $3N+1$, siendo N el número de átomos), la diagonalización de la matriz de covarianza genera un conjunto de $3N$ vectores propios o *eigenvectors* y sus valores propios asociados o *eigenvalues*, de los cuales 6 serán representativos de movimientos rígidos (rotación y traslación) y tendrán un valor de cero. La diagonalización de matrices es un cálculo computacional muy costoso tanto en términos de tiempo de procesador como de memoria virtual, especialmente en sistemas con más de 3000 átomos. Para solucionar este problema se suele recurrir a diversas alternativas numéricas que normalmente ahorrarán recursos computacionales. Un ejemplo serían los métodos de fragmentación de matriz de covarianza (Amadei et al., 1996) o los relacionados con diagonalización iterativa, poseyendo estos últimos la ventaja que el usuario puede elegir el número de vectores propios a calcular. Existen también rutinas iterativas de análisis que permiten calcular exclusivamente los valores propios sin necesidad de obtener los vectores propios (Press et al., 1992). En nuestra experiencia, los métodos iterativos suelen ser más eficientes que los directos para el estudio de trayectorias de macromoléculas.

Los vectores propios que se obtienen con el análisis de componentes principales:

i) Son capaces de explicar toda la varianza del sistema, ii) son ortogonales y iii) la fracción de la varianza total explicada por uno de ellos nos viene dada por su correspondiente valor propio. En la figura 2.7 se muestra la representación espacial de un *eigenvector* extraído de una trayectoria de dinámica molecular de ADN.

El análisis de los valores propios obtenidos de la diagonalización de la matriz de covarianza cartesiana permite estudiar la flexibilidad del ADN en términos armónicos. Es posible, por ejemplo, derivar constantes de fuerza asociadas a la deformación de un modo (ecuación 2.17).

$$K_i = k_B T / \lambda_i \quad (2.17)$$

donde k_B la constante de Boltzmann (en unidades de kcal/mol K), T es la temperatura y λ_i es el *eigenvalue* asociado al modo i (en \AA^2).

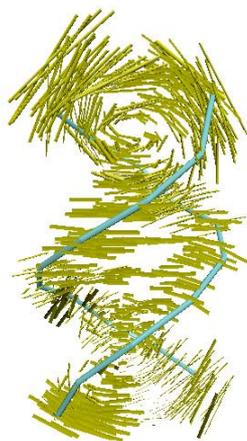


Figura 2.7 1er *eigenvector* (amarillo) de una MD de un dodecámero de Dickerson en modo *porcupine plot*. Las posiciones atómicas de la estructura promedio son nuestro origen para cada componente del vector (dado que el vector es unitario, se ha escalado por un factor para la correcta visualización de las direcciones). Se muestra el movimiento en ambos sentidos.

Una vez conocida la constante, es posible reproducir la energía de deformación a lo largo de un modo mediante la ecuación 2.18.

$$E_i = \frac{K_i}{2} (\Delta X_i)^2 \quad (2.18)$$

donde ΔX es la deformación cartesiana a lo largo del modo i .

Los valores propios representan el desplazamiento a lo largo de un modo a una determinada temperatura. Por tanto, multiplicándolos se puede obtener un estimado del volumen configuracional accesible (ecuación 2.19). Desafortunadamente, el uso directo de este valor es ambiguo ya que depende del número de valores propios utilizados, tendiendo a cero al aproximarse a $3N-6$ (siendo N el número de átomos). Es por ello que es posible referirse al espacio conformacional haciendo uso del *volumen configuracional generalizado* y la *dimensionalidad*. El primero es obtenido como una modificación *ad hoc* del volumen configuracional que tiene propiedades de límite (ecuación 2.20). La dimensionalidad es una medida conformacional que se refiere al número de modos esenciales que se necesitan para describir el espacio de flexibilidad importante de una molécula (Perez et al., 2005) y se obtiene como el índice del primer modo que contribuye en menos de 1 \AA^2 al volumen.

$$Vol = \prod_{i=1}^m (\lambda_i) \quad (2.19)$$

donde m es el número de valores propios utilizados.

$$Vol_{gen} = \prod_{i=1}^m (1 + \lambda_i)^{1/2} \quad (2.20)$$

Los vectores propios dan información sobre el tipo de movimiento de la biomolécula. Una primera aproximación al estudio de dichos movimientos sería calcular los desplazamientos de unas coordenadas promedio a lo largo de los vectores más representativos en términos de varianza, pudiendo generar así una nueva trayectoria. Esta *pseudo*-trayectoria puede analizarse numéricamente o simplemente visualizada.

En el caso de que se necesite comparar la naturaleza de varios vectores o, es decir, comparar *espacios esenciales* se puede recurrir al producto escalar entre los diferentes vectores (ecuación 2.21) (Hess, 2000; Hess, 2002). Esto puede ser útil como medida de la convergencia en el espacio muestreado por una molécula o para comparar movimientos esenciales entre simulaciones.

$$\gamma_{AB} = \frac{1}{n} \sum_{j=1}^n \sum_{i=1}^n (v_i^A \cdot v_j^B)^2 \quad (2.21)$$

donde n es el número mínimo de eigenvectors que se necesitan para definir un conjunto importante de vectores (que expliquen por ejemplo el 80% o 90% de la varianza); valores de 0 indican movimientos ortogonales entre las trayectorias mientras que valores cercanos a 1 indican alto grado de semejanza.

La variabilidad conformacional que suelen poseer las trayectorias hace necesario el uso de índices relativos. Se han propuesto diferentes alternativas como sería calcular la proyección cuadrada media de todos los vectores de una simulación frente a otra (Amadei et al., 1993; Amadei et al., 1996) o el uso de la variabilidad intrínseca de cada simulación (ecuación 2.22) (Perez et al., 2005).

$$\kappa_{AB} = 2 \frac{\gamma_{AB}}{(\gamma_{AA}^T + \gamma_{BB}^T)} \quad (2.22)$$

donde γ_{XX}^T es el índice a auto-similitud absoluto (ecuación 2.20) para la trayectoria X obtenido mediante la comparación de la primera y segunda mitades de la misma.

Si bien los índices de similitud señalados son medidas robustas para el estudio de las similitudes entre espacios conformacionales, presentan limitaciones. Por ejemplo, asumen que todos los modos contribuyen de igual manera al producto escalar, con lo que es imposible detectar permutaciones de los mismos. El uso de índices más complejos que tienen en cuenta los valores propios asociados a cada movimiento puede ser útil en estos casos (véase ecuación 2.23) (Perez et al., 2005).

$$\xi_{AB} = \frac{2 \sum_{i=1}^{i=z} \sum_{j=1}^{j=z} \left[\left(v_i^A \circ v_j^B \right) \frac{\exp \left\{ -\frac{(\Delta x)^2}{\lambda_i^A} - \frac{(\Delta x)^2}{\lambda_j^B} \right\}}{\sum_{i=1}^{i=z} \exp \left\{ -\frac{(\Delta x)^2}{\lambda_i^A} \right\} \sum_{j=1}^{j=z} \exp \left\{ -\frac{(\Delta x)^2}{\lambda_j^B} \right\}} \right]^2}{\sum_{i=1}^{i=z} \left(\frac{\exp \left\{ -2 \frac{(\Delta x)^2}{\lambda_i^A} \right\}}{\left(\sum_{i=1}^{i=z} \exp \left\{ -\frac{(\Delta x)^2}{\lambda_i^A} \right\} \right)^2} \right)^2 + \sum_{j=1}^{j=z} \left(\frac{\exp \left\{ -2 \frac{(\Delta x)^2}{\lambda_j^B} \right\}}{\left(\sum_{j=1}^{j=z} \exp \left\{ -\frac{(\Delta x)^2}{\lambda_j^B} \right\} \right)^2} \right)^2} \quad (2.23)$$

donde λ_i es el valor propio (en \AA^2) asociado al vector propio μ_i , con vector unitario v_i . La suma puede ser extendida a todos ($z=m$) o a un número importante de vectores propios ($z=n$).

3-Cálculos de estabilidad

La información derivada de las simulaciones de MD puede ser procesada para estimar la estabilidad absoluta o relativa de un sistema. Esto se puede llevar a cabo partiendo la energía libre del sistema entre *intramolecular* y de *solvatación*. La energía libre intramolecular se puede dividir entonces en contribuciones entrópicas y entálpicas. La entalpía se obtiene directamente a través del campo de fuerzas en las propias estructuras recogidas durante la simulación, mientras que para el cálculo de la entropía pueden usarse diferentes aproximaciones. Siguiendo la definición de Shannon, la entropía puede ser calculada en un espacio configuracional finito definido por M mediante la ecuación 2.24. Desgraciadamente, el uso de esta ecuación es complejo para macromoléculas, ya que implica un muestreo completo de todos los M estados. Por este motivo se suele recurrir a aproximaciones de tipo cuasi-armónico, como las propuestas por Andricioaei y Karplus (Andricioaei & Karplus, 2001) o Schlitter (Schlitter, 1993). Ambas están basadas en el método de componentes principales ya comentado en el apartado anterior, haciendo uso de la matriz de covarianza cartesiana ponderada por masas para el cálculo de los valores propios, que se relacionan directamente con las frecuencias.

En el método de Andricioaei y Karplus (ecuación 2.25) se recurre a un modelo de oscilador armónico cuántico, mientras que en el de Schlitter (ecuación 2.26) se utiliza un modelo híbrido cuántico-clásico. Ambos métodos son capaces de transformar los valores propios en entropía, sugiriendo además resultados similares en numerosos experimentos. También son parecidas las limitaciones que poseen. Por ejemplo, el tratamiento armónico que emplean es útil sólo en casos de estructuras fluctuando sobre un sólo mínimo. Otro factor a tener en cuenta es que la entropía es dependiente del tiempo de simulación. Para corregir esta dependencia temporal se han sugerido diferentes aproximaciones, la más popular fue la propuesta por Harris (Harris et al., 2001) y se basa en realizar el cálculo de la entropía a diferentes intervalos temporales para posteriormente ajustarlos a una función (ecuación 2.27 y figura 2.8).

$$S = -k_B \sum_{i=1}^M P_i \ln P_i \quad (2.24)$$

$$S = k \sum_i \frac{\alpha_i}{e^{\alpha_i} - 1} - \ln(1 - e^{-\alpha_i}) \quad (2.25)$$

$$S \approx 0.5k \sum_i \ln \left(1 + \frac{e^2}{\alpha_i^2} \right) \quad (2.26)$$

donde $\alpha_i = \hbar \omega_i / kT$, ω son los valores propios obtenidos por diagonalización de la matriz de covarianza pesada por masas.

$$S(t) = S_\infty - \frac{\alpha}{t^\beta} \quad (2.27)$$

donde α y β son los parámetros a ajustar y t es el tiempo de simulación usado para el cálculo de la entropía.

Finalmente, para el cálculo de la energía libre de solvatación se utilizan típicamente modelos *continuos* aplicados sobre una serie de estructuras significativas recogidas durante la simulación (ecuación 2.28). Dentro de estos métodos, GB/SA (Generalized Born-Solvent Accesible) y PB/SA (Poisson-Boltzmann/Solvent Accesible) son los más usados, generando ambos resultados similares (Cubero et al., 2001b; Wang et al., 2001). Existe una aproximación alternativa que hace uso la *teoría de respuesta lineal*

conectada con superficie accesible al solvente (Morreale et al., 2004; Morreale et al., 2005). La ventaja de este método es que no necesita modelos continuos, ya que la energía de solvatación se calcula usando moléculas de solvente explícito (ecuación 2.29).

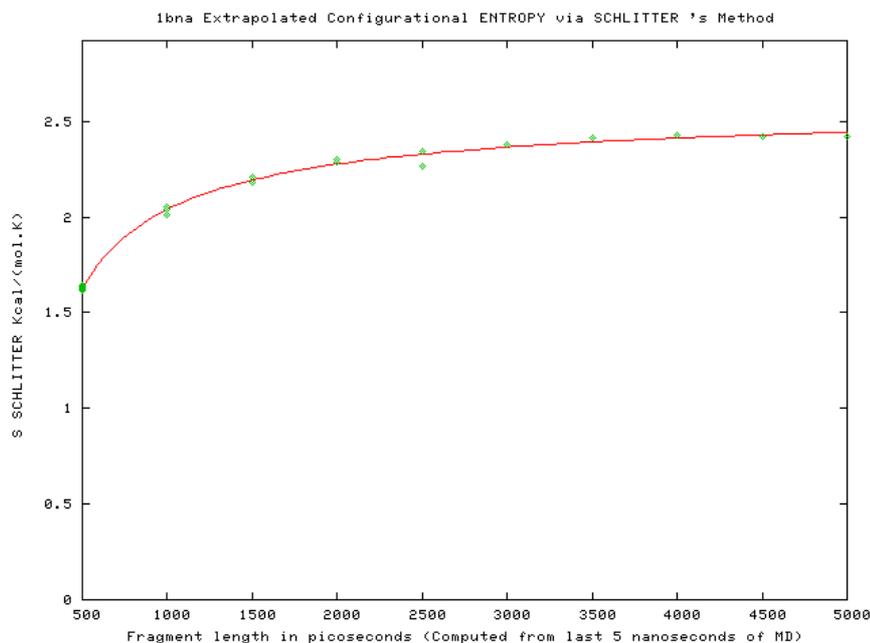


Figura 2.8 Ajuste temporal de valores de entropía configuracional obtenidos por el método de Schlitter a una función. Los valores se han obtenido de fragmentos de una simulación de un dodecámero de ADN.

$$G_{solv} = \langle G_{solv}(X) \rangle \quad (2.28)$$

donde X es una conformación de MD

$$G_{solv}(X) = \sum_{i=1}^N \xi_i SAS_i + \sum_{i=1}^N \sum_{j=1}^M \frac{q_i q_j}{r_{ij}} \quad (2.29)$$

donde ξ_i es el factor de tensión, SAS es la superficie accesible al solvente asignada a los átomos de soluto, M es el número de átomos del solvente, q se corresponde a las cargas atómicas y r_{ij} son las distancias atómicas.

2.5 Referencias bibliográficas

- Allen, M. P. & Tildesley, D. J. (1989). *Computer Simulation of Liquids*, Carendon Press.
- Amadei, A., Linssen, A. B. & Berendsen, H. J. (1993). Essential dynamics of proteins. *Proteins* 17, 412-425.
- Amadei, A., Linssen, A. B. & Berendsen, H. J. (1996). A comparison of Techniques for Calculating Protein Essential dynamics. *Journal of Computational Chemistry* 18, 169-181.
- Andricioaei, I. & Karplus, M. (2001). On the calculation of entropy from covariance matrices of the atomic fluctuations. *Journal of Chemical Physics* 115, 6289-6292.
- Auffinger, P. & Westhof, E. (1998). Simulations of the molecular dynamics of nucleic acids. *Curr Opin Struct Biol* 8, 227-236.
- Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F., DiNola, A. & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *Journal of Chemical Physics* 81, 3684-3690.
- Beveridge, D. L., Barreiro, G., Byun, K. S., Case, D. A., Cheatham, T. E., 3rd, Dixit, S. B., Giudice, E., Lankas, F., Lavery, R., Maddocks, J. H., Osman, R., Seibert, E., Sklenar, H., Stoll, G., Thayer, K. M., Varnai, P. & Young, M. A. (2004). Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. Research design and results on d(CpG) steps. *Biophys J* 87, 3799-3813.
- Beveridge, D. L. & McConnell, K. J. (2000). Nucleic acids: theory and computer simulation, Y2K. *Curr Opin Struct Biol* 10, 182-196.

- Bhandarkar, M., Brunner, R., Chipot, C., Dalke, A., Dixit, P., Grayso, P., Gullinsgrud, J., Gursov, A., Humfrey, W., Hurwitz, D., Krawetz, N., Nelson, M., Phillips, J., Shinozaki, A. & Zhu, F. (2005). *NAMD User's Guide*, Urbana-Champaign.
- Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham, T. E., III, Ross, W. S., Simmerling, C. L., Darden, T. L., Marz, K. M., Stanton, R. V., Cheng, A. L., Vincent, J. J., Crowley, M., Tsui, V., Radmer, R. J., Duan, Y., Pitera, J., Massova, I., Seibel, G. L., Singh, U. C., Weiner, P. K. & Kollman, P. A. (2004a). *AMBER8*. University of California, San Francisco.
- Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham, T. E., III, Ross, W. S., Simmerling, C. L., Darden, T. L., Marz, K. M., Stanton, R. V., Cheng, A. L., Vincent, J. J., Crowley, M., Tsui, V., Radmer, R. J., Duan, Y., Pitera, J., Massova, I., Seibel, G. L., Singh, U. C., Weiner, P. K. & Kollman, P. A. (2004b). *AMBER8. User's Manual*. University of California, San Francisco.
- Chandler, D. (1987). *Introduction to Modern Statistical Mechanics*, Oxford University Press.
- Cheatham, T. E., 3rd. (2004). Simulation and modeling of nucleic acid structure, dynamics and interactions. *Curr Opin Struct Biol* 14, 360-367.
- Cheatham, T. E., 3rd & Kollman, P. A. (2000). Molecular dynamics simulation of nucleic acids. *Annu Rev Phys Chem* 51, 435-471.
- Cheatham, T. E. & Young, M. A. (2001). Molecular Dynamics Simulation of Nucleic Acids: Successes, Limitations and Promise. *Biopolymers* 56, 232-256.
- Cubero, E., Abrescia, N. G., Subirana, J. A., Luque, F. J. & Orozco, M. (2003). Theoretical study of a new DNA structure: the antiparallel Hoogsteen duplex. *J Am Chem Soc* 125, 14603-14612.

- Cubero, E., Avino, A., de la Torre, B. G., Frieden, M., Eritja, R., Luque, F. J., Gonzalez, C. & Orozco, M. (2002). Hoogsteen-based parallel-stranded duplexes of DNA. Effect of 8-amino-purine derivatives. *J Am Chem Soc* 124, 3133-3142.
- Cubero, E., Guimil-Garcia, R., Luque, F. J., Eritja, R. & Orozco, M. (2001a). The effect of amino groups on the stability of DNA duplexes and triplexes based on purines derived from inosine. *Nucleic Acids Res* 29, 2522-2534.
- Cubero, E., Laughton, C. A., Luque, F. J. & Orozco, M. (2000). Molecular Dynamics Study of Oligonucleotides Containing Difluorotoluene. *Journal of the American Chemical Society* 122, 6891-6899.
- Cubero, E., Luque, F. J. & Orozco, M. (2001b). Theoretical studies of d(A:T)-based parallel-stranded DNA duplexes. *J Am Chem Soc* 123, 12018-12025.
- Darden, T. L., York, D. & Pedersen, L. (1993). Particle Mesh Ewald: AN N-log(N) method for Ewald sums in large systems. *Journal of Chemical Physics* 98, 10089-10092.
- Ewald, P. P. (1921). Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.* 64, 253-287.
- Frenkel, D. & Smit, B. (2002). *Understanding Molecular Simulation*, Academic Press.
- Giudice, E. & Lavery, R. (2002). Simulations of nucleic acids and their complexes. *Acc Chem Res* 35, 350-357.
- Harris, S. A., Gavathiotis, E., Searle, M. S., Orozco, M. & Laughton, C. A. (2001). Cooperativity in drug-DNA recognition: a molecular dynamics study. *J Am Chem Soc* 123, 12658-12663.

- Helden, G. V., M-T., H., Gotts, N. & Bowers, M. T. (1993). Carbon Cluster Cations with up to 84 atoms: Structures, Formation Mechanism and Reactivity. *Journal of Physical Chemistry* 97, 8182-8192.
- Hess, B. (2000). Similarities between principal components of protein dynamics and random diffusion. *Phys Rev E Stat Nonlin Soft Matter Phys*, 8438-8448.
- Hess, B. (2002). Convergence of sampling in protein simulations. *Phys Rev E Stat Nonlin Soft Matter Phys* 65, 031910.
- Hoover, W. (1985). Canonical dynamics: equilibrium phase space distributions. *Phys Rev A* 31, 1695-1697.
- Hubbard, S. J., and J. M. Thornton. (1993). NACCESS Computer Program. Department of Biochemistry and Molecular Biology. University College London, London, UK.
- Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD: visual molecular dynamics. *J Mol Graph* 14, 33-38, 27-28.
- Kale, L., Skeel, R., Bhandarkar, M., Brunner, R., Gursoy, A., Krawetz, N., Phillips, J., Shinozaki, A., Varadarajan, K. & Schulten, K. (1999). NAMD2: Greater scalability for parallel molecular dynamics. *Journal of Computational Physics*, 283-312.
- Karplus, M. (1983). A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *Journal of Computational Chemistry* 4, 187-217.
- Karplus, M. (2003). Molecular dynamics of biological macromolecules: A brief history and perspective. *Biopolymers* 68, 350-358.
- Karplus, M. & McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nat Struct Biol* 9, 646-652.

- Lankas, F., Sponer, J., Langowski, J. & Cheatham, T. E., 3rd. (2003). DNA basepair step deformability inferred from molecular dynamics simulations. *Biophys J* 85, 2872-2883.
- Lavery, R. & Sklenar, H. (1988). The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J Biomol Struct Dyn* 6, 63-91.
- Leach, A. R. (2001). *Molecular Modelling: Principles and applications, 2nd edition*, Pearson Education (Prentice Hall), Essex.
- Lu, X. J. & Olson, W. K. (2003). 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res* 31, 5108-5121.
- MacKerell, A. D., Jr. & Nilsson, L. (2000). *Nucleic Acid Simulations*. University of Nottingham.
- McCammon, J. A. & Harvey, S. C. (1987). *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press, Cambridge.
- Morreale, A., de la Cruz, X., Meyer, T., Gelpi, J. L., Luque, F. J. & Orozco, M. (2004). Linear response theory: an alternative to PB and GB methods for the analysis of molecular dynamics trajectories? *Proteins* 57, 458-467.
- Morreale, A., de la Cruz, X., Meyer, T., Gelpi, J. L., Luque, F. J. & Orozco, M. (2005). Partition of protein solvation into group contributions from molecular dynamics simulations. *Proteins* 58, 101-109.
- Norberg, J. & Nilsson, L. (2002). Molecular dynamics applied to nucleic acids. *Acc Chem Res* 35, 465-472.
- Nosé, S. (1984). A molecular dynamics method for simulation in the canonical ensemble. *Molecular Physics* 52, 255-268.

- Orozco, M., Noy, A., Meyer, T., Rueda, M., Ferrer, C., Valencia, A., Perez, A., de La Cruz, X. & Luque, F. J. (2005). Data mining of molecular dynamics trajectories of nucleic acids.
- Orozco, M., Perez, A., Noy, A. & Luque, F. J. (2003). Theoretical methods for the simulation of nucleic acids. *Chem Soc Rev* 32, 350-364.
- Orozco, M., Rueda, M., Blas, J. R., Cubero, E., Luque, F. J. & Laughton, C. A. (2004). Molecular dynamics simulations of nucleic acids. *Encyclopedia of Computational Chemistry*.
- Perez, A., Blas, J. R., Rueda, M., López-Bes, J. M., de La Cruz, X., Luque, F. J. & Orozco, M. (2005). Exploring the essential dynamics of B.DNA. *Journal of Chemical Theory and Computation* 1, 790-800.
- Press, W. H., Teukolsky, W. T., Vetterling, W. T. & Flannery, B. P. (1992). *Numerical Recipes. The Art of Scientific Computing. 2nd edition*, Cambridge Univ.Press.
- Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. (1977). Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* 23, 327-341.
- Schlitter, J. (1993). Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chemical Physics Letters* 215, 617-621.
- Scott, W. R. P., Huenenberger, P. H., Tironi, I. G., Mark, A. E., Billeter, S. R., Fennen, J., Torda, A. E., Huber, T., Krueger, P. & van Gunsteren, W. F. (1999). The GROMOS Biomolecular Simulation Program Package. *Journal of Physical Chemistry A* 103, 3596-3607.
- Shvartsburg, A. A. & Jarrold, M. F. (1996). An Exact Hard Spheres Scattering Model for the Mobilities of Polyatomic Ions. *Chemical Physics Letters* 261, 86-91.

- Soliva, R., Guimil Garcia, R., Blas, J. R., Eritja, R., Asensio, J. L., Gonzalez, C., Luque, F. J. & Orozco, M. (2000). DNA-triplex stabilizing properties of 8-aminoguanine. *Nucleic Acids Res* 28, 4531-4539.
- Soliva, R., Laughton, C. A., Luque, F. J. & Orozco, M. (1998). Molecular Dynamics Simulations in Aqueous Solution of Triple Helixes Containing d(G-C·C) Trios. *Journal of the American Chemical Society* 120, 11226-11233.
- Swarnalatha, R., Leclerc, F. & Karplus, M. (2002). DNA Polymorphism: A comparison of Force Fields for Nucleic Acids. *Biophys J* 84, 1421-1449.
- van der Spoel, D., Lindahl, E., Hess, B., van Buuren, A. R., Apol, E., Meulenhoff, P. J., Tieleman, D. P., Sijbers, A., Feenstra, K. A., van Drunen, R. & Berendsen, H. J. (2004). Gromacs User Manual version 3.2.
- van Gunsteren, W. F., Billeter, S. R., Eising, A. A., Hünenberger, P. H., Krüger, P., Mark, A. E., Scott, A. M. & Tironi, I. G. (1996). Biomolecular simulations: The GROMOS96 manual and user's guide. In *Hochschulverlag Ag an der ETH Zürich, Zurich*.
- von Helden, G., Ming-Teh, H., Kemper, P. R. & Bowers, M. T. (1991). Structures of carbon cluster ions from 3 to 60 atoms: Linears to rings to fullerenes. *Journal of Chemical Physics* 95, 3835-3837.
- Wang, A. H., Donini, O., Reyes, C. & Kollman, P. A. (2001). Biomolecular simulations: Recent developments in force fields, simulations of enzyme catalysis and protein-ligand, protein-protein and protein-nucleic acid non-covalent interactions. *Annu Rev Biophys Biomol Struct* 30.

Weiner, S. J., Kollman, P. A., Nguyen, D. T. & Case, D. A. (1986). An all atom force field for simulations of proteins and nucleic acids. *Journal of Computational Chemistry* 7, 230-252.

Capítulo III

ESTRUCTURA Y DINÁMICA DE LOS ÁCIDOS NUCLEICOS

Durante los primeros años del siglo XX, los biólogos reconocieron que los *genes* estaban localizados en los *cromosomas*, conocidos entonces como estructuras lineales presentes en el núcleo de las células eucariotas (Alberts et al., 1999). Se pensaba que el ADN era una molécula de gran sencillez química, al contrario que las proteínas, por lo que se aceptaba con dificultad su papel primordial como material genético. Sólo desde los trabajos estructurales de Watson y Crick (Watson & Crick, 1953) se conoce que el ADN es una entidad estructural con unas capacidades ideales para mantener y expresar la herencia genética.

3.1 Estructura Primaria

Los *nucleótidos* son la unidad básica que compone la estructura molecular de los ácidos nucleicos. Dichas subunidades están compuestas de un azúcar cíclico de tipo furanósico (β -D-ribosa en el ARN o β -D-2'-desoxiribosa en el ADN) fosforilado en su posición 5' y sustituido en C1' por una base nitrogenada unida covalentemente por un enlace β -glucosídico. En sistemas vivos se encuentran dos tipos fundamentales de bases nitrogenadas: las *púricas* y las *pirimidínicas*. De ellas, cinco aparecen como bases codificantes portadoras de información genética: adenina (A), guanina (G), citosina (C), timina (T) y uracilo (U), encontrándose esta última sólo presente en el ARN (figura 3.1).

Los nucleótidos están unidos entre sí covalentemente a través de enlaces fosfodiéster 5'-3' que conectan el oxígeno O5' de un nucleótido con el O3' del siguiente, formando así una cadena que constituye el esqueleto de azúcar-fosfato (figura 3.2). Este tipo de unión confiere a la molécula *polaridad* química donde el fosfato formará una pequeña protuberancia tridimensional negativa y la pentosa estará formando un orificio entre ellos. Los extremos serán fácilmente diferenciables dado que se hallará una protuberancia en uno y un orificio en el otro (Alberts et al., 1999).

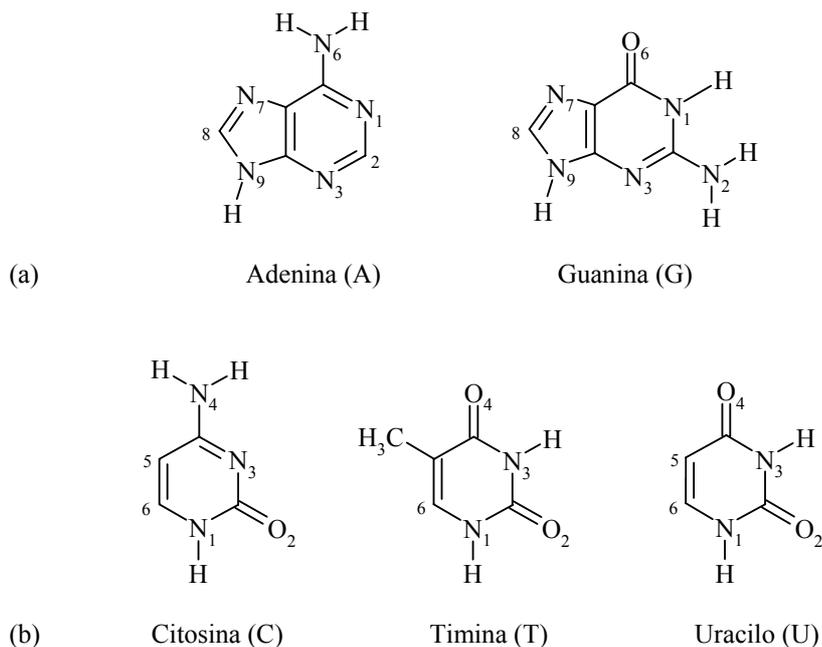


Figura 3.1 Estructura química de las bases nitrogenadas (a) púricas y (b) pirimidínicas. En su nomenclatura se numeran los átomos pesados correspondientes mientras que los hidrógenos adoptan el mismo número que el átomo al que están unidos (Saenger, 1984).

3.1.1 Propiedades físico-químicas

Debido a su naturaleza polianiónica los ácidos nucleicos son solubles en medio acuoso. Sus propiedades físico-químicas hacen que sus soluciones sean de tipo viscoso y precipitables mediante la adición de alcohol (Blackburn & Gait, 1990). Las bases son neutras en el rango fisiológico de pH y lo mismo sucede para las pentosas. En estas últimas, el 2',3'-diol podrá ceder un protón por encima de pH 12 mientras que los grupos hidroxilo se ionizarán sólo por encima de pH 15. Los fosfatos, sin embargo, pierden un protón a pH 1 y otro (en monoésteres) a pH 7 siendo los primeros en participar en el equilibrio de protones en rangos de pH cercanos a los fisiológicos (Blackburn & Gait, 1990). Las reacciones de protonación pueden tener trascendencia funcional a nivel estructural, enzimático o incluso en procesos relacionados con la mutagénesis de bases.

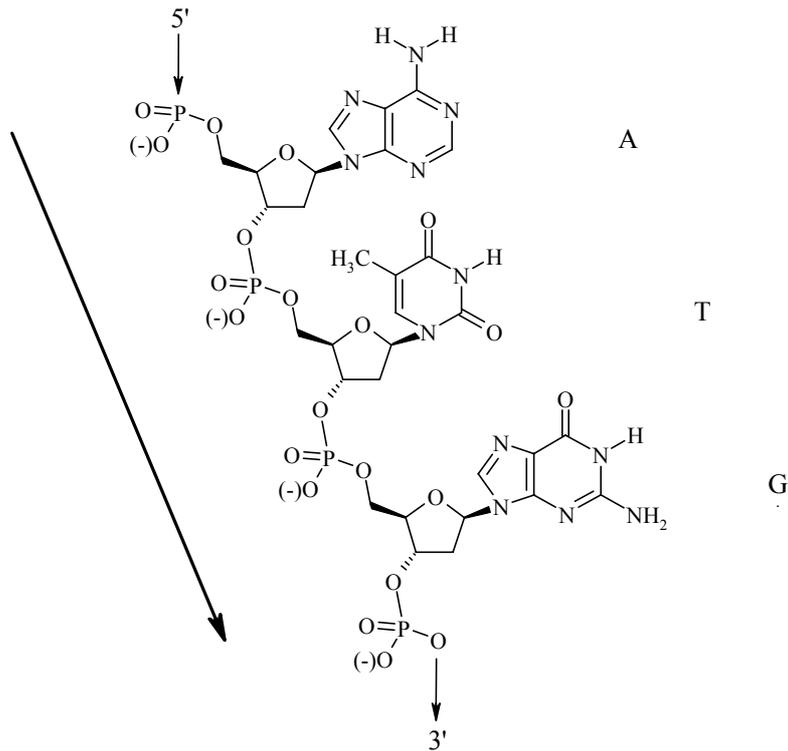


Figura 3.2 Fragmento de un ácido desoxirribonucleico (ADN) con la secuencia adenina, timina y guanina unido por enlaces 3', 5'-fosfodiéster. La dirección de la cadena viene mostrada por la flecha.

La estructura covalente de las nucleobases puede verse modificada por fenómenos de *tautomería*. En los rangos fisiológicos las bases son más estables en las formas tautoméricas mayoritarias *ceto* y *amino*, aunque se pueden encontrar formas *enol* o *imino* por debajo de un 0.1 por ciento en abundancia. Este porcentaje puede variar si se consideran bases modificadas químicamente o medios no acuosos (Colominas et al., 1996; Hernandez et al., 2000).

3.1.2 Conformación

La estructura tridimensional de los ácidos nucleicos viene modulada por rotaciones de grupos de átomos sobre enlaces sencillos (figura 3.3). De todos estos grados de libertad, los más influyentes son los que marcan la rotación respecto el enlace

glucosídico, el enlace exocíclico C4'-C5', el empaquetamiento de la ribosa y el grupo fosfato

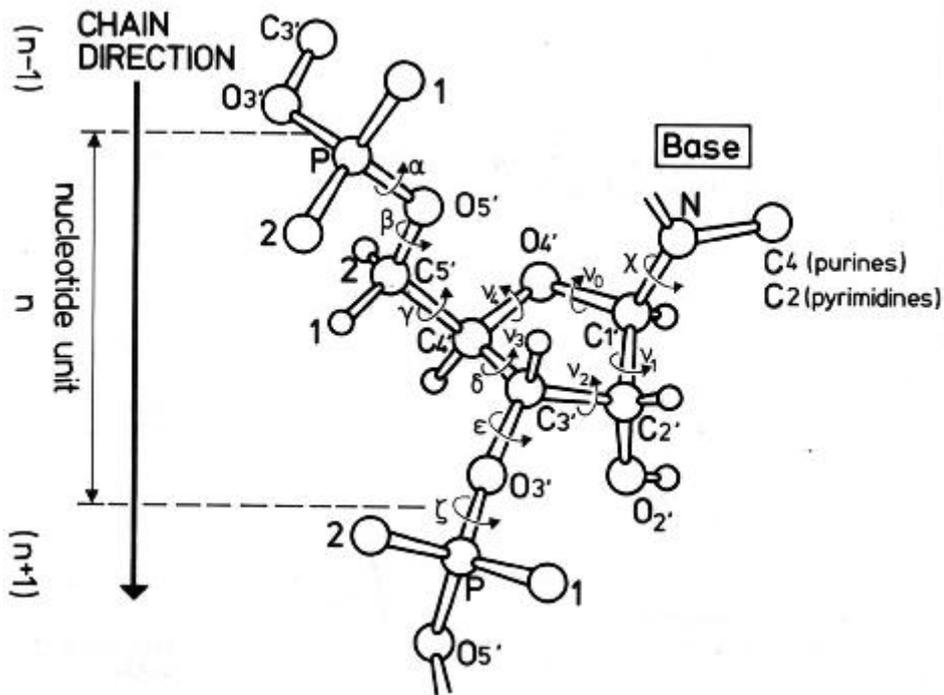


Figura 3.3 Numeración atómica y nomenclatura para las torsiones de un nucleótido (Saenger, 1984).

i) Modos de empaquetamiento o *puckering* del azúcar

El anillo furanósico de cinco átomos adopta siempre una estructura de no planaridad que pueden ser del tipo *sobre* (*E*), formada por cuatro átomos dentro de un mismo plano y un quinto en el exterior a 0.5 Å, o de tipo *girado* (*T*), donde dos átomos adyacentes son desplazados a caras opuestas de un plano formado por otros tres. Los átomos desplazados fuera de los planos en la misma cara del átomo C5' se denominan *endo*, mientras que los que se encuentran en la cara opuesta se denominan *exo* (figura 3.4).

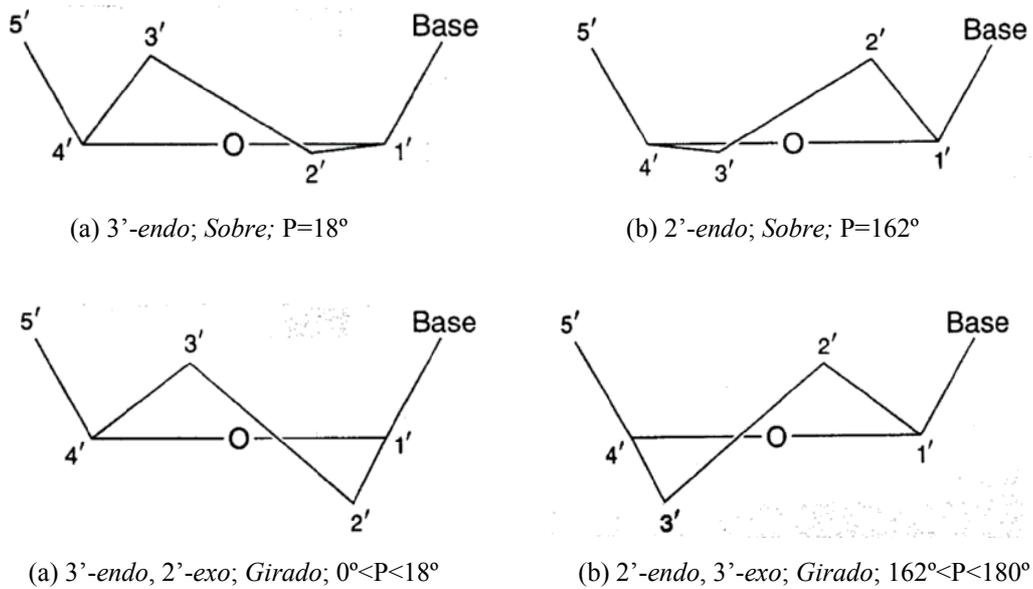
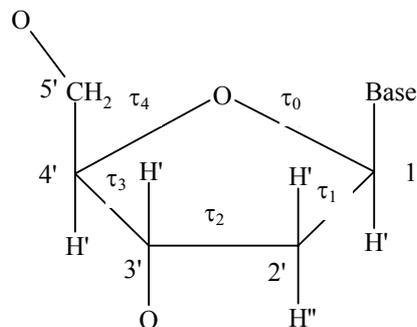


Figura 3.4 Conformaciones del azúcar encontradas en los nucleótidos. (a) Confórmeros Norte que se encuentran en la doble hélice de la forma A y (b) confórmeros Sur que se encuentran en la doble hélice de la forma B (Bloomfield et al., 2000).

En la forma *T* suele ocurrir que uno de los átomos de fuera del plano se encuentre más alejado que el otro recibiendo así la nomenclatura de *twist asimétrico*, donde el desplazamiento del átomo más alejado recibe el nombre *puckering mayor* y el del átomo más cercano se denomina *puckering menor* (Saenger, 1984).

Los métodos mencionados para definir la conformación del azúcar son aproximados y en ocasiones donde los valores de *twist* son intermedios se recurre a dos parámetros de naturaleza analítica. El primero de ellos es el ángulo de pseudorotación o *ángulo de fase* (P) y el segundo es denominado como *amplitud de puckering* (τ_m), de acuerdo con las siguientes expresiones (3.1 y 3.2).



$$P = ATAN \left[\frac{\tau_1 + \tau_4 - \tau_0 - \tau_3}{2\tau_2 (\sin 36^\circ + \sin 72^\circ)} \right] \quad (3.1) \quad \tau_m = \frac{\tau_2}{(\cos P)} \quad (3.2)$$

El ángulo de fase genera información sobre todos los posibles tipos de *puckerings* de la ribosa, pudiendo oscilar éste entre 0 y 360 grados. En función del mismo es posible distinguir cuatro zonas principales y sus combinaciones: Norte (N, $P \approx 0^\circ$), Sur (S, $P \approx 180^\circ$), Este (E, $P \approx 90^\circ$) y Oeste (W, $P \approx 270^\circ$).

Se conoce experimentalmente que existen dos familias preferidas para los nucleótidos, la C2'-*endo* con P en el rango de 144-190° o sur (S), y la C3'-*endo* con un ángulo de fase de 0-36° o norte (N). Para nucleótidos en solución ambos estados están en equilibrio, pero en polinucleótidos, éste puede estar totalmente desplazado a una región. Es posible encontrar transiciones entre ambas únicamente a través de la zona *Este* pasando por el estado intermedio O4'-*endo*. La barrera energética no es excesiva, siendo cercana a 4 Kcal/mol.

ii) Orientación *syn/anti* respecto el enlace glucosídico

La base puede adoptar dos orientaciones principales alrededor del azúcar en función del enlace glucosídico C1'-N: la forma *syn* y la *anti*. La torsión de este enlace χ se define mediante los átomos O4'-C1'-N9-C4 en el caso de bases púricas y por O4'-C1'-N1-C2 para las pirimidínicas. En la conformación *syn* la ribosa se encuentra eclipsando la base, mientras que en la *anti* el azúcar está en posición opuesta (véase figura 3.5). A las posiciones intermedias se las conoce como *high syn* y *high anti*.

En fase acuosa la conformación preferida de los nucleósidos suele ser la *anti* frente a la *syn* debido a la menor repulsión estérica entre la base y la ribosa. La preferencia por *anti* puede revertirse, por ejemplo, añadiendo un grupo voluminoso en posición 8. Otros factores que pueden influir en el equilibrio conformacional son los puentes de hidrógeno, como por ejemplo los formados entre el grupo O5' y los nitrógenos de las bases. La única excepción remarcable a la preferencia *anti* de los nucleósidos es la guanosina, donde la interacción del grupo exoamino con O5' estabiliza la conformación

syn. En el caso de los nucleótidos o polinucleótidos se encuentra mayoritariamente la forma *anti*.

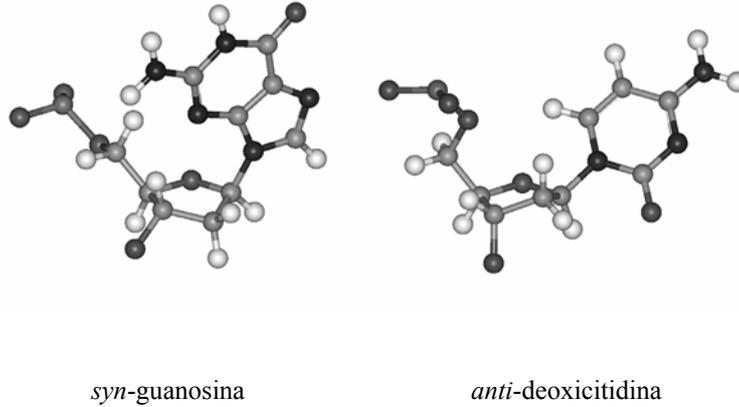


Figura 3.5 Conformaciones *syn* y *anti* de dos nucleósidos.

iii) Orientación respecto al enlace exocíclico C4'-C5'

La rotación respecto el enlace exocíclico C4'-C5' permite al átomo O5' asumir diferentes posiciones relativas al azúcar, siendo posible tres conformaciones principales (figura 3.6). Es común usar o bien dos ángulos de torsión para describir esas tres conformaciones, ϕ_{oo} (O5'-C5'-C4'-O4') y ϕ_{oc} (O5'-C5'-C4'-C3'), o bien usar los rangos angulares + (*gauche*) o - (*trans*).

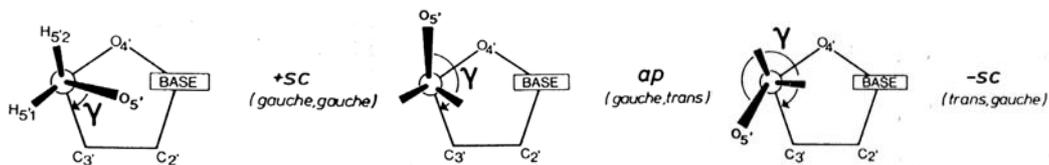


Figura 3.6 Definición de los rangos del ángulo diedro del enlace C4'-C5' (Saenger, 1984).

Este tipo de rotación es importante para el correcto posicionamiento de los fosfatos en nucleótidos. La forma *gauche, gauche* es la más abundante tanto en purinas como en pirimidinas. Generalmente se encuentra estabilizada por puentes de hidrógeno entre la base nitrogenada y el O5'.

iv) Rotación respecto a los enlaces C-O y P-O

La flexibilidad conformacional de los ácidos nucleicos puede ser ampliada por los enlaces con capacidad de rotar dentro del esqueleto fosfato. Por lo que respecta a los enlaces C-O, se ha visto en cristales que en oligonucleótidos el enlace C5'-O5' (medido por el ángulo β , véase figura 3.3) suele estar entorno a 180° mientras que el enlace C3'-O3' (dado por ϵ , véase figura 3.3) suele adoptar conformaciones en zona *anti* (alrededor 220°). Existe una cierta correlación en nucleósidos y nucleótidos entre el empaquetamiento del azúcar en C3'-endo y el desplazamiento de ϵ hacia la región *ap* (180°), mientras que si es C2'-endo se da un desplazamiento hacia la zona *-sc*. En general, en polinucleótidos se suele encontrar β entorno a $180 \pm 40^\circ$ y ϵ en rango 180 - 270° . Dos diedros definen la rotación respecto los enlaces P-O, ζ y α (véase figura 3.3), encontrándose ambos preferentemente en la zona *-sc*, si bien ζ se presenta a menudo en la conformación *-ap*.

3.2 Estructura Secundaria

Los nucleótidos se agrupan formando estructuras complejas la mayoría de las cuales se caracteriza por tener una *naturaleza helicoidal intrínseca*. La helicidad se genera, entre otras causas, en un intento de la molécula de minimizar la fuerte repulsión de carga que se origina entre sus fosfatos. Los ácidos nucleicos pueden estar formados tanto de una hebra, como de dos, tres e incluso cuatro de las mismas.

Las fuerzas estabilizantes de las conformaciones nativas de los ácidos nucleicos han de ser suficientemente fuertes como para mantener su integridad estructural, pero a la vez deben ser suficientemente lábiles como para permitir su flexibilidad conformacional. Además del requerimiento de mantener el esqueleto covalente, la estructura de los ácidos nucleicos viene dominada por diversas interacciones, las más importantes son:

-Las interacciones por *punte de hidrógeno*.

- Las interacciones de apilamiento o *stacking*.
- El efecto del solvente (agua e iones) al favorecer los apilamientos hidrófobos y apantallar la repulsión entre fosfatos.

3.2.1 Fuerzas estabilizantes de la asociación entre bases: Puentes de hidrógeno y apilamiento

- **Puentes de hidrógeno**

Las bases se reconocen específicamente mediante los puentes de hidrógeno. Cada base nitrogenada dispone de un perfil único de los mismos, que funciona a modo de huella dactilar frente al resto de moléculas. A nivel químico los puentes de hidrógeno son un tipo de interacciones con una marcada naturaleza electrostática entre un heteroátomo electronegativo poseedor de pares libres o *acceptor* y otro heteroátomo, el llamado *dador*, el cuál tiene unido un hidrógeno ácido. En el caso de los nucleótidos los grupos N-H son buenos dadores, mientras que los electrones con hibridación sp^2 de los C=O, o los nitrógenos del anillo son buenos aceptores.

En muchas ocasiones se suele recurrir a criterios geométricos de distancia para definir si existen este tipo de interacciones entre bases. Por ejemplo, se puede considerar que tres átomos $X-H\cdots Y$ están formando puente de hidrógeno si la distancia entre los heteroátomos está comprendida entre 2.6-3.5 Å y el ángulo formado por los enlaces *heteroátomo:hidrógeno-heteroátomo:heteroátomo* es inferior a 30°. En el caso de los ácidos nucleicos, la distancia entre los heteroátomos involucrados en el enlace suele oscilar entre 2.8 y 3.2 Å.

Además de servir para el reconocimiento horizontal entre bases, los puentes de hidrógeno son claves en muchas de las funciones que desempeñan los ácidos nucleicos. La unión específica de proteínas al surco mayor se realiza mayoritariamente a través de puentes de hidrógeno entre las bases y las cadenas laterales de los aminoácidos.

También se utilizan puentes de hidrógeno en la unión de ligandos pequeños al surco menor (figura 3.7).

Debido a esta diversidad de patrones de unión, se pueden encontrar hasta 28 tipos de apareamientos distintos de alta eficiencia (al menos dos puentes de hidrógeno lineales) entre las bases. El modo más común de reconocimiento en el ADN biológico es el propuesto por Watson y Crick (WC) (Watson & Crick, 1953), donde guanina se enfrenta con citosina mediante tres puentes de hidrógeno y adenina lo hace con timina mediante dos puentes de hidrógeno (figura 3.7). Los carbonos C1' están separados una distancia de $10.60 \pm 0.15 \text{ \AA}$ y los enlaces glucosídicos forman un ángulo de $68 \pm 2^\circ$.

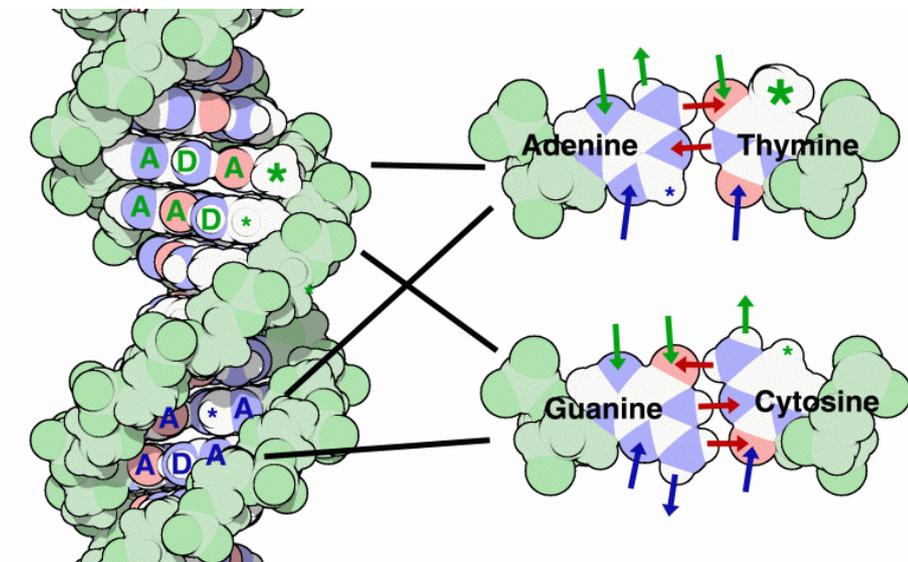


Figura 3.7 Posibilidades de puente de hidrógeno en una hebra de B-DNA. En rojo están los puentes de hidrógeno entre bases, en azul los puentes de hidrógeno en surco menor y en verde los del surco mayor. A=aceptor, D=donador (extraído de www.rcsb.org).

Otros modos de apareamiento detectados en los ácidos nucleicos son los pares *Hoogsteen* (en forma normal y reversa) y los *wobble* (figura 3.8). Este último fue propuesto por Crick para explicar la degeneración del código genético y desde entonces se ha encontrado en diferentes estructuras de ácidos nucleicos. En ellos se suelen perder

puentes de hidrógeno pero se favorecen las interacciones de apilamiento (Blackburn & Gait, 1990).

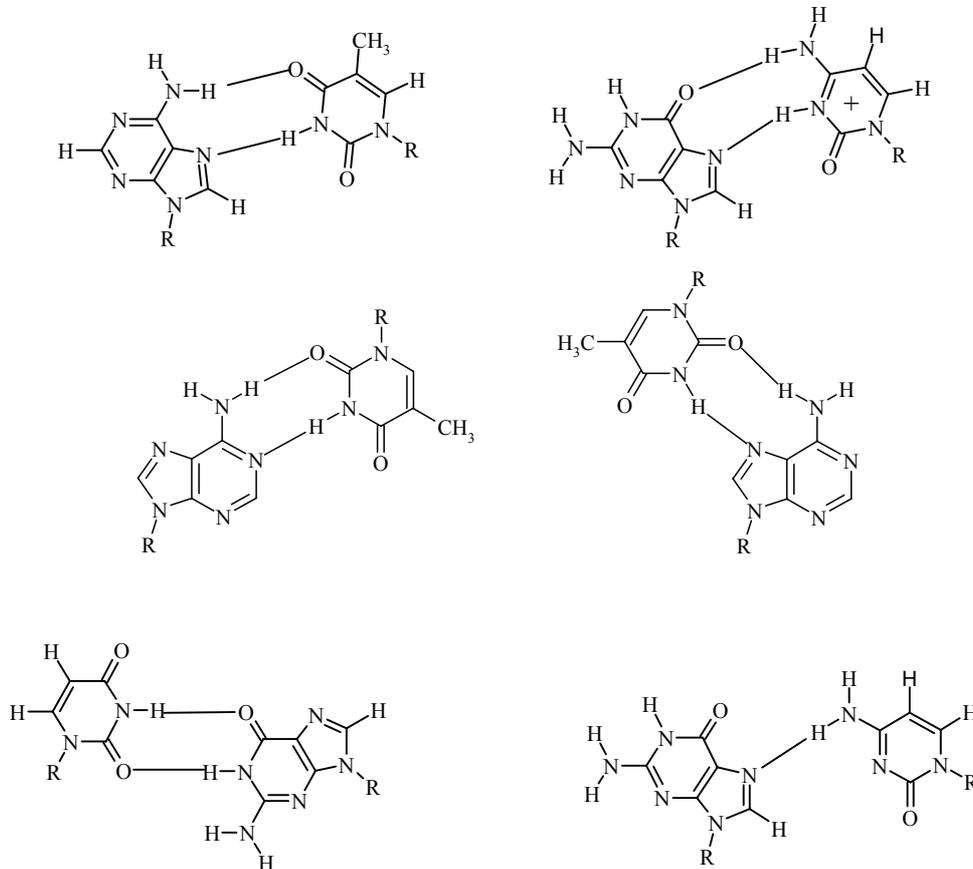


Figura 3.8 Apareamientos tipo *Hoogsteen* entre pares A·T y entre G·C protonada (arriba), apareamientos entre A·T de tipo rWC y de tipo rH (en medio) y apareamiento tipo ‘wobble’ entre U·G y G·C (*Hoogsteen*) (abajo).

- **Interacciones de apilamiento**

La interacción de apilamiento o *stacking* en la que dos bases se apilan de manera vertical es un fenómeno complejo donde participan varias contribuciones diferentes. Existe una parte de contribución electrostática, así como interacciones de van der Waals muy intensas (Weiner et al., 1986) y, finalmente, en medios acuosos, el efecto hidrófobo derivado de la desempaquetación de moléculas de agua cercanas a los anillos aromáticos.

Se ha demostrado experimentalmente que las bases tienden a agregarse en solución acuosa, siendo el apilamiento entre las mismas la fuerza primordial que genera este fenómeno. La energética del apilamiento ha sido estudiada por métodos de RMN e infrarrojo pudiendo determinar que, en general, en solución acuosa la estabilidad sigue esta pauta: purina:purina > pirimidina:purina > pirimidina:pirimidina (Saenger, 1984). La interacción de bases libres por puentes de hidrógeno en agua está desfavorecida, ya que existe competencia con la formación de puentes de hidrógeno con las moléculas de agua (Norberg & Nilsson, 1998). Por ello, a pesar de que en fase gas la interacción de puente de hidrógeno es más intensa que la de apilamiento, en muchas ocasiones se considera que el apilamiento es la fuerza directora que guía el plegamiento de los ácidos nucleicos.

3.2.2 Estructuras canónicas de los ácidos nucleicos

Al contrario de lo que originalmente se creyó, el ADN no es una estructura rígida y constante, ya que dependiendo de la composición de bases y del entorno puede adoptar distintas conformaciones. Dicho polimorfismo estructural se relaciona con implicaciones biológicas, ya que modificaciones locales o globales pueden afectar al modo de unión de proteínas posibilitando así alteraciones, por ejemplo, en la expresión génica. Hoy, más de cincuenta años después de la descripción de la estructura de la doble hélice, es posible dividir las estructuras secundarias de los ácidos nucleicos en dos grandes familias: las formas canónicas, que implican la asociación de dos cadenas de manera antiparalela y, las no canónicas, las cuales permiten otros tipos de asociaciones. Todas ellas serán helicoidales y con mayor o menor dificultad se clasificarán con ayuda de parámetros helicoidales como los mostrados en la figura 3.9.

El ADN canónico de doble hélice puede ser dextrógiro o levógiro en función del sentido de giro de las cadenas. Las formas dextrógiras poseen dos grandes familias: la A y la B, mientras que las levógiras poseen una sola familia, la Z. Todas las familias tienen un único tipo, a excepción de la B que cuenta con 4 tipos: B, C, D y T. Las

dobles hebras de ARN suelen estar confinadas a las formas A, por lo que su geometría será semejante al A-ADN.

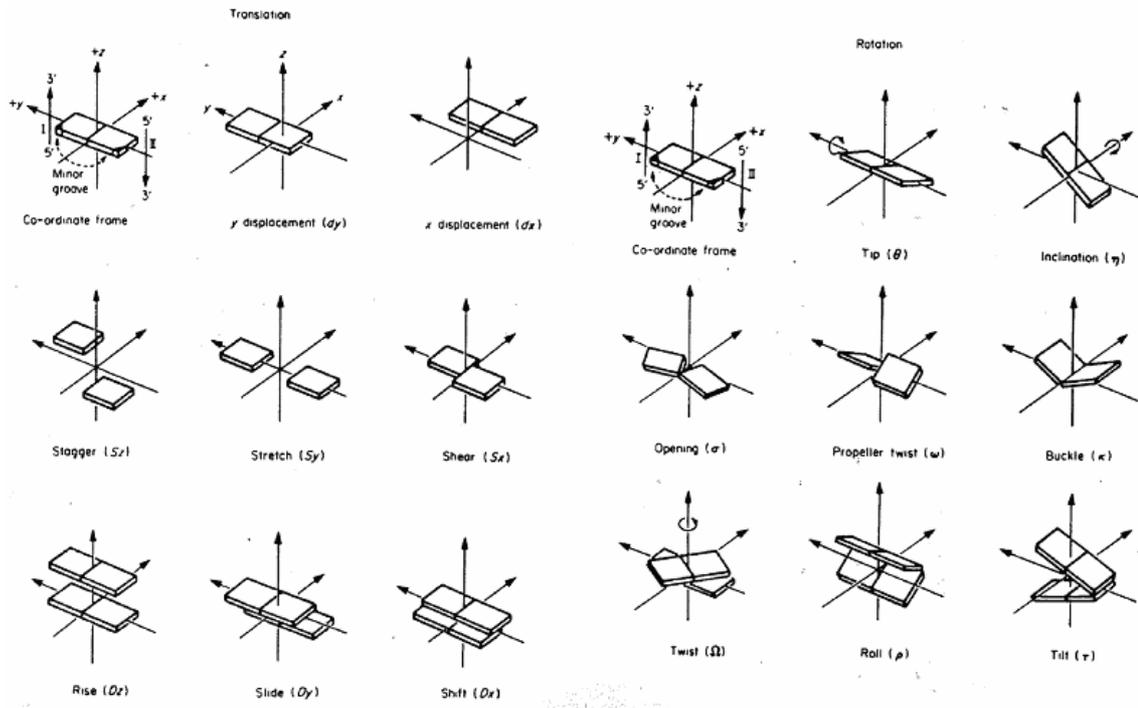


Figura 3.9 Definición de los parámetros conformacionales de translación y de rotación para describir el ADN (Dickerson, 1989).

- Formas dextrógiras

Las dos grandes familias de este grupo son la A y la B (figura 3.10), siendo esta última la más común en condiciones fisiológicas. Las características principales de ambas se recogen en la tabla 3.1.

i) Familia B

Es la conformación más común *in vivo* y la biológicamente activa en la mayoría de los casos. En ella existen dos surcos, el surco ancho o *major groove* y el surco estrecho o *minor groove*. La familia B se caracteriza por los pares de bases perpendiculares al eje de la hélice, su forma estilizada y su elevado grado de compactación. El resto de tipos

de ADN dentro de la familia B son bastante similares aunque con algunas peculiaridades.

Tabla 3.1 Características estructurales principales del A-ADN, B-ADN y Z-ADN.

| Propiedad | A-DNA | B-DNA | Z-DNA |
|--|----------------|----------------|-----------------------|
| Giro de la hélice | Dextrógiro | Dextrógiro | Levógiro |
| Unidad repetitiva | 1 par de bases | 1 par de bases | 2 pares de bases |
| Rotación por par de bases | 32.7° | 34.6° | 30° |
| Pares de bases por vuelta | ~11 | 10.4 | 12 |
| Inclinación de un par respecto al eje de la hélice | 19° | 1.2° | 9° |
| Avance por par de bases al lo largo de la hélice | 2.3 A | 3.4 A | 3.8 A |
| Paso de la hélice | 24.6 A | 34.0 A | 45.6 A |
| Diámetro | 25.5 A | 23.7 A | 18.4 A |
| Conformación del enlace glucosídico | Anti | Anti | Anti en C Syn en G |

El C-ADN aparece en condiciones de humedad relativamente bajas (44% a 66%) en presencia de iones litio o cuando se incrementa la concentración de iones sodio. Presenta unos surcos más estrechos que la forma B con 9.3 residuos por vuelta de hélice y un ángulo de *twist* superior al B-ADN. (Saenger, 1984). Esta conformación también puede lograrse a través de ácidos nucleicos sintéticos.

El D-ADN se ha observado sólo en ácidos nucleicos sintéticos, no encontrándose de forma natural en la célula. Posee solamente ocho residuos por vuelta de hélice y un *tilt* de -16°, además de un surco menor estrecho y profundo.

El T-ADN se ha encontrado en el fago T2, en el que muchas de las citosinas están 5-OH-metiladas y pueden tener glucosilaciones. Es muy parecido al D-ADN, con ocho

residuos por vuelta de hélice. Le caracteriza un surco menor estrecho y profundo, un surco mayor ancho y superficial y un *tilt* es de -6° .

ii) Familia A

Cuando se secan los cristales de B-ADN o cuando se disminuye el contenido en sales del cristal la estilizada molécula del B-ADN se transforma en otra más corta y gruesa, el A-DNA (figura 3.10). A diferencia del B-ADN donde los pares de bases están apilados paralelamente, en esta conformación los pares de bases se encuentran desplazados hacia el saliente mayor. El surco mayor se hace más profundo y estrecho, mientras que el surco menor se hace superficial y ancho. Se ha comprobado que esta forma es favorecida en secuencias ricas en pares C·G. En ella, el número de pares por vuelta pasa de 10.4 a 11, con lo que la hélice se hace más ancha. Esto mismo se refleja en el *twist*, que pasa de valores comprendidos entre $36-38^\circ$ en la familia B a 33° en la familia A, y en el *rise*, que demuestra que las bases en la familia A se encuentran más apiladas. Otra característica diferencial es el empaquetamiento *C3'-endo* de los azúcares, mientras que en la familia B es *C2'-endo*.

- Forma Levógira

Tanto en el A-ADN como en el B-ADN el azúcar y la base se encuentran en lados opuestos respecto al enlace glucosídico (conformación *anti*), en la cual la repulsión estérica es mínima. Sin embargo, en presencia de altas concentraciones de cationes algunos nucleótidos sufren una rotación y adoptan conformaciones *syn*, en las que el azúcar y la base se eclipsan. En este contexto puede darse una conformación del ADN sorprendentemente diferente, el Z-ADN (Z proviene de *zig-zag*, figura 3.10).

La forma Z da lugar a una estructura más larga y estrecha que las anteriores presentando 12 pares de bases por vuelta y una vuelta de hélice de 45.6 \AA , frente a los 34.0 \AA del B-DNA. El diámetro de hélice del Z-ADN es de 18.4 \AA , mientras que en el B-ADN es de 23.7 \AA . Se ha clasificado el Z-ADN en cuatro subfamilias, la Z, Z_I , Z_{II} y Z' , el equilibrio entre las cuáles viene dado por diferentes condiciones ambientales. En

general, el Z-ADN puede ser estabilizado por iones divalentes, por esperminas o por modificaciones químicas (Saenger, 1984).

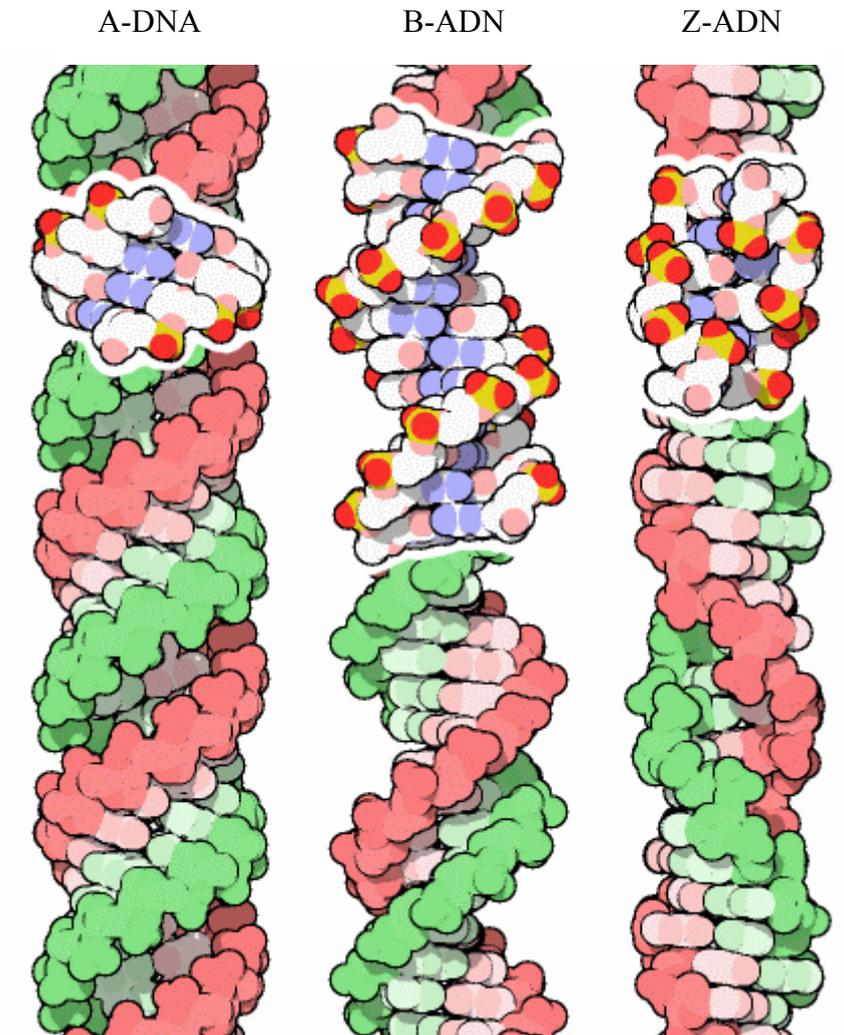


Figura 3.10 Modelos de doble hélice del ácido desoxirribonucleico (ADN). Se observan claramente el surco mayor y el surco menor formado por el esqueleto de azúcares unidos a través de fosfatos. Las bases nitrogenadas aparecen en el interior de la hélice (extraído de www.rcsb.org).

En las células, la mayor parte del ADN se encuentra en la conformación B aunque las regiones ricas en pares de bases GC pueden adoptar la conformación Z (Saenger, 1984). Debido a que las superficies de ambas conformaciones son muy diferentes, la actividad de las proteínas reguladoras que interaccionan con el ADN puede resultar fuertemente afectada.

3.2.3 Estructuras no canónicas de ADN: El *G-cuadrúplex*

Como ya se ha comentado, los ácidos nucleicos poseen una gran variabilidad conformacional y esto les permite no sólo formar estructuras secundarias basadas en la doble hélice antiparalela, sino que las bases pueden emparejarse de muchos modos distintos. De hecho es posible encontrar en la literatura un número elevado de estructuras alternativas en las cuales se observan apareamientos incorrectos entre bases para formar complejos de dos, tres y hasta cuatro cadenas, con frecuentes cambios en la polaridad de las mismas (Blackburn & Gait, 1990). Un ejemplo serían los dúplex de ADN de tipo paralelo donde las bases se aparejarían del modo Watson-Crick reverso (figura 3.8), o los ácidos nucleicos de tres cadenas o *tríplex* de ADN, donde una tercera cadena de nucleótidos interacciona mediante puentes de hidrógeno de tipo *Hoogsteen* a lo largo del surco mayor del ADN. Finalmente, los ácidos nucleicos de cuatro cadenas o *G-cuadrúplexes* están tomando un papel relevante en el campo de la biología molecular debido a su estrecha relación con procesos celulares básicos.

- *ADN G-cuadrúplex*

En 1988 Sen y Gilbert reportaron que una secuencia de ADN rica en guaninas podía formar estructuras estables de cuatro cadenas. Desde entonces, numerosos estudios de RMN y Rayos-X han verificado este tipo de complejos. Estas agrupaciones se conocen como G-cuadrúplex (figura 3.11) y en ellas se forman apilamientos de tétradas de cuatro guaninas que requieren la presencia de cationes en el interior del canal para ser estables.

El G-cuadrúplex puede formarse a partir de una única cadena (G-cuadrúplex intramolecular), de dos o a partir de cuatro cadenas de polinucleótidos (G-cuadrúplex intermolecular). Las secuencias suelen ser del tipo G_nT_n , $G_nT_nG_n$ o G_nA_n . A nivel conformacional estas estructuras presentan un patrón alternado *syn/anti* de los ángulos glucosídicos a lo largo de la cadena, pero también es posible encontrar un patrón todo *anti*. La disposición de las cadenas puede ser paralela o antiparalela, dependiendo de la secuencia que conecta las guaninas o de los cationes que estabilizan la estructura.

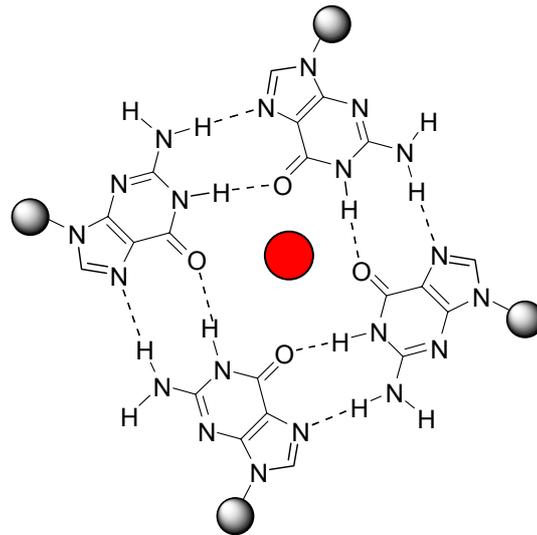


Figura 3.11 Tétrada de guaninas en un *G-cuadrúplex* de ADN. Las cuatro guaninas se emparejan mediante puentes de hidrógeno tipo *Hoogsteen*. La presencia de cationes coordinados en el interior estabiliza el complejo.

La gran estabilidad de los G-cuadrúplexes se justifica por diversos factores. A nivel intrínseco, existen fuertes interacciones de *apilamiento* y puente de hidrógeno entre las bases. Además, los iones presentes en el interior del canal estabilizan el complejo a través de interacciones electrostáticas con los grupos carbonilos de las guaninas. Finalmente la repulsión originada en los fosfatos del esqueleto fosfodiéster es reducida en medio acuoso por efecto del solvente.

A nivel biológico, estas estructuras se presentan con frecuencia en los telómeros de los cromosomas eucariotas, contribuyendo así a su estabilidad durante la replicación y previniendo su degradación. Los telómeros son mantenidos por la enzima *telomerasa*, una transcriptasa reversa específica capaz de añadir patrones repetitivos a la zona 3' de los cromosomas. El acortamiento de estas secuencias se ha relacionado en numerosas ocasiones con el envejecimiento y la muerte celular, mientras que el alargamiento no controlado de las mismas se ha asociado con cáncer. De hecho, la telomerasa está sobreexpresada en tumores (Ambrus et al., 2005; Seenisamy et al., 2004) y se postula que puede estar relacionada con la proliferación celular mediante su contribución a la

conservación de los telómeros. Esta implicación biológica hace que dicha enzima sea motivo de estudio como posible diana anticancerígena. Al parecer, la unión de agentes externos a los G-cuadrúplex impide la unión de la enzima, imposibilitando así que ésta realice su función reparadora.

3.3 Efectos del entorno en la estructura y dinámica del ADN.

La primera estructura cristalográfica de la forma B del ADN generó una serie de cuestiones fundamentales sobre cómo la secuencia afecta a la conformación de los ácidos nucleicos (Beveridge et al., 2004). Desde entonces, numerosos estudios han mostrado que el ADN es una molécula muy flexible, capaz de sufrir tanto cambios conformacionales pequeños a nivel de par de bases como cambios macroscópicos en sus fibras. De entre sus movimientos de corta escala destaca la respiración o *breathing*, que implica la apertura de las bases hacia el exterior durante un breve periodo de tiempo. Esto puede ser clave, por ejemplo, en procesos de intercalación por fármacos o en fenómenos de verificación de lectura y eliminación de apareamientos incorrectos. A nivel macroscópico, el movimiento más destacado es el de curvatura o *bending*, el cuál a su vez puede estar relacionado con niveles de empaquetamiento de orden superior. En ciertos casos esta curvatura es espontánea, mientras que en otras ocasiones es debida a tensión en las fibras causado por el superenrollamiento de las mismas. En todo caso, tradicionalmente se ha supuesto que las propiedades conformacionales del ADN dependientes de secuencia vienen determinadas a nivel de par de base (10 alternativas únicas). Esta hipótesis está actualmente en revisión y se piensa que el modelo de tetrámero (136 posibilidades) es más realista (Beveridge et al., 2004).

3.3.1 El agua

El agua, junto con los iones, es uno de los grandes contribuyentes al mantenimiento de la estructura tridimensional de los ácidos nucleicos, ya que gracias a ella se apantallan las repulsiones de carga entre fosfatos, que de otra manera destrozarían la estructura de la macromolécula.

La solvatación juega un papel muy importante en los cambios conformacionales, como se demuestra en las transiciones desde la forma B (de alta hidratación) a la A (de baja hidratación) (Saenger, 1984). La hidratación alrededor de las dobles hélices de ADN no es homogénea, sino que se diferencian dos *capas de hidratación* (Saenger, 1984). En la *primera capa de hidratación* existen entre once y veinte moléculas de agua por nucleótido. Estas moléculas se agrupan en tres tipos en función de su afinidad decreciente por el fosfato, enlace fosfodiéster más azúcar y base. Dependiendo de la concentración de agua, se observa que la proporción de aguas unidas a las diferentes regiones del ADN varía. La *segunda capa de hidratación* es menos estructurada y en muchas ocasiones indistinguible del resto del medio acuoso por lo que respecta a permeabilidad de iones y movilidad de las aguas, pero es clave para estabilizar la estructura.

El plegamiento de los ácidos nucleicos en estructuras ordenadas diferentes conlleva que algunos átomos queden más expuestos al solvente que otros, lo cuál se puede monitorizar mediante el cálculo de la *superficie accesible al solvente* (SAS). Así, mientras que la SAS de las formas A y B son prácticamente equivalentes cuando se usa el radio de una molécula de agua (1.4 Å), al incrementar el valor del radio (típico de iones hidratados o cadenas laterales de aminoácidos) se observa que los fosfatos de la forma B están más expuestos que los de la A, al contrario de lo que sucede con los carbonos.

3.3.2 Los iones

El ADN es un polielectrolito en condiciones fisiológicas y por ello normalmente está rodeado de iones que neutralizan parcialmente sus cargas. Estos cationes suelen ser inorgánicos e hidratados, pero también es posible encontrar cadenas laterales de aminoácidos o poliaminas.

Los iones inorgánicos presentes en el interior celular en concentraciones micromolares son Mg^{++} , Ca^{++} , Na^+ y K^+ y por ello se hallan normalmente unidos al ADN. En general, van a tener propiedades estabilizadoras globales, aunque pueden

producir el efecto contrario si sus concentraciones son muy elevadas. Los iones son capaces de discriminar la estructura tridimensional del ADN a través del potencial electrostático molecular y del entorno de puentes de hidrógeno. Atraídos por la carga del fosfato, los cationes están normalmente ligados a sus oxígenos, aunque también pueden asociarse con los átomos de las bases. Mientras que los iones monovalentes interactúan de manera poco específica con la secuencia, los divalentes utilizan las aguas coordinadas para favorecer zonas específicas de interacción con el ácido nucleico. Esta relativa especificidad de secuencia hace que los iones divalentes afecten en mayor medida a la estructura de la molécula, pudiendo producir cambios en el *bending* del ADN, tal y como se ha visto para Ca^{++} y Mg^{++} en algunos cristales (Dickerson, 1998).

Estudios sobre diferentes estructuras cristalográficas del dodecámero de Dickerson d(CGCGAATTCGCG) han puesto de manifiesto que el surco menor del fragmento central AATT es ligeramente más estrecho que el resto, asociándose este fenómeno local a rigidez conformacional y curvatura del ADN. Para explicar este suceso existen dos teorías, la primera de ellas propone que la información que define la estructura esta implícita en la secuencia, mientras que la segunda propone que son las interacciones con el solvente las responsables de la menor amplitud del surco. En el caso de esta última, la amplitud del surco se ha relacionado con la existencia de una *espina de hidratación* en el interior o incluso con la posible presencia de iones *estructurales* en el interior del surco estrecho. La influencia de los iones en la conformación del ADN es una cuestión abierta, tal y como muestran diferentes revisiones sobre el campo (Auffinger & Bielecki, 2004; Bonvin, 2000; Chandrasekaran et al., 1995; Denisov & Halle, 2000; Egli, 2002; Feig & Pettitt, 1999; Gearheart et al., 2003; Hamelberg et al., 2000; Hamelberg et al., 2001; Hud & Polak, 2001; Korolev et al., 2002; Laughton et al., 1995; Madhumalar & Bansal, 2003; McConnell & Beveridge, 2000; McConnell & Beveridge, 2001; Mocci & Saba, 2003; Ponomarev et al., 2004; Shui et al., 1998; Sines et al., 2000; Stefl & Koca, 2000; Stellwagen et al., 2001; Subirana & Soler-López, 2003; Subramanian et al., 1988; Varnai & Zakrzewska, 2004; Young et al., 1997).

3.3.3 Interacción del ADN con ligandos: *minor groove binders*

Durante las dos décadas pasadas las técnicas de biología estructural han generado información precisa sobre cómo diferentes moléculas son capaces de unirse específicamente a los ácidos nucleicos. En el caso de las proteínas, los primeros estudios sobre represores de fagos mostraban un motivo de unión relativamente simple, donde dímeros de proteína posicionaban sus hélices alfa de manera que éstas encajasen entre dos surcos anchos consecutivos de la doble hebra. Esta unión apenas generaba modificación de las estructuras originales del complejo. A medida que se han obtenido más estructuras de complejos ADN-proteína, se ha visto que los modos de reconocimiento son muchos y diferentes. Por ejemplo, algunas proteínas distorsionan el ácido nucleico para su correcto reconocimiento, a partir de uno o los dos surcos. Otras sufren un proceso de estructuración después de la unión al ADN y algunas son incluso capaces de combinar partes rígidas con colas flexibles que tienen la capacidad de entrelazarse al ADN.

Por lo que respecta a ligandos químicos de pequeño tamaño, también existen modelos de reconocimiento a través de ambos surcos, la mayoría de ellos reversible. En el caso de que exista especificidad de secuencia, suele haber preferencia por el surco estrecho con respecto al ancho. Como sucede con las proteínas, en algunos casos el ligando se une de manera rígida sin modificar la geometría del ADN mientras que en otras es capaz de distorsionarlo, pero normalmente no de la misma manera que las proteínas.

Existen dos tipos principales de modos fundamentales por los que los ligandos químicos pueden unirse de manera no covalente al ADN: la *intercalación* y la *unión al surco estrecho*.

Los *intercaladores* poseen en su estructura química diversos anillos aromáticos que les brindan la posibilidad de introducirse entre dos bases adyacentes de la hélice de ADN. La unión se realiza principalmente por interacciones asociativas de *apilamiento* que pueden complementarse con interacciones de largo alcance del surco estrecho. En el

transcurso de esta unión, los intercaladores suelen generar fuertes perturbaciones en el esqueleto del ácido nucleico debido al movimiento que se genera en las bases para permitir su intercalación. Los intercaladores suelen tener preferencia por zonas citosina-guanina (o guanina-citosina) frente a zonas adenina-timina. Tienen además preferencia en general por pasos d(5'CG3') frente a pasos d(5'GC3'), aunque las razones fisicoquímicas que se esconden detrás de estas preferencias no están claras. En cualquier caso, y aún cuando la preferencia es por CG (o GC) se pueden dar intercalaciones en cualquier otra secuencia, especialmente si el intercalador cuenta con un brazo que se dispone a lo largo del surco estrecho.

Los intercaladores presentan especificidad limitada de secuencia debido a su dificultad para reconocer más de un par de bases. Por ello normalmente se le aplican modificaciones químicas en la estructura con el fin de ampliar su capacidad de reconocimiento. En cualquier caso, la especificidad conseguida parece ser capaz de interferir en el reconocimiento específico entre, por ejemplo, ADN y factores de transcripción y ejercer con ello acciones terapéuticas como es el caso de la *nogalomicina* (Wemmer, 2000).

Diferentes características hacen al surco estrecho del B-ADN idóneo para el reconocimiento específico por parte de fármacos. Se encuentran átomos dadores y aceptores de puente de hidrógeno, variaciones en el potencial electrostático molecular, cambios en las interacciones de van der Waals y en el perfil de interacción con el agua. Los ligandos de unión al surco menor o *minor groove binders* poseerán pues unas características químicas únicas que les permitirán aprovechar las características que les brinda este surco del ADN (figuras 3.12 y 3.13). Estos compuestos poseen anillos aromáticos capaces de adaptarse a la curvatura del surco estrecho, cargas positivas en los extremos que les procuran afinidad hacia el túnel electronegativo en el interior del surco y además muchos de ellos poseen dadores o aceptores de puente de hidrógeno complementarios a los que las bases del ADN disponen por su surco estrecho. Su unión al ADN apenas causa distorsión en el esqueleto fosfodiéster, aunque en ocasiones son capaces de estabilizar la estructura de B-ADN a través de su rigidificación (Harris et al., 2001; Wemmer, 2000). Los *minor groove binders* tienen una pauta de lectura más larga

que los intercaladores y por ello son capaces de reconocer un número mayor de pares de bases, obteniendo una mayor especificidad de secuencia. La mayoría de los *minor groove binders* se unen de manera selectiva a las secuencias ricas en AT, donde la amplitud del surco suele ser menor, lo que favorece la inserción de los anillos aromáticos tanto por interacciones de van der Waals como electrostáticas. La interacción se estabiliza con la formación de puentes de hidrógeno con grupos aceptores, como el átomo O2 de la Timina y N3 de la Adenina, por la interacción con alguna molécula de agua y por el término entrópico ligado a la desestructuración de la espina de hidratación (Wemmer, 2000).

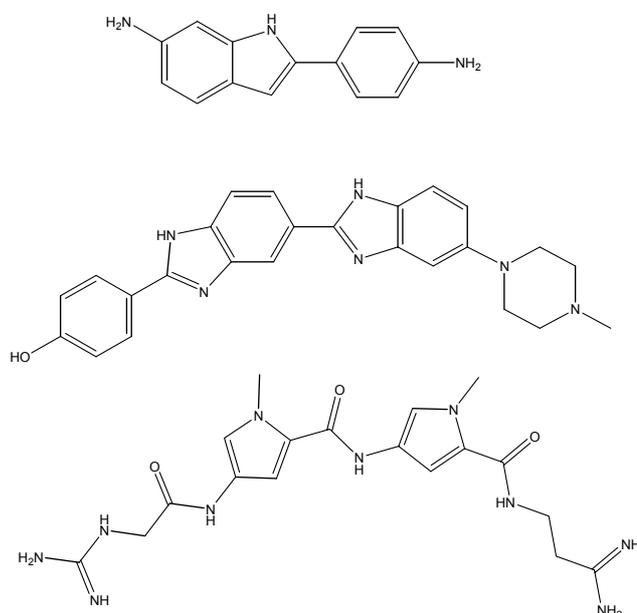


Figura 3.12 Estructura química de tres conocidos *minor groove binders*. De arriba hacia abajo; DAPI, Hoechst-33258 y Netropsina.

Finalmente, cabe señalar que el surco estrecho en regiones AT es tan buen receptor para este tipo de compuestos que es posible la inserción de dos de ellos en *tándem*, con lo que se consiguen complejos muy estables y específicos (Wemmer, 2000).

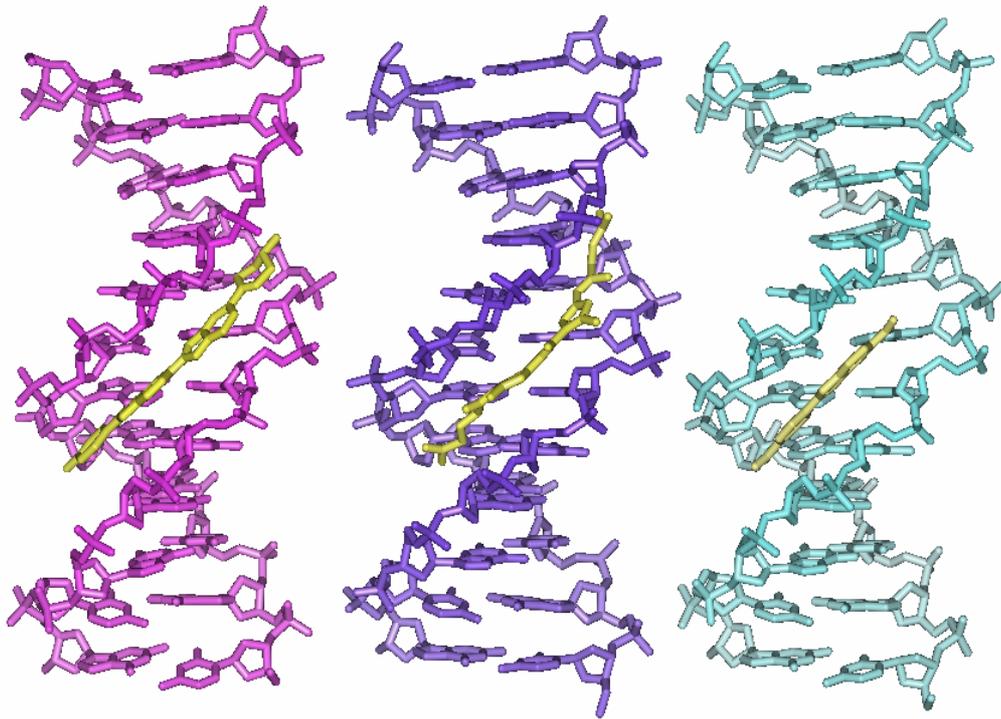


Figura 3.13 Estructura cristalográfica de un dodecámero de Dickerson unido a tres *minor groove binders*. De izquierda a derecha; Hoechst-33258 (1d43), Netropsin (1d86) y DAPI (1d30).

Muchos *minor groove binders* actúan como fármacos. Por ejemplo, el berenilo es un fármaco antiparasitario, la pentamidina y el bisguanilfenilfurano son activos frente el patógeno asociado al SIDA, mientras que la netropsina y la distanomicina son conocidos fármacos antitumorales (Wemmer, 2000). Los mecanismos de acción a nivel molecular suelen estar relacionados con la inhibición de la expresión de ciertos genes, pero distan mucho de estar descritos en detalle.

3.4 Métodos experimentales para el estudio de los Ácidos Nucleicos en fase gas

Las estructuras tridimensionales de los nucleósidos, nucleótidos y ácidos nucleicos en solución han sido elucidadas, principalmente, por las técnicas de cristalografía de Rayos-X y la Resonancia Magnética Nuclear. Ambos métodos han sido muy potentes para describir la estructura de los ácidos nucleicos en medios altamente hidratados, pero son inadecuados para interpretar los datos en fase gas proporcionados por las modernas técnicas de espectrometría de masas por ionización suave, como son la ionización por electrospray (ESI-MS) y la desorción/ionización asistida por láser a partir de matrices (MALDI) (Hoffmann & Stroobant, 2003; Larsen & McEwen, 1998).

Las técnicas de espectrometría de masas permiten determinar la identidad y abundancia de diferentes especies moleculares de manera directa a través de sus masas, las cuáles sirven de *etiquetas* moleculares (Hofstadler & Griffey, 2001). El primer paso en un análisis de este tipo es la generación de *iones* (moléculas cargadas), a partir de una muestra original, los cuáles podrán ser posteriormente detectados de manera inequívoca por su relación *masa-carga* (m/z) mediante un espectrómetro de masas. Existen diferentes métodos para la generación de los iones. En los casos donde las moléculas puedan ser vaporizadas sin descomposición se suele elegir la técnica de *electron ionization* (*EI*). En ella se bombardea la muestra con un flujo de electrones de alta energía, los cuales son capaces, además de ionizar, de fragmentar los enlaces y generar nuevas especies. Una alternativa menos agresiva en términos energéticos sería la ionización por métodos químicos o *chemical ionization* (*CI*). Esta técnica está basada en la generación de iones mediante colisiones de la muestra con iones presentes en el medio. En ella, la muestra es expuesta a un exceso de gas ionizante, el cual le podrá ceder protones. También es posible la generación de iones negativos mediante la transferencia de protones al gas o por adición de electrones (Hoffmann & Stroobant, 2003). Tanto en el caso de *EI* como *CI* se requiere que las moléculas sujetas a estudio estén en fase gas, por lo que su utilidad está limitada a compuestos volátiles o a muestras con posibilidad de ser volatilizadas.

Con la llegada de la ionización por *Field Desorption (FD)* se consiguió la ionización de pequeñas moléculas no volátiles, aunque la técnica presentaba la limitación que la masa de las mismas debía ser inferior a 5 kDa (kilodaltons) (Hoffmann & Stroobant, 2003). La ionización por desorción describe el proceso por el cual una molécula es a la vez ionizada y evaporada de una superficie. El desarrollo de los métodos de ionización por desorción basados en la emisión de iones pre-existentes en un líquido o una superficie sólida, como *Plasma Desorption (PD)*, *Fast Atom Bombardement (FAB)* y *Laser Desorption* (ver Tabla 3.3), permitió el acceso de las técnicas espectrométricas al estudio de biomoléculas. El problema no fue entonces la generación de iones, sino el análisis de iones de carga única de elevado peso molecular (Hoffmann & Stroobant, 2003).

Al inicio de los noventa, irrumpieron dos nuevos métodos de ionización especialmente adecuados para el estudio de biomoléculas: la desorción/ionización asistida por láser a partir de matrices (*matrix-assisted laser desorption/ionization; MALDI*) y la ionización por electrospray (*electrospray ionization mass spectrometry; ESI-MS*).

En la primera, las matrices dispersan la muestra previniendo la fragmentación del compuesto y potenciando la ionización del mismo. La técnica genera iones predominantemente cargados de manera única donde la señal suele depender de la morfología de la molécula (Jarrold, 2000). Por el contrario, la técnica de ESI-MS genera iones con diferentes estados de carga, los cuáles pueden ser fácilmente diferenciados por su relación masa-carga (m/z). Este hecho, junto con su gran sensibilidad, la hace idónea para el estudio de complejos no covalentes de biomoléculas cargadas, como es el caso de los ácidos nucleicos y proteínas.

Tal y como se esquematiza en la figura 3.14, en la técnica de ESI una muestra en solución se hace pasar por un capilar minúsculo movido por un electrodo a varios kilovoltios. El alto campo eléctrico al final del capilar empuja a la solución a un cono de Taylor. Si el campo es suficientemente fuerte para generar tensión superficial, el cono es llevado a un filamento que genera pequeñas gotas (*droplets*) cargadas. La generación

de estas gotas es un proceso que aún hoy está en debate. En cualquier caso, estas gotas cargadas son transferidas de presión ambiental a las condiciones de vacío y alta temperatura, lo que promueve su deshidratación (Jarrold, 2000). En el caso de los ácidos nucleicos, este proceso genera un rango de iones cargados preferentemente de manera negativa, con estados de carga inferiores a los del ácido nucleico en solución. (Hofstadler & Griffey, 2001). Las abundancias relativas de estos estados de carga podrán ser detectadas fácilmente mediante el espectrómetro de masas (véase figura 3.15), aunque la localización exacta de los sitios de protonación no es conocida (Gidden et al., 2004). En este sentido, estudios espectrométricos en dúplexes apuntan a que la carga se reparte uniformemente (entre fosfatos adyacentes) a lo largo de ambas cadenas (Gabelica & Pauw, 2001; Gabelica et al., 2000), mientras que datos obtenidos para cadenas sencillas indican dispersión de las cargas a lo largo de los grupos fosfato del esqueleto (Hoaglund et al., 1997).

Tabla 3.3 Métodos espectrométricos de ionización usados en el estudio de oligonucleótidos

| Técnica | Medio de ionización |
|---|--|
| Fast Atom Bombardment (FAB) | Impacto de átomos a gran velocidad en una muestra disuelta en una matriz líquida. |
| Secondary Ion MS (SIMS) | Impacto de iones alta velocidad en una pequeña película de muestra en sustrato metálico o líquido. |
| Plasma Desorption | Impacto de fragmentos de fisión nuclear (^{235}Cf) en una muestra sólida depositada en metal. |
| Matrix Assisted Laser Desorption Ionization (MALDI) | Impacto de fotones de alta energía en una muestra embebida en una matriz orgánica sólida. |
| Field Desorption | Imposición de un gradiente de campo eléctrico elevado en una muestra depositada en soporte sólido. |
| Electrospray | Formación de <i>droplets</i> líquidos cargados, a partir de los cuales son desolvatados los iones. |

El estudio de oligonucleótidos por espectrometría de masas presenta algunos inconvenientes. El primero de ellos es que se pueden formar fuertes enlaces entre los grupos fosfodiéster e iones alcalinos presentes en la solución original donde se sumerge la muestra. Esto puede producir que se generen complejos que posteriormente

dificultarán la correcta interpretación de los resultados. Por ello, normalmente se suelen introducir iones amonio (NH_4^+) en el medio en vez de iones sodio o potasio. Una segunda dificultad radica en el hecho de que los oligonucleótidos se suelen fragmentar con facilidad, siendo esto aparentemente problemático cuando se utiliza la técnica de MALDI para la ionización. Por último, a medida que se incrementa el tamaño del ácido nucleico, más difícil resulta la ionización y por tanto su detección, lo que limita el estudio a oligonucleótidos de tamaño pequeño o medio (Hoffmann & Stroobant, 2003).

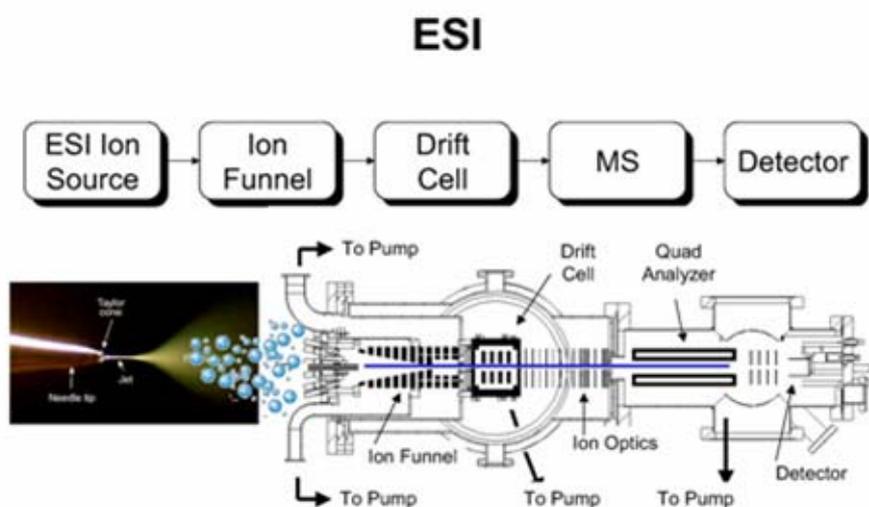


Figura 3.14 Esquema básico del funcionamiento de la técnica de *ionización por electrospray* (ESI-MS).

La técnica de ESI-MS ha sido muy aplicada para la determinación estequiométrica de interacciones entre proteínas y ADN, proteínas y ARN, estructuras no covalentes de ADN como dúplexes, tríplexes, G-cuadrúplexes y ADN unido a ligandos (Hanson & Robinson, 2004; Hofstadler & Griffey, 2001). No obstante, asociada a estos estudios existe la cuestión de hasta qué punto las estructuras detectadas en fase gas son representativas de las estructuras en solución o son simplemente artefactos creados durante la evaporación del solvente. Este tema ha suscitado un notable interés en los últimos años, desarrollándose técnicas que permiten obtener información estructural

indirecta sobre los ácidos nucleicos en fase gas. Entre ellas, es posible incluir las basadas en métodos químicos y las basadas en medidas de tamaño (Jarrold, 2000). En las primeras se encuentran normalmente acopladas dos fases de espectrometría de masas o lo que se conoce como (*MS/MS*). Dentro de este grupo existen las técnicas basadas en disociación como son la *disociación inducida por colisión* (CID), que informa sobre estabilidades cinéticas, la *disociación radiativa por infrarrojos de cuerpo negro* (BIRD), que mide energías de activación, y las basadas en intercambios atómicos como el *recambio Hidrógeno/Deuterio*, que permite relacionar velocidades de intercambio atómicas con estructura secundaria (Gidden et al., 2005).

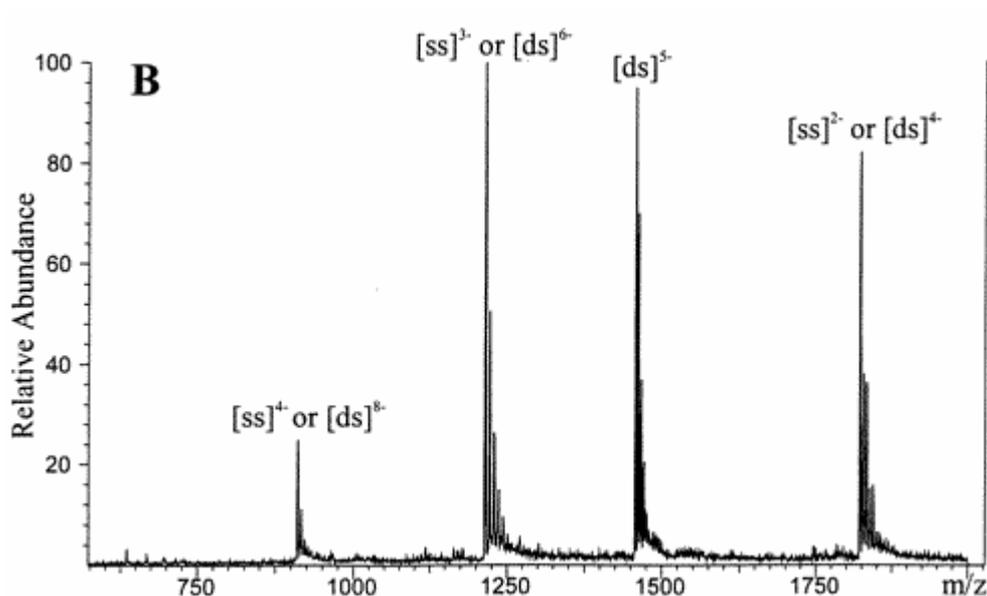


Figura 3.15 Espectro obtenido por ESI-MS sobre un dúplex de ADN de 12 pares de bases d(CGCGAATTCGCG)₂. En abscisas se encuentra la relación masa-carga (m/z) y en ordenadas la abundancia relativa (Wan et al., 2000).

Los métodos físicos son una alternativa a los métodos químicos para la obtención de evidencias estructurales en fase gas. La mayoría de ellos están basados en medidas de tamaño, siendo el más usado el de *movilidad de iones* (Gabelica et al., 2005). La *movilidad iónica* (K) es una medida experimental que genera información sobre la velocidad de desplazamiento de un ión a través de un tampón gaseoso bajo la influencia de un campo eléctrico débil. La velocidad de desplazamiento del ión dependerá del tamaño de la muestra (y por tanto de las colisiones con el gas del medio), de la

intensidad del campo y de la fuerza de fricción. A nivel experimental, la movilidad iónica se deriva a partir de las *distribuciones de tiempo de llegada* de los iones (ATD).

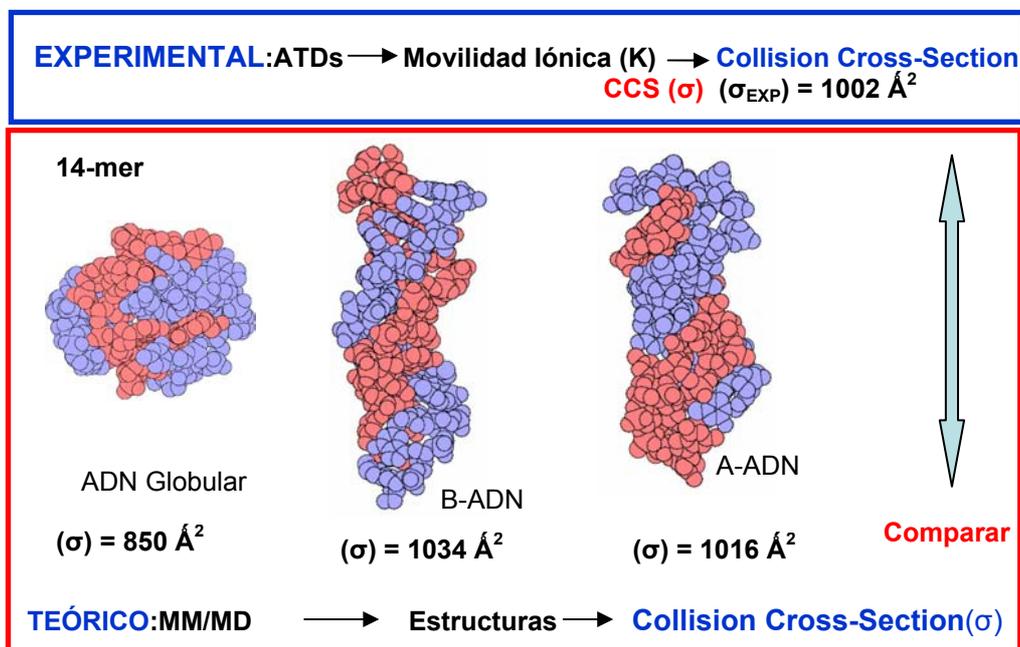


Figura 3.16 Esquema de la comparación entre métodos físicos basado en *movilidad iónica* y su relación con el parámetro *CCS (σ)*. En la figura se muestran tanto valores teóricos para diferentes conformaciones de un 14-mer (globular, formas B y A) como el valor experimental.

La movilidad iónica es fuertemente dependiente de la morfología de la molécula, por lo que se pueden establecer relaciones empíricas entre ellas. Por ejemplo, a partir de la movilidad es posible derivar el factor de sección de colisión (*CCS, Collision Cross Section*), un parámetro de baja resolución que genera información sobre el perfil general de la molécula, pero que tiene la ventaja de que puede ser directamente comparado con cálculos teóricos (véase figura 3.16) (Jarrold, 2000). De hecho es posible calcular la *CCS* teóricamente a partir de un conjunto de coordenadas conocidas (Helden et al., 1993; Mesleh et al., 1996; Shvartsburg & Jarrold, 1996; von Helden et al., 1991; Wyttenbach et al., 1997). Los dos algoritmos más utilizados para el cálculo de la *CCS* son los desarrollados por los grupos de Bowers y Jarrold. El método del grupo de Jarrold se basa en evaluar las colisiones entre varias orientaciones de la molécula y los átomos de un gas en una trayectoria mediante un potencial de interacción simple. En el

caso del grupo de Bowers, la CCS se obtiene a partir de un proceso iterativo donde se evalúan contactos entre puntos de una malla bidimensional (correspondientes al gas) y diferentes orientaciones de la molécula respecto su centro de masas. En nuestra experiencia, ambos métodos proponen resultados similares, siendo el método de Bowers mucho más eficiente en términos computacionales.

3.5 Los ácidos nucleicos en fase gas

Ha sido la norma pensar que el solvente juega un papel esencial en la estabilidad de las estructuras secundarias de los ácidos nucleicos debido a que, entre otras razones, contribuye a disminuir la intensa repulsión intra e intercatenaria que generan los fosfatos. La mayoría de las aportaciones experimentales sobre caracterización estructural de ácidos nucleicos provienen de datos en solución o de fase condensada, y de hecho, la obtención de datos sobre estabilidad de polinucleótidos u otras biomoléculas en condiciones de *fase gas* es difícil, debido en gran medida al desconocimiento por parte de la comunidad científica de los sucesos que se originan en las biomoléculas bajo esas condiciones extremas.

En el caso de bases aisladas, es posible estudiar experimentalmente sus interacciones en *fase gas* mediante métodos espectroscópicos. En uno de los primeros estudios realizados con esta técnica (Yanson et al., 1979), se consiguieron determinar con éxito entalpías para algunos pares de bases. En ellos sin embargo, no se mostraba ninguna información estructural de cómo eran estas interacciones. Algunos estudios teóricos posteriores sugieren que pueden coexistir diferentes poblaciones de estructuras, donde con frecuencia se encuentran bases apiladas, junto con toda una gama de dímeros estabilizados por puentes de hidrógeno (Kabelac et al., 2000; Kratochvil et al., 2000).

Para profundizar en el estudio de la naturaleza intrínseca de las interacciones entre bases aisladas se pueden utilizar métodos teóricos basados en la mecánica cuántica. Estas técnicas poseen la ventaja que no se necesita un conocimiento previo del sistema, poseyendo en muchos casos una precisión comparable a las aproximaciones experimentales. En algunos de estos estudios se ha demostrado que el equilibrio entre

dímeros apilados y formadores de puente de hidrógeno depende del tipo de nucleobase y del estado de micro-hidratación, ya que incluso unas pocas moléculas de agua pueden desestabilizar los puentes de hidrógeno entre pares de bases (Sponer et al., 2001; Sponer et al., 2002).

El avance en los métodos de espectrometría de masas ha hecho también posible el estudio de algunas propiedades de los ácidos nucleicos en condiciones de *vacío*. Haciendo uso de estas técnicas, diversos autores (Ganem et al., 1993; Light-Wahl et al., 1993) postularon en el año 1993 que los dúplexes de ADN podían ser transferidos de manera intacta de solución a fase gas, usando ESI-MS. Desde entonces, una gran cantidad de artículos y revisiones han sido publicados describiendo la utilidad de estas técnicas para describir propiedades del ADN en estas condiciones. El tipo de resultados derivados de estos análisis van desde la composición de bases a la identificación de sitios de unión por ligandos no covalentes o a la caracterización de modificaciones post-traduccionales (Hofstadler & Grifffey, 2001). Es importante destacar, no obstante, que la *conservación de la estructura* de los complejos en fase gas *no es necesaria* para el recuento de las abundancias de iones en los espectros de ESI-MS, siempre y cuando *las cadenas permanezcan unidas de alguna manera* en el transcurso del experimento. Consecuentemente, las buenas correlaciones entre los datos en solución y los extraídos en fase gas no pueden ser considerados una prueba de que las estructuras en vacío son las mismas que en fase condensada (Gabelica et al., 2005). De hecho, el desconocimiento sobre cuál es la estructura real de los ácidos nucleicos en fase gas es muy grande (Gidden et al., 2005). En este contexto, algunos autores han intentado recientemente obtener información experimental indirecta. Por ejemplo, Gabelita y De Pauw realizaron experimentos de colisión inducida por disociación (CID) para examinar una serie de dúplexes de dieciséis pares de bases (Gabelica & De Pauw, 2002; Gabelica & Pauw, 2001) observando que las estabildades relativas de las estructuras en fase gas correlacionaban bien con las encontradas en solución. Ellos atribuían esos resultados a la retención de los puentes de hidrógeno y a las interacciones de apilamiento después de la transferencia a fase gas. Schnier et al. (Schnier et al., 1998) estudiaron la cinética de disociación de diferentes dúplexes complementarios y no complementarios usando disociación radiativa por infrarrojos de cuerpo negro (BIRD) y observaron que las

energías de activación para la disociación de los dúplexes complementarios eran más altas que para los no complementarios. Estos datos apuntaban en la dirección de que algunos pares WC podían quedar retenidos en la transferencia a fase gas. Diversos autores han examinado la fragmentación de dúplexes con apareamientos incorrectos a través de experimentos CID, observando fragmentación preferente en los apareamientos erróneos. Este hecho sugiere que los pares Watson y Crick se mantienen en el resto del dúplex (Gabelica & De, 2002; Griffey et al., 1999). Se han realizado también algunos estudios de ESI-MS de dobles hebras de ADN unidas a ligandos de manera no covalente sugiriendo que los compuestos ternarios sobreviven a la transferencia a fase gas y con abundancias relativas de los mismos similares a las encontradas en solución (Gabelica et al., 1999; Gabelica et al., 2000; Guittat et al., 2003; Rosu et al., 2002b; Wan et al., 2000). Estos descubrimientos son muy sorprendentes, especialmente en el caso de los *minor groove binders*, dado que su interacción con el ADN es fuertemente dependiente de la conformación de ADN.

Además de con dobles hebras de ADN, la ESI-MS se ha utilizado para estudiar tríplexes y G-cuadrúplexes (Hofstadler & Griffey, 2001). En estos casos, la concentración de carga negativa es muy elevada lo que debería inducir la disrupción completa de la estructura después de su transferencia a fase gas. No obstante, la mayoría de los estudios indican que existe conservación de las propiedades de solución, siendo más evidente en el caso de los G-cuadrúplex (Baker et al., 2005; Gabelica et al., 2005; Goodlett et al., 1993; Rosu et al., 2002a; Rosu et al., 2003; Vairamani & Gross, 2003).

3.6 Referencias bibliográficas

- Alberts, B., Bray, D., Johnson, A., Lewis, J., Raff, M., Roberts, K. & Walter, P. (1999). *Biología Molecular de la Célula*. 3rd edit, Ediciones Omega, Barcelona.
- Ambrus, A., Chen, D., Dai, J., Jones, R. A. & Yang, D. (2005). Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. *Biochemistry* 44, 2048-2058.
- Auffinger, P. & Bielecki, L. (2004). Anion Binding to Nucleic Acids. *Structure* 12, 379-388.
- Baker, E. S., Bernstein, S. L. & Bowers, M. T. (2005). Structural Characterization of G-Quadruplexes in Deoxyguanosine Clusters Using Ion Mobility Mass Spectrometry. *Journal of the American Society of Mass Spectrometry* 16, 989-997.
- Beveridge, D. L., Barreiro, G., Byun, K. S., Case, D. A., Cheatham, T. E., 3rd, Dixit, S. B., Giudice, E., Lankas, F., Lavery, R., Maddocks, J. H., Osman, R., Seibert, E., Sklenar, H., Stoll, G., Thayer, K. M., Varnai, P. & Young, M. A. (2004). Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. Research design and results on d(CpG) steps. *Biophys J* 87, 3799-3813.
- Blackburn, G. M. & Gait, M. J. (1990). *Nucleic Acids in Chemistry and Biology*, IRL Press., New York.
- Bloomfield, V. A., Crothers, D. M. & Tinoco, I. (2000). *Nucleic Acids. Structures, Properties and Functions*, University Science Books, Sausalito.

- Bonvin, A. M. J. J. (2000). Localization and dynamics of sodium counterions around DNA in solution from molecular dynamics simulation. *European Biophysics Journal* 29, 57-60.
- Chandrasekaran, R., Radha, A. & Park, H. S. (1995). Sodium ions and water molecules in the structure of poly(dA)-poly(dT). *Acta Crystallographica* D51, 1025-1035.
- Colominas, C., Luque, F. J. & Orozco, M. (1996). Tautomerism and Protonation of Guanine and Cytosine. Implications in the Formation of Hydrogen-Bonded Complexes. *Journal of the American Chemical Society* 118, 6811-6821.
- Denisov, V. & Halle, B. (2000). Sequence-specific binding to counterions to B-DNA. *Proc Natl Acad Sci U S A* 97, 629-633.
- Dickerson, R. E. (1989). Definitions and nomenclature of nucleic acid structure components. *Nucleic Acids Res* 17, 1797-1803.
- Dickerson, R. E. (1998). DNA bending: the prevalence of kinkiness and the virtues of normality. *Nucleic Acids Res* 26, 1906-1926.
- Egli, M. (2002). DNA-Cation Interactions: Quo Vadis? *Chemistry and Biology* 9, 277-286.
- Feig, M. & Pettitt, B. M. (1999). Sodium and chlorine ions as part of the DNA solvation shell. *Biophysical Journal* 77, 1769-1781.
- Gabelica, V. & De Pauw, E. (2002). Collision-induced dissociation of 16-mer DNA duplexes with various sequences: evidence for conservation of the double helix conformation in the gas phase. *International Journal of Mass Spectrometry* 219, 151-159.

- Gabelica, V., De Pauw, E. & Rosu, F. (1999). Interaction between antitumor drugs and a double-stranded oligonucleotide studied by electrospray ionization mass spectrometry. *J Mass Spectrom* 34, 1328-1337.
- Gabelica, V. & De, P. E. (2002). Comparison of the collision-induced dissociation of duplex DNA at different collision regimes: evidence for a multistep dissociation mechanism. *J Am Soc Mass Spectrom* 13, 91-98.
- Gabelica, V. & Pauw, E. D. (2001). Comparison between solution-phase stability and gas-phase kinetic stability of oligodeoxynucleotide duplexes. *J Mass Spectrom* 36, 397-402.
- Gabelica, V., Rosu, F., Houssier, C. & De Pauw, E. (2000). Gas phase thermal denaturation of an oligonucleotide duplex and its complexes with minor groove binders. *Rapid Commun Mass Spectrom* 14, 464-467.
- Gabelica, V., Rosu, F., Witt, M., Baykut, G. & De Pauw, E. (2005). Fast gas-phase hydrogen/deuterium exchange observed for a DNA G-quadruplex. *Rapid Commun Mass Spectrom* 19, 201-208.
- Ganem, B., Yu-Tsyr. Li, Hsieh, Y.-L. & Henion, J. (1993). Detection of oligonucleotide duplex forms by ion spray mass spectrometry. *Tetrahedron Letters* 34, 1445-1448.
- Gearheart, L. A., Somoza, M. M., Rivers, W. E., Murphy, C. J., Coleman, R. S. & Berg, M. A. (2003). Sodium-Ion Binding to DNA: Detection by Ultrafast Time-Resolved Stokes-Shift Spectroscopy. *Journal of the American Chemical Society* 125, 11812-11813.

- Gidden, J., Baker, E. S., Ferzoco, A. & Bowers, M. T. (2005). Structural motifs of DNA complexes in the gas phase. *International Journal of Mass Spectrometry* 240, 183-193.
- Gidden, J., Ferzoco, A., Baker, E. S. & Bowers, M. T. (2004). Duplex formation and the onset of helicity in poly d(CG)_n oligonucleotides in a solvent-free environment. *J Am Chem Soc* 126, 15132-15140.
- Goodlett, D. R., Camp, D. G., Hardin, C. C., Corregan, M. & Smith, D. (1993). Direct observation of a DNA quadruplex by electrospray ionization mass spectrometry. *Biol. Mass Spectrom* 22, 181-183.
- Griffey, R. H., Grieg, M. J., An, H., Sasmor, H. & Manalili, S. (1999). Targeted Site-Specific Gas-Phase Cleavage of Oligoribonucleotides. Application in Mass Spectrometry-Based Identification of Ligand Binding Sites. *Journal of the American Chemical Society* 121, 474-475.
- Guittat, L., Alberti, P., Rosu, F., Van Miert, S., Thetiot, E., Pieters, L., Gabelica, V., De Pauw, E., Ottaviani, A., Riou, J. F. & Mergny, J. L. (2003). Interactions of cryptolepine and neocryptolepine with unusual DNA structures. *Biochimie* 85, 535-547.
- Hamelberg, D., McFail-Isom, L., Williams, L. D. & Wilson, W. D. (2000). Flexible Structure of DNA: Ion Dependence of Minor-Groove Structure and Dynamics. *Journal of the American Chemical Society* 122, 10513-10520.
- Hamelberg, D., Williams, L. D. & Wilson, W. D. (2001). Influence of the Dynamic Positions of Cations on the Structure of the DNA Minor Groove: Sequence-Dependent Effects. *Journal of the American Chemical Society* 123, 7745-7755.

- Hanson, C. L. & Robinson, C. (2004). Protein-Nucleic Acid Interactions and the Expanding Role of Mass Spectrometry. *Journal of Biological Chemistry* 279, 24907-24910.
- Harris, S. A., Gavathiotis, E., Searle, M. S., Orozco, M. & Laughton, C. A. (2001). Cooperativity in drug-DNA recognition: a molecular dynamics study. *J Am Chem Soc* 123, 12658-12663.
- Helden, G. V., M-T., H., Gotts, N. & Bowers, M. T. (1993). Carbon Cluster Cations with up to 84 atoms: Structures, Formation Mechanism and Reactivity. *Journal of Physical Chemistry* 97, 8182-8192.
- Hernandez, B., Soliva, R., Luque, F. J. & Orozco, M. (2000). Misincorporation of 2'-deoxyoxanosine into DNA: a molecular basis for NO-induced mutagenesis derived from theoretical calculations. *Nucleic Acids Res* 28, 4873-4883.
- Hoaglund, C. S., Liu, Y., Ellington, A. D., Pagel, M. & Clemmer, D. E. (1997). Gas-Phase DNA: Oligothymidine Ion Conformers. *Journal of the American Chemical Society* 119, 9051-9052.
- Hoffmann, E. d. & Stroobant, V. (2003). *Mass Spectrometry: Principles and Applications*. Second edit (Sons, J. W., Ed.), John Wiley & Sons, New York.
- Hofstadler, S. A. & Griffey, R. H. (2001). Analysis of noncovalent complexes of DNA and RNA by mass spectrometry. *Chem Rev* 101, 377-390.
- Hud, N. V. & Polak, M. (2001). DNA-cation interactions: the major and minor grooves are flexible ionophores. *Curr Opin Struct Biol* 11, 293-301.
- Jarrold, M. F. (2000). Peptides and proteins in the vapor phase. *Annu Rev Phys Chem* 51, 179-207.

- Kabelac, M., Ryjacek, F. & Hobza, P. (2000). Already Two Water Molecules Change Planar H-bonded Structure of the Adenine...Thymine Base Pair to the Stacked Ones: a Molecular Dynamics Simulation Study. *Physical Chemistry Chemical Physics* 2, 4906-4909.
- Korolev, N., Lyubartsev, A. P., Laaksonen, A. & Nordenskiöld, L. (2002). On the competition between water, sodium ions, and spermine in binding to DNA: a molecular dynamics computer simulation study. *Biophysical Journal* 82, 2860-2875.
- Kratochvil, M., Spomer, J. & Hobza, P. (2000). Global Minimum of the Adenine...Thymine Base Pair Corresponds Neither to Watson-Crick Nor to Hoogsteen Structures. Molecular Dynamic/Quenching/AMBER and ab Initio beyond Hartree-Fock Studies. *Journal of the American Chemical Society* 122, 3495-3499.
- Larsen, B. S. & McEwen, C. N. (1998). *Mass Spectrometry of Biological Materials*. Second edit (Marcel Dekker, I., Ed.), Marcel Dekker, Inc., New York.
- Laughton, C. A., Luque, F. J. & Orozco, M. (1995). Counterion Distribution around DNA Studied by Molecular Dynamics and Quantum Mechanical Simulations. *Journal of Physical Chemistry* 99, 11591-11599.
- Light-Wahl, K. J., Springer, D. L., Winger, B. E., Edmonds, C. G., Camp, D. G., Thrall, B. D. & Smith, R. D. (1993). Observation of a small oligonucleotide duplex by electrospray ionization mass spectrometry. *Journal of the American Chemical Society* 115, 803-804.

- Madhumalar, A. & Bansal, M. (2003). Structural insights into the effect of hydration and ions on A-tract DNA: A molecular dynamics study. *Biophysical Journal* 85, 1805-1816.
- McConnell, K. J. & Beveridge, D. L. (2000). DNA structure: what's in charge? *J Mol Biol* 304, 803-820.
- McConnell, K. J. & Beveridge, D. L. (2001). Molecular dynamics simulations of B'-DNA: sequence effects on A-tract-induced bending and flexibility. *J Mol Biol* 314, 23-40.
- Mesleh, M. F., Hunter, J. M., Shvartsburg, A. A., Schatz, G. C. & Jarrold, M. F. (1996). Structural Information from Ion Mobility Measurements: Effects of the Long Range Potentia. *Journal of Physical Chemistry* 100, 16082-16086.
- Mocci, F. & Saba, G. (2003). Molecular dynamics simulations of A·T-rich oligomers: Sequence-specific binding of Na⁺ in the minor groove of B-DNA. *Biopolymers* 68, 471-485.
- Norberg, J. & Nilsson, L. (1998). Solvent influence on base stacking. *Biophys J* 74, 394-402.
- Ponomarev, S. Y., Thayer, K. M. & Beveridge, D. L. (2004). Ion motions in molecular dynamics simulations on DNA. *Proc Natl Acad Sci U S A* 101, 14771-14775.
- Rosu, F., Gabelica, V., Houssier, C., Colson, P. & Pauw, E. D. (2002a). Triplex and quadruplex DNA structures studied by electrospray mass spectrometry. *Rapid Commun Mass Spectrom* 16, 1729-1736.
- Rosu, F., Gabelica, V., Houssier, C. & De Pauw, E. (2002b). Determination of affinity, stoichiometry and sequence selectivity of minor groove binder complexes with

- double-stranded oligodeoxynucleotides by electrospray ionization mass spectrometry. *Nucleic Acids Res* 30, e82.
- Rosu, F., Gabelica, V., Shin-ya, K. & De Pauw, E. (2003). Telomestatin-induced stabilization of the human telomeric DNA quadruplex monitored by electrospray mass spectrometry. *Chem Commun (Camb)*, 2702-2703.
- Saenger, W. (1984). *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
- Schnier, P. D., Klassen, J. S., Strittmatter, E. F. & Williams, E. R. (1998). Activation energies for dissociation of double strand oligonucleotide anions: Evidence for Watson-Crick base pairing in vacuo. *Journal of the American Chemical Society* 120, 9605-9613.
- Seenisamy, J., Rezler, E. M., Powell, T. J., Tye, D., Gokhale, V., Joshi, C. S., Siddiqui-Jain, A. & Hurley, L. H. (2004). The dynamic character of the G-quadruplex element in the c-MYC promoter and modification by TMPyP4. *J Am Chem Soc* 126, 8702-8709.
- Shui, X., McFail-Isom, L., Hu, G. G. & Williams, L. D. (1998). The B-DNA Dodecamer at High Resolution Reveals a Spine of Water on Sodium. *Biochemistry* 37, 8341-8355.
- Shvartsburg, A. A. & Jarrold, M. F. (1996). An Exact Hard Spheres Scattering Model for the Mobilities of Polyatomic Ions. *Chemical Physics Letters* 261, 86-91.
- Sines, C. C., McFail-Isom, L., Howerton, S. B., VanDerveer, D. & Williams, L. D. (2000). Cations Mediate B-DNA Conformational Heterogeneity. *Journal of the American Chemical Society* 122, 11048-11056.
- Sponer, J., Leszczynski, J. & Hobza, P. (2001). Hydrogen Bonding, Stacking, and Cation Binding of DNA Bases. *Journal of Molecular Structure* 573, 42-53.

- Sponer, J., Leszczynski, J. & Hobza, P. (2002). Electronic Properties, Hydrogen Bonding, Stacking, and Cation Binding of DNA and RNA Bases. *Biopolymers* 61, 3-31.
- Stefl, R. & Koca, J. (2000). Unrestrained Molecular Dynamic Simulations of [d(AT)₅]₂ Duplex in Aqueous Solution: Hydration and Binding of Sodium in the Minor Groove. *Journal of the American Chemical Society* 122, 5025-5033.
- Stellwagen, N. C., Magnusdottir, S., Gelfi, C. & Righetti, P. G. (2001). Preferential Counterion binding to A-tract DNA oligomers. *Journal of Molecular Biology* 305, 1025-1033.
- Subirana, J. A. & Soler-López, M. (2003). Cations as Hydrogen Bond Donors: A view of Electrostatic Interactions in DNA. *Annual Review of Biophysics and Biomolecular Structure* 32, 27-45.
- Subramanian, P. S., Ravishanker, G. & Beveridge, D. L. (1988). Theoretical considerations on the "spine of hydration" in the minor groove of d(CGCGAATTCGCG).d(GCGCTTAAGCGC): Monte Carlo computer simulation. *Proc Natl Acad Sci U S A* 85, 1836-40.
- Vairamani, M. & Gross, M. L. (2003). G-Quadruplex Formation of Thrombin-Binding Aptamer Detected by Electrospray Ionization Mass Spectrometry. *Journal of the American Chemical Society* 125, 42-43.
- Varnai, P. & Zakrzewska, K. (2004). DNA and its counterions: a molecular dynamics study. *Nucleic Acids Res* 32, 4269-4280.
- von Helden, G., Ming-Teh, H., Kemper, P. R. & Bowers, M. T. (1991). Structures of carbon cluster ions from 3 to 60 atoms: Linears to rings to fullerenes. *Journal of Chemical Physics* 95, 3835-3837.

- Wan, K. X., Shibue, T. & Gross, M. L. (2000). Non-covalent complexes between DNA-binding drugs and double-stranded oligodeoxynucleotides: a study by electrospray ionization mass spectrometry. *Journal of the American Chemical Society* 122, 300-307.
- Watson, J. D. & Crick, F. H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 171, 737-738.
- Weiner, S. J., Kollman, P. A., Nguyen, D. T. & Case, D. A. (1986). An all atom force field for simulations of proteins and nucleic acids. *Journal of Computational Chemistry* 7, 230-252.
- Wemmer, D. E. (2000). Designed sequence-specific minor groove ligands. *Annu Rev Biophys Biomol Struct* 29, 439-461.
- Wytenbach, T., von Helden, G., Batka, J. J., Carlat, J., Douglas, & Bowers, M. T. (1997). Effect of the Long-Range Potential on Ion Mobility Measurements. *Journal of the American Society for Mass Spectrometry* 8, 275-282.
- Yanson, I. K., Teplitsky, A. B. & Sukhdub, L. F. (1979). Experimental studies of molecular interactions between nitrogen bases of nucleic acids. *Biopolymers* 18, 1149-1160.
- Young, M. A., Jayaram, B. & Beveridge, A. (1997). Intrusion of Counterions into the Spine of Hydration in the Minor Groove of B-DNA: Fractional Occupancy of Electronegative Pockets. *Journal of the American Chemical Society* 119, 59-69.

Capítulo IV
RESULTADOS

En este capítulo se muestran los resultados obtenidos en los diferentes trabajos realizados durante esta tesis. Se han agrupado los mismos en dos bloques, el primero de ellos relacionado con el estudio de la atmósfera de iones que rodea a una secuencia de ADN, mientras que el segundo está dedicado al estudio de diferentes formas de ADN en vacío, condiciones similares a las que se utilizan en los experimentos de espectrometría de masas por ionización con electrospray (ESI-MS).

4.1 Exploración de la atmósfera de iones alrededor del ADN

Desde los primeros estudios experimentales sobre la estructura del ADN se conoce que el solvente (agua e iones) juega un papel primordial en la estabilización de la doble hélice y en el mantenimiento de su conformación global. Recientemente, han aparecido un número elevado de trabajos teóricos y experimentales que intentan relacionar interacciones específicas ión-ADN con cambios conformacionales locales, como por ejemplo la menor amplitud del surco estrecho en secuencias ricas en pares A·T (Auffinger & Bielecki, 2004; Bonvin, 2000; Chandrasekaran et al., 1995; Denisov & Halle, 2000; Egli, 2002; Feig & Pettitt, 1999; Gearheart et al., 2003; Hamelberg et al., 2000; Hamelberg et al., 2001; Hud & Polak, 2001; Korolev et al., 2002; Laughton et al., 1995; Madhumalar & Bansal, 2003; McConnell & Beveridge, 2000; McConnell & Beveridge, 2001; Mocci & Saba, 2003; Ponomarev et al., 2004; Shui et al., 1998; Sines et al., 2000; Stefl & Koca, 2000; Stellwagen et al., 2001; Subirana & Soler-López, 2003; Subramanian et al., 1988; Varnai & Zakrzewska, 2004; Young et al., 1997). Muchos de estos estudios tenían por objeto corroborar resultados obtenidos por Dinámica Molecular en el grupo de D. Beveridge (Young et al., 1997), los cuales proponían que algunos Na^+ podían permanecer cercanos a las zonas más electronegativas del surco estrecho de la región AATT del dodecámero de Dickerson-Drew durante tiempos de residencia largos, poniendo en cuestión la integridad de la espina de hidratación de Dickerson.

En este trabajo se retoma la investigación sobre interacciones entre ADN y Na⁺ mediante técnicas de simulación basadas en Dinámica Molecular, con el objetivo doble de: i) racionalizar cómo los contraiones modulan las propiedades estructurales y dinámicas del dodecámero de Dickerson-Drew (5'-CGCGAATTCGCG-3), ii) determinar cuales son los límites de las simulaciones actuales para el análisis de procesos poco frecuentes como, por ejemplo, el intercambio Na⁺/agua en el interior del surco estrecho.

4.1.1 Exploring the Counterion Atmosphere around DNA: What Can Be Learned from Molecular Dynamics Simulations?

Manuel Rueda, Elena Cubero, Charles A. Laughton y Modesto Orozco

Biophysical Journal. 2004, 87, 800-811

4.2 La estructura y dinámica del ADN en fase gas

Podría pensarse que la transferencia de la molécula de ADN desde entorno acuoso a fase gas produciría la disociación de las cadenas, debido a la fuerte repulsión de carga no apantallada que se origina en los fosfatos. No obstante, datos recientes obtenidos mediante espectrometría de masas por ionización con electrospray (ESI-MS) muestran que los dúplexes de ADN sobreviven a la fuerte deshidratación que se produce en estos experimentos (Ganem et al., 1993; Light-Wahl et al., 1993). A pesar de estas evidencias, se desconoce cuál es la naturaleza del dúplex en vacío y cuáles son los factores que influyen en su estructura y dinámica. De hecho, sólo existen datos *indirectos* obtenidos a partir de técnicas espectrométricas, que apuntan en la dirección de que los dúplexes son capaces de retener algunos apareamientos WC, lo que impediría su colapso hacia formas globulares (Gabelica & De Pauw, 2002; Gabelica & De, 2002; Gabelica & Pauw, 2001; Griffey et al., 1999; Schnier et al., 1998). Para suplir esta carencia de información se puede recurrir a las técnicas de simulación por ordenador, las cuales brindan la posibilidad de describir a nivel atómico el comportamiento estructural, energético y dinámico del ADN bajo estas condiciones extremas.

En el apartado 4.2.1 se estudian mediante Dinámica Molecular dos secuencias de ADN (12 y 16 pb) en condiciones similares a las de un experimento de espectrometría de masas por ionización con electrospray (ESI-MS).

Experimentos de ESI-MS sugieren también que el ADN puede interactuar de manera no covalente con pequeños ligandos (mG-binders) en fase gas, conservando además las abundancias relativas encontradas en solución (Gabelica et al., 1999; Gabelica et al., 2000; Guittat et al., 2003; Rosu et al., 2002b; Wan et al., 2000). Estos descubrimientos resultan sorprendentes, ya que previsiblemente los cambios estructurales que se generan en el ADN después de la deshidratación deberían afectar a su capacidad para acomodar ligandos en el surco estrecho. En el apartado 4.2.2 se presenta un estudio sistemático mediante Dinámica Molecular sobre la estructura y

dinámica de complejos ADN-ligando (mG-binders: DAPI, Hoechst 33258 y Netropsina) bajo las condiciones experimentales existentes en los experimentos de ESI-MS.

Finalmente, diversos autores han encontrado evidencias de que los G-cuadrúplexes de ADN pueden mantener en fase gas parte de su estructura original en solución (Baker et al., 2005 ; Gabelica et al., 2005; Goodlett et al., 1993; Rosu et al., 2002a; Rosu et al., 2003; Vairamani & Gross, 2003). En el apartado 4.2.3 se estudian diferentes motivos basados en el G4-ADN mediante simulaciones de Dinámica Molecular en vacío y con diferentes iones monovalentes (Na^+ , K^+ y Li^+).

4.2.1 The Structure and Dynamics of DNA in the Gas Phase

Manuel Rueda, Susana Kalko, F. Javier Luque y Modesto Orozco

J. Am. Chem. Soc. **2003**, *125*, 8007-8014

4.2.2 Nature of Minor-Groove Binders-DNA Complexes in the Gas Phase

Manuel Rueda, F. Javier Luque y Modesto Orozco

J. Am. Chem. Soc. 2005, *127*, 11690-11698

4.2.3 G-Quadruplexes Can Maintain Their Structure in the Gas Phase

Manuel Rueda, F. Javier Luque y Modesto Orozco

J. Am. Chem. Soc. 2006, 128, 3608-3619

Referencias bibliográficas

- Auffinger, P. & Bielecki, L. (2004). Anion Binding to Nucleic Acids. *Structure* 12, 379-388.
- Baker, E. S., Bernstein, S. L. & Bowers, M. T. (2005). Structural Characterization of G-Quadruplexes in Deoxyguanosine Clusters Using Ion Mobility Mass Spectrometry. *Journal of the American Society of Mass Spectrometry* 16, 989-997.
- Bonvin, A. M. J. J. (2000). Localization and dynamics of sodium counterions around DNA in solution from molecular dynamics simulation. *European Biophysics Journal* 29, 57-60.
- Chandrasekaran, R., Radha, A. & Park, H. S. (1995). Sodium ions and water molecules in the structure of poly(dA)·poly(dT). *Acta Crystallographica* D51, 1025-1035.
- Denisov, V. & Halle, B. (2000). Sequence-specific binding to counterions to B-DNA. *Proc Natl Acad Sci U S A* 97, 629-633.
- Egli, M. (2002). DNA-Cation Interactions: Quo Vadis? *Chemistry and Biology* 9, 277-286.
- Feig, M. & Pettitt, B. M. (1999). Sodium and chlorine ions as part of the DNA solvation shell. *Biophysical Journal* 77, 1769-1781.
- Gabelica, V. & De Pauw, E. (2002). Collision-induced dissociation of 16-mer DNA duplexes with various sequences: evidence for conservation of the double helix conformation in the gas phase. *International Journal of Mass Spectrometry* 219, 151-159.

- Gabelica, V., De Pauw, E. & Rosu, F. (1999). Interaction between antitumor drugs and a double-stranded oligonucleotide studied by electrospray ionization mass spectrometry. *J Mass Spectrom* 34, 1328-1337.
- Gabelica, V. & De, P. E. (2002). Comparison of the collision-induced dissociation of duplex DNA at different collision regimes: evidence for a multistep dissociation mechanism. *J Am Soc Mass Spectrom* 13, 91-98.
- Gabelica, V. & Pauw, E. D. (2001). Comparison between solution-phase stability and gas-phase kinetic stability of oligodeoxynucleotide duplexes. *J Mass Spectrom* 36, 397-402.
- Gabelica, V., Rosu, F., Houssier, C. & De Pauw, E. (2000). Gas phase thermal denaturation of an oligonucleotide duplex and its complexes with minor groove binders. *Rapid Commun Mass Spectrom* 14, 464-467.
- Gabelica, V., Rosu, F., Witt, M., Baykut, G. & De Pauw, E. (2005). Fast gas-phase hydrogen/deuterium exchange observed for a DNA G-quadruplex. *Rapid Commun Mass Spectrom* 19, 201-208.
- Ganem, B., Yu-Tsyr. Li, Hsieh, Y.-L. & Henion, J. (1993). Detection of oligonucleotide duplex forms by ion spray mass spectrometry. *Tetrahedron Letters* 34, 1445-1448.
- Gearheart, L. A., Somoza, M. M., Rivers, W. E., Murphy, C. J., Coleman, R. S. & Berg, M. A. (2003). Sodium-Ion Binding to DNA: Detection by Ultrafast Time-Resolved Stokes-Shift Spectroscopy. *Journal of the American Chemical Society* 125, 11812-11813.

- Goodlett, D. R., Camp, D. G., Hardin, C. C., Corregan, M. & Smith, D. (1993). Direct observation of a DNA quadruplex by electrospray ionization mass spectrometry. *Biol. Mass Spectrom* 22, 181-183.
- Griffey, R. H., Grieg, M. J., An, H., Sasmor, H. & Manalili, S. (1999). Targeted Site-Specific Gas-Phase Cleavage of Oligoribonucleotides. Application in Mass Spectrometry-Based Identification of Ligand Binding Sites. *Journal of the American Chemical Society* 121, 474-475.
- Guittat, L., Alberti, P., Rosu, F., Van Miert, S., Thetiot, E., Pieters, L., Gabelica, V., De Pauw, E., Ottaviani, A., Riou, J. F. & Mergny, J. L. (2003). Interactions of cryptolepine and neocryptolepine with unusual DNA structures. *Biochimie* 85, 535-547.
- Hamelberg, D., McFail-Isom, L., Williams, L. D. & Wilson, W. D. (2000). Flexible Structure of DNA: Ion Dependence of Minor-Groove Structure and Dynamics. *Journal of the American Chemical Society* 122, 10513-10520.
- Hamelberg, D., Williams, L. D. & Wilson, W. D. (2001). Influence of the Dynamic Positions of Cations on the Structure of the DNA Minor Groove: Sequence-Dependent Effects. *Journal of the American Chemical Society* 123, 7745-7755.
- Hud, N. V. & Polak, M. (2001). DNA-cation interactions: the major and minor grooves are flexible ionophores. *Curr Opin Struct Biol* 11, 293-301.
- Korolev, N., Lyubartsev, A. P., Laaksonen, A. & Nordenskiöld, L. (2002). On the competition between water, sodium ions, and spermine in binding to DNA: a molecular dynamics computer simulation study. *Biophysical Journal* 82, 2860-2875.

-
- Laughton, C. A., Luque, F. J. & Orozco, M. (1995). Counterion Distribution around DNA Studied by Molecular Dynamics and Quantum Mechanical Simulations. *Journal of Physical Chemistry* 99, 11591-11599.
- Light-Wahl, K. J., Springer, D. L., Winger, B. E., Edmonds, C. G., Camp, D. G., Thrall, B. D. & Smith, R. D. (1993). Observation of a small oligonucleotide duplex by electrospray ionization mass spectrometry. *Journal of the American Chemical Society* 115, 803-804.
- Madhumalar, A. & Bansal, M. (2003). Structural insights into the effect of hydration and ions on A-tract DNA: A molecular dynamics study. *Biophysical Journal* 85, 1805-1816.
- McConnell, K. J. & Beveridge, D. L. (2000). DNA structure: what's in charge? *J Mol Biol* 304, 803-820.
- McConnell, K. J. & Beveridge, D. L. (2001). Molecular dynamics simulations of B'-DNA: sequence effects on A-tract-induced bending and flexibility. *J Mol Biol* 314, 23-40.
- Mocci, F. & Saba, G. (2003). Molecular dynamics simulations of A-T-rich oligomers: Sequence-specific binding of Na⁺ in the minor groove of B-DNA. *Biopolymers* 68, 471-485.
- Ponomarev, S. Y., Thayer, K. M. & Beveridge, D. L. (2004). Ion motions in molecular dynamics simulations on DNA. *Proc Natl Acad Sci U S A* 101, 14771-14775.
- Rosu, F., Gabelica, V., Houssier, C., Colson, P. & Pauw, E. D. (2002a). Triplex and quadruplex DNA structures studied by electrospray mass spectrometry. *Rapid Commun Mass Spectrom* 16, 1729-1736.

- Rosu, F., Gabelica, V., Houssier, C. & De Pauw, E. (2002b). Determination of affinity, stoichiometry and sequence selectivity of minor groove binder complexes with double-stranded oligodeoxynucleotides by electrospray ionization mass spectrometry. *Nucleic Acids Res* 30, e82.
- Rosu, F., Gabelica, V., Shin-ya, K. & De Pauw, E. (2003). Telomestatin-induced stabilization of the human telomeric DNA quadruplex monitored by electrospray mass spectrometry. *Chem Commun (Camb)*, 2702-2703.
- Schnier, P. D., Klassen, J. S., Strittmatter, E. F. & Williams, E. R. (1998). Activation energies for dissociation of double strand oligonucleotide anions: Evidence for Watson-Crick base pairing in vacuo. *Journal of the American Chemical Society* 120, 9605-9613.
- Shui, X., McFail-Isom, L., Hu, G. G. & Williams, L. D. (1998). The B-DNA Dodecamer at High Resolution Reveals a Spine of Water on Sodium. *Biochemistry* 37, 8341-8355.
- Sines, C. C., McFail-Isom, L., Howerton, S. B., VanDerveer, D. & Williams, L. D. (2000). Cations Mediate B-DNA Conformational Heterogeneity. *Journal of the American Chemical Society* 122, 11048-11056.
- Stefl, R. & Koca, J. (2000). Unrestrained Molecular Dynamic Simulations of [d(AT)₅]₂ Duplex in Aqueous Solution: Hydration and Binding of Sodium in the Minor Groove. *Journal of the American Chemical Society* 122, 5025-5033.
- Stellwagen, N. C., Magnusdottir, S., Gelfi, C. & Righetti, P. G. (2001). Preferential Counterion binding to A-tract DNA oligomers. *Journal of Molecular Biology* 305, 1025-1033.

-
- Subirana, J. A. & Soler-López, M. (2003). Cations as Hydrogen Bond Donors: A view of Electrostatic Interactions in DNA. *Annual Review of Biophysics and Biomolecular Structure* 32, 27-45.
- Subramanian, P. S., Ravishanker, G. & Beveridge, D. L. (1988). Theoretical considerations on the "spine of hydration" in the minor groove of d(CGCGAATTCGCG).d(GCGCTTAAGCGC): Monte Carlo computer simulation. *Proc Natl Acad Sci U S A* 85, 1836-40.
- Vairamani, M. & Gross, M. L. (2003). G-Quadruplex Formation of Thrombin-Binding Aptamer Detected by Electrospray Ionization Mass Spectrometry. *Journal of the American Chemical Society* 125, 42-43.
- Varnai, P. & Zakrzewska, K. (2004). DNA and its counterions: a molecular dynamics study. *Nucleic Acids Res* 32, 4269-4280.
- Wan, K. X., Shibue, T. & Gross, M. L. (2000). Non-covalent complexes between DNA-binding drugs and double-stranded oligodeoxynucleotides: a study by electrospray ionization mass spectrometry. *Journal of the American Chemical Society* 122, 300-307.
- Young, M. A., Jayaram, B. & Beveridge, A. (1997). Intrusion of Counterions into the Spine of Hydration in the Minor Groove of B-DNA: Fractional Occupancy of Electronegative Pockets. *Journal of the American Chemical Society* 119, 59-69.

Capítulo V

DISCUSIÓN DE RESULTADOS

En este apartado se comentarán los resultados más relevantes obtenidos en las publicaciones del capítulo anterior.

5.1 Exploración de la atmósfera de iones alrededor del ADN

Dado que los iones juegan un papel muy importante en la estructura y función del ADN, es de gran interés aplicar las técnicas de Dinámica Molecular para examinar en detalle su distribución estructural y dinámica alrededor de la molécula. En este trabajo se ha explorado la atmósfera de iones Na^+ que rodea una secuencia de 12 pares de bases (DDD, 5'-CGCGAATTCGCG-3'), haciendo hincapié en la posibilidad de que un ión pueda reemplazar a una molécula de agua dentro de la *espina de hidratación* del surco estrecho. Para dicho propósito se han realizado tres tipos diferentes de simulaciones:

- Simulaciones de Dinámica Molecular sin restricciones.
- Simulaciones de Dinámica Molecular acoplada a Monte Carlo.
- Simulaciones de Dinámica Molecular guiadas por información.

Las simulaciones sin restricciones generan trayectorias estables, capaces de reproducir datos experimentales con notable precisión, tal y como muestran los datos de RMSD respecto a la estructura del cristal (1.8-1.9 Å), generando en todos los casos estructuras de ADN muy similares, independientemente de la fuerza iónica o la configuración inicial de Na^+ (RMSD cruzados sobre las diferentes estructuras promedio de 1.3-1.8 Å).

Cuando se revisan las poblaciones de iones alrededor del B-ADN, se puede concluir que existe un sitio de unión profundo en la parte final de la secuencia AATT de esta secuencia. Este sitio posiblemente sea el mismo que han sugerido otros autores en base a sus simulaciones (Hamelberg et al., 2000) y probablemente sea también el mismo que se ha asignado a tiempos de residencia de 50 ns a partir de datos de NMR (Denisov & Halle, 2000). El largo tiempo de residencia del Na^+ en la parte final del surco estrecho,

así como el largo tiempo necesario para que el ión acceda a dicha posición hace que posiblemente se necesiten tiempos de simulación mayores para estudiar el intercambio Na^+ /agua en el surco, tal y como indican los experimentos de NMR (Denisov & Halle, 2000), los cuales sugieren que el agua puede penetrar en esta región con una eficiencia 1000 veces superior a la del Na^+ . Las posiciones de alta residencia son difíciles de alcanzar por los iones, pero una vez lo hacen, pueden permanecer en ellas durante tiempos de residencia superiores a los 10 ns. Los resultados sobre MD de otros grupos (Hamelberg et al., 2000; McConnell & Beveridge, 2000; Mocci & Saba, 2003) sugieren que si las simulaciones *IIII* (con iones iniciales en el surco) se extendiesen por periodos más largos, finalmente se obtendría la salida del Na^+ hacia el exterior del surco, pero también previenen sobre problemas de equilibrado en los tiempos de simulación considerados estándar actualmente en Dinámica Molecular de ácidos nucleicos.

El análisis de las trayectorias *WWWW* sugiere una población de estructuras conteniendo Na^+ cerca del surco por debajo del 10%, en concordancia con trabajos anteriores (McConnell & Beveridge, 2000). Sin embargo, estas configuraciones corresponden a Na^+ unidos de manera débil y no son equiparables a las posiciones de larga residencia en el interior del surco.

La falta de convergencia por lo que respecta a la distribución de iones entre las configuraciones sin ión inicial en el surco (*WWWW*) y con ión (*IIII*) ha sido útil dado que ha permitido la caracterización del B-ADN tanto en ausencia como en presencia de Na^+ en su interior. La presencia de un ión en esta posición no produce cambios significativos en la estructura, ni a nivel general ni en la región AT, tal y como muestran los análisis sobre distancia entre fosfatos o de superficie accesible al solvente (SAS). Por lo que respecta a la dinámica, los movimientos esenciales de todas las simulaciones presentan un comportamiento parecido, con índices de similitud muy altos (>0.92), similitud que se mantiene cuando el estudio se focaliza en la región central AATT. La única diferencia remarcable debida a la presencia de iones Na^+ en el interior del surco sería una ligera reducción en la flexibilidad de la zona cercana al ión en simulaciones *IIII*, la cuál se percibe también por el menor rango de fluctuaciones en la zona AT de

parámetros como la SAS o la distancia entre fosfatos, así como por unas entropías ligeramente menores.

Las simulaciones basadas en Monte-Carlo acoplada a MD generan información sobre la estabilidad relativa de las diversas configuraciones conteniendo Na^+ /agua en las posiciones propuestas por Williams (Shui et al., 1998). Claramente, el estado más estable es el que corresponde a 4 aguas en el interior del surco, mientras que el más inestable sería el que posee 4 iones Na^+ . En general, a medida que se incrementa el número de iones en estas posiciones hay un descenso en la estabilidad global, confirmando la poca probabilidad de tener un surco estrecho poblado de iones Na^+ .

Finalmente, las simulaciones de Dinámica Molecular dirigida por información han servido para reproducir la salida del ión desde el interior al exterior sin la necesidad de añadir ninguna perturbación externa al cálculo de fuerzas. De las ocho simulaciones realizadas (~25 ns en total), sólo en dos fue posible reproducir la salida del ión al exterior, lo que vuelve a confirmar la dificultad de este proceso. El análisis de las trayectorias pone de manifiesto que la salida del ión es un fenómeno complejo, que se produce en un periodo corto de tiempo, durante el cual se han de producir fluctuaciones en los esqueletos del ácido nucleico.

5.2 La estructura y dinámica del ADN en fase gas.

En el primero de los tres trabajos sobre esta temática, se realizó el estudio mediante Dinámica Molecular de dos dúplexes de ADN *desnudo* en condiciones de fase gas, propias de los experimentos de ESI-MS. En él, se observa en primer lugar que las simulaciones control en fase acuosa para ambas secuencias (12 y 16 bp) generan estructuras estables, muy cercanas a las originales, tal y como indican los valores de RMSD respecto las mismas (1.8-2.4 Å) o respecto sus estructuras promedio (1.7-1.8 Å). El resto de parámetros estructurales, energéticos y dinámicos estudiados también ponen de manifiesto la presencia de una estructura relacionada con la forma canónica de B-ADN, confirmando la capacidad, tanto de los protocolos como de los campos de fuerzas

actuales, para una descripción detallada de la estructura y dinámica del ADN en solución.

Para las simulaciones en fase gas, se asignó a los dúplexes el estado de carga más abundante en los experimentos de ESI-MS, siendo éste de -6 para la secuencia de 12 pb y de -8 para la de 16 pb. Para dichos estados, se observa una transición rápida inicial debida a la vaporización, pero después de ésta, la molécula apenas explora una pequeña parte de todo el espacio conformacional accesible. Adicionalmente, simulaciones realizadas con otros estados de carga ponen de manifiesto que, por ejemplo, dúplexes neutros (carga total 0) en condiciones de fase gas producen estructuras globulares muy distorsionadas, o que dúplexes sin neutralizaciones de carga no son posibles debido a la fuerte repulsión entre los fosfatos (las cadenas se separan). Es interesante remarcar el hecho de que, aunque las mayores abundancias relativas encontradas en ESI-MS para dúplexes de 12 pb pertenecen a estados de carga cercanos a -6, la molécula es capaz de soportar cargas mayores sin disociación de las cadenas. Además, en base a nuestros datos, parece poco probable que cadenas sencillas separadas (con estados de carga similares a los de los experimentos, por ejemplo, -3 para 12 pb) se vuelvan a unir de manera espontánea.

Las estructuras obtenidas en fase gas para los dúplexes, aunque distorsionadas, retienen cierta memoria de su conformación en agua, independientemente del protocolo de neutralización y de la temperatura. En general, a menor temperatura y carga localizada en los fosfatos, mayor semejanza con los patrones de interacción de fase acuosa. El perfil geométrico de la molécula en estas condiciones extremas es de tipo alargado, formando una hélice con esqueletos empaquetados donde los surcos no quedan bien definidos. En estas estructuras se pierden muchos puentes de hidrógeno canónicos, especialmente los A·T, pero el acercamiento entre bases posibilita la formación de nuevas interacciones, muchas de ellas no canónicas, tanto por puente de hidrógeno como por apilamiento.

Los movimientos esenciales de los dúplexes en fase gas son parecidos entre las diferentes distribuciones de carga y temperaturas. La concordancia está por encima del

30% cuando se comparan los valores de fase gas con los de solución, lo que indica una marcada conservación del tipo de flexibilidad de la molécula tras su deshidratación. Lo mismo sucede para la entropía, siendo esto indicativo de que aunque la vaporización genera cambios estructurales importantes, la molécula de ADN no adopta una forma desordenada, sino que se organiza en forma de dúplex, conservando las características básicas que posee en solución acuosa.

Una vez conocido el comportamiento de las dobles hebras de ADN en condiciones de fase gas, se procedió al estudio en dichas condiciones de complejos ternarios formados por pequeños ligandos (mG-binders) unidos de manera no covalente a dúplex de ADN.

Las simulaciones control en fase acuosa para los tres complejos ligando-ADN (los ligandos utilizados fueron: Hoechst-33258, Netropsina y DAPI) son capaces de reproducir con gran fidelidad los valores experimentales, con pequeñas desviaciones respecto a la estructura del cristal (RMSD entre 1.6 y 1.8 Å) o respecto a estructuras promedio, indicando además que las moléculas exploran una pequeña parte del espacio conformacional accesible. La unión de los ligandos no introduce grandes distorsiones en la geometría de los dúplexes, como se muestra en los RMSD cruzados entre simulaciones (valores entre 2.7 y 3.0 Å). Como ya sucedía en trabajos anteriores (Harris et al., 2001; Singh & Kollman, 1999; Spackova et al., 2003), los ligandos apenas se desplazan de su posición original a lo largo de la simulación. La unión del ligando parece limitar la flexibilidad de los dúplexes (Harris et al., 2001), como se demuestra en la reducción del volumen configuracional generalizado y la entropía (en 0.1-0.2 Kcal/mol K). Debido a esto, las constantes de fuerza asociadas a la deformación de los primeros modos se incrementan. En cualquier caso, la naturaleza de las deformaciones es parecida, tal y como muestran los altos índices de semejanza (correspondencia del 90%)

Los complejos mG-binder:ADN aparecen firmemente unidos después de entre 100 y 500 nanosegundos en fase gas, siendo esto indicativo de que los mismos son estables para los estados de carga estudiados (representativos de abundancias en ESI-MS), al

menos en la escala del μs . Estos datos sugieren además que el complejo ternario ADN-ligando queda atrapado en una conformación que evita la disociación de sus tres componentes. La transición conformacional se suele producir al inicio, tal y como se refleja en los valores de RMSD con respecto a las estructuras promedio en solución (valores entre 6.1 y 9.7 Å) y respecto a las estructuras del cristal (valores entre 5.6 y 10.0 Å). En algunos casos se puede ver una pequeña transición más tardía, pero nunca después de 50 ns.

A pesar del gran cambio conformacional producido por la deshidratación, el ADN mantiene propiedades geométricas que recuerdan a una hélice. Este cambio conformacional va a dirigir en la mayoría de los casos el posicionamiento del ligando, el cual a su vez también estará influido por las diferentes condiciones experimentales utilizadas (temperatura y estado de carga), haciendo posible distinguir varios modos de unión para las mismas. En general, existe un acercamiento del ligando al ADN después de la transferencia desde solución a fase gas, aunque la orientación relativa al eje de la hélice no varía en exceso. Este acercamiento facilita que el ligando interactúe con más residuos, aunque siempre con una clara preferencia por aquellos con los que interactúa en solución. Simulaciones adicionales realizadas en nuestro laboratorio indican que en el caso remoto de que un ligando (con carga neta) se separe varios angstroms del ADN durante el transcurso de un experimento de ESI-MS, siempre tendrá la posibilidad de volver a unirse al complejo *por el exterior* del mismo.

Los complejos ternarios están más comprimidos en fase gas que en agua, por lo que la SAS se reduce entre un 10-30%. En estas circunstancias, se incrementan las interacciones de puente de hidrógeno no canónicos y apilamientos no paralelos al eje de la hélice. La compresión del complejo se refleja también en los valores del factor de sección de colisión (CCS), los cuales son menores en fase gas que en solución. Mientras que los valores de dúplexes desnudos son cercanos a formas B del ADN, la presencia del ligando suele confinarlos a valores más cercanos a una forma A. Los movimientos esenciales confirman que después de la vaporización el ADN no se desordena, sino que incluso pierde algo de entropía respecto a las simulaciones en solución. Por último, la

dinámica esencial refleja el mantenimiento de una memoria significativa (20%) del tipo de movimientos que se realizan en vacío con respecto los que tienen lugar en solución.

Además de para dobles hélices, es posible encontrar en la bibliografía datos experimentales de ESI-MS sobre estructuras no canónicas de ADN como son los G-cuadrúplexes (Baker et al., 2005 ; Gabelica et al., 2005; Goodlett et al., 1993; Rosu et al., 2002; Rosu et al., 2003; Vairamani & Gross, 2003). La alta concentración de carga que poseen estas estructuras debería provocar la fragmentación del complejo en condiciones de vacío, no obstante, los datos experimentales apuntan hacia cierta conservación de las propiedades de solución. Con el objetivo de estudiar el comportamiento de este tipo de estructuras en fase gas, en el último trabajo se realizaron simulaciones de Dinámica Molecular de varias estructuras de G-cuadrúplexes (con y sin *loops*) basadas en dos motivos (paralelo y antiparalelo) y con tres tipos de iones (Na^+ , K^+ y Li^+) en el interior de sus canales.

Las simulaciones control en agua de los G-cuadrúplex con iones iniciales en el interior son estables. Esto es así tanto para las estructuras antiparalelas (148D y 156D, códigos PDB) como para la estructura paralela (352D, código PDB), como muestran los valores de RMSD frente a la estructura experimental en las simulaciones sin *loops* (entre 0.7 y 1.2 Å) y frente a estructuras promedio (entre 0.5 y 0.6 Å). Los diferentes parámetros asociados a estructura indican que apenas existe distorsión en la geometría global la molécula, excepto un pequeño incremento en la superficie de la estructura paralela, la cuál además presenta intrínsecamente mayor flexibilidad que las estructuras antiparalelas consideradas en este estudio. En las simulaciones realizadas con *loops* conectores se observa que éstos son bastante móviles, incrementándose los valores de RMSD frente a la estructura experimental y promedio (2.5 y 2.0 Å respectivamente), así como el resto de valores asociados a flexibilidad. Sin embargo, este movimiento de las bases conectoras apenas genera ningún cambio estructural en la porción central del G4-ADN (RMSD de 1 Å respecto la estructura de RMN y 0.6 Å frente a la promedio), siendo esto indicativo de que la porción central del G-cuadrúplex es prácticamente la misma independientemente de la presencia o ausencia de *loops* en la simulación, y validando el uso de la porción central del G4-ADN como sistema de estudio modelo. En

concordancia con las evidencias experimentales, los iones Na^+ y K^+ parecen ser más eficientes que los iones Li^+ estabilizando este tipo de estructuras. De hecho, los dos primeros apenas fluctúan de sus posiciones originales en el caso del aptámero, mientras que el Li^+ se desplaza hacia el centro de una de las tétradas. En la estructura paralela dos de los cationes permanecen en las posiciones originales y el tercero se desplaza a una posición más terminal en el caso del Na^+ e incluso llega a abandonar el G4-ADN en las simulaciones con Li^+ y K^+ . En las simulaciones sin ión inicial en el interior, se observa que la estructura tiende a capturarlos del medio. Así sucede en el caso del aptámero antiparalelo, el cual queda estabilizado, mientras que en el caso de la estructura paralela es necesaria la entrada de más de 1 ión, como ya se había apuntado con anterioridad (Fadrna et al., 2004; Hazel et al., 2004; Spackova et al., 1999; Spackova et al., 2004; Spackova et al., 2001; Stefl et al., 2003; Stefl et al., 1999; Stefl et al., 2001). En base a nuestros datos parece que el tiempo de captura del catión es mayor que el tiempo de colapso del G-cuadrúplex. Por último, mencionar que aunque en este último caso el complejo está fuertemente desestructurado, los valores de SAS y de CCS son análogos a los de un G4-ADN canónico.

En las simulaciones en fase gas de las estructuras antiparalelas no se ha detectado separación de las cadenas o difusión del catión después de la transferencia desde solución a fase gas, lo que concuerda con las abundancias relativas de los complejos correspondientes a los G-cuadrúplexes en los experimentos de ESI-MS. Las simulaciones sin iones en el interior del canal están muy distorsionadas debido a la fuerte repulsión no apantallada entre fosfatos. No obstante, para la simulación a temperatura ambiente algunos parámetros como el CCS y la SAS están en el rango de lo esperado para un G4-ADN canónico, generando nuevamente dudas sobre la validez de los mismos como únicos descriptores de propiedades estructurales. A pesar de la fuerte repulsión entre fosfatos, las simulaciones con iones en el interior del canal son estables durante 1 μs , aunque con algunas diferencias entre ellas dependiendo de la temperatura y de la naturaleza del catión. En general los iones Li^+ generan G-cuadrúplexes distorsionados y globulares, mientras que en el caso de estructuras conteniendo Na^+ y K^+ sólo se encuentren distorsiones cuando se eleva la temperatura. Estos colapsos hacen posible que aparezcan nuevas interacciones entre las bases, favoreciéndose el

apilamiento. La dinámica esencial del aptámero muestra que los cambios asociados a la transferencia desde solución a fase gas son moderados, provocando un incremento en el espacio conformacional en las simulaciones donde se mantiene la estructura (simulaciones con Na^+ y K^+). No obstante, la naturaleza de los movimientos esenciales está altamente conservada después de la vaporización (60-70%). En el caso de la estructura antiparalela de mayor tamaño (156D), la dinámica esencial muestra una reducción de la flexibilidad después de la vaporización. En base a los resultados, la presencia de los *loops* en 148D no introduce grandes cambios conformacionales en la molécula (RMSD de la región central frente a la estructura experimental de 1.9 Å) y prácticamente sucede lo mismo cuando se analiza la estructura antiparalela de mayor tamaño. Aunque en solución acuosa los *loops* son muy flexibles, en fase gas éstos se vuelven rígidos, provocando a su vez un incremento en la rigidez de la porción central del G-cuadrúplex.

Las estructuras paralelas sin cationes en el interior sólo resisten a estados de carga bajos (-9 y -8) en fase gas, dado que cargas mayores no apantalladas provocan la disociación de los complejos. Simulaciones adicionales indican que en el caso de que los componentes se separen en el transcurso de los experimentos de ESI-MS (cadenas y/o cationes), parece poco probable que éstos se reagrupen. Todas las demás simulaciones sobre G-cuadrúplexes realizadas con iones en el interior de los canales son estables (al menos en la escala del μs) e intentan generar una visión aproximada de cómo sería la estructura de los G4-ADN en las condiciones de los experimentos de ESI-MS. La mayoría de estas estructuras corresponden a G-cuadrúplexes reales, aunque con algunas excepciones (Ej.: simulaciones con Na^+ y Li^+ a alta temperatura). A pesar de esta diversidad, los valores de CCS son prácticamente idénticos entre todas las estructuras, volviendo a poner en duda la validez de dicho parámetro como medida de conservación estructural. La naturaleza del catión está muy vinculada con el mantenimiento de la estructura, de hecho, cuatro de las simulaciones realizadas con K^+ y tres con Na^+ generan estructuras casi idénticas a un G-cuadrúplex canónico después de 500 nanosegundos de simulación (RMSD inferiores a 3 Å). En estas estructuras se conserva tanto el patrón de interacciones entre bases como entre el catión y la base. En general, las simulaciones realizadas con Li^+ son menos estables y poseen patrones de

interacción complejos, aunque en ellas se conservan un gran número de interacciones intra e intercatenarias, donde se suelen incrementar las interacciones de apilamiento. A temperatura ambiente, la dinámica esencial muestra que la estructura paralela es más rígida en fase gas que en solución, como ya sucedía en la estructura antiparalela de mayor tamaño. La rigidez del sistema debe básicamente a la fuerte interacción con el catión y a una ligera compresión de las estructuras. Como también sucedía en el aptámero, en los casos donde se mantiene la estructura del G4-ADN se encuentra que la naturaleza de los modos de deformación es muy parecida (60-70%), mientras que las estructuras colapsadas parecen explorar un espacio conformacional mayor y sus movimientos esenciales difieren más del resto (semejanza del 40-50%).

5.3 Referencias bibliográficas

- Baker, E. S., Bernstein, S. L. & Bowers, M. T. (2005). Structural Characterization of G-Quadruplexes in Deoxyguanosine Clusters Using Ion Mobility Mass Spectrometry. *Journal of the American Society of Mass Spectrometry* 16, 989-997.
- Denisov, V. & Halle, B. (2000). Sequence-specific binding to counterions to B-DNA. *Proc Natl Acad Sci U S A* 97, 629-633.
- Fadrna, E., Spackova, N., Stefl, R., Koca, J., Cheatham, T. E., 3rd & Sponer, J. (2004). Molecular dynamics simulations of Guanine quadruplex loops: advances and force field limitations. *Biophys J* 87, 227-42.
- Gabelica, V., Rosu, F., Witt, M., Baykut, G. & De Pauw, E. (2005). Fast gas-phase hydrogen/deuterium exchange observed for a DNA G-quadruplex. *Rapid Commun Mass Spectrom* 19, 201-208.
- Goodlett, D. R., Camp, D. G., Hardin, C. C., Corregan, M. & Smith, D. (1993). Direct observation of a DNA quadruplex by electrospray ionization mass spectrometry. *Biol. Mass Spectrom* 22, 181-183.
- Hamelberg, D., McFail-Isom, L., Williams, L. D. & Wilson, W. D. (2000). Flexible Structure of DNA: Ion Dependence of Minor-Groove Structure and Dynamics. *Journal of the American Chemical Society* 122, 10513-10520.
- Harris, S. A., Gavathiotis, E., Searle, M. S., Orozco, M. & Laughton, C. A. (2001). Cooperativity in drug-DNA recognition: a molecular dynamics study. *J Am Chem Soc* 123, 12658-12663.
- Hazel, P., Huppert, J., Balasubramanian, S. & Neidle, S. (2004). Loop-length-dependent folding of G-quadruplexes. *J Am Chem Soc* 126, 16405-16415.

- McConnell, K. J. & Beveridge, D. L. (2000). DNA structure: what's in charge? *J Mol Biol* 304, 803-820.
- Mocci, F. & Saba, G. (2003). Molecular dynamics simulations of A·T-rich oligomers: Sequence-specific binding of Na⁺ in the minor groove of B-DNA. *Biopolymers* 68, 471-485.
- Rosu, F., Gabelica, V., Houssier, C., Colson, P. & Pauw, E. D. (2002). Triplex and quadruplex DNA structures studied by electrospray mass spectrometry. *Rapid Commun Mass Spectrom* 16, 1729-1736.
- Rosu, F., Gabelica, V., Shin-ya, K. & De Pauw, E. (2003). Telomestatin-induced stabilization of the human telomeric DNA quadruplex monitored by electrospray mass spectrometry. *Chem Commun (Camb)*, 2702-2703.
- Shui, X., McFail-Isom, L., Hu, G. G. & Williams, L. D. (1998). The B-DNA Dodecamer at High Resolution Reveals a Spine of Water on Sodium. *Biochemistry* 37, 8341-8355.
- Singh, S. B. & Kollman, P. A. (1999). Calculating the absolute free energy of association of netropsin and DNA. *Journal of the American Chemical Society* 121, 3267-3271.
- Spackova, N., Berger, I. & Sponer, J. (1999). Nanosecond Molecular Dynamics Simulations of Parallel and Antiparallel Guanine Quadruplex DNA Molecules. *Journal of the American Chemical Society* 121, 5519-5534.
- Spackova, N., Cheatham, T. E., 3rd, Ryjacek, F., Lankas, F., Van Meervelt, L., Hobza, P. & Sponer, J. (2003). Molecular dynamics simulations and thermodynamics analysis of DNA-drug complexes. Minor groove binding between 4',6-

- diamidino-2-phenylindole and DNA duplexes in solution. *J Am Chem Soc* 125, 1759-69.
- Spackova, N., Cubero, E., Sponer, J. & Orozco, M. (2004). Theoretical study of the guanine --> 6-thioguanine substitution in duplexes, triplexes, and tetraplexes. *J Am Chem Soc* 126, 14642-50.
- Spackova, N. a., Berger, I. & Sponer, J. (2001). Structural Dynamics and Cation Interactions of DNA Quadruplex Molecules Containing Mixed Guanine/Cytosine Quartets Revealed by Large-Scale MD Simulations. *Journal of the American Chemical Society* 123, 3295-3307.
- Stefl, R., Cheatham, T. E., 3rd, Spackova, N., Fadrna, E., Berger, I., Koca, J. & Sponer, J. (2003). Formation pathways of a guanine-quadruplex DNA revealed by molecular dynamics and thermodynamic analysis of the substates. *Biophys J* 85, 1787-804.
- Stefl, R., Fadrna, E. & Koca, J. (1999). Differences in conformational behavior of ATA and TAT sequences in single strand DNA trimer. *J Biomol Struct Dyn* 16, 1087-95.
- Stefl, R., Spackova, N., Berger, I., Koca, J. & Sponer, J. (2001). Molecular dynamics of DNA quadruplex molecules containing inosine, 6-thioguanine and 6-thiopurine. *Biophys J* 80, 455-68.
- Vairamani, M. & Gross, M. L. (2003). G-Quadruplex Formation of Thrombin-Binding Aptamer Detected by Electrospray Ionization Mass Spectrometry. *Journal of the American Chemical Society* 125, 42-43.

Capítulo VI
CONCLUSIONES

1. La Dinámica Molecular clásica sugiere que existe una región al final del surco estrecho de la región AATT de difícil acceso para iones del exterior, pero en la que una vez instalado el Na^+ , puede permanecer durante más de 10 ns.
2. Las configuraciones donde el surco estrecho está ocupado sólo por aguas parecen más estables que aquellas en las que aparecen iones. En cualquier caso no es posible encontrar un cambio significativo en la estructura del ADN relacionado con la presencia de Na^+ dentro del surco estrecho. La única alteración posible encontrada está relacionada con la flexibilidad, dado que la presencia del Na^+ limita la flexibilidad de esta zona.
3. En condiciones similares a las usadas en experimentos de espectrometría de masas por electrospray (ESI-MS), la vaporización no genera separación de las dos cadenas de ADN. De hecho, éstas permanecen fuertemente unidas al menos en la escala del microsegundo.
4. El análisis de las diferentes simulaciones sobre dúplexes de ADN en condiciones de fase gas apuntan a que el ADN retiene una memoria de las propiedades estructurales, energéticas y dinámicas después de la transferencia desde fase acuosa.
5. La vaporización no produce la disociación ni de las cadenas de ADN, ni de pequeños ligandos de unión no covalente (mG-binders) del ADN después de la transferencia desde solución a fase gas, al menos en la escala del microsegundo.
6. En los complejos ternarios, el ADN mantiene una estructura pseudo-helicoidal con propiedades estructurales y dinámicas similares a las de solución, mientras que el ligando no sufre desplazamientos dramáticos respecto a la posición de unión al ADN en medio acuoso.

7. En concordancia con los datos experimentales, los iones Na^+ y K^+ estabilizan los G-cuadrúplexes con mayor eficiencia que Li^+ , tanto en entorno acuoso como en fase gas.

8. En presencia de iones en el canal, la estabilidad del G4-DNA en fase gas es mayor que la de otros ácidos nucleicos y de hecho en numerosas simulaciones el cuadrúplex aparece inalterado en vacío. No se ha detectado separación de las cadenas o difusión de los iones después de la transferencia del G4-ADN desde solución a fase gas. En los casos donde se ha perdido la estructura, la molécula aún retiene muchas interacciones, incrementándose el apilamiento debido a la pérdida de puentes de hidrógeno.

Capítulo VII
OTRAS PUBLICACIONES

-
- Meyer, T., Ferrer, C., Pérez, A., **Rueda, M.**, Bidón-Chanal, A., Luque, F. J., Laughton, C. & Orozco, M. (2006). Essential Dynamics: A tool for efficient trajectory compression and management. *Journal of Chemical Theory and Computation*. In press.
- Lopez de la Osa, J., Gonzalez, C., Gargallo, R., **Rueda, M.**, Cubero, E., Orozco, M., Avino, A. & Eritja, R. (2006). Destabilization of Quadruplex DNA by 8-Aminoguanine. *ChemBiochem* 7, 46-48.
- Noy, A., Meyer, T., **Rueda, M.**, Ferrer, C., Valencia, A., Perez, A., de La Cruz, X., López-Bes, J. M., Poupiana, R., Fernandez-Recio, J., Luque, F. J. & Orozco, M. (2006). Data mining of molecular dynamics trajectories of nucleic acids. *Journal of Biomolecular Structure & Dynamics* 23, 357-484.
- Orozco, M., Noy, A., Meyer, T., **Rueda, M.**, Ferrer, C., Valencia, A., Perez, A., Carrillo, O., Fernandez-Recio, J., de La Cruz, X. & Luque, F. J. (2005). Data mining of molecular dynamics trajectories of nucleic acids. *Book Chapter*.
- Perez, A., Blas, J. R., **Rueda, M.**, López-Bes, J. M., de La Cruz, X., Luque, F. J. & Orozco, M. (2005). Exploring the essential dynamics of B.DNA. *Journal of Chemical Theory and Computation* 1, 790-800.
- Orozco, M., **Rueda, M.**, Blas, J. R., Cubero, E., Luque, F. J. & Laughton, C. A. (2004). Molecular dynamics simulations of nucleic acids. *Encyclopedia of Computational Chemistry. Book Chapter*.
- Trygubenko, S. A., Bogdan, T. V., **Rueda, M.**, Orozco, M., Luque, F. J., Sponer, J., Slavicek, P. & Hobza, P. (2002). Correlated ab initio study of nucleic acid bases and their tautomers in the gas phase, in a microhydrated environment and in

aqueous solution. Part 1.Cytosine. *Physical Chemistry Chemical Physics* 4, 4192-4203.

Rueda, M., Lopez, J. M., Luque, F. J. & Orozco, M. (2001a). Amino-Imino Tautomerism in Derivatives of Cytosine: Effect on Hydrogen-Bonding and Stacking Properties. *Journal of Physical Chemistry A* 105, 6575-6580.

Rueda, M., Luque, F. J. & Orozco, M. (2001b). A theoretical investigation on the effect of remote amino groups in hydrogen bonding of nucleic acids. *Biopolymers* 61, 52-60.

Agradecimientos

Después de estos 6 *primeros* años dentro del mundo de la investigación me gustaría dar las gracias a varias personas:

En primer lugar, al Prof. Modesto Orozco y al Prof. F. Javier Luque por brindarme la oportunidad de introducirme en este campo, tan fascinante como complejo, que es el Modelado Molecular. Uno nunca sabrá lo suficiente de él, ni aún viviendo varias vidas. También agradezco de ellos su comprensión y paciencia frente a los desastres.

Al recién “Doctor” Carles Ferrer por todos los buenos ratos que me ha hecho pasar, tanto en *Químicas* como en el Parc. Espero podamos seguir arreglando muchos *she bangs* en el futuro ;)

A Alberto Pérez, por sus consejos, su humor y por su afán en enseñarme cuán útiles pueden ser las matemáticas y las fórmulas en Bioquímica ☺.

A José Ramón Blas por su ayuda sincera frente a las adversidades computacionales. Dos biólogos en un mismo pajar.

Al Dr. Josep Lluís Gelpí por su conocimiento infinito de *todas* las materias. Gracias por alumbrar mi ofuscación en tantas ocasiones.

Al Dr. Xavier de La Cruz y al Dr. Juan F. Recio por sus sabios consejos, tanto laborales como personales.

A Ivan Párraga y Tim Meyer por la simbiosis de conocimientos y dolores de cabeza conjuntos que hemos sufrido durante el último año con el proyecto MoDEL.

A los Drs. Elena Cubero, Begoña Hernández, Xavier Barril, Carles Curutchet, y Robert Soliva por su ayuda inicial, así como los actuales miembros del grupo *mmb*. Gracias por mantener ese clima agradable de trabajo dentro del laboratorio.

Al Prof. Charlie Loughton y al Prof. Klaus Schulten por acogerme durante mis estancias breves en sus respectivos grupos de investigación en Inglaterra y Estados Unidos.

A mi familia por el apoyo incondicional que me da en todo.

A Nuria, por soportarme *incluso* durante la escritura de esta tesis. Si existe el cielo, sin duda hay un sitio para ti en él.

El presente trabajo se ha llevado a cabo gracias a una beca predoctoral del Ministerio de Educación y Ciencia.

Castellbisbal, Octubre de 2005.