

Departament de Química Física
Institut de Recerca Biomédica
Universitat de Barcelona

LA FLEXIBILIDAD EN LOS ÁCIDOS NUCLEICOS:
UN ESTUDIO DE DINÁMICA MOLECULAR

Agnès Noy Freixa

Programa de Doctorado de Química Teórica y Computacional

Departament de Química Física

Universitat de Barcelona

Bienio 2002-2004

Esta tesis ha sido realizada por la licenciada Agnès Noy Freixa
bajo la dirección de los Drs. Modesto Orozco López y F. Javier Luque Garriga,
y con la tutoría de Juan J. Novoa, para optar al título de doctor por la
Universitat de Barcelona

ÍNDICE

| | |
|--|----|
| 1. <u>INTRODUCCIÓN</u> | 1 |
| 2. <u>MÉTODOS TEÓRICOS</u> | 7 |
| 2.1 La superficie de energía potencial. | 10 |
| 2.1.1 La aproximación de Born-Oppenheimer. | 10 |
| 2.1.2 La ecuación de Schrödinger electrónica. | 11 |
| 2.1.2.1 El método Hartree-Fock. | 11 |
| 2.1.2.2 El método Møller-Plesset. | 12 |
| 2.1.3 El campo de fuerza. | 13 |
| 2.1.3.1 Términos de enlace. | 13 |
| 2.1.3.2 Términos de no enlace. | 15 |
| 2.1.4 La parametrización. | 16 |
| 2.1.4.1 Términos de enlace. | 16 |
| 2.1.4.2 Términos de no enlace. | 16 |
| 2.2 Métodos de exploración. | 18 |
| 2.2.1 La mecánica molecular. | 18 |
| 2.2.1.1 <i>Steepest Descent</i> | 18 |
| 2.2.1.2 <i>Conjugate Gradient</i> | 19 |
| 2.2.2 La dinámica molecular. | 19 |
| 2.2.2.1 Las ecuaciones y métodos de integración. | 20 |
| 2.2.2.2 Aspectos prácticos de las simulaciones por ordenador. | 21 |
| 2.2.2.3 La preparación del sistema. | 25 |
| 2.2.2.4 Limitaciones y comparaciones. | 26 |
| 2.2.3 Cálculos de energía libre aplicados a grandes transiciones conformacionales. | 27 |
| 2.2.3.1 <i>Umbrella Sampling</i> | 28 |
| 2.2.3.2 WHAM. | 29 |
| 2.2.3.3 <i>Targeted Molecular Dynamics</i> | 30 |

| | |
|---|----|
| 2.2.4 Métodos de Monte Carlo. | 31 |
| 2.3 Análisis de resultados. | 33 |
| 2.3.1 Promedio y convergencia. | 33 |
| 2.3.1.1 La hipótesis ergódica. | 33 |
| 2.3.1.2 Convergencia. | 33 |
| 2.3.1.3 el RMSd. | 34 |
| 2.3.1.4 Potencial de interacción molecular. | 34 |
| 2.3.2 Análisis de dinámica esencial. | 35 |
| 2.3.2.1 Análisis de componentes principales. | 35 |
| 2.3.2.2 Índices de similitud. | 36 |
| 2.3.2.3 Distancia de Mahalanobis. | 36 |
| 2.3.3 Análisis de elasticidad. | 38 |
| 2.3.3.1 La aproximación cuasi-harmónica. | 38 |
| 2.3.3.2 La elasticidad de los modos esenciales. | 38 |
| 2.3.3.3 Índices de similitud ponderados. | 39 |
| 2.3.3.4 Los ácidos nucleicos como polímeros. | 40 |
| 2.3.4 La entropía configuracional. | 41 |
| 2.3.4.1 El método de Schlitter. | 41 |
| 2.3.4.2 El método de Andricioaei y Karplus. | 42 |
| 2.3.4.3 Extrapolación a tiempo infinito. | 42 |
| 3. <u>LOS ÁCIDOS NUCLEICOS</u> | 45 |
| 3.1 Estructura primaria. | 48 |
| 3.1.1 Componentes estructurales. | 48 |
| 3.1.2 Conformación. | 49 |
| 3.1.2.1. El anillo furanósico | 50 |
| 3.1.2.2 El enlace glucosídico (χ) | 52 |
| 3.1.2.3 El enlace exocíclico C4'-C5' (γ). | 53 |
| 3.1.2.4 Los enlaces C-O (β, ϵ) y P-O (α, ζ) | 53 |
| 3.1.3 Propiedades fisicoquímicas. | 54 |
| 3.1.3.1 Ionización. | 54 |
| 3.1.3.2 Densidad de carga. | 54 |
| 3.1.3.3 Tautomería. | 55 |

| | |
|---|----|
| 3.2 Estructura secundaria. | 57 |
| 3.2.1 Interacciones no covalentes. | 57 |
| 3.2.1.1 Puentes de hidrógeno. | 57 |
| 3.2.1.2 Interacciones de apilamiento o <i>stacking</i> | 60 |
| 3.2.1.3 El efecto del solvente. | 61 |
| 3.2.2 Parámetros helicoidales. | 62 |
| 3.2.2.1 Parámetros entre pares de bases. | 62 |
| 3.2.2.2 Parámetros entre bases complementarias. | 64 |
| 3.2.2.3 Parámetros globales. | 66 |
| 3.2.2.4 El modelo <i>Worm-like Chain</i> (WLC) | 67 |
| 3.2.3 Conformaciones canónicas. | 67 |
| 3.2.3.1 Formas dextrógiras. La forma A y B. | 67 |
| 3.2.3.2 Forma Levógira. La forma Z. | 70 |
| 3.2.4 Conformaciones no canónicas. | 71 |
| 3.3 Estructura terciaria y función biológica. | 75 |
| 3.3.1 El dogma de la biología molecular. | 75 |
| 3.3.1.1 La replicación. | 76 |
| 3.3.1.2 La transcripción. | 77 |
| 3.3.1.3 La traducción. | 77 |
| 3.3.1.4 Los retrovirus. | 79 |
| 3.3.2 La terapia antisentido. | 79 |
| 3.3.2.1 La RNasa H. | 81 |
| 3.3.2.2 Oligonucleótidos antisentido de 1ª generación. | 82 |
| 3.3.2.3 Oligonucleótidos antisentido de 2ª generación. | 82 |
| 3.3.2.4 El ARN de interferencia. | 84 |
| 3.3.2.5 Las ribozimas. | 85 |
| 3.3.3 Curvatura y flexibilidad del ADN. | 85 |
| 3.3.3.1 Dependencia de estructura y flexibilidad con la secuencia. | 87 |
| 3.3.3.2 La transición B↔A. | 89 |
| 3.3.4 Técnicas experimentales. | 91 |
| 3.3.4.1 Métodos de difracción de rayos X. | 92 |

| | | |
|---------|--|-----|
| 3.3.4.2 | La resonancia magnética nuclear (RMN). | 94 |
| 3.3.4.3 | Métodos de baja resolución. | 95 |
| 4. | <u>RESULTADOS</u> . | 97 |
| 4.1 | <i>Relative Flexibility of DNA and RNA:</i> | |
| a | <i>Molecular Dynamics Study.</i> | 99 |
| 4.2 | <i>Structure, Recognition Properties,</i> | |
| and | <i>Flexibility of the DNA·RNA Hybrid.</i> | 113 |
| 4.3 | <i>Theoretical Analysis of Antisense Duplexes:</i> | |
| D | <i>eterminants of the RNase H Susceptibility.</i> | 127 |
| 4.4 | <i>Theoretical Study of Large Conformational Transitions in DNA:</i> | |
| the | <i>B↔A Conformational Change in Water and Ethanol/water.</i> | 165 |
| 5. | <u>DISCUSIÓN</u> . | 179 |
| 5.1 | Estudio Comparativo de la Flexibilidad en las | |
| M | oléculas de ADN y ARN. | 181 |
| 5.1.1 | Motivación. | 181 |
| 5.1.2 | Plan de trabajo. | 182 |
| 5.1.3 | Discusión de los principales resultados. | 183 |
| 5.1.4 | Conclusión. | 189 |
| 5.2 | Estructura, Propiedades de Reconocimiento y | |
| F | lexibilidad del Híbrido ADN·ARN. | 190 |
| 5.2.1 | Motivación. | 190 |
| 5.2.2 | Plan de trabajo. | 190 |
| 5.2.3 | Discusión de los principales resultados. | 191 |
| 5.2.4 | Conclusión. | 196 |
| 5.3 | Análisis Teórico de Dúplexes Antisentido: | |
| D | eterminantes de la Susceptibilidad hacia la RNasa H. | 198 |
| 5.3.1 | Motivación. | 198 |
| 5.3.2 | Plan de trabajo. | 198 |
| 5.3.3 | Discusión de los principales resultados. | 201 |
| 5.3.4 | Conclusión. | 206 |
| 5.4 | Estudio Teórico de Grandes Transiciones Conformacionales | |
| en | el ADN: el Cambio B↔A en Agua y Etanol/Agua. | 207 |

| | |
|--|-----|
| 5.4.1 Motivación. | 207 |
| 5.4.2 Plan de trabajo. | 208 |
| 5.4.3 Discusión de los principales resultados. | 208 |
| 5.4.4 Conclusión. | 213 |
| 6. <u>CONCLUSIONES</u> | 215 |
| 7. <u>BIBLIOGRAFÍA</u> | 219 |
| 8. <u>APÉNDICES</u> | 231 |
| A. Abreviaturas. | 233 |
| B. Otras publicaciones. | 237 |
| C. Agradecimientos. | 239 |

1. INTRODUCCIÓN

La química teórica aplicada a la bioquímica

La química teórica se empezó a desarrollar hace más de un siglo, pero para pasar de la mera representación formal a la aplicación práctica en algoritmos capaces de afrontar sistemas reales, se ha tenido que esperar al desarrollo, mucho más tardío, de la tecnología computacional. De esta manera, el vertiginoso aumento de la potencia de cálculo de los ordenadores en los últimos años ha convertido la química computacional en una disciplina de amplio impacto en la investigación química y biológica.

El objetivo de las técnicas computacionales aplicadas a la bioquímica es proporcionar una comprensión de los procesos que tienen lugar en los seres vivos siguiendo los principios básicos de la química y de la física. La aplicación de estos fundamentos a los sistemas macromoleculares que caracterizan la célula presenta una dificultad añadida derivada de su enorme tamaño. Por esto y a pesar del prometedor desarrollo de los ordenadores, es necesario simplificar el formalismo teórico. Así, del uso de la teoría cuántica propio de los sistemas químicos sencillos, se pasa a algoritmos clásicos en los que el hamiltoniano se simplifica por un campo de fuerza, donde solo se consideran explícitamente los términos nucleares introduciendo los efectos electrónicos en forma paramétrica.

A pesar de las limitaciones evidentes de las técnicas clásicas para describir el enlace químico, se puede obtener información muy relevante a nivel conformacional y termodinámico de las grandes moléculas biológicas. Así, se pueden calcular geometrías moleculares y sus energías relativas, barreras torsionales, energías de interacción, frecuencias vibracionales, calores de formación, muestreos en transiciones temporales, se pueden refinar estructuras obtenidas mediante rayos-X y RMN (Resonancia Magnética Nuclear) etc... No obstante su eficiencia computacional, existe aun una limitación importante en los sistemas clásicos en cuanto al tamaño del modelo simulado (1). Si bien se han publicado simulaciones masivas con más de un millón de átomos (2), a la práctica son raras simulaciones con más de 10^5 átomos, lo que a efectos de los ácidos nucleicos corresponde a unos dos giros de ADN estando muy lejos de fragmentos de interés como un promotor o todo un gen. También, la longitud de las simulaciones en el caso de la dinámica molecular (DM) (3) está limitada mayoritariamente a los 10-100 nanosegundos (ns), escala de tiempo corta comparada con

muchas de las transiciones temporales de importancia biológica, aunque algunas simulaciones de longitud superior han sido publicadas (4).

Un episodio de la evolución de los ordenadores se detecta en el desarrollo de mi propia tesis. En los primeros años, la longitud de las trayectorias estaba alrededor de 10 ns, mientras que al final de la misma, la longitud promedio ha aumentado a 100 ns. Teniendo en cuenta esta tendencia, podemos esperar que las simulaciones en el rango del microsegundo sean lo habitual en un plazo muy corto. Así, aunque en la actualidad estemos limitados a la hora de estudiar moléculas de interés biológico, hemos aumentado nuestra capacidad de aportar información relevante de manera vertiginosa y siguiendo la misma tendencia, se pueden prever estudios muy interesantes en un futuro muy cercano.

La química computacional en los ácidos nucleicos

La genética, la bioquímica y la biología molecular tal y como las entendemos hubieran sido imposibles sin el conocimiento de la estructura de los ácidos nucleicos y en concreto la de la doble hélice del ADN que permite una conexión directa entre la química y la biología al explicar la transmisión de la información genética a partir de simples principios químicos (5). El perfeccionamiento constante de las técnicas experimentales basadas en rayos X y RMN ha revelado que el ADN contiene no solo información en la lectura directa de su secuencia, sino además en su estructura tridimensional (6). Se ha visto que tanto la configuración de los surcos como la facilidad de deformarse dependen de la secuencia y son responsables de modular las interacciones con moléculas que regulan la funcionalidad del ADN, principalmente las proteínas (7,8). Así, podemos decir que el ADN tiene un código primario o de lectura de secuencia que corresponde al código genético y uno secundario mediante el cual se regula la funcionalidad de los genes y que va ligado a un mecanismo indirecto de lectura basado en su estructura tridimensional.

Asimismo es importante considerar el resto de ácidos nucleicos de la célula. Si el ADN es el portador de la información, el ARN es el ácido nucleico responsable de su descodificación transformándola a código proteico e incluso de controlar parte de la expresión de los genes. A diferencia del ADN, el ARN presenta una estructura mucho más compleja, aunque

habitualmente construida a partir de la doble hélice. El ARN lo encontramos básicamente como ARN mensajero (código ARN), ARN de transferencia y ribosomal (encargados de la síntesis de proteínas), y ARN no codificante de control de expresión (9). A parte de los homopolímeros, se sintetiza también el dúplex ADN·ARN o híbrido a lo largo de la vida celular en los procesos de replicación y transcripción del ADN (9).

Los objetivos de esta tesis

Una cuestión no resuelta hasta ahora es como la célula distingue estas moléculas que en muchos casos se presentan en forma de doble hélice. Años de estructuras experimentales no han resuelto este enigma que probablemente requiera un análisis dinámico el cual a día de hoy solamente es posible en estudios teóricos. Precisamente, la DM confiere la posibilidad de extender la información experimental dando un conjunto de configuraciones representativas a temperatura ambiente y así, poder estudiar la información contenida en la estructura tridimensional y en la flexibilidad de los ácidos nucleicos (10).

En esta tesis se han usado las técnicas clásicas de la DM con la finalidad global de estudiar las propiedades estructurales y dinámicas de los ácidos nucleicos en forma de doble hélice antiparalela presentes en la célula e intentar relacionarlas con las características funcionales de las mismas. De este modo los objetivos son:

1. Estudio comparativo de la flexibilidad en las moléculas de ADN y ARN.

Experimentos tanto de baja resolución como de RMN y rayos X han dado la idea a la comunidad científica que la molécula de ADN es más flexible que la de ARN (11,12), pero viendo la anisotropía y riqueza conformacional de las mismas esto parece cuestionable. Nuestro objetivo ha sido estudiar la flexibilidad de ambas estructuras y determinar hasta que punto la célula puede usar esta diferencia para marcar las distintas funciones de estos dos ácidos nucleicos.

2. Estructura, propiedades de reconocimiento y flexibilidad del híbrido de ADN·ARN

Nos planteamos estudiar la molécula híbrida de ADN·ARN con la finalidad de describir sus propiedades físicas y ver en que grado son parecidas a las de los homopolímeros de ADN

o ARN para así entender el mecanismo de especificidad de la RNasa H, enzima que reconoce el híbrido de ADN·ARN (no ADN₂ o ARN₂) cortando la cadena de ARN (13) y que es clave en la terapia antisentido.

3. Análisis teórico de los dúplexes antisentido: determinantes en la susceptibilidad de la RNasa H

Hemos estudiado las propiedades de un conjunto representativo de híbridos (AXN·ARN) desarrollados como candidatos para la terapia antisentido. En concreto, hemos analizado en detalle que hace que, con una estructura promedio similar, unos sean reconocidos y degradados por la RNasa H y los otros no.

4. Estudio teórico de grandes transiciones conformacionales en el ADN: el cambio de la conformación B↔A en agua y etanol/agua.

En este trabajo se explora la posibilidad de utilizar algoritmos de DM sesgada para dirigir la trayectoria según una coordenada de reacción que nos permita estudiar grandes transiciones conformacionales. Concretamente, con el uso de los métodos de *Umbrella Sampling* (US) (14) y *Weighted Histogram Analysis Methods* (WHAM) (15) acoplados por primera vez a técnicas de *Targeted Molecular Dynamics* (TMD) (16) pretendemos calcular la diferencia de energía libre, así como averiguar cuales son los grados de libertad determinantes. En concreto, hemos analizado la transición de las conformación B↔A (en agua y etanol) debido a su importancia biológica (16,17) y a lo bien caracterizadas que están experimentalmente. La forma B es la más estable en agua, mientras que la A aparece cuando el ADN está unido a algunas proteínas tan importantes como polimerasas (17) y en ambientes apolares (18).

2. MÉTODOS TEÓRICOS

La gran variedad de métodos de la química teórica permite abordar una amplia gama de sistemas bioquímicos. Generalmente, a medida que aumenta el nivel teórico, aumenta la calidad de los resultados, así como su coste computacional, lo cual a su vez limita el tamaño del sistema susceptible de estudio. Así, la elección del método teórico depende de las propiedades que son objeto de interés y por lo tanto del nivel de detalle que se quiere alcanzar, del modelo molecular empleado para simular el sistema, y finalmente del coste computacional.

La mayoría de los métodos de la química teórica se sustenta en la aproximación de Born-Oppenheimer (BO) (19), que permite dividir el estudio de la estructura molecular en dos etapas. La primera consiste en estudiar la problemática electrónica para diferentes geometrías nucleares fijas y, junto a la repulsión nuclear, definir la superficie de energía potencial (SEP). La segunda trata de explorar esta superficie para encontrar puntos críticos y determinar parámetros macroscópicos como la energía libre, entropía o entalpía de un proceso químico. Dentro de la aproximación BO, podemos adoptar dos estrategias para determinar la energía del sistema según i) una representación cuántica y ii) en base a una aproximación clásica. En el primer caso se plantea la ecuación de Schrödinger electrónica para cada configuración nuclear, y se trata posteriormente el movimiento cuántico de éstos. En el segundo caso, el movimiento nuclear se simplifica mediante el campo de fuerza, donde el efecto electrónico queda recogido implícitamente mediante parámetros ajustados a datos experimentales o mecano-cuánticos.

2.1 LA SUPERFICIE DE ENERGÍA POTENCIAL

2.1.1 La aproximación de Born-Oppenheimer

Siguiendo los postulados de la mecánica cuántica, los estados estacionarios de una molécula (Ψ) vienen definidos por la ecuación de Schrödinger:

$$\hat{H}\Psi(x, z) = E\Psi(x, z) \quad [1]$$

donde x es el conjunto de coordenadas nucleares, z las electrónicas, E indica los valores de energía de los diferentes estados, y \hat{H} es el operador de energía o hamiltoniano que para una molécula tiene la forma:

$$\hat{H} = \hat{T}_{nuc} + \hat{T}_{ele} + \hat{V}_{nuc} + \hat{V}_{ele} + \hat{V}_{nuc-ele} \quad [2]$$

donde \hat{T} representa la energía cinética de núcleos (*nuc*) y electrones (*ele*), y \hat{V} la energía potencial coulombiana de repulsión de núcleos, electrones y atracción núcleo-electrón (*nuc-ele*).

Debido a la gran diferencia de masa entre núcleos y electrones, se puede asumir que la relajación electrónica respecto a los movimientos nucleares es instantánea. Ello permite separar el movimiento nuclear y electrónico, dando lugar a la formulación de las ecuaciones de Schrödinger electrónica [3,4] y nuclear [5,6]:

$$\hat{H}_{ele}\Psi_{ele}(z) = E_{ele}\Psi_{ele}(z) \quad [3]$$

donde Ψ_{ele} es la función de onda electrónica, E_{ele} designa la energía electrónica, y \hat{H}_{ele} es el hamiltoniano electrónico, que incluye paraméricamente las coordenadas nucleares:

$$\hat{H}_{ele} = \hat{T}_{ele} + \hat{V}_{ele} + \hat{V}_{nuc-ele} \quad [4]$$

$$\hat{H}_{nuc}\Psi_{nuc}(x) = E_{nuc}\Psi_{nuc}(x) \quad [5]$$

donde Ψ_{nuc} es la función de onda nuclear, E_{nuc} es la energía, y \hat{H}_{nuc} es el hamiltoniano nuclear, donde $U(x)$ es la energía potencial efectiva [7].

$$\hat{H}_{nuc} = \hat{T}_{nuc} + U(x) \quad [6]$$

$$U(x) = \hat{V}_{nuc} + E_{ele,x} \quad [7]$$

donde $E_{ele,x}$ es la energía electrónica para unas determinadas coordenadas nucleares.

2.1.2 La ecuación de Schrödinger electrónica

La resolución exacta de la ecuación de Schrödinger electrónica no es posible para los sistemas moleculares, por lo que usualmente su resolución se realiza con una serie de simplificaciones. Revisaremos aquí algunas de las más usadas, si bien se puede encontrar una presentación más detallada y extensa en otras fuentes (20-22).

2.1.2.1 El método Hartree-Fock

Por su simplicidad es uno de los métodos más comunes para resolver de manera aproximada la ecuación de Schrödinger. Tiene la ventaja de permitir acoplarlo a métodos más sofisticados como es el caso del método de Møller-Plesset (MP). Su principal aproximación consiste en no considerar explícitamente la correlación electrónica, ya que cada electrón se encuentra sujeto a la repulsión efectiva creada por el resto de electrones. Además, describe la función de onda como un único determinante de Slater, si bien proporciona una vía sistemática para encontrar la solución más parecida al estado fundamental gracias al método variacional (20).

Para una molécula, el hamiltoniano electrónico se puede expresar:

$$\hat{H} = \sum_{i=1}^n \hat{h}_i + \sum_{i=1}^{n-1} \sum_{j>i}^n \frac{1}{r_{ij}} \quad [8]$$

donde i es el número de electrones, r_{ij} la distancia entre ellos, y \hat{h}_i el hamiltoniano monoeléctrico:

$$\hat{h}_i = \hat{T}_i + \hat{V}_{nuc-i} \quad [9]$$

donde \hat{T}_i y \hat{V}_{nuc-i} son los términos de energía cinética y atracción núcleo-electrón, respectivamente. La repulsión interelectrónica hace que el hamiltoniano no sea separable en términos monoeléctricos. No obstante, ello se solventa definiendo el operador de Fock:

$$\hat{f}_i = \hat{h}_i + \widehat{v}_{re,i} \quad [10]$$

donde $\widehat{v}_{re,i}$ es un operador monoeléctrico que representa la repulsión efectiva promedio de la interacción entre un electrón y el resto de electrones.

La función de onda polieletrónica queda definida por:

$$\Psi(z_1, z_2, \dots, z_n) = |\phi_1(z_1), \phi_2(z_2), \dots, \phi_n(z_n)| \quad [11]$$

donde z son las coordenadas electrónicas, ϕ_i representa funciones monoeléctricas u orbitales moleculares y las barras indican el determinante de Slater. Los orbitales moleculares se expresan como combinación lineal de orbitales atómicos (aproximación LCAO; *Linear Combination of Atomic Orbitals* (23)):

$$\phi_i(z) = \sum_{m=1}^N c_{i,m} \chi_m \quad [12]$$

donde χ_m es el conjunto de funciones de base y $c_{i,m}$ son los coeficientes. Una buena aproximación es utilizar funciones parecidas a los orbitales hidrogenoides o denominadas funciones de Slater (STO) (24). Sin embargo, por razones de eficiencia computacional, las funciones de base se construyen como combinaciones lineales de funciones gaussianas (GTO) (25).

2.1.2.2 El método Møller-Plesset

El método MP mejora la descripción Hartree-Fock (HF) introduciendo la correlación electrónica. Este método se basa en la teoría de perturbaciones (26), según la cual el hamiltoniano del sistema se divide en dos partes. La primera es el hamiltoniano de un sistema de referencia, mientras que la segunda corresponde al término de perturbación, el cual permite introducir los efectos de la correlación electrónica (27) expresando el hamiltoniano del sistema como:

$$\widehat{H}_\lambda = \widehat{H}^0 + \lambda \widehat{H}' \quad [13]$$

donde \widehat{H}_λ es el hamiltoniano global que depende de λ , que es un parámetro adimensional que varía entre 0 y 1, \widehat{H}^0 es el hamiltoniano de HF, y \widehat{H}' es el operador de perturbación que tiene la forma precisa para obtener el hamiltoniano exacto:

$$\widehat{H}' = \sum_{i=1}^{n-1} \sum_{j>i}^n \frac{1}{r_{ij}} - \sum_{i=1}^n \widehat{v}_{re,i} \quad [14]$$

donde $\widehat{v}_{re,i}$ es el operador de repulsión efectiva definido por el método HF.

La función de onda (Ψ_λ) y la energía (E_λ) global del sistema pueden expresarse como una expansión en serie según:

$$\Psi_\lambda = \Psi^{(0)} + \lambda\Psi^{(1)} + \lambda^2\Psi^{(2)} + \dots \quad [15]$$

$$E_\lambda = E^{(0)} + \lambda E^{(1)} + \lambda^2 E^{(2)} + \dots \quad [16]$$

donde los diferentes términos $\Psi^{(n)}$ y $E^{(n)}$ corresponden a las correcciones de orden-n de la función de onda y energía. La energía HF es la suma de los términos 0 y 1, de manera que las correcciones para la correlación electrónica empiezan en el orden 2, dando lugar a cálculos MP2, MP3, MP4...

2.1.3 El campo de fuerza

En los métodos clásicos el hamiltoniano se expresa mediante ecuaciones simples que dependen de las posiciones nucleares formando el campo de fuerza. Este se podría definir como la dependencia empírica de la energía potencial respecto a la geometría de los núcleos, donde el efecto de los electrones se introduce implícitamente mediante un conjunto de parámetros. Estos potenciales efectivos permiten estudiar sistemas de gran tamaño siempre que no haya alteración de enlaces covalentes ni cambios drásticos en la distribución electrónica. En el caso de los ácidos nucleicos, los campos de fuerza más utilizados son AMBER (28) y CHARMM (29), que han sido capaces de reproducir estructuras y propiedades macroscópicas tal y como muestran revisiones recientes sobre el campo (10,30-32). A nuestra experiencia, los mejores resultados para el estudio de formas canónicas e inusuales de los ácidos nucleicos se obtienen con el campo de fuerza de AMBER (33-35), que ha sido el utilizado en esta tesis. Así, pues, nos centraremos en explicar este campo de fuerza.

2.1.3.1 Términos de enlace

Longitud del enlace

La curva de energía potencial que describe un enlace típico tiene la forma de la figura 1. No obstante, por simplicidad se considera la aproximación armónica:

$$U = K_{AB}(l - l_0)^2 \quad [17]$$

donde l_0 es el valor de equilibrio de la distancia entre los átomos A y B, y K_{AB} es la constante de fuerza de *stretching*. Como se puede ver en la figura 1, la expresión armónica sólo da buen resultado en valores cercanos al equilibrio.

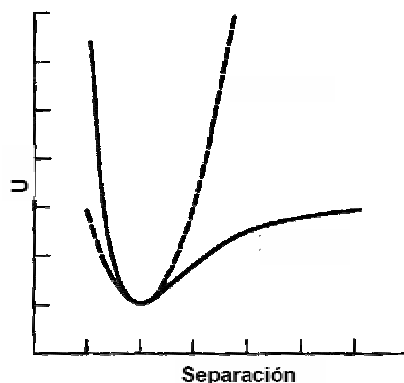


Figura 1. Variación en la energía de enlace en función de la separación interatómica (línea continua).
Aproximación obtenida con una expresión armónica (línea discontinua)

Ángulos de enlace

La dependencia de la energía con la variación del ángulo de enlace se describe también mediante la aproximación armónica:

$$U = K_{ABC}(\theta - \theta_0)^2 \quad [18]$$

donde θ_0 es el ángulo de equilibrio entre los átomos A, B y C, y K_{ABC} es la constante de fuerza de *bending*.

Ángulos diedros o torsiones

La energía asociada a la rotación respecto a un enlace químico se describe mediante series de Fourier:

$$U = \sum_{n=1}^N V_n / 2 [1 + \cos(n\omega - \partial)] \quad [19]$$

donde V_n es la barrera torsional, n es la periodicidad, ω es el ángulo de torsión y ∂ es el ángulo de fase.

Torsiones impropias

Este término se encarga de describir la relación geométrica entre cuatro átomos no enlazados, normalmente para mantenerlos en un mismo plano (como es el caso de los sistemas aromáticos) o para tener el isómero correcto en átomos asimétricos (como es el caso del carbono $C\alpha$ en aminoácidos). Dichas torsiones se tratan como ángulos diedros aunque no estén enlazados, usándose series de Fourier.

2.1.3.2 Términos de no enlace

La energía electrostática

Dado que la distribución de carga molecular se describe mediante cargas atómicas parciales, la interacción electrostática se reduce a una sumatoria de términos coulómicos:

$$U = \frac{q_A q_B}{\epsilon R_{AB}} \quad [20]$$

donde q_A y q_B son las cargas parciales sobre los átomos A y B respectivamente, ϵ es la constante dieléctrica y R_{AB} es la distancia entre ellos.

La energía de van der Waals

Las fuerzas de dispersión-repulsión se describen mediante un potencial de Lennard-Jones, que tiene la siguiente forma para la interacción entre dos átomos:

$$U = 4\epsilon \left[\left(\frac{\sigma}{R_{AB}} \right)^{12} - \left(\frac{\sigma}{R_{AB}} \right)^6 \right] \quad [21]$$

donde ϵ es la profundidad del pozo, y σ el diámetro de colisión o el punto en que la energía es cero (figura 2).

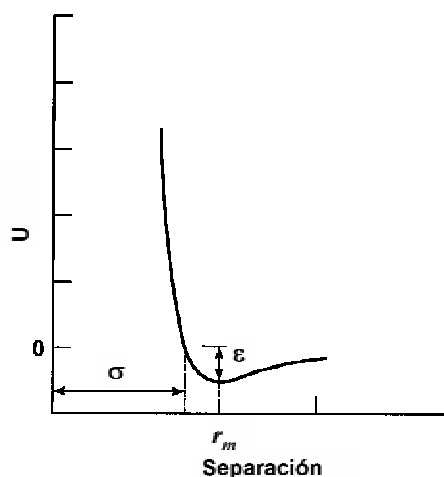


Figura 2. Representación energética del potencial de Lennard-Jones.

En el campo de fuerza de AMBER, la ecuación 21 se expresa en función del parámetro r_m que corresponde a la distancia de mínima energía (figura 2).

$$U = \varepsilon \left[\left(\frac{r_m}{R_{AB}} \right)^{12} - \left(\frac{r_m}{R_{AB}} \right)^6 \right] \quad [22]$$

donde $r_m = 2^{1/6} \sigma$.

2.1.4 La parametrización

Los parámetros que aparecen en las ecuaciones anteriores se determinan en base a valores experimentales u obtenidos a partir de cálculos mecano-cuánticos. Para ello, se asume cierta transferencia de los parámetros obtenidos a partir de modelos moleculares pequeños a moléculas más grandes. En esta tesis se han usado los parámetros parm99 (36,37) y parm-*bsc0* (38) del campo de fuerza de AMBER, los cuales están especializados en proteínas y ácidos nucleicos. Además, para obtener los parámetros asociados a determinadas modificaciones, se ha utilizado el campo de fuerza GAFF (*General Amber Force Field*) (39) que presenta un carácter más global abarcando la mayoría de moléculas orgánicas.

2.1.4.1 Términos de enlace

Se acostumbran a utilizar datos de espectroscopia IR, Raman, RMN, rayos X o difracción de neutrones. En los términos de torsión es habitual usar una parametrización cuántica debido a la ausencia de valores experimentales.

2.1.4.2 Términos de no enlace

La energía electrostática

Las cargas parciales se obtienen de una parametrización cuántica. En las primeras versiones del campo de fuerza AMBER (40) se usó el método ESP, que consiste en el ajuste del potencial electrostático determinado cuánticamente en un conjunto de puntos alrededor de la molécula. No obstante, se usaron niveles de cálculo cuántico poco precisos (HF/STO-3G),

que daban resultados mediocres sobre todo en ácidos nucleicos. En versiones posteriores (36), se usaron cálculos HF/6-31G(d) como referencia, que tienden a sobrestimar los momentos dipolares y por ello simulaban el efecto polarizante del agua. Paralelamente, se introdujo la metodología RESP (41) para solucionar ciertas desviaciones presentes en cargas de átomos enterrados. En las parametrizaciones más actuales de AMBER (2002), se usan niveles muy altos de cálculo (B3LYP/cc-pVTZ), con representaciones multicéntricas de carga (42) incorporando el efecto de polarización permanente, además de introducir la polarización inducida (43). Estos campos de fuerzas aún no han sido ampliamente aplicados al estudio de ácidos nucleicos (44).

La energía de van der Waals

Los parámetros pueden obtenerse tanto de datos mecano-cuánticos como experimentales. Teóricamente se tendrían que obtener las constantes para cada par distinto de átomos. No obstante, se aplican las normas combinatorias de Lorentz-Berthelot (45), donde el diámetro de colisión (σ_{AB}) sale de la media aritmética y la profundidad del pozo (ϵ_{AB}) de la media geométrica para los átomos interaccionantes:

$$\sigma_{AB} = \frac{1}{2}(\sigma_{AA} + \sigma_{BB}) \quad [23]$$

$$\epsilon_{AB} = \sqrt{\epsilon_{AA}\epsilon_{BB}} \quad [24]$$

El solvente

En las simulaciones realizadas en esta tesis, el agua se ha representado de forma explícita usando el modelo TIP3P (46) que es, junto al SPC (47), el más usado en la actualidad. En ambos la molécula de agua es rígida, se derivan las cargas de forma empírica y se considera un solo punto de interacción de van der Waals (vdW). A pesar de ser modelos simples, describen con elevada precisión las propiedades del agua en simulaciones clásicas.

2.2 MÉTODOS DE EXPLORACIÓN

En este apartado nos centraremos en las técnicas que usan las leyes de la mecánica clásica para tratar el movimiento de los núcleos. La SEP para biomoléculas es realmente compleja debido al gran número de grados de libertad que tienen y a su riqueza topológica, con gran cantidad de mínimos y barreras, siendo un caso especialmente difícil para los métodos de exploración.

2.2.1 La mecánica molecular

Es el cálculo de optimización que permite encontrar distintos puntos estacionarios en la SEP. En biomoléculas complejas resulta impracticable obtener y caracterizar todos estos puntos, aunque sí es conveniente obtener una estructura estable propia de un mínimo local para empezar una trayectoria de DM. Para ello hemos aplicado algoritmos iterativos de cálculo de derivada primera.

2.2.1.1 *Steepest descent*

Consiste en seguir el vector de máximo gradiente para definir la geometría en la siguiente iteración.

$$x_k = x_{k-1} + \lambda_{k-1} s_{k-1} \quad [25]$$

donde x_k es el vector que representa la configuración en la iteración k , λ_{k-1} es un escalar que indica la magnitud del paso y s_{k-1} es el vector unitario de búsqueda:

$$s_{k-1} = -\frac{g(x_{k-1})}{|g(x_{k-1})|} \quad [26]$$

donde $g(x_{k-1})$ es el gradiente en la conformación de la iteración $k-1$.

Es un método robusto para acercarse al mínimo, aunque cuesta llegar a la convergencia. Por ello, suele usarse en los primeros pasos de minimización, cambiando después al algoritmo de *Conjugate gradient*

2.2.1.2 Conjugate gradient

En este caso el vector de búsqueda tiene memoria de las direcciones exploradas en pasos iterativos anteriores:

$$s_{k-1} = -\frac{g(x_{k-1})}{|g(x_{k-1})|} + b_{k-1}s_{k-2} \quad [27]$$

donde x_k es el vector que representa la configuración en la iteración k , s_{k-1} es el vector unitario de búsqueda, $g(x_{k-1})$ es el gradiente en la conformación de la iteración $k-1$ y b_{k-1} es el peso relativo:

$$b_{k-1} = \frac{|g_{k-1}|^2}{|g_{k-2}|^2} \quad [28]$$

Esta técnica es menos eficiente que *Steepest descent* cuando la geometría inicial es poco precisa, pero más robusta en los últimos pasos de minimización.

2.2.2 La dinámica molecular

Hace casi medio siglo se realizó la primera DM de un sistema de fase condensada (2), pero no fue hasta mediados de los 70 que se realizó la primera simulación sobre una biomolécula (48) y la primera proteína (49), y ya en los ochenta el primer ácido nucleico (50). Estos constituyeron los pasos iniciales en la creciente aplicación de la DM al estudio de sistemas biológicos. Concretamente, desde los años noventa, el uso de la DM para el estudio de los ácidos nucleicos se ha incrementado notablemente, tal como se puede ver en la figura 3.

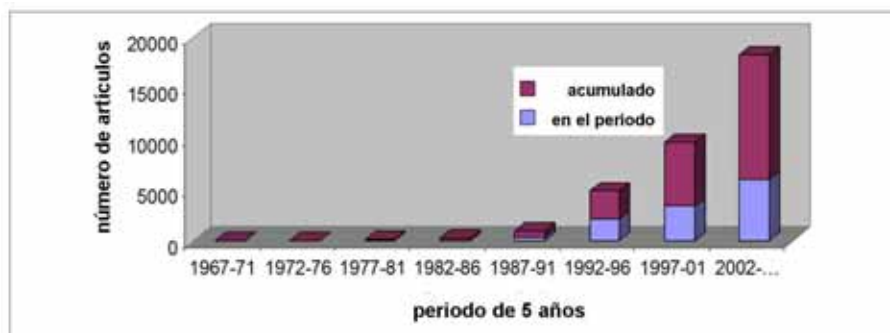


Figura 3. Evolución en el número de citaciones de artículos según ISI

(<http://portal.isiknowledge.com/portal.cgi>) con las palabras clave “(molecular dynamics or simulation or modelling or theoretical) and (nucleic acids or DNA or RNA or ribosome or ribozyme or triplex or tetraplex)”.

2.2.2.1 Las ecuaciones y métodos de integración

La DM explora la superficie conformacional con un algoritmo determinista que consiste en seguir la evolución a lo largo de una trayectoria temporal usando las ecuaciones de la mecánica clásica de Newton:

$$-\frac{dU}{dx} = m \frac{d^2x}{dt^2} \quad [29]$$

donde x son las coordenadas nucleares y U la energía del sistema.

Estas ecuaciones no se pueden resolver analíticamente debido a que la fuerza ejercida sobre un átomo depende de la posición de los otros, dando un movimiento acoplado de partículas. Por esto, se usa el método de diferencias finitas, donde en base a la fuerza que actúa en cada átomo y junto a las coordenadas y velocidades se calculan las futuras coordenadas, asumiendo que dentro este intervalo temporal la fuerza será constante. Hay diversos algoritmos para integrar las ecuaciones dinámicas dentro del método de las diferencias finitas. El utilizado en el caso de AMBER y que además es muy común es el algoritmo de *leap-frog* (51), que es una modificación del algoritmo de Verlet (52).

El algoritmo de Verlet

Se caracteriza por producir trayectorias estables tras largos tiempos de iteración. Considera las expansiones de Taylor alrededor de las configuraciones del paso posterior ($t+\Delta t$) y el anterior ($t-\Delta t$):

$$x(t + \Delta t) = x(t) + \frac{dx}{dt} \Delta t + \frac{1}{2} \frac{d^2x}{dt^2} \Delta t^2 + \dots \quad [30]$$

$$x(t - \Delta t) = x(t) - \frac{dx}{dt} \Delta t + \frac{1}{2} \frac{d^2x}{dt^2} \Delta t^2 + \dots \quad [31]$$

de manera que truncando en el segundo término tenemos:

$$x(t + \Delta t) = 2x(t) - x(t - \Delta t) + a(t)\Delta t^2 \quad [32]$$

donde x y a son los vectores de posición y aceleración respectivamente.

Se puede ver que el vector de velocidad no se determina directamente, dificultando las simulaciones a temperatura constante. También, la primera iteración resulta problemática, ya que normalmente no se tienen dos conformaciones para empezar la iteración.

El algoritmo de *leap-frog*

Se calculan las coordenadas a partir de las anteriores, junto con las velocidades evaluadas a medio paso de integración:

$$x(t + \Delta t) = x(t) + v(t + \frac{1}{2} \Delta t) \Delta t \quad [33]$$

$$v(t + \frac{1}{2} \Delta t) = v(t - \frac{1}{2} \Delta t) + a(t) \Delta t \quad [34]$$

donde v es el vector de velocidades.

El algoritmo es suficientemente robusto, si bien coordenadas y velocidades se deducen de forma no sincronizada (es decir, no se puede calcular la influencia de la energía cinética al mismo tiempo que la potencial), lo que puede generar algunos problemas de inestabilidad en la trayectoria. En la primera iteración, las velocidades son asignadas al azar restringiendo que globalmente se ajusten a una distribución de Boltzmann para una temperatura determinada.

El paso de integración

El paso de integración tiene que ser menor que el movimiento más rápido del sistema. En el caso de macromoléculas, esto supone una restricción muy severa debido a las rápidas vibraciones en la longitud de los enlaces. No obstante, la aplicación del método SHAKE (53), que usa multiplicadores de Lagrange para congelar estos grados de libertad, permite ampliar el paso de 1 a 2 femtosegundos (fs).

2.2.2.2 Aspectos prácticos de las simulaciones por ordenador

Límites del sistema simulado

La metodología clásica permite incorporar un número de aguas relativamente grande para solvatar el sistema. Aún así, no son suficientes para conseguir un solvente infinito y poder derivar propiedades macroscópicas. Por ello, es importante solucionar el problema de los límites del sistema. Las condiciones periódicas de contorno (PBC, *Periodic Boundary Conditions*) es el método que mejor lo consigue a base de introducir réplicas del sistema en

las tres dimensiones del espacio, de forma que una molécula pueda moverse de manera infinita entre la caja original de solvente y sus réplicas.

“Truncamiento del potencial” y la “Convención de imagen mínima”

El cálculo de las interacciones no enlazantes es computacionalmente la etapa limitante de la simulación. Dichas interacciones aumentan de forma cuadrática con el número de átomos a diferencia del resto, que lo hacen de forma lineal. Ello se puede paliar con la “Convención de imagen mínima”, que consiste en que cada átomo interactúa como máximo con la réplica más cercana de todos los átomos del sistema; y con una distancia de *cutoff* o de “Truncamiento de potencial”, a partir de la cual las interacciones se anulan. Esta aproximación se puede realizar con las interacciones de vdW, ya que decaen rápidamente con la distancia, pero puede ser arriesgado con el término coulombico, que decae mucho más lentamente.

Particle Mesh Ewald

Es el método utilizado para calcular la energía electrostática y solucionar así los problemas de las interacciones de largo alcance tan importantes en el caso de los ácidos nucleicos (54). En vez de usar un *cutoff*, cada átomo interactúa con los demás y todas sus réplicas intentando simular un sistema infinito. El método, originalmente propuesto por Ewald (55), sustituye el sumatorio de todas las interacciones en el sistema periódico por una suma de dos términos que convergen más rápidamente:

$$U_{ele} = U_{e_{dir}} + U_{e_{rec}} + U_0 \quad [35]$$

donde el primer y segundo término se refieren, respectivamente, al espacio directo y recíproco, y el término U_0 es una constante.

El espacio directo es de corto alcance y consiste en las interacciones directas de las cargas puntuales apantalladas por funciones gaussianas (figura 4). Por el otro lado, el espacio recíproco es de largo alcance, siendo una transformada de Fourier, y recoge la corrección derivada de las funciones gaussianas opuestas a las del primer término. Normalmente en el espacio directo se tienen en cuenta las interacciones en la caja original, mientras que en el espacio recíproco se incluyen las propias de las réplicas del sistema.

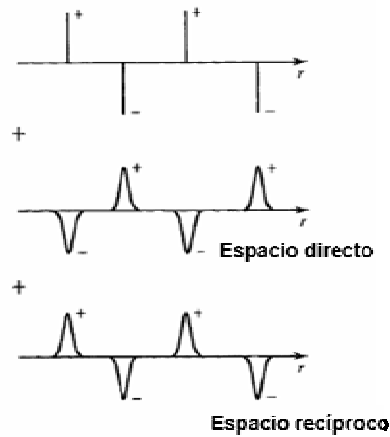


Figura 4. En el método del sumatorio de Ewald el conjunto de cargas iniciales son rodeadas por una distribución gaussiana (calculada en el espacio directo), para la cual se tiene que añadir una distribución de carga opuesta (calculada en el espacio recíproco)

En el método del sumatorio de Ewald la evaluación del espacio directo crece linealmente, mientras que la del espacio recíproco lo hace cuadráticamente. Para intentar optimizar el coste computacional de este último, se desarrolló el método *Particle Mesh Ewald* (PME) (56), donde se utiliza el procedimiento FFT (*Fast Fourier Transform*), que consigue escalar a $N \ln N$. La implementación de este método ha significado una mejora considerable en las simulaciones de DM aplicadas a sistemas biológicos, siendo su uso crucial en el caso de los ácidos nucleicos debido a su alta densidad de carga. No obstante, este método, que se desarrolló originalmente para sistemas cristalinos, pone de relieve la periodicidad derivada del tratamiento de los límites del sistema, que en una solución pueden ser erróneos (57).

Temperatura y presión constante

Debido a que la mayoría de medidas experimentales se realizan a temperatura y presión constante, es mejor realizar la simulación en el colectivo isotérmico-isobárico. Por esto, se usan las ecuaciones de Langevin, donde el sistema está acoplado a un baño térmico que actúa como reservorio suministrando o eliminando energía térmica, de forma que el sistema mantiene su temperatura constante. A la práctica, ello se consigue mediante el algoritmo de Berendsen (58), donde el cambio de temperatura es proporcional a las diferencias entre las temperaturas del sistema y del baño:

$$\frac{dT(t)}{dt} = \frac{1}{\tau_T} (T_0 - T(t)) \quad [36]$$

donde τ_T es el parámetro de acoplamiento y T_0 la temperatura del baño. Las velocidades de los átomos se escalan según:

$$\lambda^2 = 1 + \frac{\Delta t}{\tau_T} \left(\frac{T_0}{T(t)} - 1 \right) \quad [37]$$

Para la presión se actúa de forma análoga, acoplando el sistema a un baño de presión y escalando las coordenadas para cambiar el volumen:

$$\frac{dP(t)}{dt} = \frac{1}{\tau_p} (P_0 - P(t)) \quad [38]$$

donde τ_p es el parámetro de acoplamiento y P_0 la presión del baño. Las coordenadas de los átomos se escalan según:

$$\lambda = 1 - \kappa \frac{\Delta t}{\tau_p} (P(t) - P_0) \quad [39]$$

donde κ es la compresibilidad isotérmica:

$$\kappa = -\frac{1}{V} \left(\frac{\delta V}{\delta P} \right)_T \quad [40]$$

De esta manera el esquema de la DM se puede resumir en la figura 5:

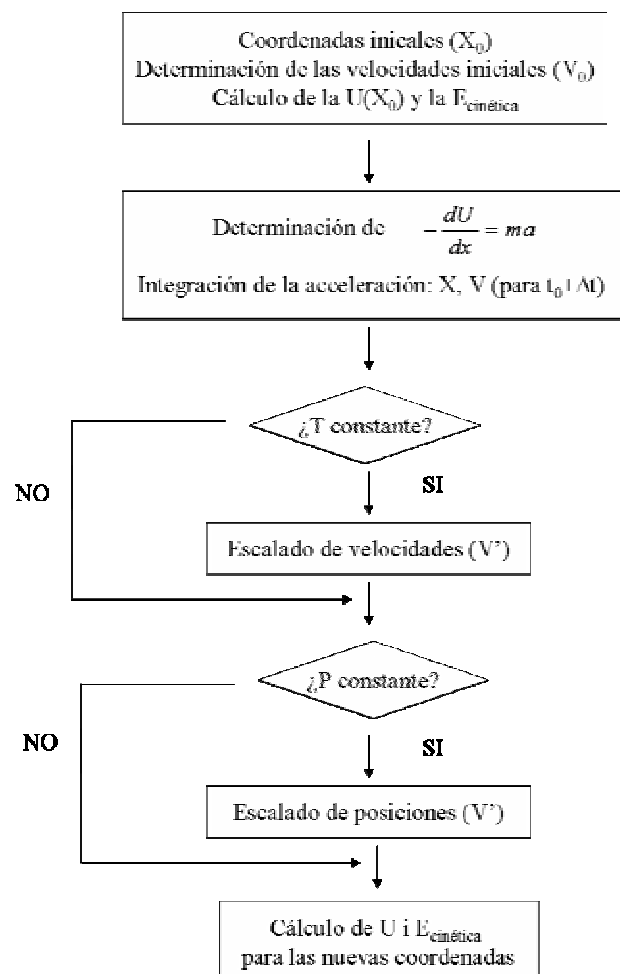


Figura 5. Algoritmo básico de los métodos de DM.

2.2.2.3 La preparación del sistema

Estructura de inicio

Se rodea el soluto de una caja de aguas preequilibradas eliminando las que están demasiado cerca. Los iones se ponen en los mejores sitios siguiendo típicamente el potencial electrostático de Poisson-Boltzmann.

Minimización

Es conveniente empezar la DM con una estructura correcta físicamente con el fin de evitar problemas de inestabilidad en la trayectoria. Por ello, se usan estructuras de rayos X o RMN, muchas veces minimizadas. En los primeros ciclos de minimización se acostumbra a usar el algoritmo de *Steepest descent* debido a su eficacia para acercarse al mínimo, seguido de *Conjugate Gradient*, que es más rápido en obtener la convergencia.

Equilibrado y termalización

Se tienen que asignar unas velocidades al azar a los átomos, que se seleccionan de manera que en promedio cumplan la distribución de Maxwell-Boltzmann:

$$T = \frac{1}{3Nk_B} \sum_{i=1}^N m_i v_i^2 \quad [41]$$

donde T es la temperatura inicial de la dinámica, N el número de partículas, k_B la constante de Boltzmann y m_i y v_i la masa y la velocidad de la partícula i .

Las primeras velocidades sirven para iniciar la trayectoria por lo que tienen poco sentido físico, haciéndose necesario este periodo de equilibrado. Es común generar las velocidades a una temperatura más baja para evitar así desbalances energéticos del sistema. Posteriormente, se aumenta la temperatura mediante pequeños incrementos hasta alcanzar el valor deseado (termalización), lo que conduce a una trayectoria mucho más estable.

Producción

Es la etapa dónde se recogen resultados. Su duración es un compromiso entre la propiedad de interés, los recursos computacionales y el espacio de disco. En general, se considera que la trayectoria tiene que ser 10 veces más larga que la escala temporal del proceso a estudiar.

2.2.2.4 Limitaciones y comparaciones

Como limitaciones fundamentales, podemos destacar que la mayoría de los campos de fuerza actuales no consideran el efecto de la polarización de manera explícita ni tampoco el de transferencia de carga. También pueden aparecer errores derivados de una parametrización incorrecta con datos de baja cualidad u obtenidos a partir de sistemas muy diferentes a los estudiados. Asimismo, debido a que el tiempo de simulación es limitado, la DM es incapaz de explorar totalmente el espacio conformacional. Aún así, ha permitido estudios muy interesantes mediante cálculos de geometrías moleculares, visiones promediadas de sistemas, muestreos de transiciones temporales, etc...

Los campos de fuerza de la familia AMBER son los más populares para el estudio de los ácidos nucleicos. Así, Hobza y colaboradores (59) calcularon las energías de interacción nucleobase-nucleobase (*stacking* y puentes de hidrógeno), concluyendo que este campo de fuerza tenía una capacidad sorprendente de reproducir los mejores datos cuánticos. Diversos autores han realizado multitud de estudios comparando los resultados obtenidos con datos experimentales, encontrando buena concordancia (54,60-65), incluso en el análisis de detalles de propiedades conformacionales dependientes de secuencia (60). El acoplamiento de estos campos de fuerza a algoritmos de DM ha permitido la representación de transiciones básicas entre conformaciones de los ácidos nucleicos, como la transición de las formas A y B en dúplex o en tríplex de ADN (34,64). También se ha demostrado su capacidad para reproducir datos directos de RMN con buena precisión, como son los volúmenes de NOESY y los ángulos diedros (66).

A pesar de su popularidad, el campo de fuerza AMBER (parm94, parm99) presenta ciertos problemas visible en simulaciones muy largas (>10 ns), tal como puso de relieve las simulaciones realizadas por el “*Ascona B-DNA Consortium*” en 2004, mostrando unas transiciones irreversibles en los ángulos diedros alpha/gamma que producían una

conformación irreal del ADN (67). Estos ángulos diedros fueron reparametrizados en nuestro grupo (38) elaborando la última versión del campo de fuerza (parm-bsc0) capaz de reproducir la estructura del dúplex de ADN hasta 1 microsegundo (3). Futuras mejoras pasaran por el uso de campos de fuerza polarizable (44), aunque este es un tema aun no muy maduro en el campo.

2.2.3 Cálculos de energía libre aplicados a grandes transiciones conformacionales

Actualmente las simulaciones MD de biomoléculas cubren un rango de 10-100 ns, lo que permite observar procesos que ocurren en la escala del sub-microsegundo. No obstante, muchos procesos precisan una escala temporal más grande, siendo necesario realizar cálculos más extensos o recurrir a métodos de sesgo de los muestreos, como los que se comentan a continuación dirigidos al cálculo de la energía libre.

La energía libre es la función termodinámica más fundamental, pues establece las estabilidades relativas de diferentes sistemas y sirve de base para deducir otras propiedades. Su cálculo es extremadamente delicado, ya que las zonas del espacio conformacional altas en energía son muy poco probables de muestrear (véase eq. 42) en una trayectoria de DM.

$$P(x) \sim \exp\left(-\frac{U(x)}{k_B T}\right) \quad [42]$$

donde $U(x)$ es la energía del sistema, x las coordenadas nucleares y k_B la constante de Boltzmann.

Por ello, generalmente los estudios se limitan a la evaluación de diferencias de energía libre entre estados. Para ello, suele recurrirse al potencial de fuerza media (PMF, *Potencial of Mean Force*), que proporciona el perfil de energía libre a lo largo de una coordenada privilegiada ($\xi(x)$) identificada como el cambio relevante asociado al proceso:

$$\omega(\xi) = \omega(\xi^*) - k_B T \ln \left[\frac{\langle \rho(\xi) \rangle}{\langle \rho(\xi^*) \rangle} \right] \quad [43]$$

donde $\omega(\xi^*)$ y $\langle \rho(\xi^*) \rangle$ son valores arbitrarios de referencia, y $\langle \rho(\xi) \rangle$ es la función de distribución promedio (FDP):

$$\langle \rho(\xi) \rangle = \frac{\int dx \delta(\xi'[x] - \xi) e^{(-U(x)/k_B T)}}{\int dx e^{(-U(x)/k_B T)}} \quad [44]$$

donde $\delta(\xi'[x] - \xi)$ es la función delta de Dirac para la coordenada ξ . Gracias al PMF el problema del muestreo se reduce a estimar una fracción del espacio conformacional: la FDP a lo largo de una determinada variable, forzando para ello el muestreo a lo largo de dicha coordenada. No obstante, ello puede ser difícil si existen barreras elevadas, y por eso hemos usado técnicas de sesgo como *Umbrella Sampling* (US) (14) y WHAM (15,68,69).

En esta tesis se ha aplicado el cálculo del PMF para estudiar grandes transiciones conformacionales. A diferencia de reacciones químicas convencionales, dichas transiciones no se pueden caracterizar mediante conformaciones individuales en la SEP debido al gran número de grados de libertad y caminos por los cuales se puede producir la transición. Por ello, el problema principal es conocer la coordenada de reacción, aunque se ha intentado estimar con la técnica de *Targeted Molecular Dynamics* (TMD) (16,70).

2.2.3.1 *Umbrella Sampling*

En este método el sistema es simulado con la presencia de un sesgo artificial $(U(\xi)^s)$ en la SEP:

$$U(x)' = U(x) + U(\xi)^s \quad [45]$$

que acostumbra a tener forma harmónica:

$$U(\xi)^s = \frac{1}{2} k (\xi - \xi_0)^2 \quad [46]$$

Normalmente se realizan diferentes simulaciones (ventanas) donde el potencial de sesgo se centra en diferentes valores de ξ_0 para obtener un muestreo eficiente a lo largo de todo el rango de interés. De esta manera, se obtienen una FDP sesgada para cada ventana i :

$$\langle \rho(\xi) \rangle_i^s = \frac{\int dx \delta(\xi - [x]) e^{(-U(x)/k_B T)}}{\int dx e^{(-U(x)/k_B T)}} \quad [47]$$

Después, las FDPs asociadas a cada ventana tienen que combinarse para obtener la estimación final no sesgada del PMF ($\omega(\xi)^n$). Así, la parte correspondiente a la ventana i se define como:

$$\omega(\xi)_i^n = \omega(\xi^*) - k_B T \ln \left[\frac{\langle \rho(\xi) \rangle_i^s}{\langle \rho(\xi^*) \rangle} \right] - U(\xi)_i^s + F_i \quad [48]$$

donde los superíndices n y s indican *normal* y *sesgado* respectivamente, y F_i es una constante que representa la energía libre asociada a la introducción del potencial de sesgo:

$$e^{-F_i/k_B T} = \left\langle e^{-U(\xi)_i^s/k_B T} \right\rangle \quad [49]$$

Estas constantes (F_i) se obtienen ajustando el PMF de ventanas contiguas ($\omega(\xi)_i^n$), siendo necesario que exista un solapamiento significativo entre ventanas, lo que constituye la principal dificultad del método.

2.2.3.2 WHAM

Este método es una evolución del anterior y soluciona algunos de sus problemas. En concreto, propone una manera sistemática de obtener las constantes F_i sin descartar las regiones de solapamiento, y además es fácilmente extensible al cálculo de PMF multidimensionales (69).

La FDP total ($\langle \rho(\xi) \rangle^n$) se obtiene a partir de las FDP sesgadas para cada ventana ($\langle \rho(\xi) \rangle_i^s$):

$$\langle \rho(\xi) \rangle^n = \sum_{i=1}^N n_i \langle \rho(\xi) \rangle_i^s \times \left[\sum_{j=1}^N n_j e^{-(U(\xi)_j^s - F_j)/k_B T} \right]^{-1} \quad [50]$$

donde N es el número de ventanas y n_i el número de datos (pasos de DM) por ventana. Las constantes F_i se calculan mediante la fórmula [49] y con la ayuda de la FDP total ($\langle \rho(\xi) \rangle^n$) para obtener:

$$e^{-F_i/k_B T} = \int d\xi e^{-U(\xi)_i^s/k_B T} \langle \rho(\xi) \rangle^n \quad [51]$$

Debido a que la FDP total depende de las constantes F_i , las ecuaciones [50,51] se tienen que resolver iterativamente hasta llegar a la consistencia. Un procedimiento común es empezar con una primera aproximación para las constantes y parar cuando el cambio de estas sea menor a un determinado umbral.

2.2.3.3 Targeted Molecular Dynamics

Esta técnica de sesgo fue desarrollada con el objetivo de encontrar caminos potenciales de transiciones conformacionales en proteínas (16). Consiste en utilizar la desviación cuadrática media o RMSd (ver apartado 2.3.1.3) de la trayectoria (x) respecto a la conformación diana o *target* (x_T) como coordenada de reacción, para así dejar el resto de dimensiones libres de adaptarse de la mejor manera para alcanzar el objetivo.

Concretamente al principio se implementó con una restricción holonómica dependiente del tiempo:

$$U^s = |x - x_T|^2 - \eta^2 \quad [52]$$

donde η^2 es la distancia deseada que a tiempo cero es:

$$\eta^2 = |x_I - x_T|^2 \quad [53]$$

donde x_I es la conformación inicial y se va reduciendo a cada paso de integración con un diferencial lo más pequeño posible hasta ser cero. Este método tiene el problema que sólo puede encontrar caminos de reacción donde se va reduciendo monótonicamente la distancia a la configuración diana, pudiendo obtener transiciones irreales donde se atraviesan barreras y por lo tanto se presentan problemas de reversibilidad (70).

En esta tesis se ha utilizado la modificación donde la restricción impuesta es harmónica, dejando así más libertad de elección (70):

$$U^s = \frac{1}{2} k (RMSd(x, x_T) - \eta)^2 \quad [54]$$

donde en este caso η es el valor de RMSd deseado que a tiempo cero es:

$$RMSd = \sqrt{\frac{1}{N} \sum_{j=1}^{3N} (x_{I,j} - x_{T,j})^2} \quad [55]$$

donde N es el número de átomos. Así, se va reduciendo discontinuamente siendo el algoritmo muy similar al método de US. Aunque este esquema es más flexible que el anterior, se tiene que ser cuidadoso para obtener transiciones reales.

Debido al gran número de dimensiones de las biomoléculas, estas restricciones [52,54] no afectarán seriamente a su movilidad interna siempre que η sea lo suficientemente grande. A medida que ésta va disminuyendo, el número de conformaciones posibles para ese determinado valor va decreciendo también, y las fluctuaciones dinámicas se reducirán considerablemente cuando η alcance un valor cercano al provocado por las fluctuaciones térmicas (figura 6).

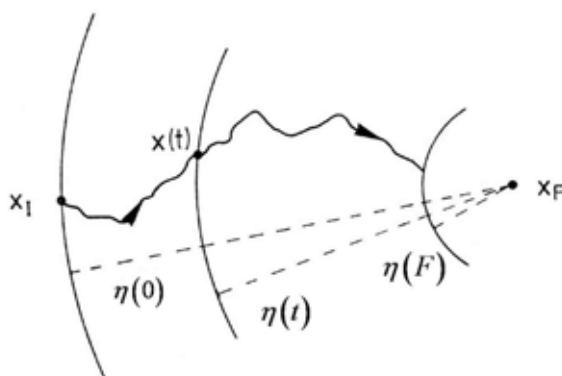


Figura 6. Proyección bidimensional del espacio conformacional

2.2.4 Métodos de Monte Carlo

Este tipo de técnicas persiguen, al igual que la DM, obtener una visión promediada del sistema. La principal diferencia entre la DM y el método de Monte Carlo (MC) es que éste último no incluye el tiempo como variable. Es esencialmente útil para analizar sistemas pequeños y de rigidez elevada en fases condensadas, mientras que los sistemas moleculares más grandes son estudiados preferiblemente por DM.

La filosofía de los métodos de MC está ligada a los formalismos de la mecánica estadística. De acuerdo con ellos, un sistema debe cumplir la ley de Boltzmann, por la cual las

configuraciones más estables estarán ocupadas por un mayor número de moléculas (71). El método de MC más común usa el algoritmo de Metrópolis (72), en el que se parte de una configuración inicial a partir de la cual se genera otra al azar cuya energía es evaluada con un campo de fuerza (figura 7). Si ésta es menor que la energía de la configuración original, entonces es aceptada, pero en caso contrario aun tendrá una probabilidad de ser aceptada, tanto menor cuanto mayor sea la diferencia con respecto a la configuración inicial. Después de varios millones de configuraciones generadas de esta manera, se consigue un muestreo similar al obtenido por DM para sistemas ergódicos.

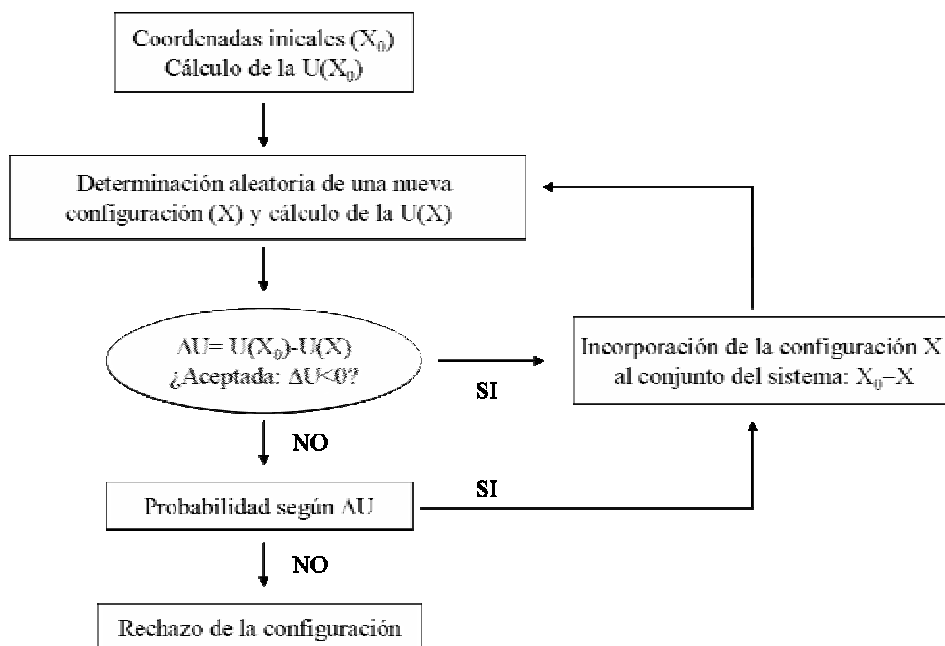


Figura 7. Algoritmo básico de los métodos de Monte Carlo

2.3 ANÁLISIS DE LAS TRAYECTORIAS

La comprensión a nivel atómico de la relación entre estructura y función de biomoléculas implica no sólo disponer de simulaciones realistas, sino también realizar un análisis riguroso de las mismas. Las metodologías empleadas en dicho análisis son muy variadas y cambian en función del interés del estudio. Este capítulo citará solamente aquellas que han sido utilizadas a lo largo de la presente tesis.

2.3.1 Promedio y convergencia

2.3.1.1 La hipótesis ergódica

La técnica de la DM proporciona una trayectoria a lo largo del tiempo constituida por una colección de conformaciones que exploran el espacio definido por la SEP. Así, se asume la hipótesis ergódica, que postula que el promedio de un parámetro explorado en función del tiempo es igual al promedio del espacio de fases, es decir, se asume que la observación de un proceso a lo largo de la trayectoria es lo mismo que hacer realizaciones independientes del mismo proceso. En nuestros estudios se han coleccionado conformaciones separadas por 1 picosegundo (ps), lo cual significa 500 pasos de integración de 2 fs.

2.3.1.2 Convergencia

Una cuestión central es saber si la trayectoria es lo suficientemente larga para estudiar las propiedades de interés con fiabilidad, es decir, si los datos están convergidos. De esta manera, los promedios serán representativos cuando el tiempo de equilibrado sea mayor y el de producción mucho mayor que el tiempo de relajación de la propiedad a estudiar. Cuando no se satisfagan estas dos condiciones, se producirá un cambio progresivo del promedio a lo largo del tiempo (proceso intrínsecamente lento) o bien un cambio errático (proceso rápido de baja probabilidad) (32,73).

Hay varias maneras de calcular la escala temporal de relajación de un proceso, examinar su evolución temporal para determinar su valor promedio y desviación estándar o empezar en

distintas condiciones iniciales y ver cuando se alcanzan valores típicos. Cuando se ha llegado a la estabilización en los valores de la propiedad, se dice que se ha alcanzado la convergencia empezando así la etapa de recogida de datos (32,73).

2.3.1.3 El RMSd

El RMSd o la desviación cuadrática media respecto a una estructura de referencia (experimental, inicial, promedio...) es una medida ampliamente utilizada como referente de la estabilidad y convergencia general de una trayectoria:

$$RMSd = \sqrt{\frac{1}{N} \sum_{j=1}^{3N} (x_j - x_{R,j})^2} \quad [56]$$

donde N es el número de átomos, x_j las configuraciones de nuestra trayectoria y $x_{R,j}$ la conformación de referencia. El procedimiento es hacer una superposición rígida de las dos estructuras para minimizar esta medida, que nos informa de la distancia promedio entre los átomos correspondientes.

2.3.1.4 Potencial de interacción molecular

Es una medida del perfil de interacción de una molécula. En él se calcula el potencial de interacción clásico entre un átomo de prueba y la estructura promedio de la simulación. Se ha realizado usando el programa CMIP (74), desarrollado en nuestro grupo, el cual define el potencial como la adición de un término electrostático (U_{ele}) y otro estérico o de vdW (U_{vdw}):

$$U = U_{ele} + U_{vdw} \quad [57]$$

La contribución estérica se calcula mediante un término de Lennard-Jones utilizando los mismos parámetros que AMBER (ver apartado 2.1.4.2), mientras que la parte electrostática se obtiene de resolver numéricamente la ecuación de Poisson-Boltzmann lineal (75):

$$\nabla[\epsilon(x)\nabla\Phi(x)] - \kappa^2\Phi(x) = -4\pi\rho(x) \quad [58]$$

simplificación de la ecuación no lineal:

$$\nabla[\epsilon(x)\nabla\Phi(x)] - \kappa^2 \sinh(\Phi(x)) = -4\pi\rho(x) \quad [59]$$

donde x son las coordenadas internas, ϵ es la constante dieléctrica, Φ el potencial electrostático, ρ la densidad de carga interna a la molécula y κ es la inversa de la longitud de Debye-Hückel (en unidades de Gauss):

$$\kappa = \sqrt{\frac{8\pi N_a e^2 I}{10^3 \epsilon_s k_B T}} \quad [60]$$

donde N_a es el número de Avogadro, e la carga del electrón, I la fuerza iónica, ϵ_s la constante dieléctrica del solvente, k_B la constante de Boltzmann y T la temperatura.

Generalmente en este trabajo se ha supuesto que la fuerza iónica es 0.145 M, la constante dieléctrica para el ácido nucleico es 2.0 y para el agua 80.0, y se ha calculado para cada punto de una malla espaciada por 0.5 Å.

2.3.2 Análisis de dinámica esencial

Como se comentó en la introducción, la DM complementa la información experimental, básicamente estática, pues las fluctuaciones térmicas son claves para la funcionalidad de las biomoléculas. Extraer los movimientos importantes de una dinámica compleja no es trivial y requiere del uso de potentes herramientas estadísticas, destacando el análisis de componentes principales.

2.3.2.1 Análisis de componentes principales

El análisis de componentes principales (ACP) es una técnica estadística que permite eliminar la redundancia de los datos y así reducir la dimensionalidad del sistema. Consiste en una transformación lineal ortogonal de cambio de base. Por esta razón, se diagonaliza la matriz de covarianzas (C) obtenida en la DM buscando la dirección que explique más variabilidad y que de esta forma contenga más información.

$$\Lambda = Y^T C Y \quad [61]$$

donde Λ y Y son las matrices de valores y vectores propios respectivamente.

Los valores propios obtenidos (λ_i) consistirán en la varianza de estas nuevas coordenadas explicadas con los coeficientes de los vectores propios correspondientes:

$$v = xY \quad [62]$$

donde v es el vector de las nuevas coordenadas y x el vector de las coordenadas atómicas primitivas.

En el caso de macromoléculas se ha visto que gran parte de la variabilidad se explica con relativamente pocos grados de libertad, que implican a la mayoría de átomos, definiéndose así el subespacio esencial (76-79), que contiene la información principal sobre la flexibilidad conformacional de los ácidos nucleicos.

2.3.2.2 Índices de similitud

Comparar la similitud de dos estructuras es sencillo, pero no lo es tanto comparar dos pautas de flexibilidad (10,34,80). Una medida es el índice de similitud absoluto (ISA), que calcula el solapamiento de los subespacios creados por los vectores propios (81):

$$\gamma_{AB} = \frac{1}{n} \sum_{j=1}^n \sum_{i=1}^n (v_i^A \cdot v_j^B)^2 \quad [63]$$

donde n es el número de vectores propios considerado en el análisis, y v_i^A y v_j^B son los vectores propios de las dos trayectorias comparadas. Este índice puede tener un valor de 1 cuando los espacios son idénticos y de 0 cuando son totalmente distintos.

Debido a que el ACP depende de la longitud de la simulación, se ha desarrollado el índice de similitud relativo (ISR), que persigue normalizar el ISA por las autosimilitudes de las trayectorias involucradas:

$$\kappa_{AB} = 2 \frac{\gamma_{AB}}{(\gamma_{AA} + \gamma_{BB})} \quad [64]$$

donde γ_{AA} y γ_{BB} se obtienen de enfrentar las dos mitades de una misma trayectoria. Este índice también varía entre 0 y 1, con la ventaja que es menos sensible a la simulación.

2.3.2.3 Distancia de Mahalanobis

Es una medida de distancia estadística usada para determinar la similitud entre dos vectores de variables que, a diferencia de la distancia euclídea (d_E), tiene en cuenta la varianza de cada grado de libertad y la correlación entre ellos (d_M) (82):

$$d_E = \left[(\Delta x)^T (\Delta x) \right]^{1/2} \quad [65]$$

$$d_M = \left[(\Delta x)^T C^{-1} (\Delta x) \right]^{1/2} \quad [66]$$

donde Δx es el vector de desplazamiento y C la matriz de covarianzas. En nuestro caso hemos aplicado la fórmula a los vectores esenciales, que al ser ortogonales, se reduce a:

$$d_M = \left[\sum_{i=1}^n \left(\frac{\Delta x_i}{\lambda_i^{1/2}} \right)^2 \right]^{1/2} \quad [67]$$

donde se ve que la distancia de Mahalanobis corresponde a una distancia euclídea normalizada.

En este estudio se ha usado la distancia de Mahalanobis en examinar la capacidad de los vectores esenciales de una simulación para aproximarse a una estructura determinada. De esta manera, se ha implementado un algoritmo de búsqueda de los desplazamientos asociados a los vectores esenciales, los cuales nos permiten reducir más el RMSd (figura 8).

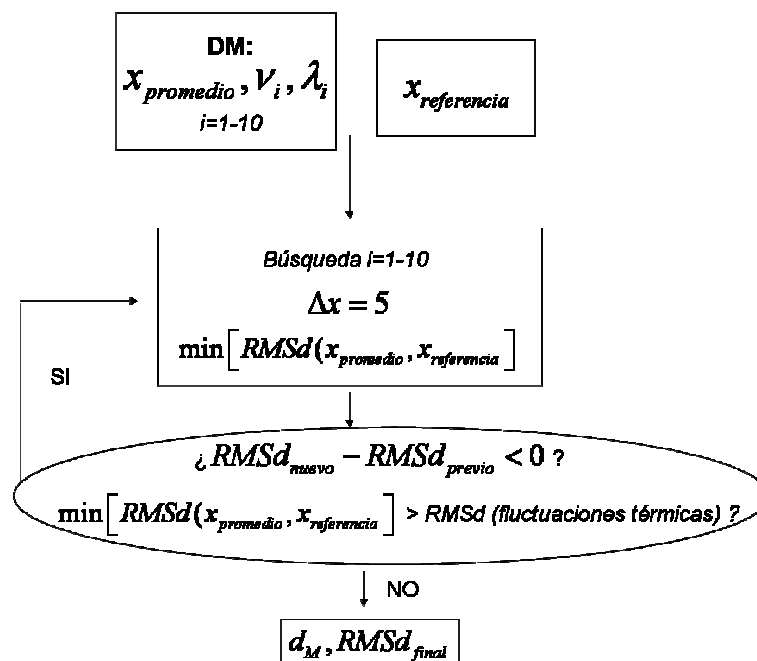


Figura 8. Algoritmo implementado de cálculo de la distancia de Mahalanobis

2.3.3 Análisis de elasticidad

En este apartado se explicará como usar la información extraída de DM para calcular las propiedades elásticas de los ácidos nucleicos, y así asignar una medida energética a los diferentes grados de libertad.

2.3.3.1 La aproximación cuasi-harmónica

Esta hipótesis supone una simplificación de la SEP, la cual pasa a tener localmente la forma de un potencial cuadrático efectivo (83):

$$U_{eff} = \frac{1}{2} x^T K x \quad [68]$$

donde x es el vector de variables y K es la matriz hessiana.

Así, las fluctuaciones de los movimientos siguen una distribución de probabilidad multivariante gaussiana:

$$P(x) \propto e^{(-x^T C^{-1} x)} \quad [69]$$

donde C es la matriz de covarianzas y K se calcula como:

$$K = k_B T C^{-1} \quad [70]$$

siendo k_B la constante de Boltzmann. Los elementos diagonales de la matriz (k_{ii}) representan las constantes elásticas, y los de fuera la diagonal (k_{ij}) los elementos cruzados:

$$U_{eff} = \sum_i \frac{k_{ii}}{2} (x_i - x_{0,i})^2 + \sum_{i \neq j} \frac{k_{ij}}{2} (x_i - x_{0,i})(x_j - x_{0,j}) \quad [71]$$

donde $x_{0,i}$ son los promedios de las variables.

2.3.3.2 La elasticidad de los modos esenciales

Un aspecto delicado del análisis anterior consiste en considerar el conjunto de variables relevantes que describen el movimiento. Los modos esenciales obtenidos a partir de la técnica

del ACP permiten eliminar la redundancia, y así se puede describir la dinámica de nuestra molécula con un número reducido de constantes elásticas (k_i). Aprovechando la ortogonalidad de los vectores propios, éstas vienen definidas por:

$$k_i = k_B T / \lambda_i \quad [72]$$

siendo k_B la constante de Boltzmann y λ_i el valor propio correspondiente al vector propio v_i .

Los primeros modos tienen el valor propio más grande y por tanto la menor constante, siendo los movimientos más fáciles. Por tanto, la SEP efectiva se puede expresar como:

$$U_{eff} = \sum_i \frac{k_B T}{2\lambda_i} (x_i - x_{0,i})^2 \quad [73]$$

donde $x_{0,i}$ son los promedios de los modos considerados.

En el caso de querer obtener la energía de elasticidad asociada a la distancia de Mahalanobis (d_M , apartado 2.3.2.3), la fórmula resultante de aplicar [67] a [73] es:

$$U_{eff} = \frac{k_B T}{2} d_M^2 \quad [74]$$

2.3.3.3 Índices de similitud ponderados

Como se puede observar en las fórmulas [63,64], los índices de similitud son insensibles a la importancia relativa que tienen los diferentes modos esenciales, así como a las permutaciones de los mismos (apartado 2.3.2.2). Por esta razón, en nuestro grupo se han desarrollado unos índices de similitud ponderados (84) donde se aprovecha la aproximación cuasi-harmónica (apartados 2.3.3.1 y 2.3.3.2) para ponderar cada vector propio (v_i) por su función de probabilidad de Boltzmann:

$$P(v_i) = \frac{e^{(-U(v_i)/k_B T)}}{\sum_{j=1}^{j=z} e^{(-U(v_j)/k_B T)}} \quad [75]$$

Con las expresiones del punto anterior [72,73], la fórmula [75] nos queda:

$$P(v_i) = \frac{e^{-(\Delta x)^2/\lambda_i}}{\sum_{j=1}^{j=z} e^{-(\Delta x)^2/\lambda_j}} \quad [76]$$

donde Δx es el desplazamiento sobre el vector propio y λ_i el valor propio correspondiente.

Así, la fórmula del ISA [63] queda modificada:

$$\xi_{AB} = \frac{2 \sum_{i=1}^{i=z} \sum_{j=1}^{j=z} \left[(v_i^A \cdot v_j^B) \frac{\exp\left\{-\frac{(\Delta x)^2}{\lambda_i^A} - \frac{(\Delta x)^2}{\lambda_j^B}\right\}}{\sum_{i=1}^{i=z} \exp\left\{-\frac{(\Delta x)^2}{\lambda_i^A}\right\} \sum_{j=1}^{j=z} \exp\left\{-\frac{(\Delta x)^2}{\lambda_j^B}\right\}} \right]^2}{\sum_{i=1}^{i=z} \left(\frac{\exp\left\{-2\frac{(\Delta x)^2}{\lambda_i^A}\right\}}{\left(\sum_{i=1}^{i=z} \exp\left\{-\frac{(\Delta x)^2}{\lambda_i^A}\right\}\right)^2} \right)^2 + \sum_{j=1}^{j=z} \left(\frac{\exp\left\{-2\frac{(\Delta x)^2}{\lambda_j^B}\right\}}{\left(\sum_{j=1}^{j=z} \exp\left\{-\frac{(\Delta x)^2}{\lambda_j^B}\right\}\right)^2} \right)^2} \quad [77]$$

donde v_i^A y v_j^B son los diferentes vectores propios de las dos trayectorias comparadas. El numerador consiste en los dos modos con sus respectivos pesos de Boltzmann, y el denominador en la normalización. De la misma forma que en el apartado 2.3.2.2, también se puede definir el correspondiente ISR, el cual es menos sensible a la longitud de la simulación:

$$\delta_{AB} = 2 \frac{\xi_{AB}}{(\xi_{AA} + \xi_{BB})} \quad [78]$$

donde ξ_{AA} y ξ_{BB} se obtienen de enfrentar las dos mitades de una misma trayectoria.

2.3.3.4 Los ácidos nucleicos como polímeros

Dada la naturaleza estructural de los ácidos nucleicos, cabe la posibilidad de cambiar las coordenadas para realizar el análisis de flexibilidad usando una serie de coordenadas internas, cuya interpretación geométrica resulte mucho más intuitiva. En concreto, hemos usado coordenadas helicoidales (ver apartado 3.2.2) que describen el ADN como una fibra. Así, las constantes elásticas asociadas a estos parámetros informan acerca de la flexibilidad desde el punto de vista de una hélice definida en cada paso por 3 rotaciones y 3 traslaciones (7):

$$K_h = k_B T C_h^{-1} \quad [79]$$

donde C_h es la matriz de covarianzas asociada a los parámetros helicoidales, y K_h la matriz hessiana que contendrá en la diagonal las constantes elásticas y fuera de la diagonal los términos cruzados.

2.3.4 La entropía configuracional

La entropía configuracional refleja el nivel de desorden intrínseco de una molécula, que en el caso de biomoléculas no se puede despreciar debido a su gran tamaño y a la complejidad de su espacio conformacional. Aquí presentaremos métodos que pretenden calcular el valor de la entropía en base a la aproximación cuasi-harmónica, utilizando así las propiedades de la distribución multivariante gaussiana que adopta la matriz de covarianzas. Esta puede ser calculada fácilmente mediante una trayectoria de DM tras eliminar los grados de traslación y rotación.

2.3.4.1 El método de Schlitter

La fórmula *ad hoc* [80] introducida por Schlitter (85) soluciona el problema de la singularidad de la matriz de covarianzas producido después de eliminar las variables de traslación y rotación:

$$S_h < S' = \frac{1}{2} k_B \ln \left[1 + \left(\frac{k_B T e^2}{\hbar} \right) \langle x^2 \rangle_c \right] \quad [80]$$

donde k_B es la constante de Boltzmann y e el número de Euler, siendo una aproximación al límite superior de la entropía de un oscilador armónico mecano-cuántico (S_h):

$$S_h = k_B \frac{\hbar \omega / k_B T}{e^{\hbar \omega / k_B T} - 1} - \ln(1 - e^{-\hbar \omega / k_B T}) \quad [81]$$

donde \hbar es la constante de Planck dividida por 2π y ω la frecuencia, que está determinada por la varianza clásica $\langle x^2 \rangle_c$ gracias al principio de equipartición:

$$m\omega^2 \langle x^2 \rangle_c = k_B T \quad [82]$$

Esta aproximación actúa correctamente en el límite cuántico, ya que se aproxima a cero cuando la temperatura baja hasta cero, y también en el límite clásico, porque la entropía es proporcional a $\ln T$ para T infinita.

Es especialmente útil para la generalización a muchas dimensiones, pues permite utilizar directamente la matriz de covarianzas extraída de una simulación de DM:

$$S_h < S' = \frac{1}{2} k_B \ln \det \left[1 + \left(\frac{k_B T e^2}{\hbar} \right) MC \right] \quad [83]$$

donde 1, M y C son las matrices unitarias, de masa y de covarianza respectivamente.

2.3.4.2 El método de Andricioaei y Karplus

Este método (86) soluciona el problema de la singularidad de la matriz de covarianzas aprovechando su propiedad de ser diagonalizable, para así, suministrar unas nuevas coordenadas no relacionadas o vectores propios con los valores propios correspondientes λ_i .

Estos últimos se pueden relacionar con las frecuencias cuasi-harmónicas:

$$\omega_i = \sqrt{\frac{k_B T}{\lambda_i}} \quad [84]$$

En consecuencia, la fórmula de la entropía del oscilador armónico cuántico se aplica directamente para aquellos vectores propios donde el valor propio es distinto de cero:

$$S = k_B \sum_i^{3n-6} \frac{\hbar \omega_i / k_B T}{e^{\hbar \omega_i / k_B T} - 1} - \ln(1 - e^{-\hbar \omega_i / k_B T}) \quad [85]$$

2.3.4.3 Extrapolación a tiempo infinito

La posibilidad de obtener la entropía a partir de simulaciones de DM mediante los métodos anteriores es muy atractiva. No obstante, no se puede ignorar las aproximaciones asumidas, como el tratamiento armónico de las vibraciones. Asimismo, la entropía refleja la extensión del espacio configuracional accesible para una macromolécula, y por ello depende de la longitud de la simulación. Esto se puede ver en la figura 9, donde se representa la entropía según el método de Schlitter de un dodecámero de ARN calculada en ventanas de 1 a 6 ns de una misma trayectoria. Afortunadamente, la relación entre los valores de entropía y la

longitud de la trayectoria se pueden ajustar en la mayoría de los casos a una relación exponencial [86] que nos permite obtener la entropía esperada de una simulación infinita (87).

$$S(t) = S_{\infty} - \frac{a}{t^b} \quad [86]$$

donde S_{∞} , a y b son los parámetros obtenidos por ajuste.

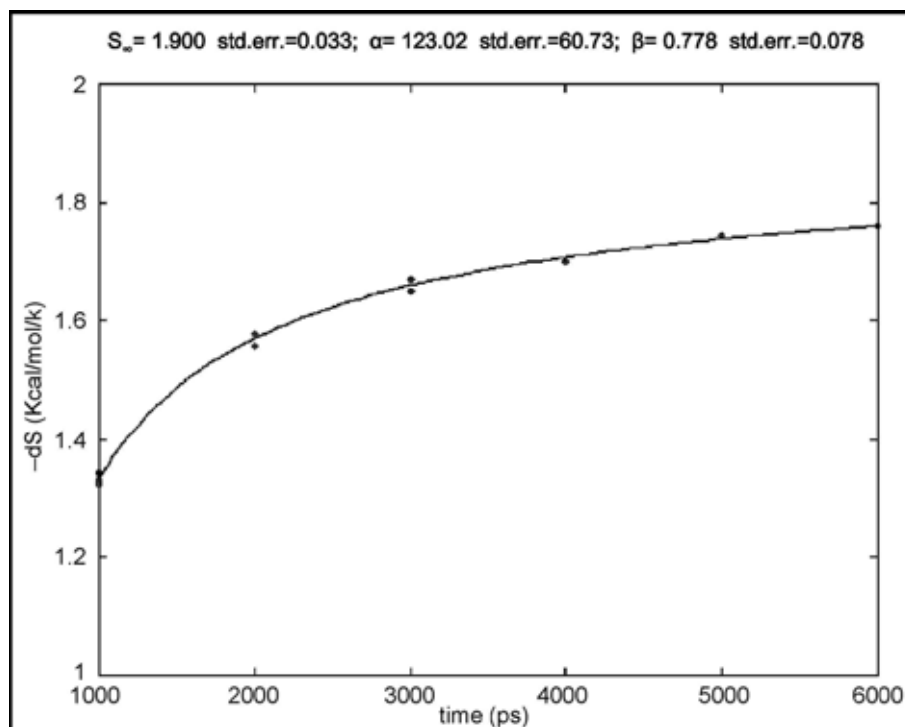


Figura 9. Estimación a tiempo infinito de la entropía de un dodecámero de ARN considerando valores obtenidos con el método de Schlitter de varias ventanas con distinta longitud a lo largo de una trayectoria común (10).

3. LOS ÁCIDOS NUCLEICOS

Existen dos tipos fundamentales de ácidos nucleicos: el ácido ribonucleico (ARN) y el ácido desoxirribonucleico (ADN). Ambos desempeñan un papel crucial en el almacenaje, la transmisión y el procesado de la información genética, siendo por ello de gran relevancia biológica. En este capítulo se describen brevemente algunos de los conceptos esenciales para la comprensión de su estructura y de sus propiedades (9,88-91).

3.1 ESTRUCTURA PRIMARIA

3.1.1 Componentes estructurales

Los nucleótidos son la unidad básica de la estructura molecular de los ácidos nucleicos. Dichas unidades constan de un azúcar cíclico furanósico (β -D-ribosa en el ARN o β -D-2'-desoxirribosa en el ADN) fosforilado en 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) encontrándose solo en el ADN, y uracilo (U) propia del ARN (figura 10).

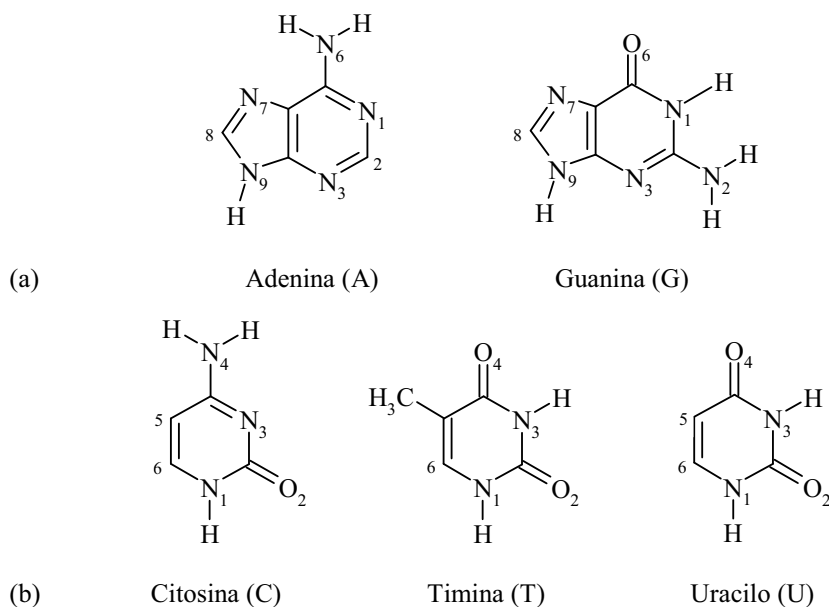


Figura 10. Estructura química y nomenclatura de las bases nitrogenadas (a) púricas y (b) pirimidínicas.

Los ácidos nucleicos son polímeros de nucleótidos, unidos a través de enlaces fosfodiéster, que por una parte enlazan la posición C3' de la pentosa de un nucleósido y por otra el C5' de la pentosa de otro nucleósido. Los polímeros forman largas cadenas con un sector idéntico en todas ellas (fosfato-pentosa-fosfato) y otro variable (las bases unidas a las pentosas) (figura 11). Las cadenas se nombran a nivel de estructura primaria por la

composición de las bases del extremo 5' al 3', donde 5' es el extremo que tiene el grupo fosfato libre unido al carbono C5'.

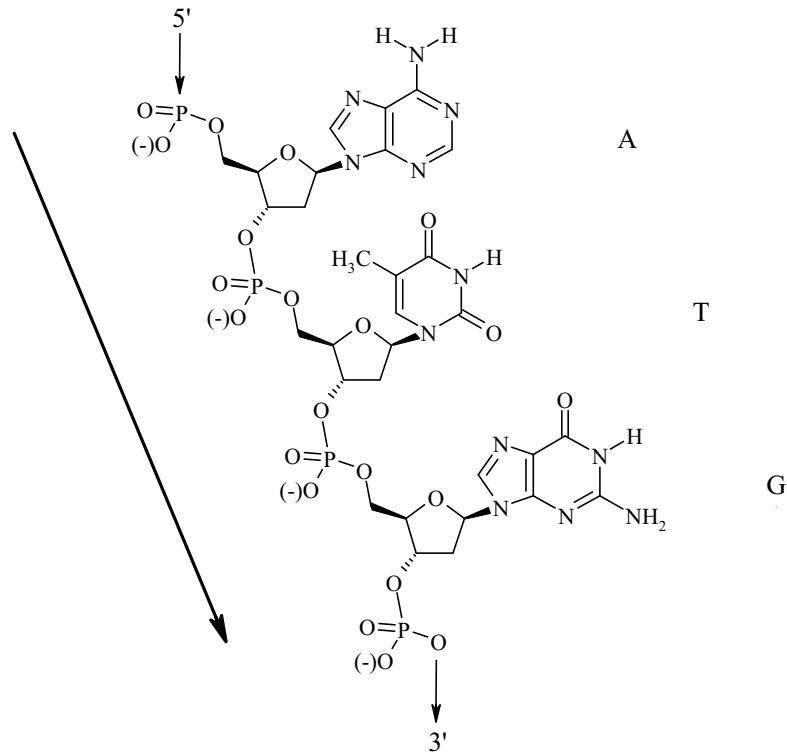


Figura 11. Fragmento de ADN con la secuencia adenina, timina y guanina unido por enlaces 3', 5'-fosfodiéster. La dirección de la cadena viene mostrada por la flecha.

3.1.2 Conformación

La estructura tridimensional de los ácidos nucleicos viene modulada por rotaciones de grupos de átomos sobre enlaces sencillos (figura 12), siendo los más relevantes los que marcan la rotación respecto el enlace glucosídico, el enlace exocíclico C4'-C5', el empaquetamiento de la ribosa y el enlace fosfodiéster.

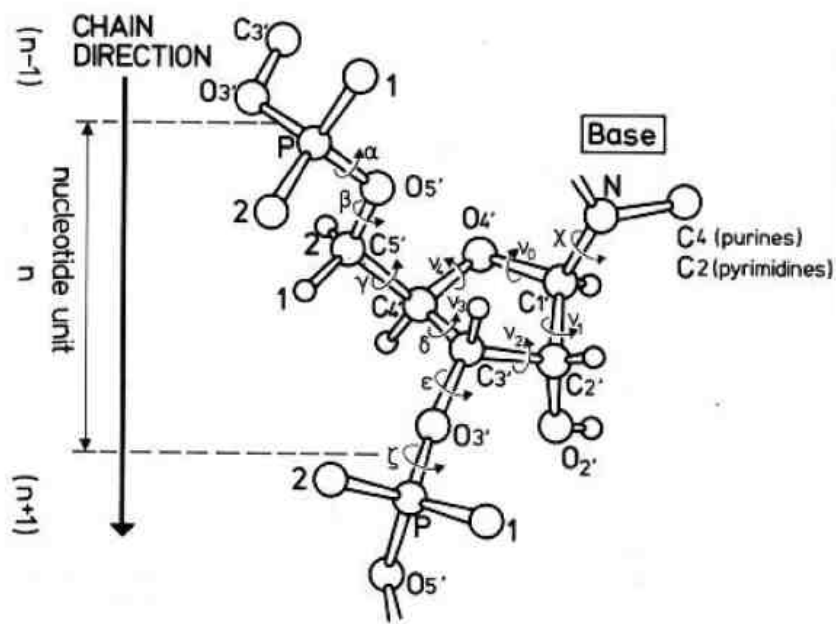


Figura 12. Numeración atómica y nomenclatura para las torsiones de un nucleótido (88).

3.1.2.1. El anillo furanósico

El anillo furanósico adopta siempre una estructura no planar que puede ser del tipo sobre (*envelope*), formada por cuatro átomos dentro de un mismo plano y un quinto en el exterior a $\sim 0.5 \text{ \AA}$, o de tipo girado (*twist*), 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 13).

En la forma girada suele ocurrir que uno de los átomos situados fuera del plano se encuentra más alejado que el otro, recibiendo así la nomenclatura de *twist* asimétrico, donde el desplazamiento del átomo más alejado recibe el nombre de *puckering* mayor y el del átomo más cercano de *puckering* menor. Dado que los procedimientos mencionados para definir la conformación del azúcar son cualitativos, se suele recurrir a dos parámetros de naturaleza analítica (92): el ángulo de pseudorrotación o ángulo de fase (P), y la amplitud de *puckering* (ν_m), que vienen definidos como:

$$\tan P = \frac{(v_4 + v_1) - (v_3 + v_0)}{2v_2 (\sin 36^\circ + \sin 72^\circ)} \quad [87]$$

$$v_m = \frac{v_2}{\cos P} \quad [88]$$

donde el conjunto v_i son los ángulos diedros que describen el azúcar (figura 12).

El ángulo de fase genera información sobre todos los posibles tipos de *puckering* de la ribosa, pudiendo oscilar entre 0 y 360 grados. En función del mismo cabe distinguir cuatro zonas principales y sus combinaciones (figura 14): Norte ($P \approx 0^\circ$), Sur ($P \approx 180^\circ$), Este ($P \approx 90^\circ$) y Oeste ($P \approx 270^\circ$). Se conoce experimentalmente que existen dos familias preferidas para los nucleótidos: la C2'-*endo*, con un ángulo de fase en el rango de 144-190° o Sur, y la C3'-*endo*, con un rango de 0 a 36 grados o Norte (figura 13). Para nucleótidos en solución ambos estados están en equilibrio, pero en polinucleótidos puede estar totalmente desplazado a una región. Es posible encontrar transiciones entre ambos estados únicamente a través de la zona Este pasando por el estado O4'-*endo*. La barrera energética no es excesiva, siendo cercana a 4 kcal/mol (93).

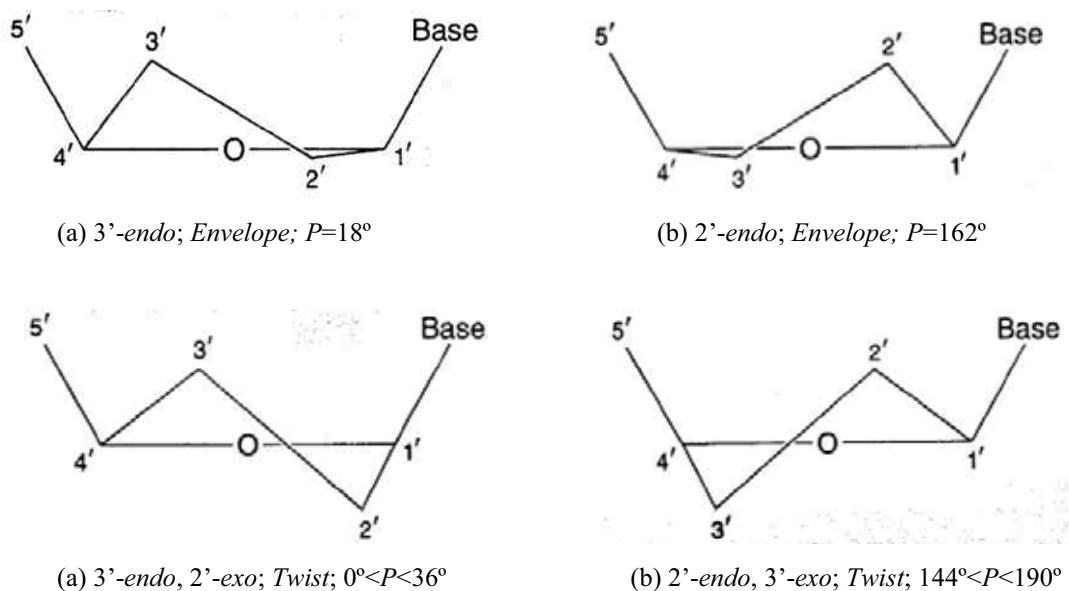


Figura 13. Conformaciones del azúcar encontradas en los nucleótidos. (a) Confórmeros Norte que se encuentran en la doble hélice de la forma A, y (b) confórmeros Sur que se encuentran en la doble hélice de la forma B (89).

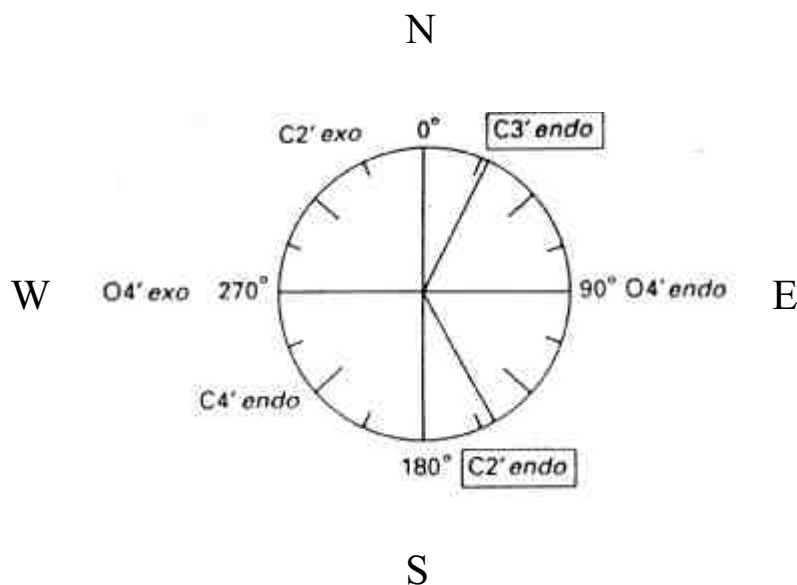
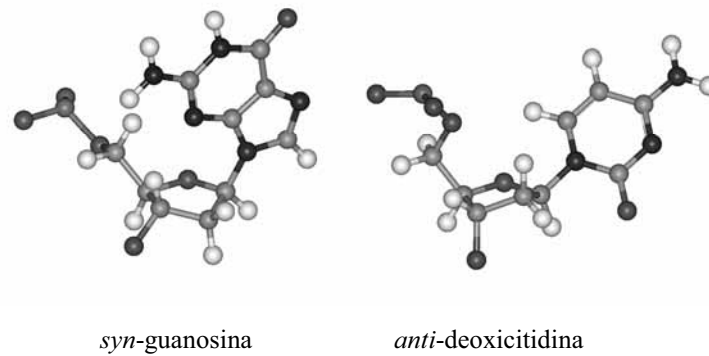


Figura 14. Ciclo de pseudorrotación (91).

3.1.2.2 El enlace glucosídico (γ)

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, 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 eclipsa la base, mientras que en la *anti* el azúcar está en posición opuesta (figura 15). 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 (94). 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 15. Conformaciones *syn* y *anti* de dos nucleósidos.

3.1.2.3 El enlace exocíclico C4'-C5' (γ)

El enlace exocíclico C4'-C5' se encuentra mayoritariamente en las tres conformaciones alteARNdas. En una proyección de Newman (se mira la posición de O5' respecto a O4') si el O5' está entre O4' y C3' es *gauche-gauche*, si está en *trans* al C3' y al lado del O4' es *gauche-trans*, y si está en *trans* al O4' y al lado de C3' es *trans-gauche* (figura 16). 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 (95). Generalmente se encuentra estabilizada por puentes de hidrógeno entre la base nitrogenada y el O5'.

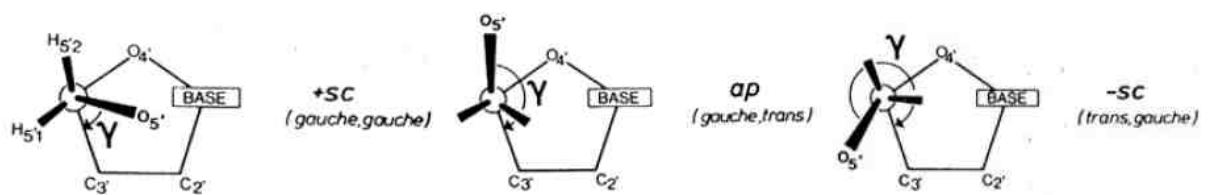


Figura 16. Definición de los rangos del ángulo diedro del enlace C4'-C5'(88).

3.1.2.4 Los enlaces C-O (β, ϵ) y P-O (α, ζ)

La flexibilidad conformacional de los ácidos nucleicos también viene dada por los enlaces rotables del esqueleto azúcar-fosfato. En el caso de los átomos situados a 5' del azúcar, α presenta su conformación canónica alrededor de 270° (rotación respecto P-O5'; figura 12)

aunque el 1% del tiempo visita la región de 100° de manera correlacionada con el ángulo γ , que cambia de 60° a 200° ($C4'-C5'$; véase punto anterior) (96). El ángulo β ($C5'-O5'$) se suele encontrar entorno a 180° . Por otro lado, los átomos situados a $3'$, los ángulos ϵ/ζ ($C3'-O3'/P-O3'$) presentan rotaciones concertadas generando dos conformaciones: BI y BII. En la primera los ángulos ϵ/ζ presentan valores de $180^\circ/270^\circ$ respectivamente (85%), siendo de $270^\circ/180^\circ$ en la segunda (15%) (97).

3.1.3 Propiedades fisicoquímicas

Las interacciones de las bases nitrogenadas entre sí y el reconocimiento de la fibra por factores externos, fenómenos cruciales en la función biológica de los ácidos nucleicos, están enormemente influenciados por sus características fisicoquímicas individuales. Propiedades como el valor de pK, la densidad de carga, o el estado tautomérico de las bases, deben ser conocidos de cara a una mejor comprensión del comportamiento de estas biomoléculas.

3.1.3.1 Ionización

El comportamiento ácido-base es una de las características más importantes, pues determina su carga, su capacidad de actuar como dador/aceptor de puente de hidrógeno y su tautomería. En el rango fisiológico de pH ($5 < \text{pH} < 9$), las bases nitrogenadas son neutras. Lo mismo ocurre con las pentosas, que sólo pueden perder el protón del grupo hidroxilo en medios extremadamente básicos ($\text{pH} > 12$). Cuando el grupo fosfato se une a dos nucleósidos, éste sólo tiene un único hidroxilo ionizable con un $\text{pK}_a \approx 1$, y sólo cuando se trata de un monoéster puede perder un segundo protón a pH 6-7. Así, pues, a pH fisiológico, los ácidos nucleicos son polianiones.

3.1.3.2 Densidad de carga

Las bases nitrogenadas son estructuras polares que presentan una distribución de carga muy localizada en determinados átomos (98), presentando momentos dipolares elevados (99). Esta distribución de carga justifica las interacciones no covalentes entre bases, tan importantes en el reconocimiento molecular, como son los puentes de hidrógeno.

Concretamente, los grupos -NH de las bases nitrogenadas (donde se concentra la carga positiva) son buenos dadores de puentes de hidrógeno, mientras que los pares libres del oxígeno de los grupos -C=O y de los nitrógenos de los anillos (donde se concentra la carga negativa) son buenos aceptores (figura 17). Por otro lado, la capacidad de formar puentes de hidrógeno también está asociada a los oxígenos de la ribosa y del fosfato.

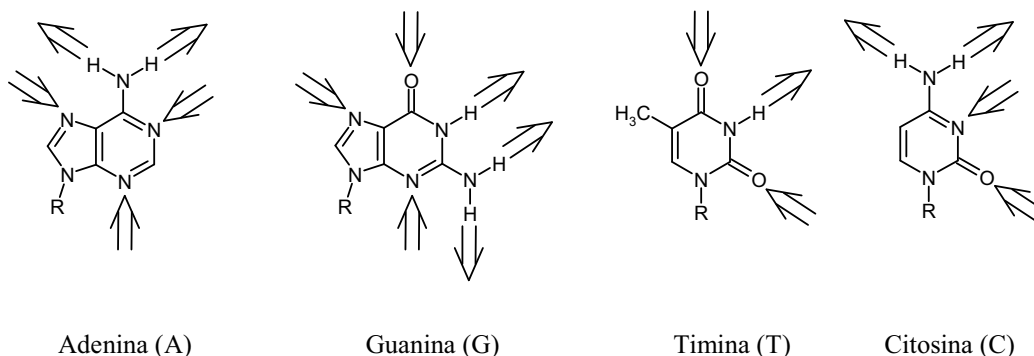
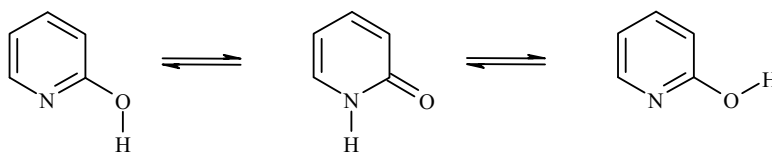


Figura 17. Patrón de centros dadores/aceptores de puentes de hidrógeno en las bases nitrogenadas.

Los centros dadores se muestran con una flecha apuntando hacia fuera; los centros aceptores se muestran con una flecha apuntando hacia dentro.

3.1.3.3 Tautomería

La tautomería es una propiedad clave para entender la funcionalidad de los ácidos nucleicos. El equilibrio tautomérico prototípico (100) implica estructuras alternativas que sólo difieren en la localización de los átomos de hidrógeno. Las bases nitrogenadas de los ácidos nucleicos disponen de un equilibrio ceto-enol y de uno amino-imino (figura 18). Las bases más comunes existen mayoritariamente en la forma tautomérica ceto-amino. Se ha establecido, mediante técnicas espectroscópicas, que dichos tautómeros son más estables que los enol-imino (101) en solución acuosa. No obstante, modificaciones en el heterociclo o cambios en el solvente pueden alterar la propensión tautomérica.



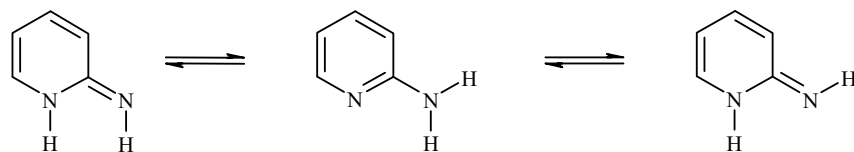


Figura 18. Modelo de la tautomería ceto-enol para 2-piridona y de la tautomería amino-imino para 2-aminopiridina.

Cabe destacar que muchas mutaciones espontáneas (102) tienen su origen en cambios temporales de la preferencia ceto-amino, ya que altera totalmente el patrón de reconocimiento afectando al tipo de afinidad de las bases y a su reactividad.

3.2 Estructura secundaria

Tras el modelo teórico de doble hélice propuesto por Watson y Crick para la estructura del ADN (5), la idea de que esta biomolécula puede explorar disposiciones tridimensionales alternativas ha ido consolidándose. De hecho, el ADN, así como el resto de ácidos nucleicos, puede adoptar una gran diversidad de formas en función de las condiciones experimentales, de la secuencia de las bases o del proceso biológico en que esté involucrado. Todas ellas presentan en la mayoría de los casos una estructura helicoidal, siendo posibles hélices intra- o intercatenarias de 2, 3 y hasta 4 cadenas. Una clasificación aceptable de estas estructuras podría ser aquella que considera por un lado las conformaciones canónicas (formadas por la unión de dos cadenas antiparalelas según los apareamientos de Watson y Crick), y por otro lado las conformaciones no canónicas (en las que no se cumple esta sencilla regla).

3.2.1 Interacciones no covalentes

Los requerimientos estructurales de los ácidos nucleicos son el fruto de la necesidad de estabilizar zonas cercanas con una elevada densidad de carga negativa (grupos fosfato), sosteniendo una estabilidad mecánica lo suficientemente intensa que no evite la flexibilidad necesaria para permitir los procesos de replicación o transcripción (véase apartado 3.3.1). La estructura y la flexibilidad de este sistema tan peculiar se modelan gracias al concurso de varias fuerzas. Principalmente, podemos distinguir tres tipos de interacciones responsables de estabilizar la conformación de los ácidos nucleicos:

3.2.1.1 Puentes de hidrógeno

Un puente de hidrógeno es aquella interacción de naturaleza electrostática entre un grupo químico dador (A-H) y un aceptor (B) que cumpla dos requisitos:

- la electronegatividad de A ha de ser elevada, de forma que pueda apropiarse parcialmente de la densidad electrónica del átomo de hidrógeno, y
- B ha de poseer pares electrónicos libres o electrones π polarizables.

El sistema de tres átomos A-H...B suele ser lineal, con una desviación no superior a los 20°, y la distancia entre heteroátomos está establecida entre 2.6 y 3.5 Å (100).

El reconocimiento entre bases se da por enlaces de puentes de hidrógeno. *A priori* existen muchos patrones de formación, pero en el ADN biológico el modo de reconocimiento más común es el propuesto por Watson y Crick (WC) (5) (figura 19), cuyas características son:

- la preferencia de coplanaridad de las dos bases
- la formación de tres puentes de hidrógeno para el apareamiento C·G y dos para T(U)·A
- el tamaño del par de bases G·C y A·T es prácticamente el mismo y el ángulo de inclinación de los dos pares de bases muy similar (geometría isomorfa), lo que permite la formación de una doble hélice regular (figura 20).

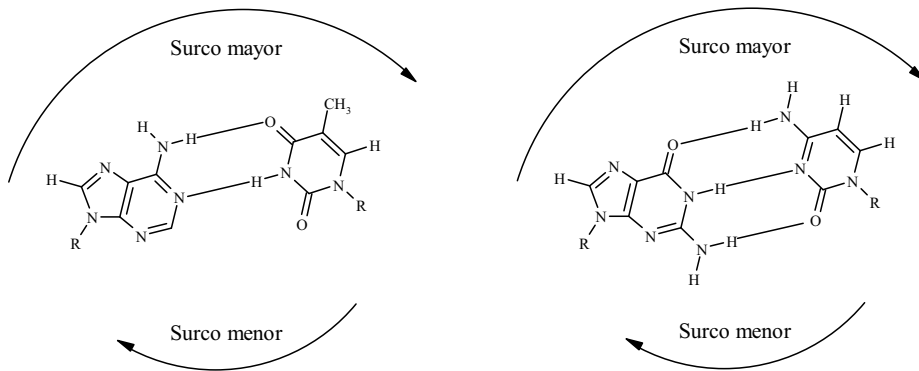


Figura 19. Pares de bases WC encontrados en las dobles hélices de ADN y de ARN. Izquierda A·T y derecha G·C. Se muestra el surco menor y el surco mayor de la hélice.

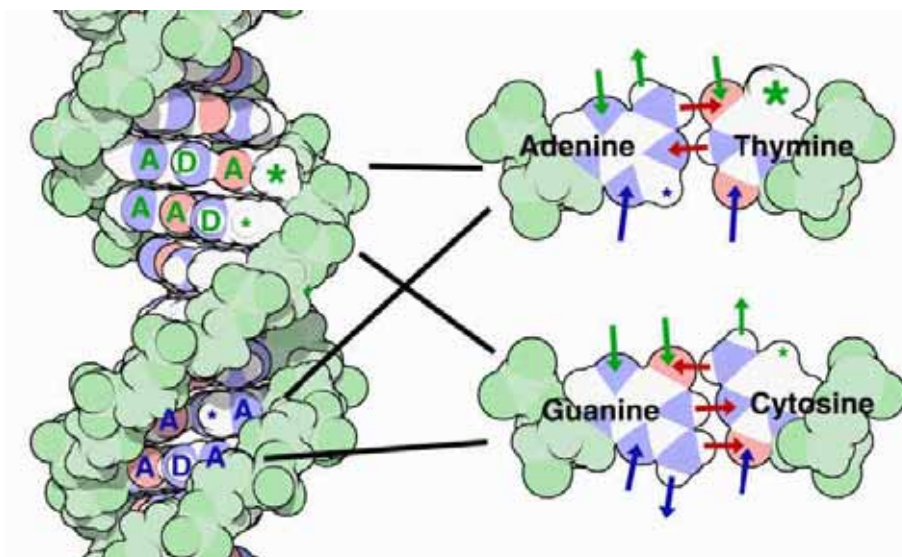
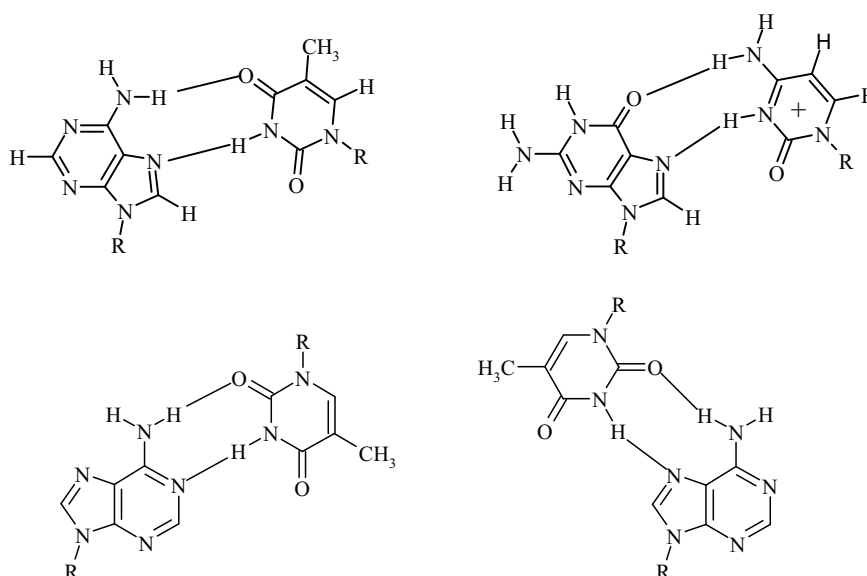


Figura 20. Posibilidades de puente de hidrógeno en una hebra de ADN. En rojo están los puentes de hidrógeno entre bases, en azul los puentes de hidrógeno en el surco menor y en verde los del surco mayor. A=aceptor, D=donador y * donadores secundarios (www.rcsb.org).

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 20).

Existen otros modos de reconocimiento diferentes al propuesto por Watson y Crick. De hecho, es posible encontrar 28 modos de aparear las bases (neutras) con alta eficiencia (al menos 2 puentes de hidrógeno lineales). Es interesante ver que es posible formar apareamientos purina·purina y pirimidina·pirimidina, además de purina·pirimidina. Si incluimos las posibilidades que proporciona el cambio en el estado de ionización de las bases o apareamientos con un solo puente de hidrógeno, las posibilidades son enormes. Algunos modos de apareamiento detectados en los ácidos nucleicos son los pares Hoogsteen (103) (en forma normal y reversa) y los *wobble* (figura 21). 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 (104) (apartado 3.3.1). En ellos se suelen perder puentes de hidrógeno, pero se favorecen las interacciones de apilamiento.



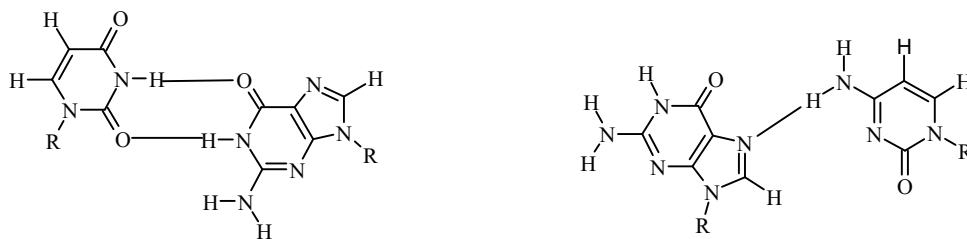


Figura 21. Arriba: apareamientos tipo Hoogsteen entre A·T y G·C protonada. Medio: apareamientos entre A·T de tipo WC reverso y Hoogsteen reverso. Abajo: apareamiento tipo *wobble* entre U·G y G·C (Hoogsteen).

3.2.1.2 Interacciones de apilamiento o *stacking*

Las bases también interactúan entre ellas por apilamiento o *stacking*. En este tipo de interacciones dos bases se disponen aproximadamente paralelas una sobre la otra a una distancia cercana a 3.4 Å, evitando contactos electrostáticos desfavorables (se evitan alineamientos paralelos de dipolos o que átomos de la misma carga se superpongan) (105,106). La interacción de *stacking* tiene una fuerte componente de interacción de van der Waals relacionada con solapamiento de los orbitales π de los anillos aromáticos. Asimismo, viene modulada por el llamado efecto hidrofóbico, asociado a la desestructuración de las moléculas de agua entorno a las zonas apolares de las bases.

Diversas técnicas experimentales, como la osmometría, la ultracentrifugación, la calorimetría, la desnaturalización térmica y la RMN, se han empleado en la descripción termodinámica de este tipo de interacciones (107-110), siendo a su vez confirmados por estudios teóricos (111-113). Según las evidencias obtenidas, los apilamientos que involucran G son los más estables, siendo los de menor estabilidad los que contienen T. Existe una disminución progresiva de la estabilidad a medida que se reduce el tamaño de los anillos implicados. De este modo, puede establecerse la siguiente gradación: purina·purina > purina·pirimidina > pirimidina·pirimidina.

En fase gas, la estabilización debida a las interacciones de puente de hidrógeno supera a la de *stacking*. Sin embargo, en solventes polares como agua, la interacción de bases por puentes de hidrógeno es desfavorable, mientras que el apilamiento se hace más favorable debido al efecto hidrofóbico comentado anteriormente. Por lo tanto, se considera que el *stacking* es el motor del empaquetamiento de la doble hélice, mientras que los puentes de hidrógeno sólo se

dan cuando se ha eliminado el agua que existía entre las bases. Una vez formados, los puentes de hidrógeno estabilizan ciertamente la estructura del ADN, pero no son la fuerza directriz que explica las razones por la que se forma un ácido nucleico (114,115).

3.2.1.3 El efecto del solvente

El entorno acuoso juega un papel decisivo en la estructura final de los ácidos nucleicos. Una de sus acciones más relevantes es el apantallamiento de las interacciones entre los grupos fosfatos del ADN, cuya repulsión electrostática tiende a desestructurar la hélice (80). Precisamente, la estructura helicoidal adoptada por los ácidos nucleicos sería una consecuencia directa de las interacciones de *stacking* entre las bases y de la minimización de la repulsión electrostática entre grupos fosfato.

Las moléculas de agua se estructuran alrededor del ADN formando dos capas de hidratación. En la capa más interna se ubican un promedio de 11 a 20 moléculas por nucleótido, que se unen directamente mediante puentes de hidrógeno a átomos del esqueleto azúcar-fosfato o de las bases, formando espinas de hidratación presentes como elemento constituyente en varios tipos de ácidos nucleicos (116). La segunda capa es menos estructurada y en muchas ocasiones indistinguible del resto del medio acuoso por lo que respecta a permeabilidad de iones y movilidad de aguas, aunque es clave para estabilizar la estructura.

El ADN es un polielectrolito en condiciones fisiológicas y por ello normalmente está rodeado de contraiones que neutralizan parcialmente su carga. Estos cationes suelen ser inorgánicos e hidratados, pero también es posible encontrar cadenas laterales de aminoácidos o poliaminas. Los iones inorgánicos presentes en la célula en concentraciones micromolares son Mg^{2+} , Ca^{2+} , Na^+ y K^+ , y por ello se hallan normalmente unidos al ADN. Atraídos por la carga del fosfato, suelen estar ligados a sus oxígenos, aunque también pueden asociarse con las bases (117). 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^{2+} y Mg^{2+} en algunos cristales (118) o incluso cambios globales en la estructura de ADN (89).

3.2.2 Parámetros helicoidales

Como hemos comentado anteriormente, los polinucleótidos suelen adoptar estructuras secundarias de hélice, la mayoría de doble hebra, que se pueden describir mediante parámetros geométricos helicoidales. Éstos tratan la molécula como una serie de pasos consecutivos y en cada uno de ellos se define un sistema de coordenadas. La relación entre estos sistemas de coordenadas locales fija una serie de parámetros capaces de describir la hélice de manera global.

3.2.2.1 Parámetros entre pares de bases

Gracias al acuerdo de Cambridge (119), se establecieron la nomenclatura y guías para calcular los parámetros helicoidales (figura 22). En nuestros estudios hemos utilizado la implementación del esquema CEHS (120) en el programa X3DNA (121), que utiliza el concepto recomendado de *mid-step triad* (MST) para asegurarse que los parámetros calculados son invariantes a la inversión del eje molecular.

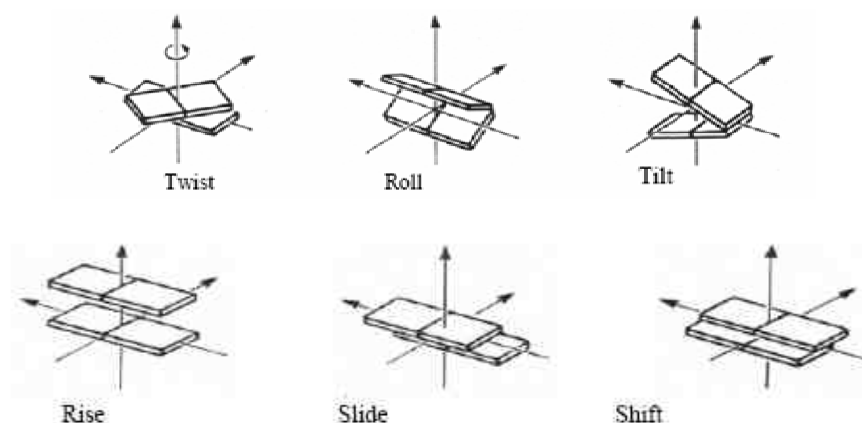


Figura 22. Nomenclatura de los parámetros entre pares de bases formados por tres rotaciones (arriba) y tres traslaciones (abajo).

La posición de cada par de bases se define por el vector r_i , que representa el origen de la tríada de referencia (x_i, y_i, z_i) usada para definir la orientación. El origen (r_i) es el punto medio de la línea que conecta el C8 de la purina con el C6 de la correspondiente pirimidina. Además se asigna arbitrariamente una cadena como I, siendo el sentido positivo del eje z el dado por la orientación $5' \rightarrow 3'$, el del eje y es hacia la cadena I, y el del eje x hacia el surco mayor (figura 23).

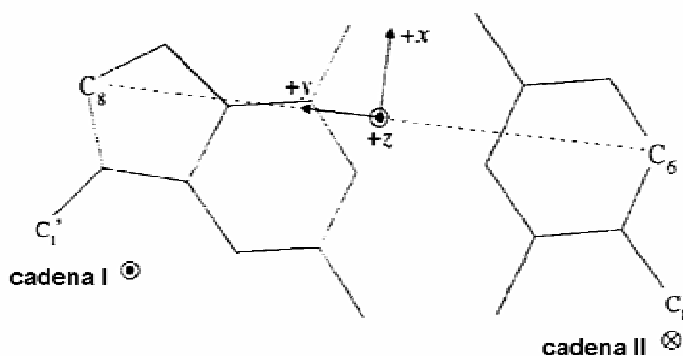


Figura 23. Esquema del vector de origen y la tríada de referencia para los pares de bases. La cadena I apunta hacia fuera mientras que la cadena II hacia dentro.

El algoritmo para definir los parámetros de la figura 22 tiene los siguientes pasos (122):

- Cálculo del ángulo *roll-tilt* (Γ) y el eje *roll-tilt* (rt):

$$\cos \Gamma = z_i \cdot z_{i+1} \quad [89]$$

$$rt = z_i \times z_{i+1} \quad [90]$$

- Rotación del paso $i, i+1$ según $+\Gamma/2, -\Gamma/2$ sobre el eje *roll-tilt* para obtener el eje z_{mst} .

Después se superponen las tríadas de los pasos consecutivos según z_{mst} para definir los ejes x_{mst}, y_{mst} como las respectivas bisectrices (figura 24)

- Definición de los tres parámetros rotacionales (figura 22):

$$\cos \Omega = y'_i \cdot y'_{i+1} \quad [91]$$

donde Ω es el *twist* (describe la rotación según el eje molecular), que se define con los vectores y'_i, y'_{i+1} de las tríadas de referencia de los pasos $i, i+1$ después de rotarlas según el eje *roll-tilt*. El sentido positivo del ángulo se establece cuando la rotación es dextrógira.

Para definir los parámetros de *roll* (ρ) y *tilt* (τ) (describen las rotaciones hacia las cadenas o los surcos respectivamente), necesitamos definir el ángulo (ϕ) entre el eje *roll-tilt* (rt) y y_{mst} (figura 24):

$$\cos \phi = rt \cdot y_{mst} \quad [92]$$

$$\rho = \Gamma \cos \phi \quad [93]$$

$$\tau = \Gamma \sin \phi \quad [94]$$

El sentido positivo del *roll* es cuando las bases se acercan en el surco mayor, y el sentido positivo del *tilt* es cuando se acercan en la cadena I.

- Definición de los tres parámetros traslacionales (figura 22):

$$\begin{bmatrix} D_x \\ D_y \\ D_z \end{bmatrix} = (r_{i+1} - r_i) T_{mst} \quad [95]$$

donde D_x, D_y, D_z son *shift, slide y rise*, respectivamente, y T_{mst} es la matriz que contiene los tres vectores que definen la MST.

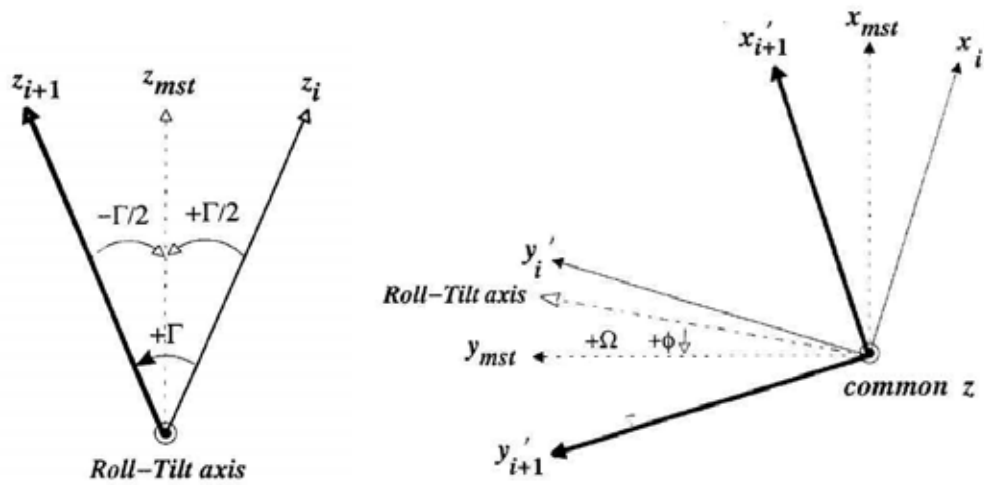


Figura 24. Descripción de los vectores que constituyen la MST, Γ , Ω y ϕ (122).

3.2.2.2 Parámetros entre bases complementarias

De una forma análoga a los parámetros entre pares de bases, los parámetros entre bases complementarias se definen con la ayuda de la *mid-base triad* (MBT) definiéndose tres rotaciones y tres traslaciones (figura 25) (122).

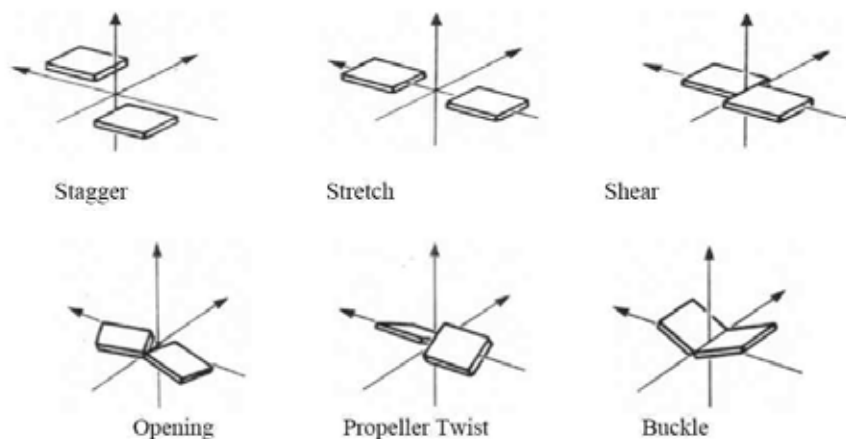


Figura 25. Nomenclatura de los parámetros entre bases complementarias formados por tres rotaciones (arriba) y tres traslaciones (abajo).

El vector de origen y la tríada de referencia correspondiente a la base tanto de la cadena I $(r_I; x_I, y_I, z_I)$ como de la cadena II $(r_{II}; x_{II}, y_{II}, z_{II})$ se asignan como se muestra en la figura 26. Los vectores de origen son el punto medio de las líneas N_1-C_4 en las purinas y N_3-C_6 en las pirimidinas, las cuales definen el eje y siendo el sentido positivo hacia la cadena I.

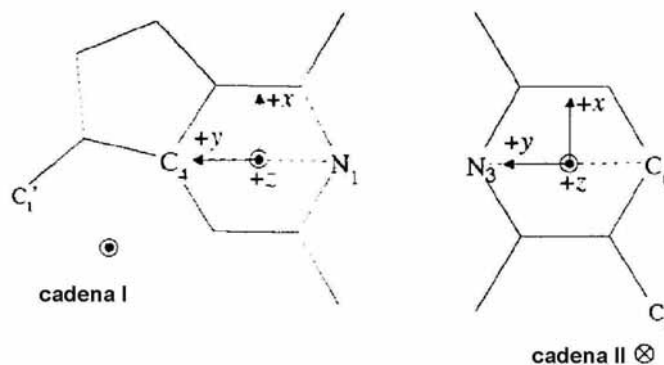


Figura 26. Esquema del vector de origen y la tríada de referencia para las bases. La cadena I apunta hacia fuera mientras que la cadena II hacia dentro.

De manera parecida al apartado anterior se calcula el ángulo *buckle-opening* (γ) y el eje *buckle-opening* (bo), los cuales permiten obtener el eje y_{mbt} , y posteriormente a la superposición de las tríadas de las dos bases, el resto de la MBT (figura 27):

$$\cos \gamma = y_I \cdot y_{II}$$

[96]

$$bo = y_I \times y_{II} \quad [97]$$

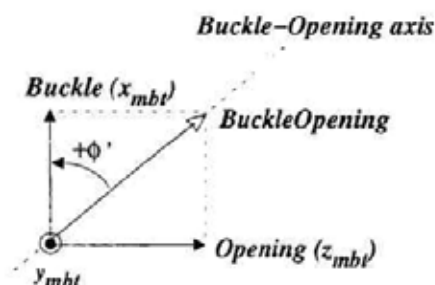


Figura 27. Descripción de los vectores que constituyen la MBT y ϕ' (122).

Entonces se calculan las rotaciones y traslaciones (figura 25):

$$\cos \omega = x'_I \cdot x'_{II} \quad [98]$$

donde ω es el *propeller twist* y x'_i los ejes después de la superposición.

$$\cos(\phi') = bo \cdot x_{mbt} \quad [99]$$

$$\kappa = \gamma \cos(\phi') \quad [100]$$

$$\sigma = \gamma \sin(\phi') \quad [101]$$

donde κ es el *buckle* y σ el *opening*.

$$[S_x S_y S_z] = (r_I - r_{II}) T_{mbt} \quad [102]$$

donde S_x, S_y, S_z son *shear, stretch* y *stagger*, respectivamente, y T_{mbt} la matriz que contiene la MBT.

3.2.2.3 Parámetros globales

Los parámetros globales utilizados en este estudio superan la escala de los pares de bases para describir fragmentos cortos de ácidos nucleicos (hasta 10 pares de bases o una vuelta entera) (123). Este método se basa en utilizar los resultados extraídos del programa X3DNA (121) para calcular el *twist, roll, tilt* y *stretch* propios de todo el fragmento:

- *twist* global: se obtiene de la adición de los valores de *twist* locales de los diferentes pasos.
- *stretch*: es la suma de las distancias entre pares de bases vecinos a lo largo de todo el fragmento.

- *roll* y *tilt* globales: se sigue un procedimiento muy parecido al del *roll* y *tilt* locales (figura 24) [92-94]. Se define una tríada intermedia equivalente a la MST, donde el eje z_{mst} se construye gracias a los ejes z de los pares de bases primero y último. Para los ejes x_{mst}, y_{mst} se utilizan los propios del par de bases central del fragmento y se proyectan al plano perpendicular al z_{mst} . Si el número de pasos de la molécula es par, los vectores x_{mst}, y_{mst} se obtienen del promedio de los dos pasos centrales.

3.2.2.4 El model *Worm-like Chain* (WLC)

Es un modelo útil para describir grandes fragmentos de ADN. Se asume que la posición de equilibrio del polímero es la recta y que por la agitación térmica tiene una capacidad de doblarse que se cuantifica en base a la *persistence length* (P) (124). Esta constante es la longitud a la cual se pierde la correlación de la dirección del polímero, y se define mediante el ángulo (θ) formado por las tangentes al polímero a la posición cero y a una distancia L :

$$\langle \cos \theta \rangle = e^{-L/P} \quad [103]$$

En este estudio hemos usado este concepto para transformar los ángulos globales (*roll*, *tilt* y *twist*) definidos en el punto anterior a distancias gracias al *stretch* o distancia total de la molécula.

3.2.3 Estructuras canónicas

Tradicionalmente se clasifican las estructuras canónicas en dos grupos, tres familias y seis tipos. Los grupos son dextrógiros y levógiros. Las familias son A, B (dextrógiros) y Z (levógira). Todas las familias tienen un único tipo, a excepción de la B que cuenta con 4 tipos: B, C, D y T.

3.2.3.1 Formas dextrógiros. La forma A y B

Se caracterizan en que la doble hélice gira en sentido horario. El aspecto de un ADN de la familia B y uno de la familia A es totalmente diferente (figura 28). Las principales características de estas familias se resumen en la tabla 1.

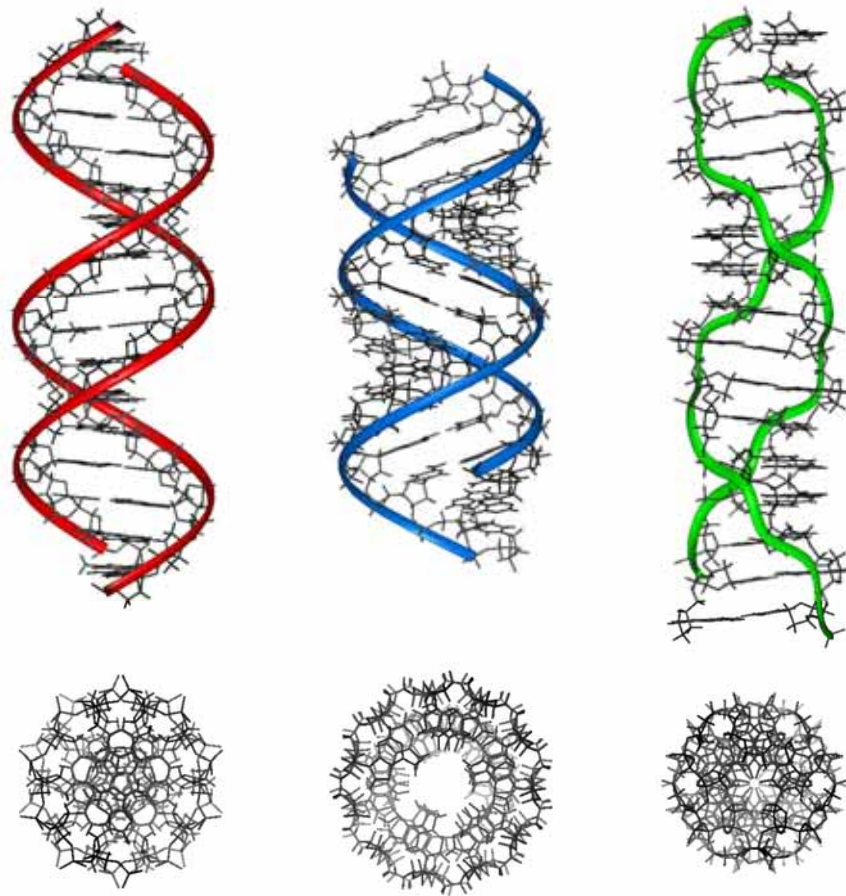


Figura 28. Modelos de doble hélice canónicos. Se observa claramente el surco mayor y el surco menor formados por el esqueleto de azúcar-fosfato. Las bases nitrogenadas aparecen en el interior de la hélice. La forma B se muestra en color rojo, la A en azul y la Z en verde.

El tipo A da lugar a una hélice más ancha al contar con más residuos por vuelta de hélice. Esto mismo se refleja en el *twist*, que pasa de 36-38° en la familia B a 33° en la familia A. Otra característica diferencial es el *puckering* de los azúcares, que en la familia B es C2'-*endo* (C3'-*exo*), mientras que en la familia A es C3'-*endo*. Ello, junto a la torsión alrededor del enlace glucosídico (χ), representan las principales diferencias en el esqueleto azúcar-fosfato. Como consecuencia, se encuentran diferencias en parámetros como la distancia entre C1' intracatenarios y el parámetro Z_p , que es la proyección del átomo de fósforo en el eje z de la MST (figura 29) (17). También son destacables las diferencias entre las características de los surcos. En la familia B el surco mayor es más ancho y un poco más profundo que el surco menor. En cambio, en la familia A el surco mayor es mucho más profundo que el surco menor, pero también más estrecho. La familia B se distingue por los pares de bases perpendiculares al eje de la hélice, la forma estilizada y por lo compacta que es ésta en su

interior. Por el contrario, la forma A aparece mucho menos estilizada, siendo visible un hueco en el interior de la hélice, así como la inclinación de los pares de bases con respecto al eje de la hélice (figura 28).

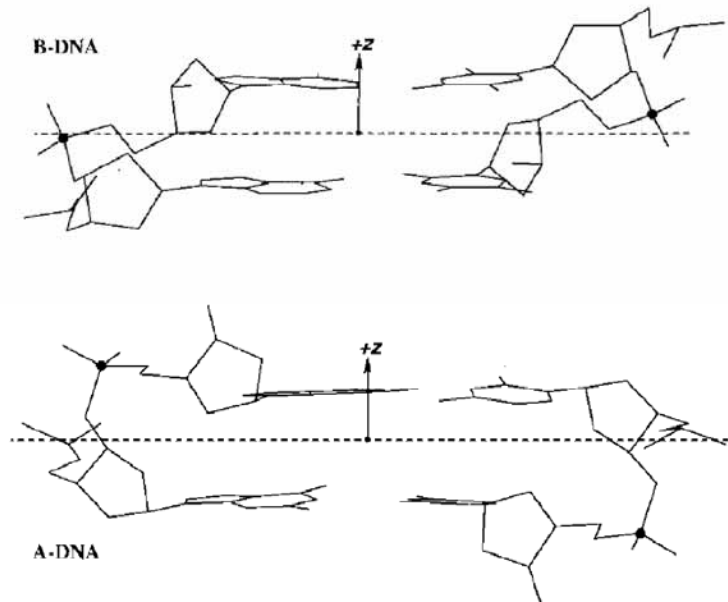


Figura 29. Ilustración del parámetro Z_p en las formas A y B. Éste parámetro corresponde al desplazamiento del fósforo fuera del plano xy del MST entre pasos vecinos (17).

| Familia B | Familia A |
|--|--|
| 8-10 nucleótidos por vuelta | 11 nucleótidos por vuelta |
| bases perpendiculares al eje de la hélice con poca inclinación | bases inclinadas 20° con respecto al eje de la hélice |
| <i>twist</i> de 33° y <i>roll</i> de 0° | <i>twist</i> de $36-38^\circ$ y <i>roll</i> de 5° |
| Hélice compacta en su interior, <i>slide</i> de 0 \AA | Existe un hueco en el interior, <i>slide</i> de -2 \AA |
| <i>puckering</i> del azúcar en zona S (C2'-endo/C3'-exo), χ de 260° | <i>puckering</i> del azúcar en zona N (C3'-endo), χ de 200° |
| $d(C1'-C1')$ de 5 \AA y Z_p -0.5 \AA | $d(C1'-C1')$ de 5.5 \AA y Z_p 2.2 \AA |
| surco mayor más ancho y algo más profundo que el surco menor | surco mayor más estrecho y profundo que el surco menor |

Tabla 1. Características de las formas dextrógiras

La familia B

La familia B es la forma más habitual del ADN, incluyendo los tipos B, C, D y T (116,125,126). Las características prototípicas las presenta el tipo B-ADN, que es el más habitual *in vivo* y el que se considera biológicamente activo. Los otros tipos de ADN dentro de la familia B son muy similares. Las características diferenciales son que el C-ADN es ligeramente más estilizado y con surcos más estrechos, y que las formas D y T presentan un surco menor profundo, pero muy estrecho. Estas formas aparecen en ADN sintéticos, en ADN naturales en diversas condiciones de humedad relativa (C-ADN) y en determinados fagos (T-ADN).

La familia A

La familia A es la forma que adopta en su mayoría un dúplex de ARN. Además, es la segunda forma más detectada en el ADN especialmente en condiciones de baja humedad (inducidas por solventes alifáticos como el etanol o por valores altos de fuerza iónica) (18,127,128), secuencias ricas en G·C (129), y formando complejo con proteínas (como diversas polimerasas o la retrotranscriptasa reversa (17)).

3.2.3.2 Forma Levógira. La forma Z

La característica más relevante es que la doble hélice gira en sentido levógiro. Presenta una estructura más larga y delgada que las anteriores. Su aspecto es inconfundible por la forma de zig-zag del esqueleto de fosfatos (figura 28). Las características principales se muestran en la tabla 2 (130,131).

| Z-ADN |
|---|
| 12 nucleótidos por vuelta |
| 1 de los 2 nucleótidos del par de bases en <i>syn</i> |
| <i>Puckering</i> del azúcar C2'- <i>endo</i> en <i>anti</i> (Sur) y C3'- <i>endo</i> en <i>syn</i> (Norte) |
| Surco menor muy profundo y estrecho. Surco mayor casi inexistente. |

Tabla 2. Características de la forma levógira.

El Z-ADN se ha detectado *in vivo* en secuencias purina-pirimidina antes G, C o en secuencias en las que alguna base se ha modificado por metilación o bromación, especialmente citosinas. La forma Z-ADN se puede interconvertir espontáneamente en formas B-ADN en presencia de fármacos o cambios del medio. Por ejemplo, un aumento de la fuerza iónica produce un cambio reversible de B a Z-ADN.

3.2.4 Conformaciones no canónicas

En condiciones normales, los ácidos nucleicos tienden a formar dúplexes, pero esta regla tiene excepciones como es el caso del ARN, pues al ser de cadena sencilla nunca se dará la complementariedad total, y junto a fragmentos dúplexes se producirán otras estructuras no usuales donde se podrán ver otros tipos de apareamientos entre bases además del de WC (apartado 3.2.1.1). No obstante, la formación de estructuras no canónicas no es patrimonio único del ARN, sino que el ADN también las puede formar (apartado 3.3).

Horquillas o *hairpins*

Son estructuras locales que se dan muy comúnmente en el ARN. La horquilla consta de un tallo y un bucle. El tallo es un fragmento de doble hélice normal, mientras que el bucle es cadena sencilla y es donde se produce el cambio de dirección que permite la formación del fragmento dúplex a partir de una sola cadena (figura 30).

Estructuras cruciformes

Se forma en dobles hélices de ADN que contienen complementariedad intracatenaria. Cada cadena forma una horquilla simétrica a la de la otra cadena dando el conjunto una estructura de cruz (figura 30).

Protuberancias o *bulges*

Se forman en estructuras dicatenarias y monocatenarias donde existe un grupo de bases que no son complementarios con la cadena vecina y por ello aparecen como protuberancias de cadena sencilla en la estructura general del dúplex (figura 30).

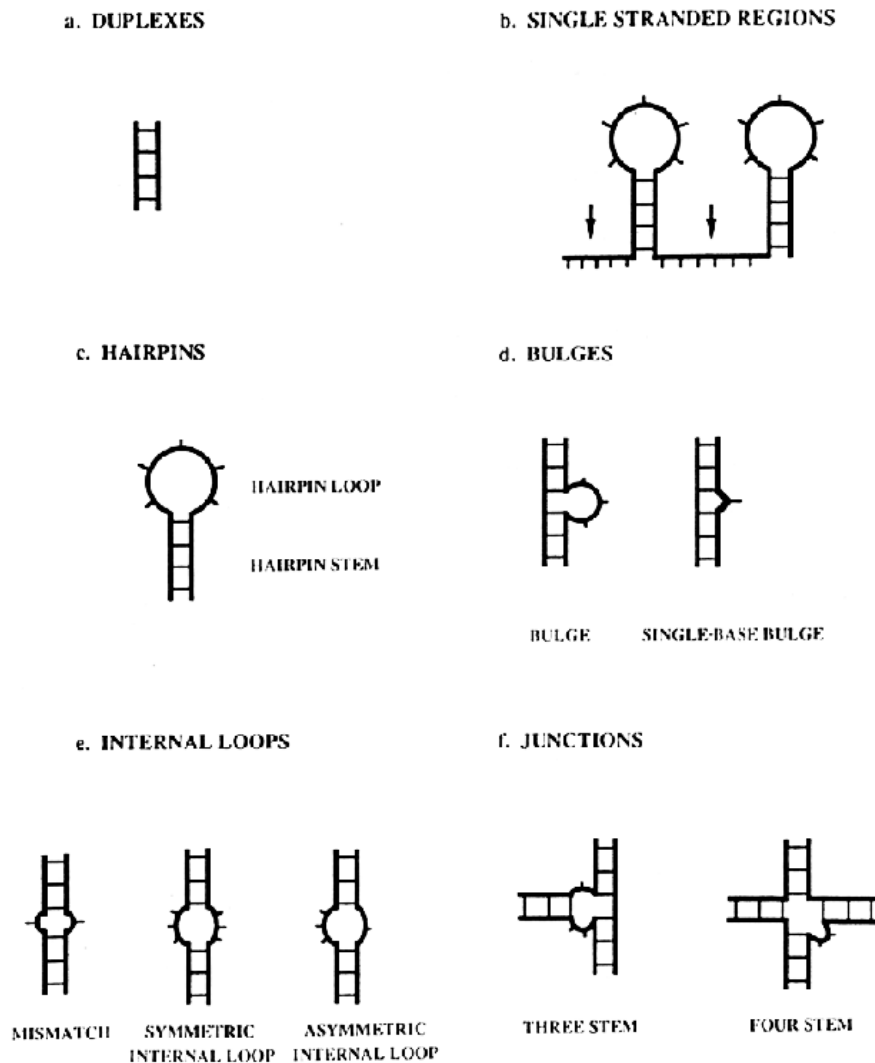


Figura 30. Diversas estructuras locales inusuales de los ácidos nucleicos (132).

Lazos o loops internos

Son una evolución de los *bulges*, que se producen cuando hay dos cadenas complementarias en toda su longitud excepto en un pequeño fragmento. Éste aparece como dos protuberancias más o menos simétricas en el dúplex (figura 30).

Cruces o junctions.

Son estructuras producidas cuando una cadena de ADN es complementaria en un fragmento no sólo con su cadena gemela, sino también con otra cadena, dando lugar a estructuras como las que se muestran en la figura 30.

Pseudoknots o nudos

Se forman en ARN cuando se produce la formación de un dúplex intracatenario con diferentes cambios en la dirección de la cadena (figura 31). Son estructuras muy importantes y conservadas en los ARN de transferencia (véase 3.3.1.3).

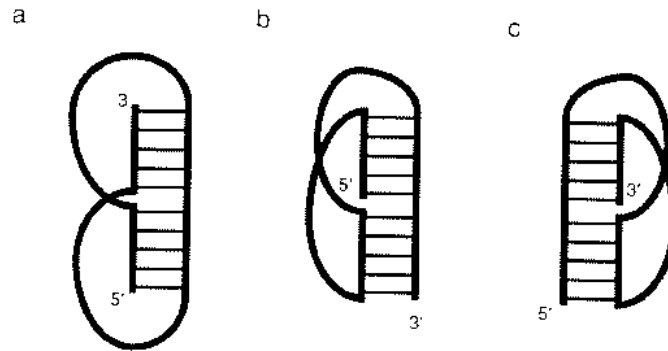


Figura 31. Estructura de 3 tipos de *pseudoknots* (132).

Doble hélice de cadenas paralelas

Pueden ser intra- o intermolecular basándose en motivos de puentes de hidrógeno tanto del tipo Hoogsteen como *reverse*-WC (figura 32).

Triples hélices

Se forman cuando en un dúplex WC de ADN una de las hebras es polipurina y es reconocida por una tercera cadena por el surco mayor mediante puentes de hidrógeno (figura 32). La tercera cadena puede ser purina o pirimidina y puede ser paralela o antiparalela a la cadena WC de purina. En general, si la tercera cadena es pirimidina los puentes de hidrógeno adicionales son Hoogsteen y la disposición es paralela. Por el contrario, si es purina los puentes de hidrógeno son *reverse*-Hoogsteen y la disposición es antiparalela.

Hélices de 4 cadenas

Se forma en cadenas simples de ADN constituidas por varias G consecutivas, que son capaces de adoptar 2 horquillas separadas y después unir las formando esta estructura (figura 32). Aunque en los últimos años se ha visto que son posibles tetraplexes con otras secuencias (133), los tetraplexes de G están presentes en los telómeros y se estabilizan gracias a iones monovalentes unidos en el canal interno de la estructura.

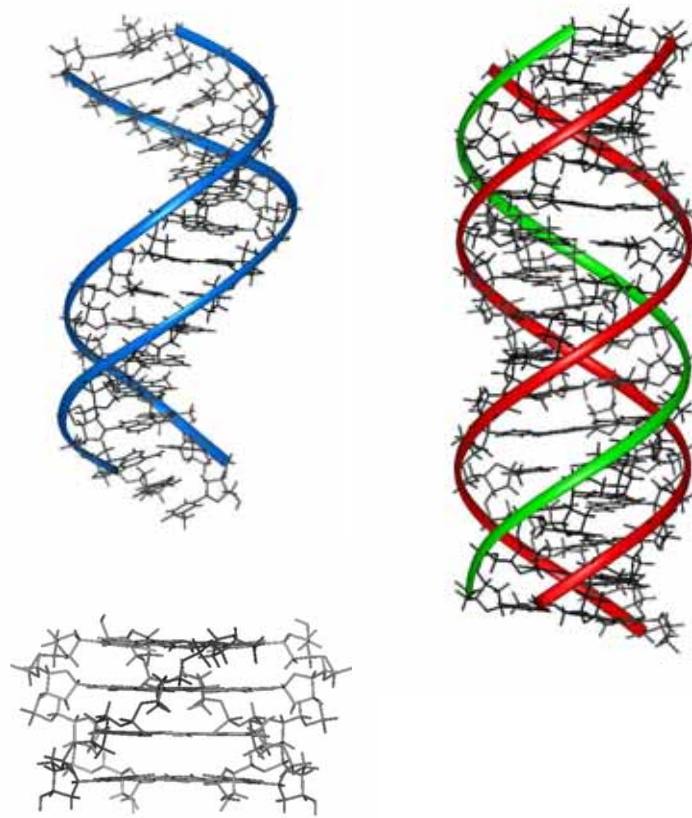


Figura 32. Modelos de doble hélice paralela (azul), de triple hélice (rojo el dúplex y verde la cadena extra unida al surco mayor) y de hélice de 4 cadenas (gris)

3.3 Estructura terciaria y función biológica

Los ácidos nucleicos realizan un amplio abanico de funciones a lo largo de la vida celular. Ya que su estructura tridimensional está íntimamente ligada a su función, en este capítulo iremos considerando ambos aspectos en paralelo (9).

3.3.1 El dogma de la biología molecular

Durante el año 1953, el modelo teórico de doble hélice propuesto por Watson y Crick para la estructura del ADN (5) permitió explicar la transmisión de la información genética a partir de simples principios químicos. En los siguientes años, se hicieron hipótesis sobre cómo la información genética se descodificaba en la célula, y en 1958 Crick formuló el dogma de la biología molecular (134) (figura 33).

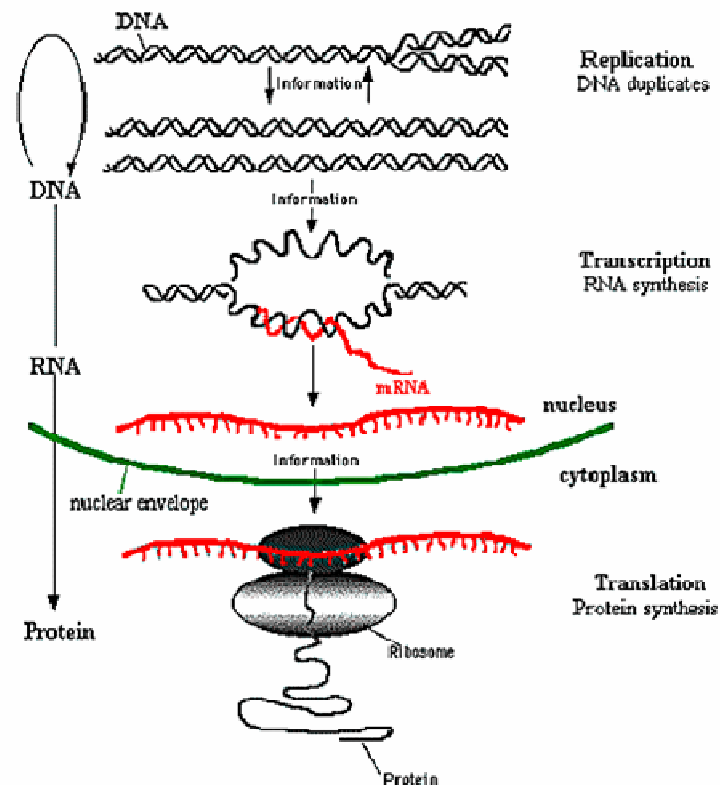


Figura 33. Dogma central de la biología molecular (www.accessexcellence.org).

Dicho dogma ilustra los mecanismos de transmisión y expresión de la herencia genética tras el descubrimiento de la codificación de ésta en la doble hélice del ADN. Propone una unidireccionalidad en la expresión de la información contenida en los genes de una célula, es decir, que el ADN es transcrito a ARN mensajero y que éste es traducido a proteína, elemento que finalmente realiza la acción celular. El dogma también postula que sólo el ADN puede replicarse y, por tanto, reproducirse y transmitir la información genética a la descendencia.

3.3.1.1 La replicación

Es el proceso mediante el cual el ADN es capaz de duplicarse a si mismo y así poder transmitir la información genética a la siguiente generación. Esto se consigue gracias a la complementariedad de bases obtenida por los apareamientos WC. Así, la información contenida en la doble hélice de ADN es redundante, pues cada cadena simple sirve de patrón para construir la complementaria y obtener de nuevo el dúplex completo (figura 33).

En este proceso se van separando las dos cadenas de la doble hélice y se van copiando las cadenas complementarias en el sentido 5'→3'. Una cadena está en el sentido correcto, mientras que la otra presenta el sentido contrario. La célula soluciona este problema utilizando pequeños fragmentos de ARN complementarios a la cadena de ADN a partir de los cuales se va sintetizando la segunda cadena de ADN (figura 34). Éstos, llamados fragmentos de Okazaki, generan una estructura de quimera (dúplex con una cadena de ADN y la otra un trozo de ADN unido a otro de ARN), sobre la cual aún hay debate acerca de si presenta una estructura de la familia A o B (135,136).

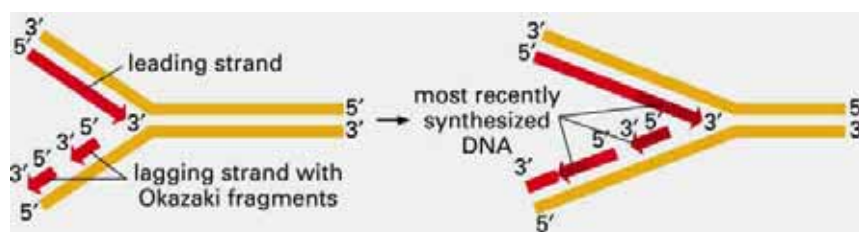


Figura 34. El proceso de la replicación. Se pueden ver los fragmentos de Okazaki (9).

La fidelidad de la replicación no es total, ya que puede presentar errores derivados por ejemplo de cambios tautoméricos, oxidaciones u otros cambios químicos en las bases. De esta

manera, la célula ha desarrollado mecanismos de reparación, que son muy sensibles a la conformación de la doble hélice siendo capaces de detectar cualquier distorsión local que pueda señalar una alteración en los apareamientos WC (137). No obstante, estos sistemas tampoco son infalibles, lo que puede conducir a cambios fenotípicos.

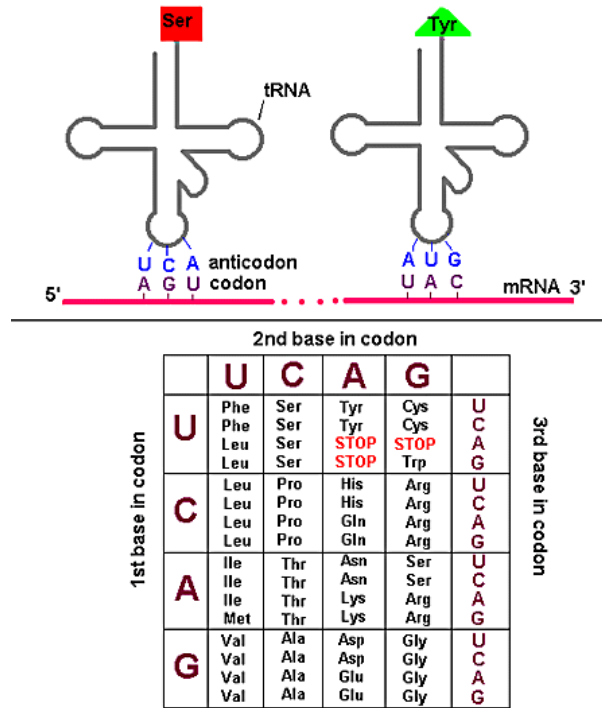
3.3.1.2 La transcripción

Es el primer proceso de expresión de un gen y consiste en generar una cadena de ARN complementaria a una cadena génica de ADN (generándose un dúplex ADN·ARN; figura 33). Este ARN (llamado mensajero, ARNm) es el encargado de salir del núcleo para llegar al citoplasma celular, donde tendrá lugar la traducción. El ARNm es de cadena simple y adopta una estructura tridimensional compleja formada con motivos canónicos y no canónicos del tipo *hairpins*, *loops*, *pseudoknots* etc. (véase apartado 3.2.4).

3.3.1.3 La traducción

Es el proceso por el cual se sintetizan las proteínas. Los diversos ARNs de transferencia (ARNt) juegan un papel clave, ya que relacionan la secuencia de tres bases consecutivas del ARNm (codón) con un aminoácido concreto conteniendo así el código genético (figura 35). Normalmente el ARNt reconoce el codón mediante apareamientos WC en las dos primeras bases, mientras que en la tercera la pauta de reconocimiento es más laxa pudiendo darse apareamientos del tipo *wobble* o Hoogsteen (apartado 3.2.1.1).

El ribosoma es la máquina que se encarga de dirigir los distintos ARNt para que lean de forma ordenada el ARNm, y así construir la proteína con la secuencia correcta de aminoácidos. Este complejo molecular está formado a partir de proteína y ARN ribosomal (ARNr), formando dos subunidades con una estructura altamente compleja (figura 36). Igual que el ARNm, el ARNt y el ARNr presentan una estructura terciaria formada a partir de elementos secundarios tanto canónicos como no canónicos (apartado 3.2.4).



The Genetic Code

Figura 35. El código genético con un esquema de la unión del ARNt con el ARNm

(www.accessexcellence.org).

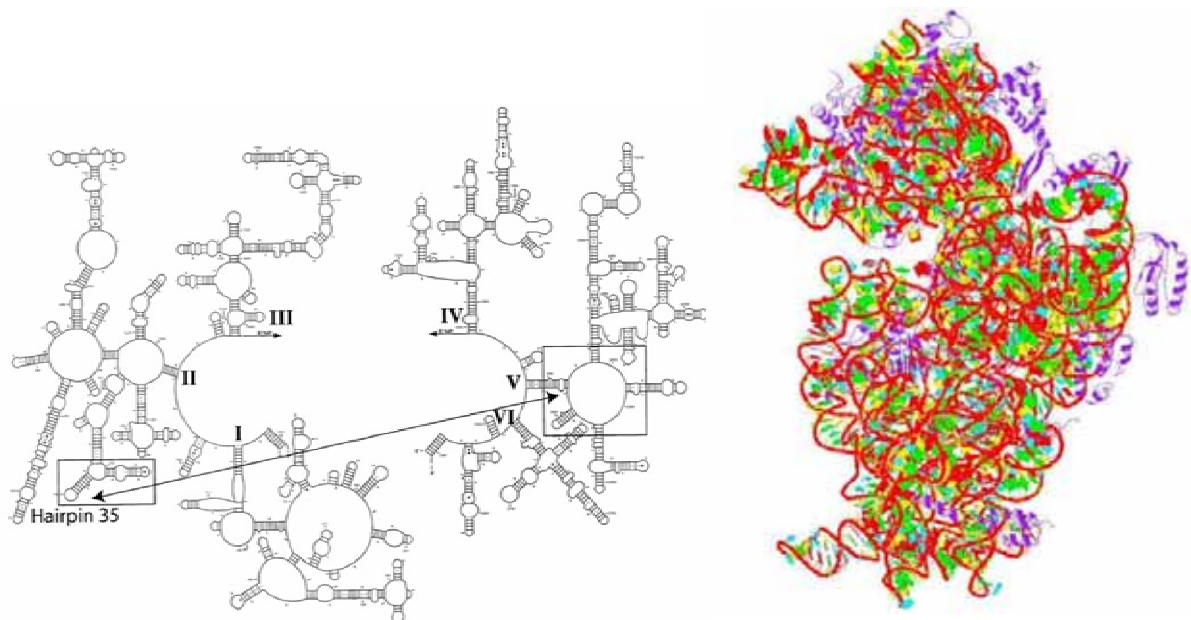


Figura 36. Subunidad pequeña del ribosoma. Izquierda: esquema de la estructura secundaria de su ARNr (138). Derecha: estructura tridimensional (139).

3.3.1.4 Los retrovirus

Los retrovirus, que provocan enfermedades tan importantes como el SIDA, fueron el primer sistema descubierto que rompía el dogma de la biología molecular. Éstos tienen el genoma en forma de cadena simple de ARN y al entrar en la célula huésped forman una doble hélice de ADN que se integra en su genoma (figura 37). Se rompe por tanto el flujo de la información, yendo al contrario de lo establecido por el dogma. Mecánicamente, el paso de ARN a ADN implica la enzima llamada retrotranscriptasa reversa, que es capaz de generar un dúplex de ADN usando una cadena simple de ARN (140).

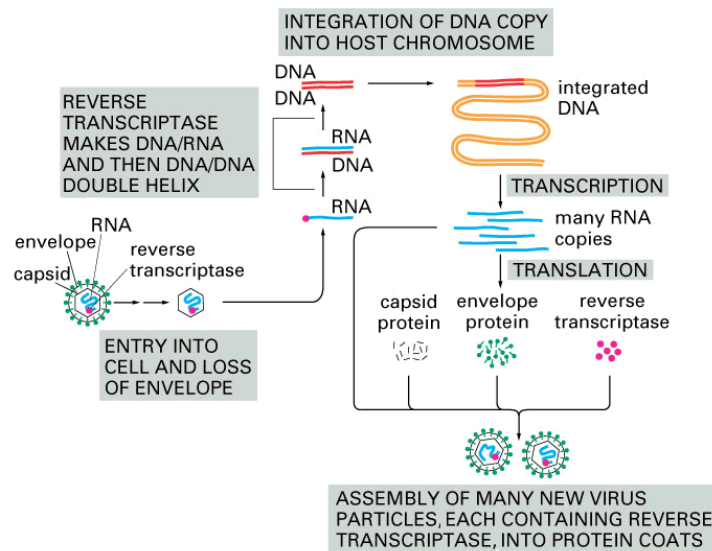


Figure 6-39 Essential Cell Biology, 2/e. (© 2004 Garland Science)

Figura 37. Ciclo de los retrovirus. De una cadena simple de ARN se forma la doble hélice de ADN por un camino intermedio ADN·ARN (9).

3.3.2 La terapia antisentido

Hace poco más de 25 años se vió la posibilidad de diseñar moléculas capaces de inhibir la expresión de un gen específico, rompiendo así el flujo de información celular (141). La idea consiste en inyectar un oligonucleótido en la célula que se una de manera específica (gracias a los apareamientos WC) a un ARNm, que codifique por una proteína cuya acción no es deseable para la célula. El dúplex resultante deja inactivo al ARNm o incluso, si la enzima

RNasa H se activa, es degradado. Esta tecnología ya ha sido capaz de lanzar los primeros fármacos al mercado (142) y de producir muchos otros en pruebas para el tratamiento de enfermedades como el cáncer, la inflamación y las infecciones virales (143). En consecuencia, se confirma el futuro prometedor de esta técnica, así como su gran impacto médico y tecnológico.

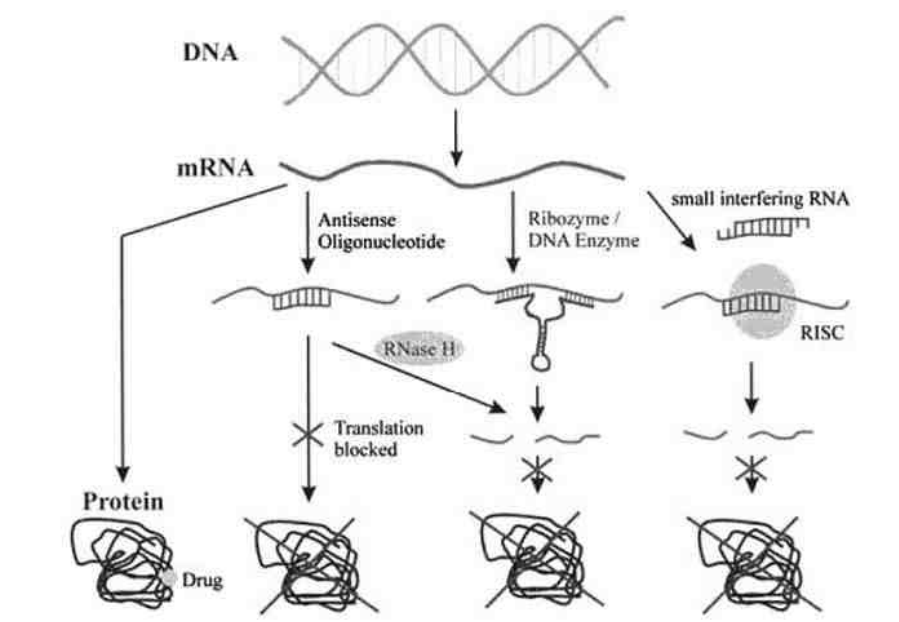


Figura 38. Comparación de las distintas estrategias antisentido o *antisense* (144).

Recientemente, se ha descubierto el mecanismo llamado ARN de interferencia o ARNi (145,146) como otra estrategia antisentido alternativa (véase 3.3.2.4), que actualmente también se está investigando con iguales finalidades médicas (147). Otro posible mecanismo es el de las ribozimas (véase 3.3.2.5), que son moléculas de ARN capaces de romper una cadena simple de ARN de forma equivalente a una enzima proteica (148), lo que también abre para ellos una vertiente como posibles agentes terapéuticos (147) (figura 38).

Parece ser que todos estos procesos se han desarrollado como mecanismos de defensa ante la entrada de ARN viral en la célula (149,150). Por ello, se requieren un conjunto de ribonucleasas cuya misión es degradar el ARN invasor, como son la RNasa III o Dicer que juntamente con el complejo RISC degrada ARNs, o la RNasa H que está especializada en cortar dúplexes ADN·ARN protegiendo así la célula de los retrovirus (151).

3.3.2.1 La RNasa H

La RNasa H es la enzima clave de la terapia antisentido, ya que su actividad, aunque no imprescindible, aumenta la eficacia de los oligonucleótidos antisentido al degradar el ARN diana y permitir el reciclaje de estas moléculas. La enzima es específica del dúplex ADN·ARN, siendo inactiva frente al ARNs y no pudiéndose unir al ADN de doble cadena (13). A pesar de la multitud de estudios bioquímicos y estructuras experimentales, las razones estructurales de su especificidad aún están sin resolver (13,152-159) (figura 39). Se ha especulado que la anchura intermedia del surco menor podría ser crucial (152,154,158), aunque otras diferencias tendrían un papel importante como el *puckering* o la proyección de grupos hacia el surco menor, siendo el sitio de unión de la enzima (155,159).

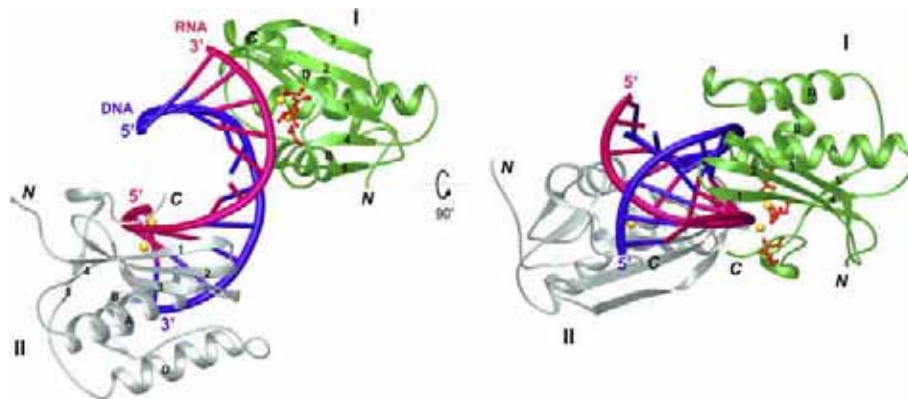


Figura 39. Dos perspectivas de la estructura de rayos X de la RNasa H unida al ADN·ARN (158).

Esta enzima ha sido identificada en organismos tan diversos como humanos y virus, como es el caso del VIH (virus de la inmunodeficiencia humana). Multitud de enzimas con actividad RNasa H han sido detectadas en procariontas, mientras que han sido descritos dos tipos en eucariotas (155). Aunque la RNasa H constituye una familia de proteínas de diversa masa molecular, la actividad nucleolítica y los requerimientos de sustrato son similares en los distintos isotipos. Por ejemplo, todas las RNasa H estudiadas hasta la actualidad no presentan especificidad de secuencia y necesitan cationes divalentes (Mg^{2+} o Mn^{2+}) para su actividad (160). Parece ser que la principal función sería eliminar los fragmentos de Okazaki durante la replicación, aunque se piensa que puede hacer otras funciones, como proteger la célula de

ataques virales o participar en la replicación del ADN mitocondrial, ya que se ha visto su presencia tanto en el núcleo como en el citoplasma (161).

3.3.2.2 Oligonucleótidos antisentido de 1ª generación

Los primeros oligonucleótidos antisentido (AXN) se diseñaron con el objetivo de aumentar la afinidad hacia el ARNm y ser resistentes a las distintas nucleasas presentes en la célula. De esta manera, se consiguió una inhibición estérica de la traducción (figura 38). La mayoría de ellos consisten en modificaciones alquilo en la posición 2' (figura 40).

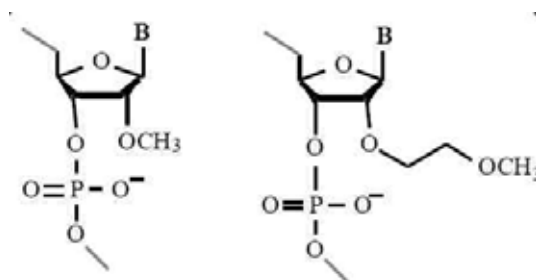


Figura 40. Los dos miembros más importantes de la 1ª generación de oligonucleótidos antisentido. A la derecha 2'-O-Metil y a la izquierda 2'-O-metoxietil (162).

Estos AXN presentan en general los azúcares en conformación Norte, lo que provoca una cierta preorganización en la conformación de la cadena simple que causa una estabilización en el dúplex AXN·ARN (159,163).

3.3.2.3 Oligonucleótidos antisentido de 2ª generación.

Tienen por objetivo mantener las propiedades de estabilidad de la primera generación y además conducir a una forma de híbrido susceptible de ser degradado por la RNasa H, aumentando así el grado de inhibición al degradar el ARN tóxico. Ejemplos de oligonucleótidos de 2ª generación son:

Los ácidos arabinonucleicos

Las modificaciones en el esqueleto azúcar-fosfato son claves para aumentar la resistencia hacia las nucleasas celulares, pero a menudo aniquilan también la susceptibilidad hacia la RNasa H. Sin embargo, los ácidos arabinonucleicos fueron las primeras manipulaciones

estructurales sobre el azúcar que mostraron buenas propiedades para la terapia antisentido (figura 41).

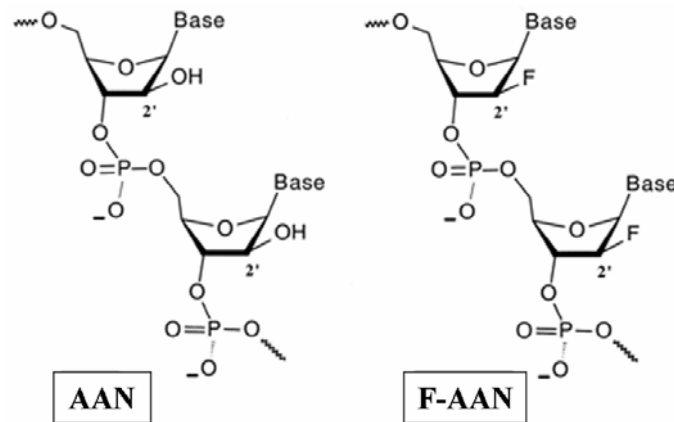


Figura 41. Estructura de la izquierda: ácido arabinonucleico (AAN), derecha: ácido 2'F-arabinonucleico (F-AAN)

Los ácidos arabinonucleicos no significan una mejora en la estabilidad de formación del AXN·ARN, probablemente por una disposición no ideal del grupo OH en el surco mayor, aunque conduce a una estructura que es reconocida como sustrato por la RNasa H (164). Por otro lado, los derivados 2'F-arabino (F-AAN) son mejores en la estabilidad del híbrido, probablemente porque al ser el F un átomo pequeño no interrumpen el *stacking* cuando se proyecta en el surco mayor (164).

Las C5-propinilo pirimidinas

La formación de un AXN mediante esta modificación en las bases pirimidínicas constituye una interesante posibilidad para la terapia antisentido por la excepcionalidad de no modificar el esqueleto azúcar-fosfato (figura 42). Estos grupos se proyectan en el surco mayor, lo que podría dar lugar a un incremento en el apilamiento. Los híbridos resultantes se mantienen como dianas de la RNasa H (165).

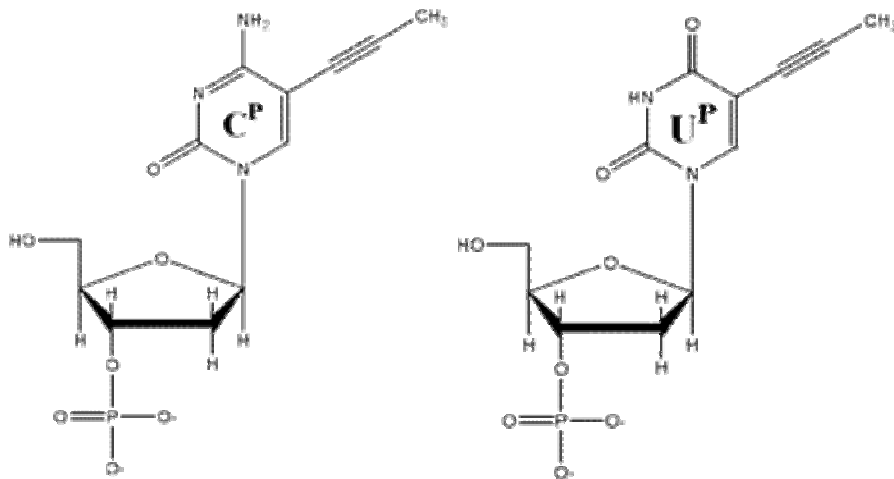


Figura 42. Estructura de la C5-propinil-desoxicitidina (C^P) y la C5-propinil-desoxiuridina (U^P) (166)

3.3.2.4 El ARN de interferencia

El descubrimiento reciente del mecanismo de ARNi ha revolucionado la investigación biológica (145) y aparece como una terapia potencial para tratar una amplia variedad de enfermedades (147). El ARNi es iniciado cuando fragmentos largos de ARN de doble cadena entran en la célula y son procesados por el complejo Dicer o RNasa III para dar los llamados ARNsi (*small interference*), que tienen 21-23 nucleótidos de largo. Estos ARNsi son incorporados al complejo RISC, que es capaz de separar las dos cadenas, unir una cadena con un ARNm complementario y finalmente degradarlo (167) (figura 43). Por ello, transfectar ARNsi o ARNi a la célula parece ser una manera sencilla de reducir la funcionalidad de un gen.

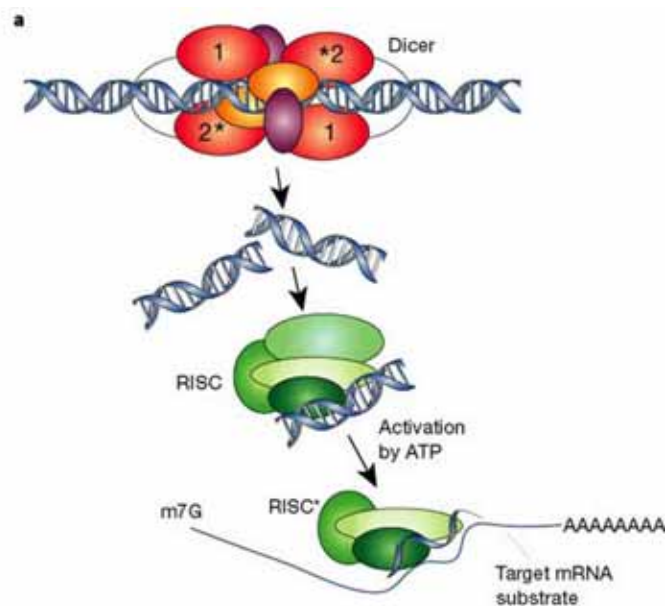


Figura 43. Esquema del complejo formado por Dicer y RISC (167)

3.3.2.5 Las ribozimas

Una ribozima es una molécula de ARN con capacidad catalítica. Al igual que las enzimas proteicas, poseen un centro activo que se une específicamente a un sustrato y que facilita su conversión en un producto. El descubrimiento de estas moléculas nos aporta luz sobre el origen de la vida biológica, ya que podemos pensar que fueron moléculas de ARN, con habilidad para autorreplicarse, las precursoras de la vida. Se ha descrito una gran variedad de ribozimas en células eucariotas, bacterias y virus. Concretamente, la ribozima de cabeza de martillo (*hammerhead rybozyme*) ha sido estudiada con gran detalle para aplicarse como agente antisentido (168) (figura 44), ya que multitud de ARNm contienen la secuencia consenso que sería reconocida por esta ribozima, aunque parece que el problema para la unión puede ser la estructura terciaria de los ARNm.

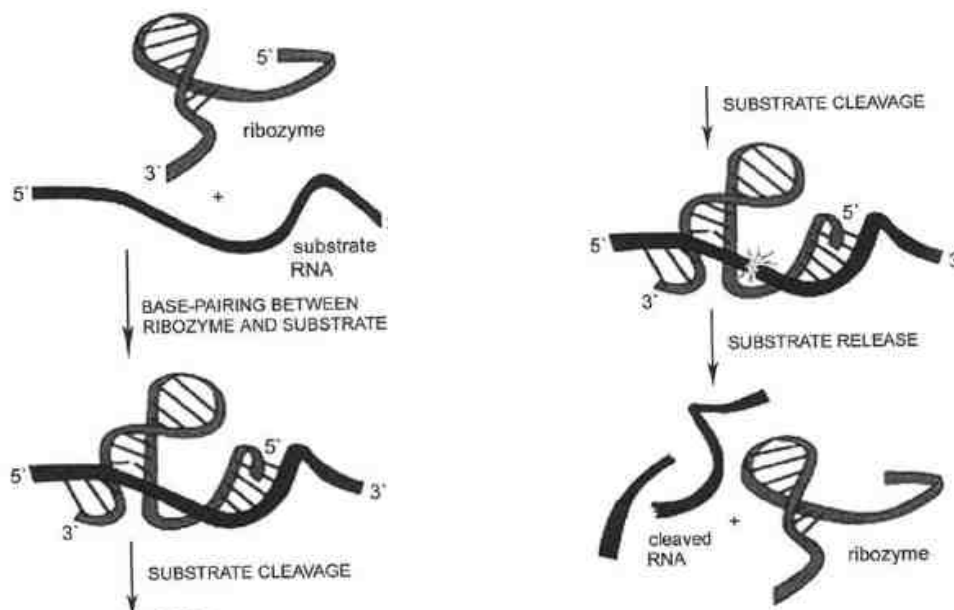


Figura 44. Catálisis de un ARNm gracias a una ribozima (162)

3.3.3 Curvatura y flexibilidad del ADN

Las primeras estructuras resueltas por rayos X indicaron que el ADN presentaba cierta curvatura en algunos pasos (116). Hoy sabemos que puede estar altamente curvado y

deformado en complejos con diversas proteínas que regulan su expresión (118,169) (figura 45 y 46). Se ha visto que la capacidad de deformación depende del tipo de nucleótidos (7,170,171), lo que sugiere la existencia de un código genético secundario donde la unión a la proteína es afectada por la secuencia de manera indirecta al modular las propiedades físicas del ADN.

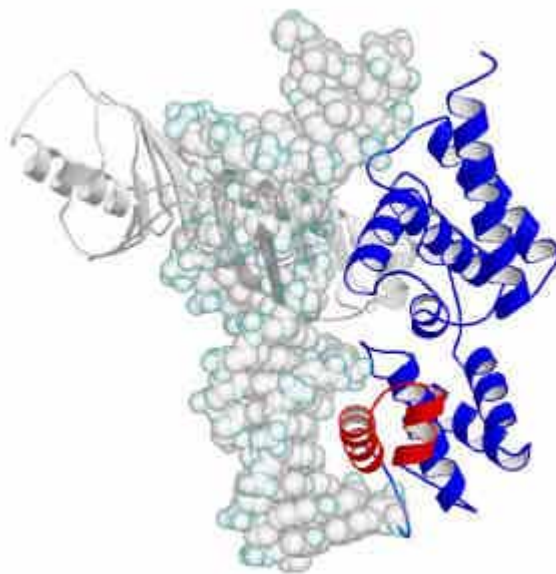


Figura 45. Ejemplo de complejo ADN·proteína donde el ADN está curvado. Se trata de la proteína de unión a la caja TATA (*TATA box binding protein*, TBP) (pdb 1YTB).

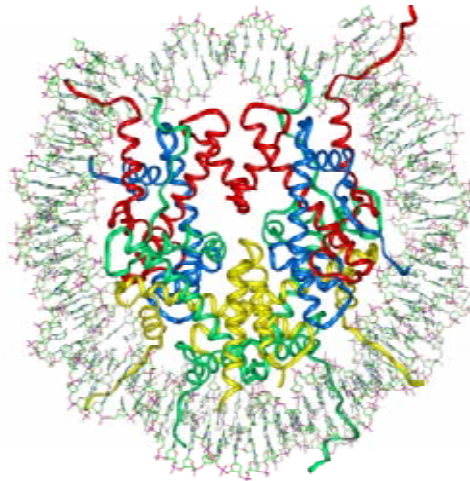


Figura 46. Estructura de un nucleosoma, unidad fundamental de empaquetamiento. (pdb 1AOI).

Además, la curvatura es clave para la compactación del ADN génico. Éste, al tener longitudes del orden de cm en la célula eucariota, tiene que adoptar niveles de estructuración

superior para ubicarse en el núcleo celular. El primer nivel de compactación se consigue con el nucleosoma, que es la estructura donde el ADN da 1 vuelta y 3/4 alrededor de un complejo de proteínas llamadas histonas (figura 46). A partir de estas unidades, se pueden establecer niveles superiores de empaquetamiento que dependerán del ciclo celular, consiguiendo el nivel máximo durante la división celular, donde se formarán los cromosomas (figura 47).

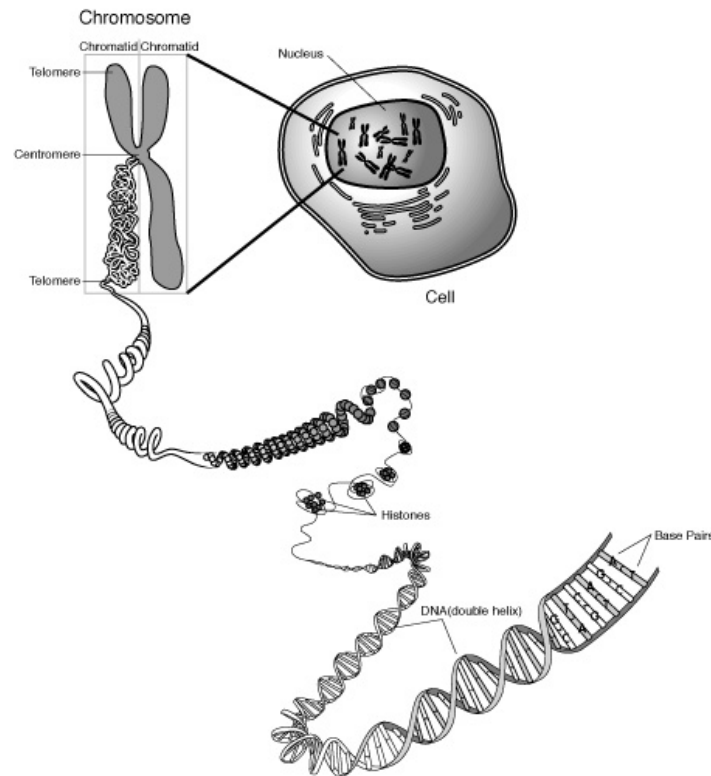


Figura 47. Niveles de empaquetamiento del ADN genético (www.accessexcellence.org).

3.3.3.1 Dependencia de estructura y flexibilidad con la secuencia

Gracias a un conjunto de estudios independientes (7,170,171), se ha visto que los pasos YpR (Y pirimidinas, R purinas, en el sentido 5'→3') acostumbra a ser los más flexibles, ya que oscilan entre dos conformaciones: i) *twist* alto, *slide* positivo y *roll* bajo parecida a la forma B, y ii) *twist* bajo, *slide* negativo y *roll* alto hacia el surco mayor parecida a la forma A. Por otro lado los pasos RpY acostumbra a ser los más rígidos, doblándose hacia el surco menor, mientras que los pasos RpR (YpY) son intermedios (figura 48).

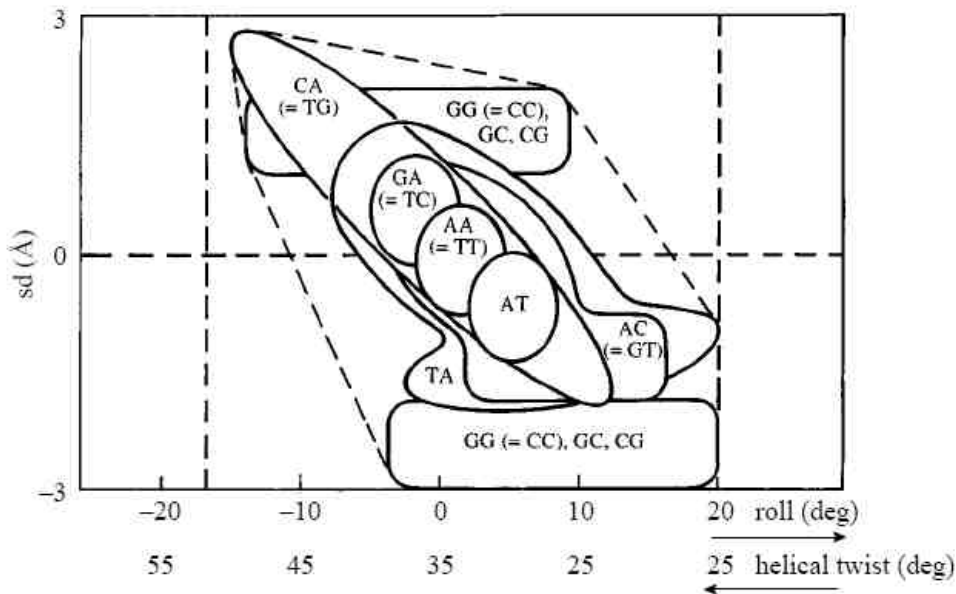
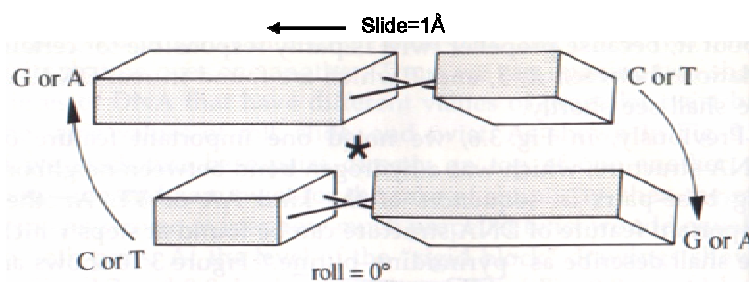


Figura 48. Mapa completo de los distintos pares de bases según una base de datos de estructuras experimentales. Nótese como el *roll*, *twist* y *slide* aparecen correlacionados (171)

Las interacciones de *stacking*

La primera explicación de las variaciones específicas de secuencia radicó en las interacciones de *stacking*, ya fuera para determinar la posición de equilibrio o para modular la flexibilidad (171,172). De esta manera, se hizo una teoría para explicar la biestabilidad de los pasos YpR, la cual considera que debido a la adopción de cierto *propeller twist* para aumentar el *stacking* aparece un choque estérico entre las purinas, que se soluciona con: i) la conformación tipo B donde el *stacking* intracatenario es mayoritario, o ii) la tipo A donde el *stacking* intercatenario adquiere un papel más importante (figura 49) (172). Además, se calculó la energía de *stacking* para todos los pares de base en función del *slide* y el *shift*, encontrándose concordancia entre los valores teóricos y los experimentales en los pasos YpR, RpY y RpR (173,174). Así, se comprobó que la repulsión electrostática debida al momento dipolar de las secuencias G·C (figura 50) ayudaba a crear su marcada biestabilidad (figura 48).



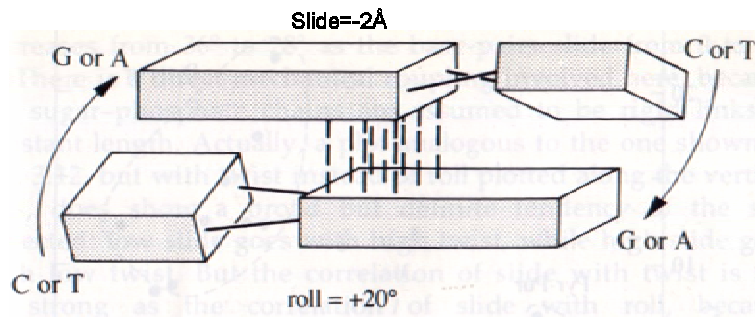


Figura 49. Arriba: *twist* alto, *slide* positivo y *roll* bajo parecido a la forma B. Abajo: *twist* bajo, *slide* negativo y *roll* alto hacia el surco mayor parecido a la forma A. Las bases adoptan el *propeller twist* para aumentar el *stacking*, lo que comporta la alternancia entre estas dos conformaciones (175)

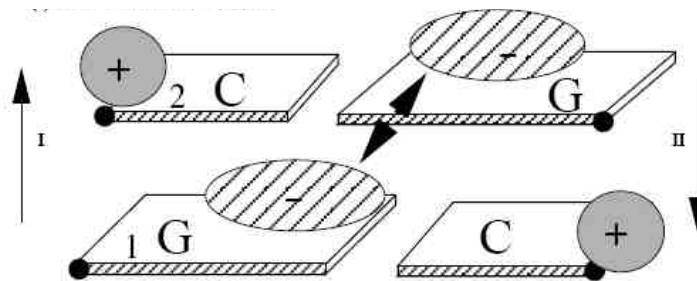


Figura 50. Esquema de la repulsión electrostática del paso GpC debido a su momento dipolar (171)

Las interacciones con cationes

Se ha propuesto que la repulsión electrostática del esqueleto azúcar-fosfato puede jugar un papel importante en modular las propiedades físicas del ADN. Así, la diferente capacidad de unir iones en los surcos de las distintas secuencias provocaría un distinto grado de apantallamiento entre las cargas negativas y, como consecuencia, distintas curvaturas (176). Concretamente, la unión de Na^+ en el paso AT en el surco menor provoca una menor anchura del mismo (4). En general, los fragmentos poliA unen cationes en el surco menor, mientras que los poliG los unen en el surco mayor (176), lo que puede explicar en parte la distinta geometría de estas dos secuencias.

3.3.3.2 La transición B↔A

Como ya se comentó en el apartado 3.2.3.1, la transición hacia la forma A es habitual cuando el ADN se une a ciertas proteínas, siendo un mecanismo ampliamente utilizado para

deformar y curvar el ADN (17,118,169) (figura 51). Además, la forma A es la segunda forma más común en el ADN, ya que se ha visto en condiciones de baja hidratación (alta concentración iónica o solventes apolares como el etanol) y especialmente en secuencias ricas en G (18,128-130). Por todo ello, la transición $B \leftrightarrow A$ ha sido objeto de estudio en esta tesis, donde hemos intentado caracterizarla desde un punto de vista molecular en diferentes solventes como son el agua y etanol/agua (85% en volumen), intentando obtener una visión global sobre la termodinámica y la cinética de la transición.

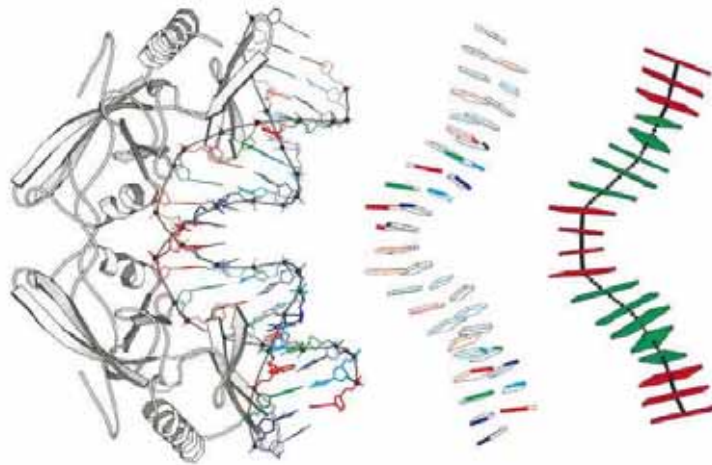


Figura 51. Imagen de la deformación $B \rightarrow A$ producida por la proteína endonucleasa I-PpoI (pdb 1A73). Derecha: en verde los fragmentos A y en rojo los B (17).

La transición en las bases nitrogenadas

En 1984, Drew y Calladine (177) conjeturaron que el cambio conformacional en las bases era lo que permitía hacer la transición y que, por tanto, el *stacking* era la fuerza conductora. Concretamente, era importante la biestabilidad de los pasos YpR y las particularidades de las secuencias C·G (véase apartado anterior). Así, podemos entender la transición como un proceso en el que a la forma B se le añade cierto *slide* y cierto *roll* para obtener la forma A (figura 52). Con esta teoría se conseguía explicar las preferencias de las secuencias G·C hacia la forma A, aunque no quedaba claro cómo las condiciones de baja hidratación provocaban la transición.

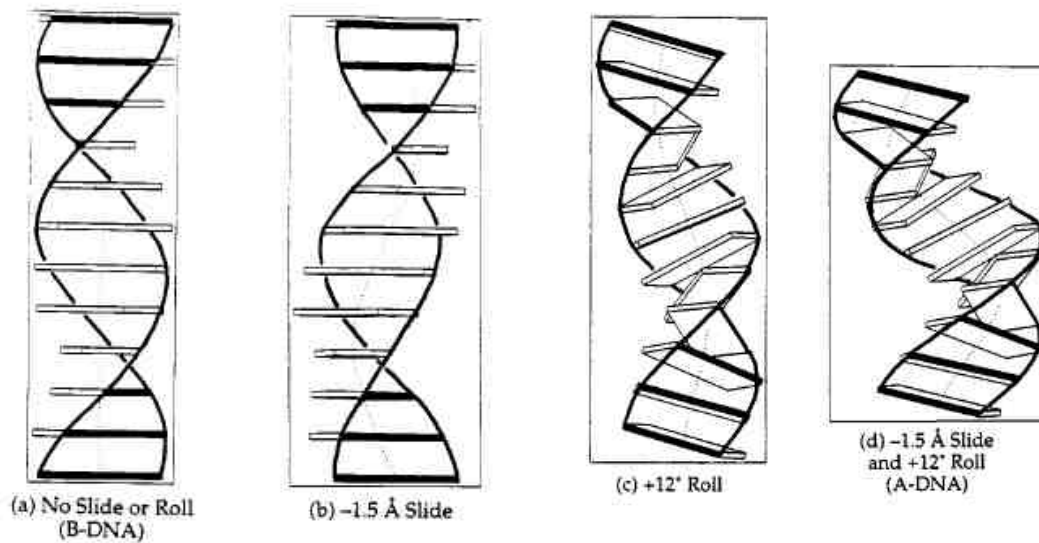


Figura 52. Esquema de la transición B→A según la conformación de las bases (178)

La transición en el esqueleto azúcar-fosfato

A medida que se han ido obteniendo las estructuras de ciertas modificaciones en el esqueleto azúcar-fosfato (como sería el caso de los AXN de 1ª generación; (159,163)), se ha visto que el *puckering* Norte es suficiente para presentar una forma A completa con unos valores correctos en los parámetros de las bases. Además, un estudio realizado en nuestro grupo en 1999 mostró que la restricción de los valores de *puckering* era suficiente para obtener una transición completa B↔A (179). Otro estudio teórico apuntó que el origen de la preferencia conformacional entre las formas A y B en agua o etanol/agua recaía en la repulsión entre fosfatos intracatenarios, la condensación de los cationes y la solvatación (180): una mayor distancia entre fosfatos favorece la forma B de manera independiente al solvente, mientras que las interacciones con cationes favorecen la forma A debido a su estructura más compacta. No obstante, la desolvatación de los iones en el caso de la forma A es más costosa que en la forma B. Por tanto, en condiciones de hidratación el balance energético es favorable a la forma B, mientras que las condiciones de baja hidratación reducen el término de desolvatación, siendo la forma A la más favorable.

3.3.4 Técnicas experimentales

Dado que los resultados teóricos siempre se tienen que contrastar con los datos experimentales, es importante tener unas pequeñas nociones de las principales técnicas que proporcionan información tridimensional de los ácidos nucleicos. En esta tesis se ha discutido en diversas ocasiones sobre la información dada por las técnicas experimentales y las posibles conclusiones que se pueden extraer. Por ello, se ha creído necesario comentar las principales técnicas experimentales, así como sus ventajas e inconvenientes (88,89,175).

En la actualidad sólo hay dos métodos capaces de determinar estructuras tridimensionales con alta resolución: la difracción de rayos X y la RMN. No obstante, existen un conjunto de técnicas de baja resolución, que dan una información muy valiosa sobre la conformación global de la molécula, como es el caso de las técnicas espectroscópicas o de la electroforesis.

3.3.4.1 Métodos de difracción de rayos X

La estructura de los ácidos nucleicos ha sido dilucidada principalmente por las técnicas de difracción de rayos X. Estos métodos proporcionan información a diferentes niveles dependiendo del peso molecular y el nivel de organización de la materia, siendo capaces de suministrar una buena resolución cuando se consigue cristalizar el material de estudio. Sin embargo, la difracción con materiales cuasi-cristalinos puede dar también información relevante, aunque de menor resolución.

Difracción de fibras

La única posibilidad de investigar la estructura de los ácidos nucleicos de gran longitud mediante difracción de rayos X es preparando fibras, que se forman a partir de soluciones altamente concentradas donde se hacen precipitar a base de añadir alcohol. Así, los largos polímeros se unen de manera paralela formando haces cuasi-cristalinos en los que sólo hay desorden rotacional a lo largo del eje molecular. En consecuencia, no se puede obtener información a nivel atómico, pero sí sobre parámetros helicoidales promedio.

La técnica de la difracción de fibras fue usada a principios de los años 50 por Franklin y Gosling sobre el ADN obteniendo dos patrones distintos de difracción, que fueron llamados A y B (181) (figura 53). El primero se obtenía en condiciones de poca humedad, y el segundo manteniendo las fibras húmedas. El patrón B, que era el más simple de los dos, fue

interpretado por Watson y Crick en su modelo de doble hélice del ADN, que recibió el nombre de B-ADN (5), mientras que más tarde se adquirió la forma A del otro patrón. En los años 80, Arnott *et al.* consiguieron patrones de difracción en diferentes condiciones que correspondían a los distintos tipos de la familia B comentados en el apartado 3.2.3.1 (182).

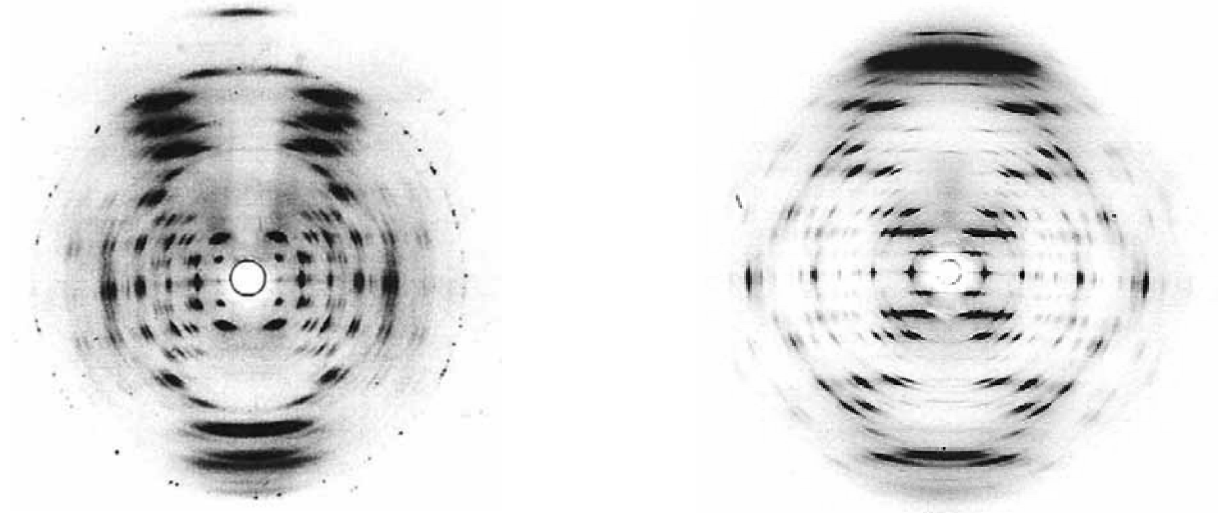


Figura 53. Patrón de difracción de la forma A (izquierda) y la forma B (derecha) del ADN (183)

Difracción de cristales

En la estructura cristalina, las moléculas están dispuestas regularmente en una matriz tridimensional. El motivo básico es la unidad asimétrica, la cual mediante operaciones de simetría forma la celda unidad. La traslación de la celda unidad en las tres direcciones del espacio es lo que forma el entramado cristalino (*crystal lattice*).

La difracción de cristales sobre fragmentos de 10-12 pares de bases puede suministrar una resolución de 2-3 Å, revelando así la estructura tridimensional más completa y precisa. No obstante, la obtención de cristales es delicada y de resultado impredecible, donde se imponen condiciones distintas a las fisiológicas como altas fuerzas iónicas o alcoholes deshidratantes. Otro inconveniente es que sólo se pueden resolver moléculas que tengan una sola conformación en equilibrio (a veces en las proteínas hay fragmentos flexibles para los cuales es imposible obtener una estructura). Finalmente en el caso de los ácidos nucleicos, los fragmentos de distintas celdas suelen interactuar entre ellos formando un polímero infinito, lo que introduce importantes distorsiones dependientes del entramado cristalino (184).

3.3.4.2 La resonancia magnética nuclear (RMN)

Constituye el otro método capaz de proveer información de la estructura tridimensional de los ácidos nucleicos con alta resolución. Consiste en aprovechar el momento magnético que presentan los núcleos de algunos átomos (^1H , ^{31}P , ^{13}C y ^{15}N) sometiendo la solución a un campo electromagnético adecuado. Posteriormente, la muestra es irradiada a ciertas frecuencias para excitar estos núcleos.

El desplazamiento químico

Los diferentes átomos absorberán radiación de distinta frecuencia debido al entorno químico particular de cada uno, de manera que se podrán identificar. Además, se obtendrá información de interacciones de puente de hidrógeno y *stacking*.

El acoplamiento J o *spin-spin*

La excitación debida al campo magnético o *spin* de un átomo está influenciada por las propiedades magnéticas de sus átomos cercanos en el espacio. Así, el acoplamiento entre *spins* da información de la disposición espacial de los átomos, en concreto de las torsiones. Un ejemplo es el *puckering* del azúcar, donde acoplamientos J entre H1' y H2' se asocian a la conformación Sur, mientras que acoplamientos J entre H3' y H4' se asocian a la Norte (185).

El tiempo de relajación

Aquí se indica en cuanto tiempo un determinado núcleo vuelve a presentar sus valores de equilibrio después de haber sido excitado. Estas medidas tienen gran importancia ya que dan información sobre la distancia que hay entre ciertos átomos (NOEs), de manera que son básicas para reconstruir la estructura tridimensional.

La principal ventaja de la RMN es que permite obtener la conformación de los ácidos nucleicos en condiciones más cercanas a las fisiológicas, o si es de interés modificándolas (por ejemplo incrementar la fuerza iónica o subir la temperatura). Además, puede dar información dinámica y así determinar la conversión o la población relativa entre conformeros (la constante de equilibrio del *puckering* $\text{S} \leftrightarrow \text{N}$ se puede calcular a partir de los acoplamientos J). Sin embargo, su principal limitación es el límite en el tamaño de las

moléculas (en la actualidad de unos 1000 átomos) debido a la complejidad de los espectros, pues su resolución suele necesitar una simulación de DM restringida por los NOEs obtenidos.

3.3.4.3 Métodos de baja resolución

Hablaremos de algunos ejemplos que han dado alguna información relevante a lo largo de esta tesis, como son la absorción ultravioleta (UV), el dicroísmo circular (CD) y la electroforesis. Estas técnicas permiten describir la estructura y las transiciones globales de una molécula, aunque no proporcionan información estructural detallada.

Espectroscopía de absorción en el ultravioleta

Los ácidos nucleicos absorben este tipo de radiación gracias a los anillos aromáticos de las bases. No obstante, el grado de absorción varía en función del grado de formación de la doble hélice. Un experimento típico es ir subiendo la temperatura para determinar el momento en que hay el 50% de cadenas unidas y el 50% de cadenas separadas, lo que se conoce como temperatura de fusión (*melting*) y se suele usar como una medida de estabilidad (figura 54).

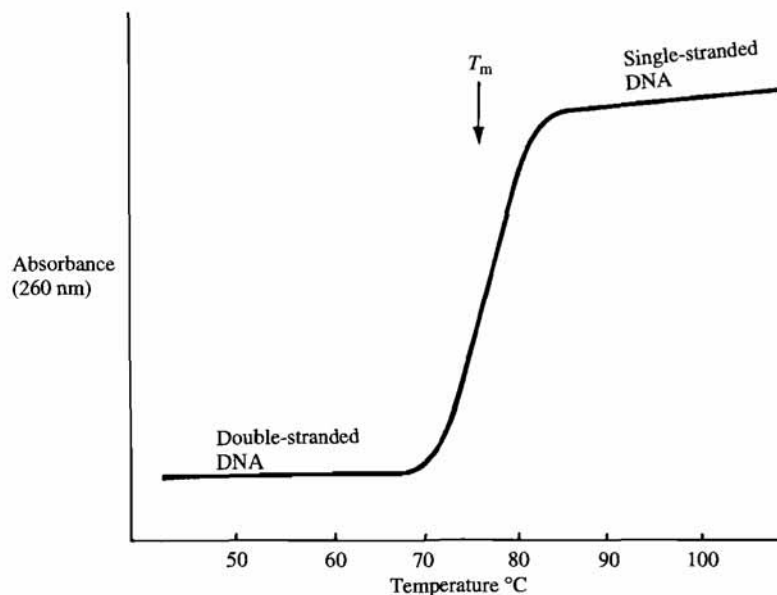


Figura 54. Curva de separación para el ADN. La temperatura de fusión (T_m) es cuando el 50% de las cadenas están unidas y el otro 50% separadas (186)

El dicroísmo circular

Consiste en irradiar la muestra con luz polarizada circular dextrógira y levógira para medir la absorción diferencial, que sólo será diferente de cero para las moléculas quirales. Se acostumbra a utilizar para detectar la estructura secundaria de las biomoléculas, como es el caso de los ácidos nucleicos con su estructura asimétrica helicoidal. Las diferentes formas canónicas presentan un espectro de absorción ligeramente distinto (figura 55)

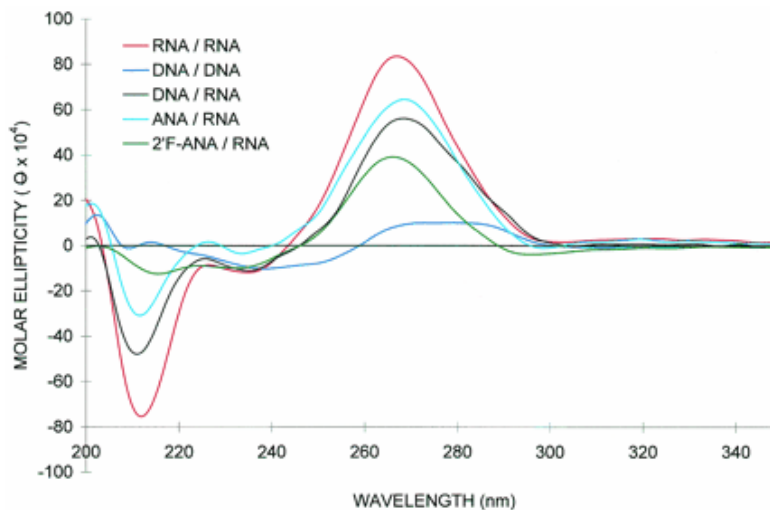


Figura 55. Dicroísmo circular de los diferentes dúplexes. ANA y 2'F-ANA se refieren a ácidos arabinonucleicos y 2'F-arabinonucleicos respectivamente (187)

Electroforesis

Es una técnica para la separación de moléculas según su movilidad en un campo eléctrico a través de una matriz porosa, la cual finalmente las separa por tamaño molecular, carga eléctrica o forma. Al poner la mezcla de moléculas y aplicar un campo eléctrico, éstas se moverán y deberán ir pasando por la malla, por la que las pequeñas se moverán mejor y avanzarán más, mientras que las más grandes quedarán cerca del lugar de partida. De la misma manera, las moléculas lineales avanzarán mejor que las globulares. Gracias a este principio, se observó por primera vez el fenómeno del ADN curvado (188) y también se hicieron experimentos clásicos de medir la flexibilidad (11,12).

4. RESULTADOS

4.1 Relative Flexibility of DNA and RNA: a Molecular Dynamics Study

Agnes Noy, Alberto Pérez, Filip Lankas, F. Javier Luque, Modesto Orozco

Journal of Molecular Biology (2004) **343**, 627-38

Relative Flexibility of DNA and RNA: a Molecular Dynamics Study

Agnes Noy¹, Alberto Pérez¹, Filip Lankas², F. Javier Luque³ and Modesto Orozco^{1,4*}

¹Molecular Modeling and Bioinformatics Unit, Institut de Recerca Biomèdica, Parc Científic de Barcelona, Josep Samitier 1-5, Barcelona 08028 Spain

²Bernoulli Institute for Mathematics, EPFL (Swiss Federal Polytechnical Institute) 1015 Lausanne, Switzerland

³Departament de Fisicoquímica Facultat de Farmàcia Universitat de Barcelona Avda Diagonal 643, Barcelona 08028, Spain

⁴Departament de Bioquímica i Biologia Molecular, Facultat de Química, Universitat de Barcelona, Martí i Franquès 1 Barcelona 08028, Spain

*Corresponding author

State of the art molecular dynamics simulations are used to study the structure, dynamics, molecular interaction properties and flexibility of DNA and RNA duplexes in aqueous solution. Special attention is paid to the deformability of both types of structures, revisiting concepts on the relative flexibility of DNA and RNA duplexes. Our simulations strongly suggest that the concepts of flexibility, rigidity and deformability are much more complex than usually believed, and that it is not always true that DNA is more flexible than RNA.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: DNA; RNA; flexibility; molecular dynamics; similarity indexes

Introduction

Under physiological conditions DNA exists as an elongated helix known as *B*-form, while if possible RNA forms a double helix more compact than that of DNA,^{1–3} which is named the *A*-form. The different puckering of sugars in *B* (South to South-East) and *A* (North) forms alters the structure of the grooves,^{1–4} which changes completely the ability of the nucleic acids to interact with other molecules, particularly with proteins.^{1–3} It seems the different helical structure of DNA and RNA might determine their different role in the cell. However, structure itself is not able to explain the incredible ability of proteins to distinguish between nucleic acid structures. For example, only a fraction of very similar structural distortions of DNA are recognized by the

UvABC excision repair mechanism.^{1–3,5} Indeed, it is known that RNase H recognizes and degrades DNA–RNA hybrids, but it is inactive against RNA–RNA duplexes and other RNA-hybrid molecules, despite the fact that all these duplexes pertain to the *A*-family.^{6,7} More surprisingly, the complex mechanism involving gene silencing by small interference RNAs is activated by duplex RNA, but not by other RNA hybrid duplexes of similar structure.^{8,9} Clearly, besides the general helical structure, other properties must be exploited by nature to distinguish between helical structures.

RNA adopts very unique conformations in the cell, like those of ribozymes, transfer or ribosomal RNAs. However, it is believed that this is mostly due to the fact that cellular RNA is single-stranded and intramolecular RNA duplexes contain unpaired bases and mismatches which favour strong twists and kinks in the helix. However, it is generally accepted in the scientific community that the DNA double helix is more flexible than the RNA one. This assumption is supported by indirect low resolution experimental data^{10,11} like that derived

Abbreviations used: MD, molecular dynamics; RMSD, root mean square deviations.

E-mail address of the corresponding author: modesto@mmb.pcb.ub.es

from electron micrography, gel electrophoresis, hydrodynamic measurements or fiber diffraction.¹² ³¹P NMR experiments of highly oriented fibers also suggest that *B*-DNA is more flexible than *A*-RNA.^{13,14} Protein NMR experiments also suggest that the RNA duplex is more rigid than the DNA one for equivalent sequences,^{13,14} a finding further supported from the comparison of experimental NMR *J*-coupling constants with the values derived from molecular dynamics (MD) simulations.¹⁵ Caution is, however, necessary, since we cannot ignore that NMR data provide information about local fluctuations in DNA microenvironments, it being less suitable to ascertain global structural changes of the duplex.

Crystal data deposited in the Protein Data Bank have been traditionally used to support the greater flexibility of *B*-DNA compared to *A*-RNA. Thus, depending on sequence, solvent composition or the presence of certain ions different *B*-like conformations (*C*-, *D*- or *T*-forms) can exist,¹⁻³ an effect that has not been detected for *A*-RNA. The larger occurrence of lattice distortions in DNA crystals^{16,17} and the ability of a given DNA structure to crystallize in different space groups¹⁵ have also been used to support the higher flexibility of *B*-DNA compared to *A*-RNA. However, we cannot ignore that similar lattice-dependence effects are detected for *A* forms.^{18,19} Furthermore, a recent extensive comparison of crystal structures of *B*-DNA and *A*-RNA²⁰ have raised doubts on the hypothesis that DNA is always more flexible than RNA. Thus, not only the number of space groups found in RNA structures is very similar to that found in *B*-DNA, but no relevant differences are found in both thermal factors and resolution between *B*-DNA and *A*-RNA crystals. In summary, the analysis of the PDB structures does not support the view of *B*-DNA as an "intrinsically" more flexible entity than *A*-RNA.

Different theoretical calculations have explored the structure and flexibility of DNA and RNA duplexes. The landmark work by Cheatham and Kollman¹⁵ provided an exhaustive MD comparison of DNA and RNA duplexes in aqueous solution. The analysis of the torsions around α , χ and γ dihedral angles and of sugar puckerings during 2 ns trajectories led the authors to conclude that *A*-RNA was more rigid than *B*-DNA.¹⁵ The authors demonstrated the complex nature of the interactions rigidifying the RNA *versus* the DNA, suggesting that perhaps high ordered water molecules around the RNA reduce the flexibility of RNA. Similar calculations were later performed by Auffinger and Westhof,^{21,22} who compared $d(\text{CG})_6 \cdot d(\text{CG})_6$ and $d(\text{TA})_6 \cdot d(\text{TA})_6$ DNAs with the corresponding RNA duplexes, $r(\text{CG})_6 \cdot r(\text{CG})_6$ and $r(\text{UA})_6 \cdot r(\text{UA})_6$. From the analysis of dihedral changes along the sequence and their time fluctuations, DNA was found to be more flexible than RNA. Only one MD simulation, performed recently by MacKerell's group has raised doubts on the

dominant hypothesis that DNA is more flexible than RNA.²³

In summary, despite the lack of definitive evidence, there is a general consensus that DNA duplex is more flexible than the RNA one under the same conditions. However, what does flexibility mean? Conceptually, it denotes the ability of a given structure to be deformed as a result of an external perturbation. For highly anisotropic macromolecules such as nucleic acids, such a definition is perhaps too ambiguous, as it would encompass a wide variety of structural deformations. For example, a molecule can be locally flexible but globally rigid when flexibility motions in different regions cancel. Here, we used state of art MD simulations to reinvestigate the relative flexibility/rigidity of *B*-DNA and *A*-RNA using a rigorous definition of these two concepts. Our purpose is then not only to obtain a qualitative picture of how "rigid" are DNA and RNA, but to describe for the first time the very complex scenario of the structural flexibility of nucleic acids.

Results and Discussion

Global structural descriptors

As it has been largely discussed,^{24,25} current force-fields provide stable trajectories for both DNA and RNA duplexes. This trend is noted in the small root mean square deviations (RMSD) determined with regard to the MD-averaged structures (see Table 1), indicating that despite local oscillations (in the sub-nanosecond time scale) the two structures sample well defined regions of the conformational space centred on their respective averaged structures. In turn, this supports the validity of simple pseudoharmonic approaches in the analysis of fluctuations detected in MD simulations (see below).

The RMSD with respect to the corresponding canonical fiber²⁶ and crystal structures (PDB entries 1BNA and 157D for DNA and RNA duplexes, respectively) are also reasonably small (see Table 1), which indicates that the sampled structures are close to the experimental ones. Interestingly, the RMSD between RNA and the *A*-form fiber structure is clearly smaller (see Table 1) than that obtained between DNA and the *B*-form fiber one. This finding, which demonstrates the suitability of Parm99 to reproduce RNA duplexes, suggests that the structure of RNA in the fiber and solution does not change remarkably, while a larger change is expected for DNA. However, care must be taken to conclude that DNA is more flexible than RNA, because we must stress that the standard deviation of the RMSD between the sampled conformations in DNA and RNA trajectories and the corresponding MD-averaged structures are similar (see Table 1), a result that does not support the assumption that RNA is more rigid than DNA.

The distance ($C1'-C1'$) between contiguous

Table 1. Average root mean square deviation (RMSD; in Å) between DNA/RNA trajectories and reference structures

| | B-form ^a | A-form ^a | Crystal ^b | MD-averaged ^c |
|-----|---------------------|---------------------|----------------------|--------------------------|
| DNA | 2.22(0.34) | 4.43(0.44) | 2.05(0.39) | 1.16(0.29) |
| RNA | 5.16(0.55) | 1.63(0.36) | 3.54(0.22) | 1.23(0.33) |

Standard deviations (in Å) are given in parentheses.

^a Canonical structures from Arnott's data.²⁶

^b Crystal structures of DNA (PDB 1BNA) and RNA (157D). Two mismatches of 157D were not considered to compute the RMSD.

^c Structures obtained by averaging the snapshots collected during the last ns of the corresponding MD simulations.

nucleotides is a simple, yet accurate general descriptor of the helical structure.²⁷ For each collected snapshot, the sequence-averaged C1'-C1' distance and its associated standard deviation were computed. This calculation was then repeated along the entire trajectory to obtain time-averaged mean C1'-C1' distances, and the associated standard deviations. As suggested by Hunter and co-workers²⁷ C1'-C1' discriminates very well between DNA and RNA (see Figure 1(top)), since RNA average distances are always larger than those of DNA, and no overlap exists between DNA and RNA C1'-C1' distance distributions. The oscillations (see Figure 1(top)) of the mean C1'-C1' distance are slightly larger for RNA (0.07 Å SD) than for DNA (0.05 Å SD). This suggests that RNA is slightly more flexible than DNA. However, the oscillations in the standard deviations associated with the sequence-averaged C1'-C1' distance in each snapshot are larger in DNA (0.36 Å SD) than in RNA (0.27 Å SD), suggesting the opposite, i.e. that DNA is more flexible than RNA (see Figure 1(bottom)). Thus, the analysis of the C1'-C1' distances and their fluctuations do not provide conclusive evidence about the relative flexibility

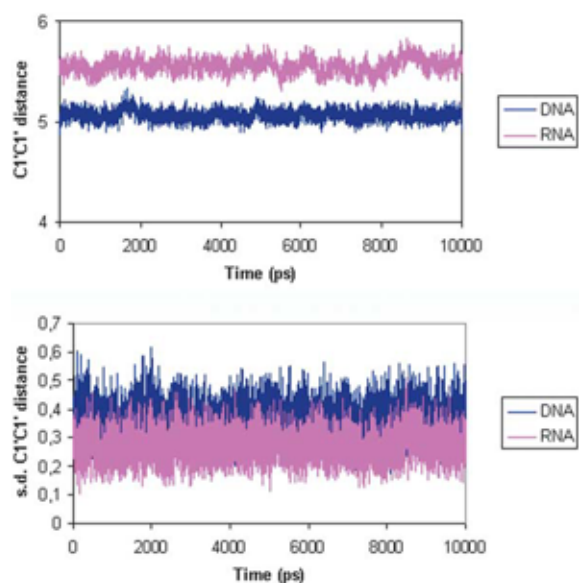


Figure 1. Top: Variation of the sequence averaged C1'-C1' distance (in Å) along DNA and RNA trajectories. Bottom: Variation of the standard deviations (in Å) associated to the sequence-averaged C1'-C1' distance (in Å) along DNA and RNA trajectories.

between DNA and RNA. The same lack of conclusions appears when the distribution of the groove widths for DNA and RNA is examined (see Figure 2), since the minor groove fluctuates more in DNA than in RNA; but the reverse trend is found concerning the fluctuations in the major groove.

In summary, the analysis of the global structural descriptors presented in this section does not convincingly support the supposed higher flexibility of DNA compared to RNA. More precise analysis needs to be made to analyze the nature of the differences in flexibility of the nucleic acids.

Entropy calculations

The intramolecular entropy can be used as a direct estimate of the global flexibility of DNA and RNA. As noted above, MD-derived entropy estimates depend on the length of the trajectory. However, the values obtained for 10 ns trajectories and those extrapolated to infinite time²⁸ are very close (see Table 2), which suggests a good convergence in the calculations, as expressed in the small

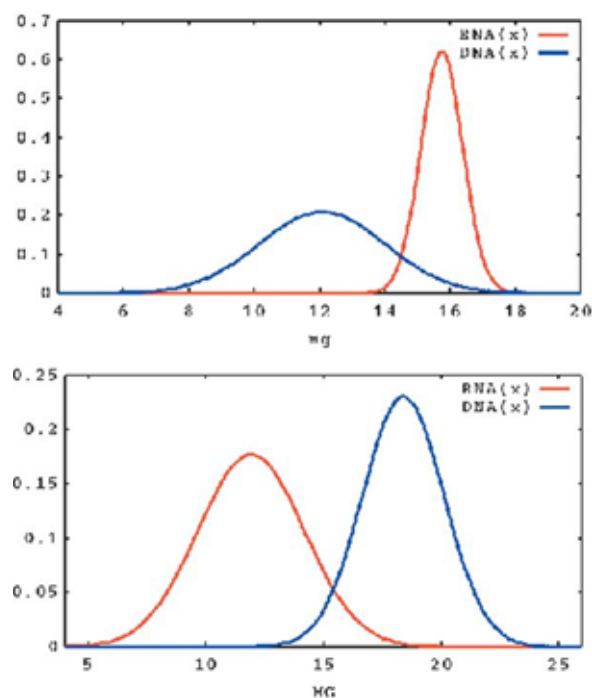


Figure 2. Distributions (fitted to normalized Gaussians) of the width (in Å) of the minor (top) and major (bottom) grooves of DNA and RNA in the MD simulations.

Table 2. Intramolecular entropies (in kcal/mol K) computed using Schlitter's (roman font) and Andreocci-Karplus's methods (italics) for DNA and RNA

| | $S(t=9 \text{ ns})$ | $S(t=\infty)$ | $S(3)$ | $S(10)$ |
|-------------------|-------------------------|---------------------------------|-------------------------|-------------------------|
| DNA (all atoms) | 2.0067 <i>1.8321</i> | 2.14(0.05) <i>1.93(0.05)</i> | 0.0328 <i>0.0327</i> | 0.0988 <i>0.0983</i> |
| RNA (all atoms) | 1.8228 <i>1.6516</i> | 1.90(0.03) <i>1.71(0.03)</i> | 0.0342 <i>0.0340</i> | 0.0976 <i>0.0971</i> |
| DNA (nucleobases) | 0.8946 <i>0.8205</i> | 0.91(0.01) <i>0.84(0.01)</i> | 0.0286 <i>0.0285</i> | 0.0849 <i>0.0845</i> |
| RNA (nucleobases) | 0.8893 <i>0.8145</i> | 0.91(0.02) <i>0.83(0.01)</i> | 0.0300 <i>0.0299</i> | 0.0860 <i>0.0858</i> |
| DNA (backbone) | 1.3424 <i>1.2395</i> | 1.43(0.07) <i>1.34(0.10)</i> | 0.0320 <i>0.0319</i> | 0.0962 <i>0.0956</i> |
| RNA (backbone) | 1.1429 <i>1.0950</i> | 1.18(0.03) <i>1.09(0.03)</i> | 0.0333 <i>0.0331</i> | 0.0944 <i>0.0940</i> |

Total intramolecular entropies were determined considering all equivalent atoms. Partial entropies were computed considering only nucleobases or backbone atoms. In all the cases, values were determined considering all the principal components, as well as only the first three and tenth ones. For the total entropy, not only the value directly obtained from the 10 ns trajectory, but also that extrapolated at infinite simulation time is displayed (see Methods for details).

standard error associated to S_∞ . This finding, in conjunction with the excellent agreement between the two methods used for the calculation of entropy, gives strong confidence on the quality of the configurational entropies determined for DNA and RNA.

The intramolecular entropy of DNA is 11–13% larger (around 0.24 kcal/mol K; see Table 2) than those of RNA. Therefore, DNA is globally more disordered and flexible than RNA, which supports the generally accepted picture of the relative DNA/RNA flexibility. However, a partitioning analysis of the entropy contributions shows that the concept of flexibility is more complex than expected. Thus, when the entropy is computed considering only the

first ten principal components, the entropy of DNA is only around 1% larger than that of RNA, and when only the first three principal components are considered (see Table 2) the RNA appears 4% more disordered than DNA. This finding suggests that disorder in RNA stems from a relatively small number of “very soft” motions, while for DNA is the result of many “soft” motions.

Diagonalization of mass-weighted covariance matrices built up by considering only certain groups of atoms allows us to examine their contribution to the global disorder in DNA and RNA. For our purposes here, we have analyzed separately the contributions of atoms in the backbone and in the nucleobases. The results in Table 2 clearly demonstrate that the larger intramolecular entropy of DNA is due to the greater disorder in the backbone. This finding is further supported from inspection of the distributions of selected backbone angles sampled along the trajectory (see Figure 3). However, the fluctuations in the backbone do not affect the nucleobases, whose entropy in DNA and RNA is not very different. Finally, it is worth noting that the addition of backbone and nucleobase entropies approaches the total entropy, confirming the relatively small coupling between backbone and nucleobases movements.

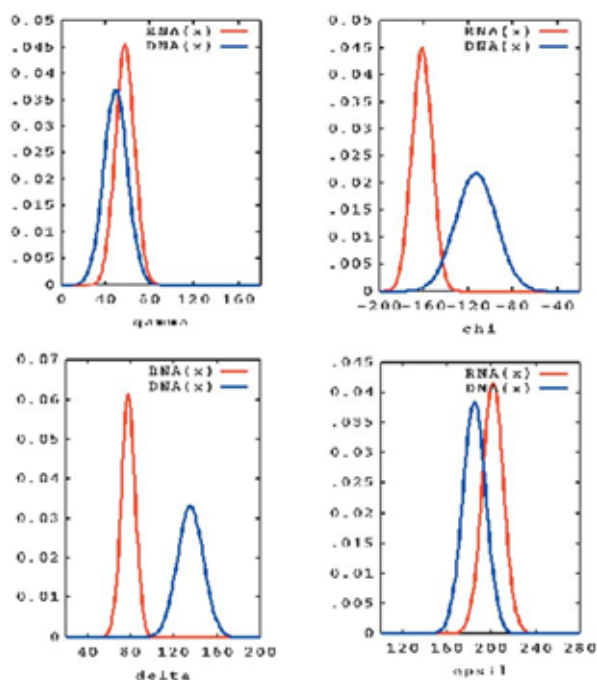


Figure 3. Distributions (fitted to normalized Gaussians) of selected backbone angles ($\gamma, \chi, \delta, \epsilon$) in DNA and RNA trajectories.

Essential dynamics analysis

Similarity analysis (see equations (1) and (2)) shows that, in general, the easiest essential motions in DNA and RNA are similar, they being associated with twisting and bending of the polymers. When similarity is computed using 500 essential movements, absolute all-atoms similarity indexes $\gamma_{\text{DNA/DNA}}$ and $\gamma_{\text{RNA/RNA}}$ are around 0.87, which are close to similarity indexes obtained considering only the backbones (see Table 3). This demonstrates the larger complexity of the essential movements of backbones compared to those of the nucleobases. Furthermore, the all-atom cross similarity index $\gamma_{\text{RNA/DNA}}$ is around 0.74 (yielding a relative

Table 3. Absolute (γ) and relative (κ) similarity indexes for DNA and RNA essential movements

| | All atoms | Nucleobases | Backbone |
|---------------------------|-----------|-------------|----------|
| <i>Ten eigenvectors</i> | | | |
| $\gamma_{\text{RNA/DNA}}$ | 0.564 | 0.616 | 0.566 |
| $\gamma_{\text{RNA/RNA}}$ | 0.860 | 0.926 | 0.865 |
| $\gamma_{\text{DNA/DNA}}$ | 0.794 | 0.829 | 0.770 |
| $\kappa_{\text{RNA/DNA}}$ | 0.682 | 0.701 | 0.692 |
| <i>500 eigenvectors</i> | | | |
| $\gamma_{\text{RNA/DNA}}$ | 0.744 | 0.938 | 0.799 |
| $\gamma_{\text{RNA/RNA}}$ | 0.873 | 0.967 | 0.919 |
| $\gamma_{\text{DNA/DNA}}$ | 0.878 | 0.967 | 0.916 |
| $\kappa_{\text{RNA/DNA}}$ | 0.850 | 0.970 | 0.870 |

Calculation is repeated considering the first ten and 500 essential movements.

$\kappa_{\text{RNA/DNA}}$ of 0.85; see Table 3), reflecting the different nature of the essential movements in DNA and RNA backbones ($\gamma_{\text{RNA/DNA}}=0.799$ and $\kappa_{\text{RNA/DNA}}=0.87$ in the backbone; see Table 3).

Absolute self-similarity indexes obtained considering only the first ten eigenvectors are much higher for RNA than for DNA (see Table 3). This finding demonstrates that more essential movements are necessary to explain the flexibility of DNA than that of RNA, which reflects the higher complexity of DNA flexibility compared to that of RNA.

In summary, despite a general similarity differences exist in the dynamics of DNA and RNA. These differences are mostly related to the higher complexity of movements in DNA, and to the nature of the essential movements of DNA and RNA backbones.

Stiffness analysis

The stiffness analysis provides useful information on the force necessary to deform a molecule along a given geometrical variable. Thus, the analysis performed using the eigenvectors obtained by diagonalization of the covariance matrix as deformation variables can be used to measure the deformability of DNA/RNA along the axis defined by their essential motions. Elastic force constants associated with the first five essential movements of

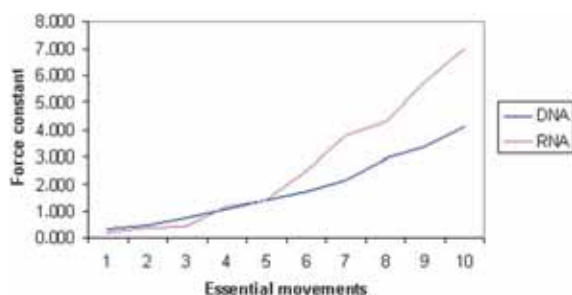


Figure 4. Force constants (in cal/mol Å²) associated to different essential movements of DNA and RNA. Essential movements are labelled in decreasing order of impact in the explanation of trajectory variance.

RNA are similar or even weaker than those of DNA, but the situation dramatically changes when further essential movements are considered (see Figure 4), and the RNA becomes stiffer than DNA. Therefore, RNA is very deformable along a small set of essential motions, but very rigid along all the others, whereas DNA has a more degenerated pattern of deformability (see Figure 4). Note that this agrees with the picture of DNA as a molecule with a much more complex flexibility pattern than RNA.

The stiffness analysis performed using essential motions as deformation variables provides information on the natural deformability of nucleic acids. However, it is difficult to interpret in term of canonical motions (see Methods). Thus, we extended the stiffness analysis using helical parameters as deformation variables. First, we used a reduced set of global parameters (tilt, roll, twist and stretch) for the 10mer central portion of DNA and RNA duplexes. The global parameters are defined essentially,²⁹ and also we describe the geometry of the fragment using the X3DNA analyzer to maintain consistency with other parts of this study. The global stretch is the sum of distances between neighbouring base-pairs along the fragment, global twist is the sum of base-pair step local twists. We define a middle frame by taking its z -axis as the average of the normal vectors of the first and last base-pair in the fragment, and its x -axis as the projection of the x -axis of the central base-pair onto the plane perpendicular to the z -axis. The y -axis complements the set to form a right-handed triad. The global roll and tilt are then defined exactly as roll and tilt in the X3DNA algorithm, using the middle frame just described. Since the x -axis of a base-pair (as defined by X3DNA) is close to the dyadic vector pointing to the major groove, our global roll measures the bending of the fragment towards the major groove in its center, and global tilt describes the bending in the perpendicular direction. If the fragment has even number of pairs, we first average the x -axes of the two central base-pairs.

The stiffness at the global level can be described by the stretch modulus and by twisting and bending persistence lengths. Since we measure bending in two perpendicular directions, we have two bending persistence lengths corresponding to the global roll (bending towards the grooves) and global tilt. In this way, we can capture a possible anisotropy in the elastic behaviour. For longer fragments (several helical turns) the anisotropy is in general supposed to disappear and the bending stiffness is then fully described by one, "isotropic" bending persistence length, which can be estimated as the harmonic average of the two anisotropic ones.²⁹

Table 4 indicates that deformation of the global twist of DNA is easier than that of RNA, but the stiffness of DNA and RNA with respect to roll is similar, and quite surprisingly DNA is stiffer than RNA in terms of global tilt and stretch. Caution is

Table 4. Elastic force constants for deformations along a reduced set of global helical parameters (angular force constants in kcal/mol degree² and displacement force constants in kcal/mol Å²) computed for the central 10mer portion of DNA and RNA duplexes (top); RNA and DNA length independent persistence lengths for tilt, roll and twist deformations (in nanometers) and stretch modulus (in picoNewtons) (bottom)

| | Global tilt ^a | Global roll ^a | Global twist ^a | Global stretch ^b |
|---------------|--------------------------|--------------------------|---------------------------|-----------------------------|
| <i>Top</i> | | | | |
| DNA | 0.0043 | 0.0059 | 0.0056 | 1.50 |
| RNA | 0.0030 | 0.0058 | 0.0129 | 0.80 |
| <i>Bottom</i> | | | | |
| DNA | 73.85 | 101.85 | 96.15 | 3269 |
| RNA | 56.01 | 107.80 | 242.00 | 1891 |

^a In kcal/mol deg².
^b In kcal/mol Å².

then necessary when general global concepts like flexibility or rigidity are applied to DNA and RNA.

Stiffness analysis performed using local helical parameters (tilt, roll, twist, shift, slide and rise) computed using X3DNA³⁰ provides a direct measure of the average local deformability of each base-pair in DNA and RNA. Fortunately for our purposes, the local helical values sampled in our MD simulations follow a normal distribution ($r^2 > 0.998$ between the real distribution and the Gaussian's predicted one (equation (7)), which supports the goodness of the harmonic analysis described below). RNA is around 190%, 75%, 35% and 12% stiffer than DNA regarding twist, slide, shift and rise deformations (see Table 5). On the contrary, DNA is stiffer than RNA in terms of roll and tilt deformations, but the differences are small (between 10% and 20% increment in force-constants; see Table 5). When cross-terms are

considered (equation (8)), qualitative results remain mostly unaltered (see Table 6), reflecting the fact that stiffness matrix is dominated by diagonal terms. The only significant difference between force constants in Table 5 and diagonal force-constants in Table 6 appears for rise, as a consequence of the strong (especially for RNA) positive coupling of rise with slide and the negative one with twist and roll.

As a whole, results in Tables 5 and 6 show that for most local motions DNA is more flexible than RNA, but it is difficult to evaluate the implication of the different force constants in the local deformability. To obtain additional information in this point we computed the local deformation energy (see equation (9)) of DNA and RNA in front of random perturbation of helical parameters (to obtain meaningful results random changes are limited to a maximum of two times the largest standard

Table 5. Elastic force constants for deformations along local helical parameters of DNA and RNA

| | Tilt ^a | Roll ^a | Twist ^a | Shift ^b | Slide ^b | Rise ^b |
|-----|-------------------|-------------------|--------------------|--------------------|--------------------|-------------------|
| DNA | 0.031 | 0.016 | 0.016 | 1.094 | 1.564 | 5.174 |
| RNA | 0.028 | 0.013 | 0.046 | 1.484 | 2.709 | 5.814 |

Couplings between different parameters are neglected.

^a In kcal/mol deg².

^b In kcal/mol Å².

Table 6. Stiffness matrix for deformations along local helical parameters of DNA and RNA

| | Tilt | Roll | Twist | Shift | Slide | Rise |
|------------|-------|-------|--------|--------|--------|--------|
| <i>DNA</i> | | | | | | |
| Tilt | 0.031 | 0.001 | 0.000 | -0.043 | -0.001 | 0.004 |
| Roll | | 0.023 | 0.004 | -0.016 | -0.011 | -0.021 |
| Twist | | | 0.014 | -0.003 | -0.060 | -0.186 |
| Shift | | | | 1.228 | 0.044 | -0.001 |
| Slide | | | | | 1.886 | 0.951 |
| Rise | | | | | | 7.217 |
| <i>RNA</i> | | | | | | |
| Tilt | 0.026 | 0.000 | -0.001 | -0.180 | -0.007 | 0.004 |
| Roll | | 0.015 | 0.001 | 0.001 | -0.007 | -0.111 |
| Twist | | | 0.052 | -0.020 | -0.141 | -0.156 |
| Shift | | | | 1.369 | -0.030 | -0.013 |
| Slide | | | | | 3.193 | 1.077 |
| Rise | | | | | | 6.183 |

All couplings (in kcal/mol deg. Å) are included. Since the Tables are symmetric, only the upper part is shown.

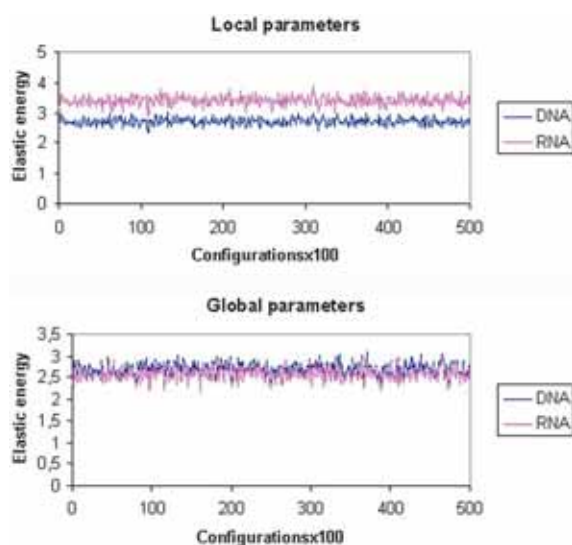


Figure 5. Elastic energy (in kcal/mol) associated to local (top) or global (bottom) distortions of DNA and RNA. In order to reduce noise, values represented here are averaged in blocks of 100 configurations.

deviation for each helical parameter). The elastic energy profiles (see Figure 5) clearly demonstrate that in terms of local deformations RNA is stiffer than DNA, despite that for a few types of special

deformations the opposite situation is found (see also Tables 5 and 6). On the contrary, when the same analysis is performed using global elastic constants, RNA is not found to be more rigid than DNA (see Figure 5), in line with the results found in essential dynamics and entropy analysis.

In summary, stiffness analysis using either essential motions or helical coordinates demonstrates that the concepts of flexibility/rigidity in DNA and RNA are complex, and that the ability of a polymer to be perturbed by an external force depends on the type of deformation. Overall, DNA is more susceptible than RNA to local deformations, but the resulting local geometrical changes are more efficiently propagated to the global level in RNA.

Analysis of interaction profiles

As noted by other authors,^{15,21,22} the relative flexibility of DNA and RNA is probably not only a direct consequence of its covalent structure, but also of its solvent environment (water and counterions). Solvation maps (see Figure 6) are clearly different for DNA and RNA because of the different groove distribution and the presence of an extra hydroxyl group, which alters the electrostatic potential around the polymer (see Figure 7). The total number of water molecules that solvate the RNA duplex is slightly larger (by around three per

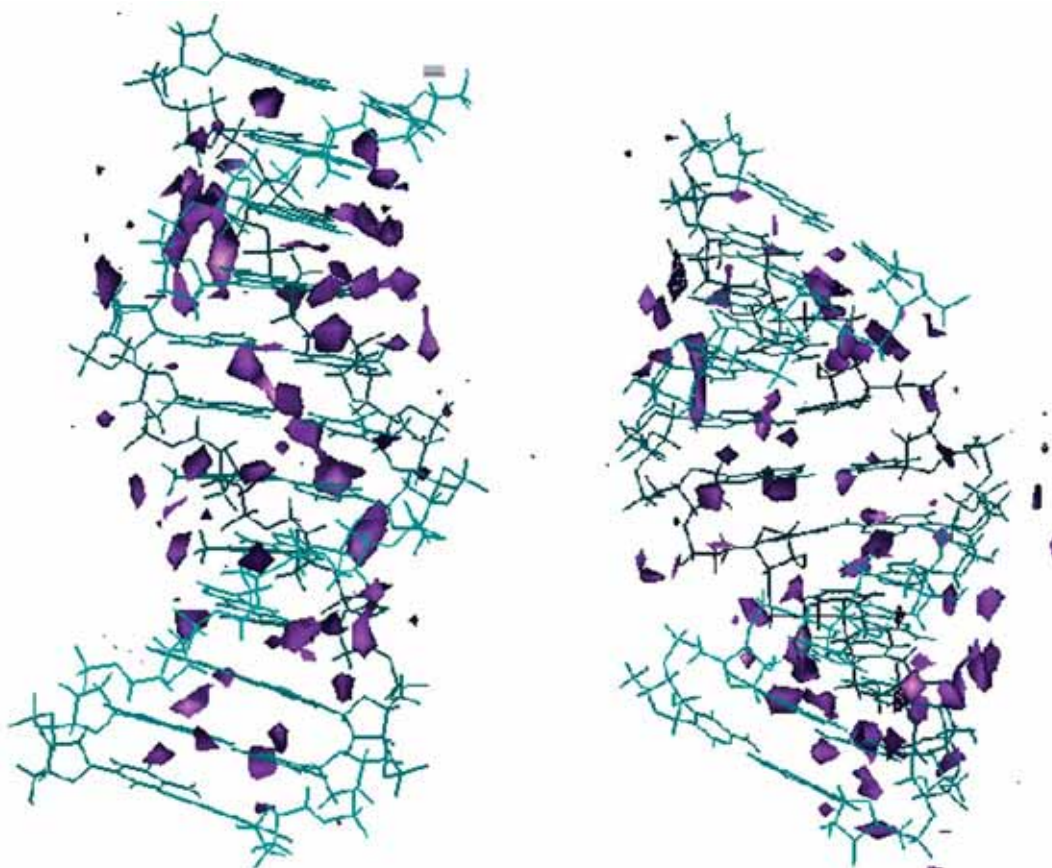


Figure 6. Solvation maps for DNA (left) and RNA (right). Water density contours correspond to an apparent density of 2 g/ml.

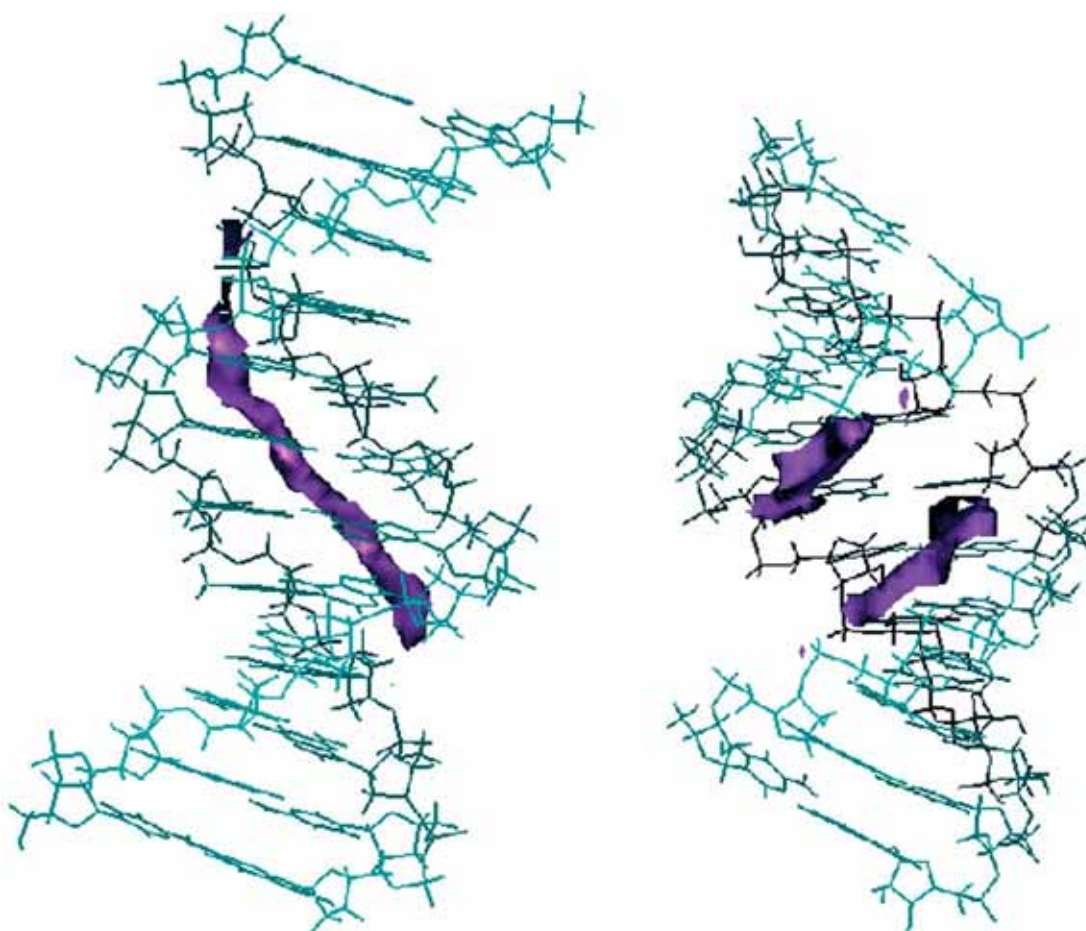


Figure 7. Classical molecular interaction potentials (in kcal/mol) for DNA (left) and RNA (right). Contour plots correspond to -4 kcal/mol for RNA and -3 kcal/mol for DNA.

base-pair in average; see Table 7) than for the DNA, the difference being mostly located in the backbone (see Table 7). However, longer residence times are found in general for DNA than for RNA, including those in the backbone (see Table 8), which is more flexible for DNA than for RNA (see above). These results strongly suggest that polymer flexibility is not closely related to specific solvation, and that the DNA/RNA motions are slow enough to allow a fast reorganization of the solvent.

The different electrostatic potential around DNA and RNA generates a different pattern of interaction with Na^+ (see Figure 7), thus justifying a different counterion distribution around DNA and RNA. Not only are Na^+ located closer to RNA than to DNA, but they have larger residence time for RNA when bound to the nucleic acid structure (up to 2 ns residence times are found for RNA in our simulations; see Table 7). Most of these “extra” bound-

Na^+ are located in the major groove of the RNA, which seems to be a better cationic binding site than the minor groove of DNA. It is worth to note that similar results were previously obtained by Westhof and co-workers^{21,22} and Cheatham & Kollman.¹⁵ Thus, though much caution is necessary in the analysis of counterion distribution due to the slow Na^+ equilibration,²⁵ MD simulations suggest that the ability of RNA to tightly capture Na^+ is larger than that of DNA. However, we must emphasize that, as shown in Figure 2, the major groove is more flexible in RNA than in DNA, which indicates that Na^+ binding has not a direct relationship with differential flexibility.

Conclusions

Essential dynamics analysis strongly suggests

Table 7. Average number of water molecules located at each base-pair of DNA and RNA

| | Backbone | m-groove | M-groove | Grooves (both) | Total |
|-----|----------|----------|----------|----------------|-------|
| DNA | 10.4 | 6.5 | 8.4 | 14.9 | 25.3 |
| RNA | 12.8 | 6.1 | 9.3 | 15.4 | 28.2 |

A contact is detected when there is a water oxygen at less than 3.5 \AA from a heteroatom of DNA or RNA.

Table 8. Maximum residence times (in ps) for water molecules (roman font) or sodium (in italics) bound to polar groups of DNA/RNA

| Atom | DNA | RNA |
|------------|---------|----------|
| O2' | – | 244 |
| O4' | 617 | 171 |
| O3' | 270 | 113 |
| O1P | 518/668 | 234/723 |
| O2P | 410/298 | 510/600 |
| O5' | 266 | 388 |
| Cyt O2 | 600/353 | 200/273 |
| Gua N3 | 438 | 145 |
| Gua N2 | 534 | 244 |
| Thy/Ura O2 | 521 | 245 |
| Ade/Ura N3 | 571 | 175 |
| Cyt N4 | 297 | 332 |
| Gua O6 | 537 | 544 |
| Gua N7 | 352/121 | 465/655 |
| Thy/Ura O4 | 402/351 | 374/449 |
| Ade N6 | 563 | 442 |
| Ade N7 | 558/415 | 500/2006 |

First block, backbone atoms; second block, atoms in the minor groove; third block, atoms in the major groove.

that the pattern of flexibility of DNA and RNA duplexes is different. In the first case there are many possible deformations of low energetic cost, while for RNA there are a few very soft deformations, but the others are very stiff.

Entropy and stiffness analysis suggest that overall the DNA duplex is more disordered and can fluctuate more than the RNA one. However, this greater flexibility of DNA is mostly local and due to backbone fluctuations that do not introduce large global changes in the helix.

Stiffness analysis using local helical parameters shows that DNA is much more flexible than RNA in terms of local twist, slide and shift and slightly more flexible for local rise. RNA is slightly more flexible in terms of local roll and tilt deformations. Considering global helical deformations the DNA is more deformable in terms of global twist, but stiffer for global tilt and stretch.

Overall, molecular dynamics simulations demonstrate that the concept of flexibility needs to be applied with caution to duplexes of RNA and DNA, since one duplex can be more flexible for some perturbations and more rigid for others. Clearly, a new, more complex view of nucleic acids dynamics seems necessary.

Methods

Molecular dynamics simulations

Two dodecamers, $d(\text{CGCGAATTCGCG})_2$ and $r(\text{CGCGAAUUCGCG})_2$, were built up using standard helical parameters²⁶ for *B*-DNA and *A*-RNA conformations. The structures were surrounded by around 4400 water molecules and 22 Na^+ placed in the regions of more electronegative potential. The hydrated systems

were then optimized, heated and pre-equilibrated using our standard multistage protocol,^{31,32} followed by 1 ns unrestrained MD simulation for equilibration. Finally, a 10 ns trajectory was collected for each structure. All MD simulations were performed in the isothermic-isobaric ensemble (1 atm, 298 K). Periodic boundary conditions and the Particle Mesh Ewald (PME³³) technique were used to account for long-range effects. All bonds were constrained using SHAKE,³⁴ which allowed us to use 2 fs time step for integration of Newton's equations. Parm99^{24,35} and TIP3P³⁶ force-fields were used to describe molecular interactions. Global translations were removed every 0.5 ns to remove erroneous partitions of the kinetic energy of the system. The analysis (see below) was performed using the last 9 ns of each trajectory and removing the base-pairs at both 5' and 3' ends. All the trajectories were obtained using the SANDER module of AMBER6.1 computer program.

Structural and energetic analysis

The helical parameters of the structures sampled along the 10 ns MD simulations were analyzed using the X3DNA program.³⁰ Additional geometrical parameters were derived with analysis modules in AMBER and with in house software. Classical molecular interaction potentials were computed using the CMIP program³⁷ with Na^+ as probe particle, and the electrostatic potentials were computed by solving numerically the Poisson-Boltzman equation.³⁸ Water densities around DNA and RNA duplexes were determined by integrating water population around polar atoms in DNA/RNA (cutoff distance of 3.5 Å). Maximum water residence times were computed by tracing all water molecules around polar groups of the nucleic acids.

Essential dynamics

This powerful analysis was used to determine which type of motions represents the most important structural fluctuations in DNA and RNA.^{25,37,38} Essential motions were determined from a principal component analysis. To this end, covariance matrices for common atoms of DNA and RNA (i.e., by excluding 5-methyl/H groups of T/U and 2'-OH/H groups in sugar moieties) were built and diagonalized. The resulting eigenvectors define the type of essential motions, and the associated eigenvalues determine how much of variance in the trajectory is explained by each eigenvector. Note that for two molecules of the same size the number of eigenvectors necessary to explain a given positional variance indicates the complexity of the molecular motions: the larger the number of essential motions, the greater the complexity.

The similarity between the essential motions of two molecules (defined by covariance matrices of the same size) can be computed using absolute (γ) and relative (κ) similarity indexes,^{25,41,42} as shown in equations (1) and (2):

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

where v_i^A stands for the unit eigenvector i of molecule A , n is the minimum number of essential motions that account for a given variance in the trajectory and the dot denotes a scalar product:

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

where the self-similarity indexes γ_{AA}^T are calculated by comparing eigenvectors obtained with the first and second parts of the same trajectory.

Entropy calculation

Since the intramolecular entropy is the best single parameter to describe the intrinsic order of a molecule, it provides a good estimate of the overall flexibility of DNA and RNA. The entropy of DNA and RNA was determined using Schiltter's⁴³ and Andreocci-Karplus⁴⁴ methods. The two models are based on the diagonalization of the mass-weighted covariance matrix, and on the use of a simple harmonic oscillator (hybrid classical-quantum (S) or purely quantum (A-K)) to link eigenvalues and entropy (see equations (3) and (4)). As noted before, intramolecular entropies were computed considering only atoms common to DNA and RNA:

$$S \approx 0.5k \sum_i \ln \left(1 + \frac{e^{\alpha_i}}{\alpha_i^2} \right) \quad (3)$$

$$S = k \sum_i \frac{\alpha_i}{e^{\alpha_i} - 1} - \ln(1 - e^{-\alpha_i}) \quad (4)$$

where $\alpha_i = \hbar\omega_i/kT$, ω denotes the eigenvalues obtained by diagonalization of the mass-weighted covariance matrix, and the sum extends to all the non-trivial vibrations.

The entropy estimates obtained by equations (3) and (4) depend on the length of the trajectory (the larger the simulation, the greater the number of visited microstates), which generates uncertainties in the calculation. To accelerate convergence, we used the extrapolation method,²⁸ which allows us to estimate the entropy at infinite simulation times from the time-evolution of the entropy estimated along the trajectory.^{25,28}

Stiffness analysis

The positional fluctuations of atoms along the trajectory were used to derive force-constants to describe the stiffness of DNA or RNA with respect to certain types of elastic deformations. This analysis provides then a unique tool for the description of the flexibility of nucleic acids.^{25,29,45-47}

We first computed the stiffness associated with the deformation of DNA and RNA along their essential modes. This analysis provides quantitative information on the deformability of both helices along the easiest motional pathways. Taking advantage of the orthogonality of eigenvectors and assuming a harmonic behaviour, force constants are easily determined from the eigenvalues by using equation (5):

$$K_i = kT/\lambda_i \quad (5)$$

where λ_i is the eigenvalue (in \AA^2) associated with the essential movement v_i determined by diagonalization of the covariance matrix, T is the temperature (in Kelvin), k is the Boltzman constant and K_i is a force constant associated to the essential motion.

Stiffness analysis applied to essential motions provides useful 3-D information on the deformability of nucleic acids. Unfortunately, it is not always simple to interpret the nature of eigenvectors, thus making it desirable to compute stiffness in a helical coordinate system closer to chemical intuition.^{45,47} Under the assumptions that the distribution of values adopted by a given variable X is fully Gaussian and that the deformation variables are

orthogonal (independent), harmonic (elastic) force constants K_X can be easily derived from the variance of the variable X during the trajectory using equation (6):

$$K_X = kT \frac{1}{\langle (X - X_0)^2 \rangle} \quad (6)$$

where X_0 is the equilibrium value of variable X and the brackets mean that fluctuations are averaged over thermodynamic ensemble.

An alternative way to compute force constants for a given deformable variable can be obtained by fitting the normalized histogram of values sampled for a given parameter along the trajectory, $\mathfrak{J}(X)$, to a Gaussian function (equation (7)). The exponent of the fitted Gaussian provides directly the force constant, and the goodness of the fitting (determined for example by the χ^2 test) gives an indication of the harmonicity of the deformation:

$$\mathfrak{J}(X) - \left[\frac{1}{\sqrt{2\pi}} \exp \frac{K_X}{2kT} (X - X_0)^2 \right] - \text{error}_X = 0 \quad (7)$$

The non-orthogonality of the helical parameters can be taken specifically into account by means of coupling terms, which requires the inclusion of force constants, K_{ij} , to account for contributions to the deformation energy arising from coupling interactions. After simple matrix algebra, equation (7) can be generalized by defining the stiffness matrix K with entries K_{ij} , which can be easily derived by inverting the covariance matrix C with entries $C_{ij} = \langle (X_i - X_{i0})(X_j - X_{j0}) \rangle$ as shown in equation (8). Diagonal elements in K represent contributions to deformation arising from individual helical variables, while off-diagonal components account for coupling terms:

$$K = kTC^{-1} \quad (8)$$

Note that when the stiffness matrix is known, the deformation energy of a given configuration can be calculated by equation (9), where the subindex 0 stands for the equilibrium value:

$$E_{\text{def}} = \sum_i \frac{K_{ii}}{2} (X_i - X_{i0})^2 + \sum_{i \neq j} \frac{K_{ij}}{2} (X_i - X_{i0})(X_j - X_{j0}) \quad (9)$$

Acknowledgements

This work has been supported by the Spanish Ministry of Science and Technology (BIO2003-06848 and SAF2002-04282), the Fundación BBVA and the VIth Framework Program of the European Union (UE project QLG1-CT-1999-00008). The computer facilities of the CIRI-CEPBA computer centre are also acknowledged. F.L. thanks for support from the Swiss National Science Foundation.

References

1. Bloomfield, V. A., Tinoco, I., Hearst, J. E. & Crothers, D. M. (2000). *Nucleic Acids: Structures, Properties, and Functions*. University Science Books, Sausalito, Calif.
2. Gait, M. J. & Blackburn, G. M. (1990). *Nucleic Acids in Chemistry and Biology*. IRL Press at Oxford University Press, Oxford.
3. Saenger, W. (1984). *Principles of Nucleic Acid Structure*. Springer, New York.
4. Soliva, R., Luque, F. J., Alhambra, C. & Orozco, M. (1999). Role of sugar re-puckering in the transition of A and B forms of DNA in solution. A molecular dynamics study. *J. Biomol. Struct. Dynam.* **17**, 89–99.
5. Sancar, A. (1996). DNA excision repair. *Annu. Rev. Biochem.* **65**, 43–81.
6. Pradeepkumar, P. I., Amirkhanov, N. V. & Chattopadhyaya, J. (2003). Antisense oligonucleotides with oxetane-constrained cytidine enhance heteroduplex stability, and elicit satisfactory RNase H response as well as showing improved resistance to both exo and endonucleases. *Org. Biomol. Chem.* **1**, 81–92.
7. Verbeure, B., Lescrinier, E., Wang, J. & Herdewijn, P. (2001). RNase H mediated cleavage of RNA by cyclohexene nucleic acid (CeNA). *Nucl. Acids Res.* **29**, 4941–4947.
8. Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E. & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, **391**, 806–811.
9. Hammond, S. M., Caudy, A. A. & Hannon, G. J. (2001). Post-transcriptional gene silencing by double-stranded RNA. *Nature Rev. Genet.* **2**, 110–119.
10. Hagerman, P. J. (1988). Flexibility of DNA. *Annu. Rev. Biophys. Biophys. Chem.* **17**, 265–286.
11. Hagerman, P. J. (1997). Flexibility of RNA. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 139–156.
12. Thomas, G. J., Jr, Benevides, J. M., Overman, S. A., Ueda, T., Ushizawa, K., Saitoh, M. & Tsuboi, M. (1995). Polarized Raman spectra of oriented fibers of A DNA and B DNA: anisotropic and isotropic local Raman tensors of base and backbone vibrations. *Biophys. J.* **68**, 1073–1088.
13. Fujiwara, T. & Shindo, H. (1985). Phosphorus-31 nuclear magnetic resonance of highly oriented DNA fibers. 2. Molecular motions in hydrated DNA. *Biochemistry*, **24**, 896–902.
14. Shindo, H., Fujiwara, T., Akutsu, H., Matsumoto, U. & Kyogoku, Y. (1985). Phosphorus-31 nuclear magnetic resonance of highly oriented DNA fibers. 1. Static geometry of DNA double helices. *Biochemistry*, **24**, 887–895.
15. Cheatham, T. E. & Kollman, P. A. (1997). Molecular dynamics simulations highlight the structural differences among DNA:DNA, RNA:RNA, and DNA:RNA hybrid duplexes. *J. Am. Chem. Soc.* **119**, 4805–4825.
16. Dickerson, R. E., Goodsell, D. S. & Neidle, S. (1994). "...the tyranny of the lattice...". *Proc. Natl Acad. Sci. USA*, **91**, 3579–3583.
17. Lipanov, A., Kopka, M. L., Kaczor-Grzeskowiak, M., Quintana, J. & Dickerson, R. E. (1993). Structure of the B-DNA decamer C-C-A-A-C-I-T-T-G-G in two different space groups: conformational flexibility of B-DNA. *Biochemistry*, **32**, 1373–1389.
18. Portmann, S., Usman, N. & Egli, M. (1995). The crystal structure of r(CCCCGGGG) in two distinct lattices. *Biochemistry*, **34**, 7569–7575.
19. Shakked, Z., Guerin-Guzikevich, G., Eisenstein, M., Frolov, F. & Rabinovich, D. (1989). The conformation of the DNA double helix in the crystal is dependent on its environment. *Nature*, **342**, 456–460.
20. Perez, A., Noy, A., Lankas, F., Luque, F. J. & Orozco, M. (2004). The relative flexibility of DNA and RNA a database analysis. *Nucl. Acid Res.* In the press.
21. Auffinger, P. & Westhof, E. (2000). Water and ion binding around RNA and DNA (C,G) oligomers. *J. Mol. Biol.* **300**, 1113–1131.
22. Auffinger, P. & Westhof, E. (2001). Water and ion binding around r(UpA)12 and d(TpA)12 oligomers—comparison with RNA and DNA (CpG)12 duplexes. *J. Mol. Biol.* **305**, 1057–1072.
23. Pan, Y. & MacKerell, A. D., Jr (2003). Altered structural fluctuations in duplex RNA versus DNA: a conformational switch involving base pair opening. *Nucl. Acids Res.* **31**, 7131–7140.
24. Cheatham, T. E., Cieplak, P. & Kollman, P. A. (1999). A modified version of the Cornell *et al.* force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dynam.* **16**, 845–862.
25. Orozco, M., Perez, A., Noy, A. & Luque, F. J. (2003). Theoretical methods for the simulation of nucleic acids. *Chem. Soc. Rev.* **32**, 350–364.
26. Arnott, S., Bond, P. J., Selsing, E. & Smith, P. J. (1976). Models of triple-stranded polynucleotides with optimised stereochemistry. *Nucl. Acids Res.* **3**, 2459–2470.
27. Packer, M. J. & Hunter, C. A. (1998). Sequence-dependent DNA structure: the role of the sugar-phosphate backbone. *J. Mol. Biol.* **280**, 407–420.
28. 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.
29. Lankas, F., Sponer, J., Hobza, P. & Langowski, J. (2000). Sequence-dependent elastic properties of DNA. *J. Mol. Biol.* **299**, 695–709.
30. Lu, X. J., Shakked, Z. & Olson, W. K. (2000). A-form conformational motifs in ligand-bound DNA structures. *J. Mol. Biol.* **300**, 819–840.
31. Shields, G. C., Laughton, C. A. & Orozco, M. (1997). Molecular dynamics simulations of the d(T center dot A center dot T) triple helix. *J. Am. Chem. Soc.* **119**, 7463–7469.
32. Shields, G. C., Laughton, C. A. & Orozco, M. (1998). Molecular dynamics simulation of a PNA center dot DNA center dot PNA triple helix in aqueous solution. *J. Am. Chem. Soc.* **120**, 5895–5904.
33. Darden, T., York, D. & Pedersen, L. (1993). Particle Mesh Ewald—an N.Log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092.
34. Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. (1977). Numerical-integration of cartesian equations of motion of a system with constraints—molecular-dynamics of N-alkanes. *J. Comput. Phys.* **23**, 327–341.
35. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M. *et al.* (1995). A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules. *J. Am. Chem. Soc.* **117**, 5179–5197.
36. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935.
37. Gelpi, J. L., Kalko, S. G., Barril, X., Cirera, J., de La Cruz, X., Luque, F. J. & Orozco, M. (2001). Classical molecular interaction potentials: improved setup procedure in molecular dynamics simulations of proteins. *Proteins: Struct. Funct. Genet.* **45**, 428–437.

38. Orozco, M. & Luque, F. J. (2000). Theoretical methods for the description of the solvent effect in biomolecular systems. *Chem. Rev.* **100**, 4187–4226.
39. Sherer, E. C., Harris, S. A., Soliva, R., Orozco, H. & Laughton, C. A. (1999). Molecular dynamics studies of DNA A-tract structure and flexibility. *J. Am. Chem. Soc.* **121**, 5981–5991.
40. Wlodek, S. T., Clark, T. W., Scott, L. R. & McCammon, J. A. (1997). Molecular dynamics of acetylcholinesterase dimer complexed with tacrine. *J. Am. Chem. Soc.* **119**, 9513–9522.
41. Cubero, E., Abrescia, N. G. A., 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.
42. Rueda, M., Kalko, S. G., Luque, F. J. & Orozco, M. (2003). The structure and dynamics of DNA in the gas phase. *J. Am. Chem. Soc.* **125**, 8007–8014.
43. Schlitter, J. (1993). Estimation of absolute and relative entropies of macromolecules using the covariance-matrix. *Chem. Phys. Letters*, **215**, 617–621.
44. Andricioaei, I. & Karplus, M. (2001). On the calculation of entropy from covariance matrices of the atomic fluctuations. *J. Chem. Phys.* **115**, 6289–6292.
45. Lankas, F., Sponer, J., Langowski, J. & Cheatham, T. E. (2003). DNA basepair step deformability inferred from molecular dynamics simulations. *Biophys. J.* **85**, 2872–2883.
46. Matsumoto, A. & Olson, W. K. (2002). Sequence-dependent motions of DNA: a normal mode analysis at the base-pair level. *Biophys. J.* **83**, 22–41.
47. Olson, W. K., Gorin, A. A., Lu, X. J., Hock, L. M. & Zhurkin, V. B. (1998). DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl Acad. Sci. USA*, **95**, 11163–11168.

Edited by P. J. Hagerman

(Received 8 March 2004; received in revised form 30 June 2004; accepted 6 July 2004)

4.2 Structure, Recognition Properties, and Flexibility of the DNA·RNA Hybrid

Agnes Noy, Alberto Pérez, Manuel Márquez, F. Javier Luque, Modesto Orozco

Journal of American Chemical Society (2005) **127**, 4910-20

Structure, Recognition Properties, and Flexibility of the DNA·RNA Hybrid

Agnes Noy,^{†,‡} Alberto Pérez,^{†,§} Manuel Márquez,[¶] F. Javier Luque,^{*,§} and Modesto Orozco^{*,†,‡}

Contribution from the Molecular Modeling and Bioinformatics Unit, Parc Científic de Barcelona, Josep Samitier 1-5, Barcelona 08028, Spain, Departament de Bioquímica i Biologia Molecular, Facultat de Química, Universitat de Barcelona, Martí i Franquès 1, Barcelona 08028, Spain, Departament de Farmàcia, Unitat de Físicoquímica, Facultat de Farmàcia, Universitat de Barcelona, Avda Diagonal 643, Barcelona 08028, Spain, and Los Alamos National Laboratory, Chemistry Division, Los Alamos, New Mexico 87545

Received November 8, 2004; E-mail: modesto@mmb.pcb.ub.es; javier@far1.far.ub.es

Abstract: Molecular dynamics is used to investigate the properties of the DNA·RNA hybrid in aqueous solution at room temperature. The structure of the hybrid is intermediate between A and B forms but, in general, closer to the canonical A-type helix. All the riboses exhibit North puckerings, while 2'-deoxyriboses exist in North, East, and South puckerings, the latter being the most populated one. The molecular recognition pattern of the DNA·RNA hybrid is a unique combination of those of normal DNA and RNA duplexes. Finally, the results obtained from essential dynamics and stiffness analysis demonstrate the large and very asymmetric flexibility of the hybrid and the strong predilection that each strand (DNA or RNA) has on the nature of their intrinsic motions in the corresponding homoduplexes. The implications of the unique structural and dynamic properties of the DNA·RNA hybrid on the mechanism of cleavage by RNase H are discussed.

Introduction

There are two major types of heterogeneous nucleic acids: hybrids and chimeras. In the former, not all the strands are of the same type (DNA or RNA), while in the latter, both DNA and RNA coexist in at least one of the strands. Hybrids and chimeras are minor species but play a key role in the cell life. Thus, transient DNA·RNA hybrids are formed as the RNA strand is created using the DNA template. Moreover, DNA replication relies on the existence of Okazaki's fragments, which are hybrid chimeras, where one strand is pure DNA and the other is an RNA–DNA chimera.¹ Furthermore, RNA viruses create DNA·RNA hybrids during retrotranscription, and the stability of these hybrid sequences is crucial in the replication cycle of these viruses.²

The large research effort spent in the past decade on the study of DNA·RNA hybrids is due not only to its biological relevance but also to its potential therapeutic application in antisense therapy. This new pharmacological strategy fights diseases by inactivation of the pathological messenger RNA (mRNA) following three main possible mechanisms:^{3–8} (i) the degrada-

tion of the mRNA by means of the RISC-DICER mechanism, (ii) the sterical interference of translation or splicing by a complementary oligonucleotide (typically a DNA derivative) bound to the target mRNA, and (iii) the degradation of the mRNA bound to a complementary DNA analogue by RNase H, which is a specific enzyme that recognizes in a catalytically active manner only DNA·RNA hybrids and degrades the RNA strand of the hybrid. At the present time, most antisense treatments in clinical trials are based on the activation of the RNase H mechanism,^{3–6} and then the design of stable and RNase H-susceptible hybrids is crucial.

The structure of DNA·RNA hybrids has been largely studied by means of experimental techniques. Early fiber diffraction data^{9,10} suggested that the structure of the hybrid was very close to the A-form. This finding was also supported by many high-resolution X-ray data. Thus, the structure of an Okazaki's fragment, r(gcg)·d(TATACGC), solved by Rich's group¹¹ showed that it was very close to a canonical A-helix, with all sugars in the North conformation. The same conclusions were found by different crystallographers in other hybrid–chimeras.^{12,13} Several pure DNA·RNA hybrids solved by X-ray

[†] Parc Científic de Barcelona.

[‡] Facultat de Química, Universitat de Barcelona.

[§] Facultat de Farmàcia, Universitat de Barcelona.

[¶] Los Alamos National Laboratory.

- (1) Meselson, M.; Stahl, F. W. *Proc. Natl. Acad. Sci. U.S.A.* **1958**, *44*, 671.
- (2) Alberts, B.; Bray, D.; Lewis, J.; Raff, M.; Roberts, K.; Watson, J. D. *Molecular Biology of the Cell*, 3rd ed.; Garland Publishing Inc.: New York, 1994.
- (3) Biroccio, A.; Leonetti, C.; Zupi, G. *Oncogene* **2003**, *22*, 6579.
- (4) Dean, N. M.; Bennett, C. F. *Oncogene* **2003**, *22*, 9087.
- (5) Wagner, W. R.; Matteucci, M. D.; Grant, D.; Huang, T.; Froehler, B. C. *Nat. Biotechnol.* **1996**, *20*, 384.

(6) Taylor, M. F.; Wiederholt, K.; Sverdrup, F. *Drug Discovery Today* **1999**, *4*, 562.

(7) Yelin, R.; Dahary, D.; Sorek, R.; Levanon, E. Y.; Goldstein, O.; Shoshan, A.; Diber, A.; Biton, S.; Tamir, Y.; Khosravi, R.; Nemzer, S.; Pinner, E.; Walach, S.; Bernstein, J.; Savitsky, K.; Rotman, G. *Nat. Biotechnol.* **2003**, *21*, 379.

(8) Carmichael, G. G. *Nat. Biotechnol.* **2003**, *21*, 371.

(9) Milman, G.; Langridge, R.; Chamberlin, M. J. *Proc. Natl. Acad. Sci. U.S.A.* **1967**, *57*, 1804.

(10) O'Brien, E. J.; MacEwan, A. W. *J. Mol. Biol.* **1970**, *48*, 243.

(11) Wang, A. H.; Fujii, S.; van Boom, J. H.; van der Marel, G. A.; van Boeckel, S. A.; Rich, A. *Nature* **1982**, *299*, 601.

crystallography exhibit a typical A-form, with all riboses in the North conformation and all¹⁴ or almost all^{15–17} 2'-deoxyriboses with North puckerings. Surprisingly, data provided by low-resolution NMR, CD, and Raman^{18–21} spectroscopies raise doubts on the X-ray structural picture of the hybrid; since all riboses are in the North conformation, a sizable portion of 2'-deoxyriboses exist in the South form, which is traditionally linked to the B-DNA. The same is found in recent NMR studies, which suggest a general structure of the hybrid intermediate between A and B canonical helices, with many characteristics of the A-form, but with important alterations in the grooves and with a large percentage of 2'-deoxyriboses in South conformations.^{22–33}

The discrepancy between high-resolution NMR and X-ray studies cannot be fully explained from the differences in base sequence between hybrids solved by X-ray (typically with polypyrimidine in the DNA strand) and by NMR (generally more heterogeneous), as demonstrated by Gyi et al. and Fedoroff et al.^{24,29,30} Clearly, experimental conditions are driving the structure toward the A and A/B conformations, suggesting a unique intrinsic structural plasticity in the DNA•RNA hybrid.

Though the discrepancy between NMR and X-ray results makes it difficult to define the major conformation of the hybrid in physiological conditions, it seems that NMR data provide an easier explanation of the fact that while DNA•RNA hybrids are degraded by RNase H, pure RNA duplexes are inhibitors.³⁴ Thus, according to NMR results, the smaller width of the minor groove in the hybrid compared to that in the pure RNA duplex has been considered to be the main structural feature to explain the different enzymatic susceptibility to RNase H.^{16,21,22,24,33,34} However, recent studies have pointed out that factors other than the shape of the minor groove must also be involved,^{32,33} thus raising doubts on the structural determinants that justify the degradation of the hybrid by RNase H.

Compared with the large amount of high-level experimental data, very few theoretical studies have examined the structure of the DNA•RNA hybrid. In a seminal paper, Cheatham and

Kollman³⁵ reported 2 ns molecular dynamics (MD) trajectories of a 10-mer DNA•RNA duplex with Cornell's force field.³⁶ They found that, irrespective of the starting structure (pure A- or B-forms), the hybrid adopts a mixed A/B conformation, whose general conformation resembles that of the A-form, but with 2'-deoxyribose puckerings in the South region, thus supporting NMR data and previous JUNMA results by Lavery's group.³⁷ Identical conclusions were reached by Venkateswarlu et al. from 1 ns MD samplings³⁸ and more recently by Lane and co-workers³⁰ in a 2 ns MD study.

In this paper, we will present extended MD simulations of DNA•RNA hybrids starting from two different X-ray structures. Present trajectories (which expand to 10 times longer simulation time than previous ones) analyzed with the help of a powerful set of datamining algorithms allowed us to obtain a complete picture of the structure and dynamics of DNA•RNA hybrids. Our trajectories clearly converge to a conformation close to that suggested by NMR techniques. The structure, molecular recognition, and especially flexibility characteristics of the hybrid are determined and compared with those obtained for pure DNA and RNA duplexes built up with the equivalent base sequence. Finally, a tentative explanation of the specificity of RNase H for the DNA•RNA hybrid is provided.

Methods

Molecular Dynamics Simulations. Dodecamers of sequence r(CGCGAAUCGCG)•d(CGCGAATTCGCG) were built using as template (i) the PDB crystal structure 1FIX³⁹ and (ii) an A-form conformation taken from Arnott's fiber diffraction data as implemented in the Biopolymer module of InsightII.⁴⁰ Choice of the Dickerson's sequence allowed us to compare the hybrid trajectories with those recently collected for the same sequence in DNA and RNA duplexes⁴¹ and which will be used as reference for pure duplexes. Moreover, the selection of these two starting structures allows us to determine whether MD simulations^{36,42} are consistent with crystal A-like structures or whether the trajectories spontaneously jump to the A/B NMR-like conformation.

The structures were first partially optimized (2000 cycles with restraints in the backbone) to avoid bad contacts emerging from changes in the sequence of the bases used in MD simulations with regard to that present in the X-ray crystallographic structure. The hybrids were surrounded by approximately 4400 water molecules and 22 Na⁺ molecules placed in the regions of more electronegative potential (this generates rectangular boxes, 62 × 62 × 57 (in angstroms)), with greater than 12 Å of waters from the DNA to the faces of the box. The hydrated systems were then optimized, heated, and equilibrated using our standard multistage protocol with length double than usual^{43,44} to prevent any equilibration problem arising from the fact that we are studying hybrids and not homoduplexes. Then, two (11 and 5 ns) unrestrained MD simulations were run (the first nanosecond was considered extra equilibration in both cases). All MD simulations were performed in

- (12) Egli, M.; Usman, N.; Zhang, S. G.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 534.
- (13) Egli, M.; Usman, N.; Rich, A. *Biochemistry* **1993**, *32*, 3221.
- (14) Xiong, Y.; Sundaralingam, M. *Nucleic Acids Res.* **2000**, *28*, 2171.
- (15) Conn, G. L.; Brown, T.; Leonard, G. A. *Nucleic Acids Res.* **1999**, *27*, 555.
- (16) Horton, N. C.; Finzel, B. C. *J. Mol. Biol.* **1996**, *264*, 521.
- (17) Katohira, M.; Lee, S. J.; Kobayashi, Y.; Sujeta, H.; Kyogoku, Y.; Iwai, S.; Ohtsuka, E.; Benevides, J. M.; Thomas, G. J. *J. Am. Chem. Soc.* **1990**, *112*, 4508.
- (18) Chou, S. H.; Flynn, P.; Reid, B. *Biochemistry* **1989**, *28*, 2422.
- (19) Chou, S. H.; Flynn, P.; Reid, B. *Biochemistry* **1989**, *28*, 2435.
- (20) Hall, K. B.; McLaughlin, L. W. *Biochemistry* **1991**, *30*, 10606.
- (21) Fedoroff, O. Y.; Salazar, M.; Reid, B. R. *J. Mol. Biol.* **1993**, *233*, 509.
- (22) Fedoroff, O. Y.; Ge, Y.; Reid, B. R. *J. Mol. Biol.* **1997**, *269*, 225.
- (23) Lane, A. N.; Ebel, S.; Brown, T. *Eur. J. Biochem.* **1993**, *215*, 297.
- (24) Salazar, M.; Fedoroff, O. Y.; Miller, J. M.; Ribeiro, N. S.; Reid, B. R. *Biochemistry* **1993**, *32*, 4297.
- (25) Gao, X.; Jeffs, P. W. *J. Biomol. NMR* **1994**, *4*, 367.
- (26) Gonzalez, C.; Stec, W.; Kobylanska, A.; Hogrefe, R. I.; Reynolds, M.; James, T. L. *Biochemistry* **1994**, *33*, 1062.
- (27) Gonzalez, C.; Stec, W.; Reynolds, M.; James, T. L. *Biochemistry* **1995**, *34*, 4969.
- (28) Gyi, J. I.; Conn, G. L.; Lane, A. N.; Brown, T. *Biochemistry* **1996**, *35*, 12538.
- (29) Gyi, J. I.; Lane, A. N.; Conn, G. L.; Brown, T. *Biochemistry* **1998**, *37*, 73.
- (30) Gyi, J. I.; Gao, D.; Conn, G. L.; Trent, J. O.; Brown, T.; Lane, A. N. *Nucleic Acids Res.* **2003**, *31*, 2683.
- (31) Cross, C. W.; Rice, J. S.; Gao, X. *Biochemistry* **1997**, *36*, 4096.
- (32) Nishizaki, T.; Iwai, S.; Ohkubo, T.; Kojima, C.; Nakamura, H.; Kyogoku, Y.; Ohtsuka, E. *Biochemistry* **1996**, *35*, 4016.
- (33) Szyperki, T.; Gotte, M.; Billeter, M.; Perola, E.; Cellai, L.; Heumann, H.; Wuthrich, K. *J. Biomol. NMR* **1999**, *13*, 343.
- (34) Salazar, M.; Fedoroff, O. Y.; Zhu, L.; Reid, B. R. *J. Mol. Biol.* **1994**, *241*, 440.

- (35) Cheatham, T. E.; Kollman, P. A. *J. Am. Chem. Soc.* **1997**, *119*, 4805.
- (36) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
- (37) Sanghani, S. R.; Lavery, R. *Nucleic Acids Res.* **1994**, *22*, 1444.
- (38) Venkateswarlu, D.; Lind, K. E.; Mohan, V.; Manoharan, M.; Ferguson, D. M. *Nucleic Acids Res.* **1999**, *27*, 2189.
- (39) Horton, N. C.; Finzel, B. C. *J. Mol. Biol.* **1996**, *264*, 521.
- (40) (a) Arnott, S.; Bond, P. J.; Selsing, E.; Smith, P. J. *Nucleic Acids Res.* **1976**, *3*, 2459. (b) *Insight II*; Accelrys Co.: San Diego, CA, 2004.
- (41) Noy, A.; Perez, A.; Lankas, F.; Luque, F. J.; Orozco, M. *J. Mol. Biol.* **2004**, *343*, 627.
- (42) Cheatham, T. E.; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845.
- (43) Shields, G. C.; Laughton, C. A.; Orozco, M. *J. Am. Chem. Soc.* **1997**, *119*, 7463.
- (44) Shields, G. C.; Laughton, C. A.; Orozco, M. *J. Am. Chem. Soc.* **1998**, *120*, 5895.

the isothermic–isobaric ensemble (1 atm, 298 K). Periodic boundary conditions and the Particle Mesh Ewald (PME⁴⁵) technique were used to treat long-range effects. All bonds were constrained using SHAKE,⁴⁶ which allowed us to use an integration time step of 2 fs. Parm99^{36,42} and TIP3P⁴⁷ force fields were used to describe molecular interactions. Global translations were removed every 0.5 ns to remove erroneous partitions of the kinetic energy of the system. Since both trajectories converged to a similar region of the configurational space (see below), the structural and flexibility analyses were performed by using only the last 10 ns of the longest trajectory. Moreover, the base pairs at both 5' and 3' ends were not considered in the analysis. All trajectories were obtained using the SANDER module of the AMBER6.1 computer program.⁴⁸

Structural and Energetic Analysis. Geometrical parameters sampled along the trajectories were studied using analysis modules in AMBER, the X3DNA program,⁴⁹ and *in house* software. Classical molecular interaction potentials were computed using the CMIP program⁵⁰ with Na⁺ as a classical probe particle. Water densities around duplexes were determined by integrating the water population around polar atoms (cutoff distance of 3.5 Å) of the nucleic acids. Water residence times were computed by tracing all water molecules around polar groups following the standard procedure in the PTRAJ module of AMBER.

Essential Dynamics. Essential motions^{41,51–53} were determined from principal component analysis (PCA) using covariance matrixes for common atoms of DNA and RNA (i.e., by excluding 5-methyl/H groups of T/U and 2'-OH/H groups in sugars). The diagonalization of the covariance matrix provided eigenvectors, which describe the nature of the essential movements, and eigenvalues, which determine the contribution of each essential movement to the positional variance of the trajectory. For two molecules of the same size, the number of eigenvectors necessary to explain a given positional variance indicates the complexity of the molecular motions; the larger the number of essential motions, the greater the complexity.

Eigenvectors of the same dimension obtained from two trajectories can be compared by means of the absolute and relative similarity indexes given in eqs 1 and 2,^{51,54,55} which measure the similarity between the essential deformation pattern of the two systems.

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

where v_i^A is the unit eigenvector i of molecule A; n is the minimum number of essential motions that account for a given variance in the trajectory, and the dot denotes a scalar product.

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

where the self-similarity indexes, γ_{AA}^T , are calculated by comparing eigenvectors obtained with the first and second parts of the same trajectory.

Entropy Calculation. Intramolecular entropies were determined using Schlitter⁵⁶ and Andreocci–Karplus⁵⁷ methods (see eqs 3 and 4) and considering only atoms common to DNA and RNA (see above). The time dependence of the entropy estimates was corrected using the standard exponential extrapolation method.^{51,58}

$$S \approx 0.5k \sum_i \ln \left(1 + \frac{e^2}{\alpha_i^2} \right) \quad (3)$$

$$S = k \sum_i \frac{\alpha_i}{e^{\alpha_i} - 1} - \ln(1 - e^{-\alpha_i}) \quad (4)$$

where $\alpha_i = \hbar\omega_i/kT$; ω denotes the eigenvalues obtained by diagonalization of the mass-weighted covariance matrix, and the sum extends to all nontrivial vibrations.

Stiffness Analysis. The positional fluctuations of atoms along the trajectory were used to derive force constants to describe the elastic deformability of the hybrid and their constituent strands.^{51,59–63} The stiffness associated with the essential movements was determined from the eigenvalues obtained by diagonalization of the Cartesian covariance matrix⁴¹ (see eq 5). Alternatively, the stiffness matrix K (with entries K_{ij}) associated with deformability along helical parameters was determined (see eq 6) by inverting the covariance matrix C (with entries $C_{ij} = \langle (X_i - X_{i_0})(X_j - X_{j_0}) \rangle$). Diagonal elements in K represent contributions to deformation arising from individual helical variables, while off-diagonal components account for coupling terms. Note that once the force matrix is known, the deformation energy of a given configuration can be calculated by eq 7, where the subindex 0 stands for the equilibrium value.

$$K_{\lambda_i} = kT/\lambda_i \quad (5)$$

where λ_i is the eigenvalue (in angstroms²) associated with the essential movement v_i determined by diagonalization of the covariance matrix; T is the temperature (in kelvin); k is the Boltzmann constant, and K_{λ_i} is a force constant associated with the essential motion.

$$K = kTC^{-1} \quad (6)$$

$$E_{\text{def}} = \sum_i \frac{K_{ii}}{2} (X_i - X_{i_0})^2 + \sum_{i \neq j} \frac{K_{ij}}{2} (X_i - X_{i_0})(X_j - X_{j_0}) \quad (7)$$

Helical force constants were determined at the local (using local helical parameters as determined by X3DNA) and global (using the parameters defined by Lankas et al.^{59,60}) levels. The local analysis was performed by using all of the snapshots collected from the MD trajectory.

Once the stiffness constants for global or local parameters are determined, the average deformation energy of the duplex can be

- (45) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
 (46) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
 (47) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
 (48) 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. *AMBER6*; University of California: San Francisco, CA, 1999.
 (49) Lu, X. J.; Shakked, Z.; Olson, W. K. *J. Mol. Biol.* **2000**, *300*, 819.
 (50) Gelpi, J. L.; Kalko, S. G.; Barril, X.; Cirera, J.; de La Cruz, X.; Luque, F. J.; Orozco, M. *Proteins* **2000**, *45*, 428.
 (51) Orozco, M.; Perez, A.; Noy, A.; Luque, F. J. *Chem. Soc. Rev.* **2003**, *32*, 350.
 (52) Wlodek, S. T.; Clark, T. W.; Scott, L. R.; McCammon, J. A. *J. Am. Chem. Soc.* **1997**, *119*, 9513.
 (53) Sherer, E. C.; Harris, S. A.; Soliva, R.; Orozco, M.; Laughton, C. A. *J. Am. Chem. Soc.* **1999**, *121*, 5981.
 (54) Cubero, E.; Abrescia, N. G. A.; Subirana, J. A.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **2003**, *125*, 14603.

- (55) Rueda, M.; Kalko, S. G.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **2003**, *125*, 8007.
 (56) Schlitter, J. *Chem. Phys. Lett.* **1993**, *215*, 617.
 (57) Andricioaei, I.; Karplus, M. *J. Chem. Phys.* **2001**, *115*, 6289.
 (58) Harris, S. A.; Gavathiotis, E.; Searle, M. S.; Orozco, M.; Laughton, C. A. *J. Am. Chem. Soc.* **2001**, *123*, 12658.
 (59) Lankas, F.; Sponer, J.; Hobza, P.; Langowski, J. *J. Mol. Biol.* **2000**, *299*, 695.
 (60) Lankas, F.; Sponer, J.; Langowski, J.; Cheatham, T. E. *Biophys. J.* **2003**, *85*, 2872.
 (61) Olson, W. K.; Gorin, A. A.; Lu, X. J.; Hock, L. M.; Zhurkin, V. B. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11163.
 (62) Matsumoto, A.; Olson, W. K. *Biophys. J.* **2000**, *83*, 22.
 (63) For comparison purposes with present Figure 9, note that Figure 4 of ref 46 has a typo error, and values in the y-axis need to be multiplied by a factor of 10.

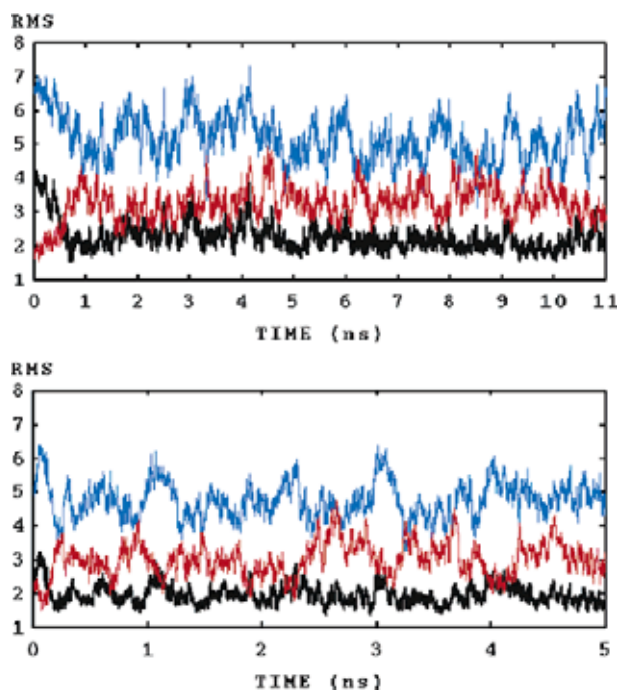


Figure 1. Root-mean-square deviations (RMSD for backbone atoms in angstroms) for the two hybrid trajectories; (top) starting 1FIX, and (bottom) from InsightII A-form, with respect to A (red), B (blue), and NMR (black; generated from 1EFS by (i) change of the sequence to the target one and (ii) restricted optimization of nucleobases to remove bad contacts) structures. Values displayed here correspond to the central 10-mer segment of the duplex.

computed considering a common limit of deformation for each helical coordinate. This limit should (i) make all helical deformation equally important in the definition of the distortion energy, and (ii) allow the comparison between DNA₂, RNA₂, and DNA-RNA stiffness. Thus, we defined the consensus perturbation for each helical parameter as twice the largest standard deviation found for this parameter in the three trajectories. Conformations are then randomly generated by moving randomly each individual helical parameter within these limits.

Results and Discussion

Convergence of the Trajectories. There is a clear displacement in the two trajectories (in less than 1 ns) from the A-form to an A/B conformation, closer to the A- than to the B-form, but clearly different from the canonical A-helix (see Figure 1). Both simulations sampled the same region of the configurational space, which is close to the conformation found in NMR experiments (1EFS⁶⁴ was used here for reference; see Figure 1). Analysis of the trajectories clearly demonstrated that, as previous MD simulations suggested,^{30,36,41} the fast repuckering of the 2'-deoxyriboses from pure North to South/South-East conformations is responsible for the structural transitions detected in the simulations.

Taking data from our longest trajectory (very similar results can be obtained considering the 5 ns one starting from pure A-form), we can quantify deviations in the converged structure with respect to reference conformations. The average RMSD (values taken for the central 10-mer backbone (including C1') in the last 10 ns of the trajectory) with respect to the starting conformation is 2.9 ± 0.5 Å, a value which is slightly larger than that found when RMSD is computed from a classical NMR

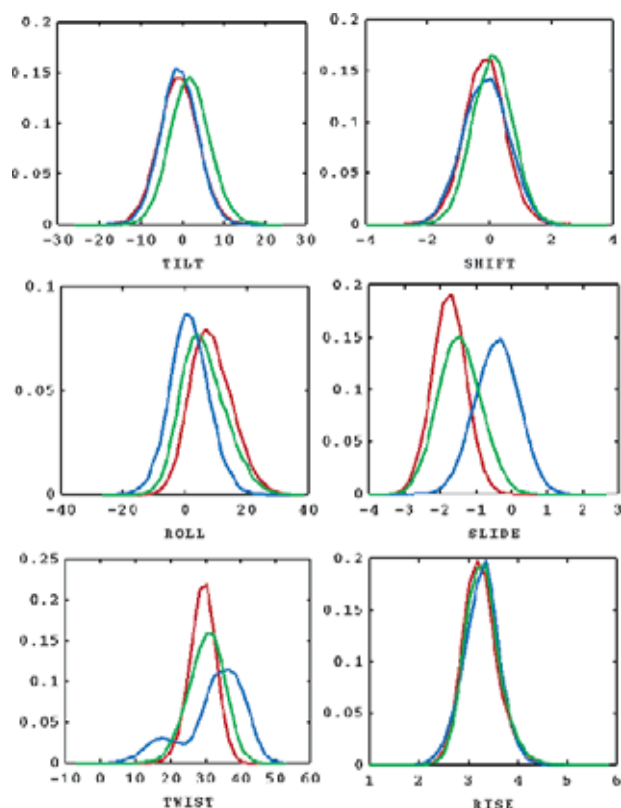


Figure 2. Distributions of selected local helical parameters in DNA₂, RNA₂, and DNA-RNA trajectories. Rotational values are in degrees, and the translational ones are in angstroms. DNA₂ (blue), RNA₂ (red), and hybrid (green).

(1EFS⁶⁴) conformation (2.2 ± 0.3 Å). The RMSD with respect to the B-form is larger (5.1 ± 0.7 Å with respect to Arnott's values and 3.5 ± 0.6 Å with respect to the MD-averaged conformation of B-DNA) than that with respect to A-form (3.3 ± 0.4 Å from Arnott's⁴⁰ values and 2.8 ± 0.5 Å from the MD-averaged conformation of the A-RNA). In summary, MD simulations drive the structure of the hybrid from fiber or crystal conformations to NMR-like conformation, which is not a pure A-form, but that, in general, is clearly closer to this conformation than to the B-form.

General Structural Characteristics. Helical analysis shows that the average twist angle is $\sim 31^\circ$, which is close to that found for the RNA duplex ($\sim 30^\circ$) and smaller than that for the DNA one ($\sim 33^\circ$) (see Figure 2). With regard to slide, the hybrid is also closer to RNA than to DNA. For the rest of the local helical parameters, the hybrid is either intermediate between DNA and RNA or does not change significantly between the three kinds of duplexes (see Figure 2). The intermediate A/B character of the hybrid becomes more evident when looking to more general descriptors, such as the shortest C1'–C1' intrastrand distance⁴¹ or the global (see Methods) helical parameters (see Figure 3). Interestingly, compared with the DNA duplex, both RNA and hybrid duplexes seem to be less sequence-dependent, as reflected in a more normal distribution of some helical parameters, such as twist (see Figure 2).

There are several differences in the distribution of the backbone dihedral angles (available upon request) with respect to pure DNA and RNA duplexes. In general, the differences are due to a strong asymmetry between the two strands in the hybrid, which seems to retain some kind of predilection for the

(64) Hantz, E.; Larue, V.; Ladam, P.; Le Moyec, L.; Gouyette, C.; Huynh Dinh, T. *Int. J. Biol. Macromol.* **2001**, *28*, 273.

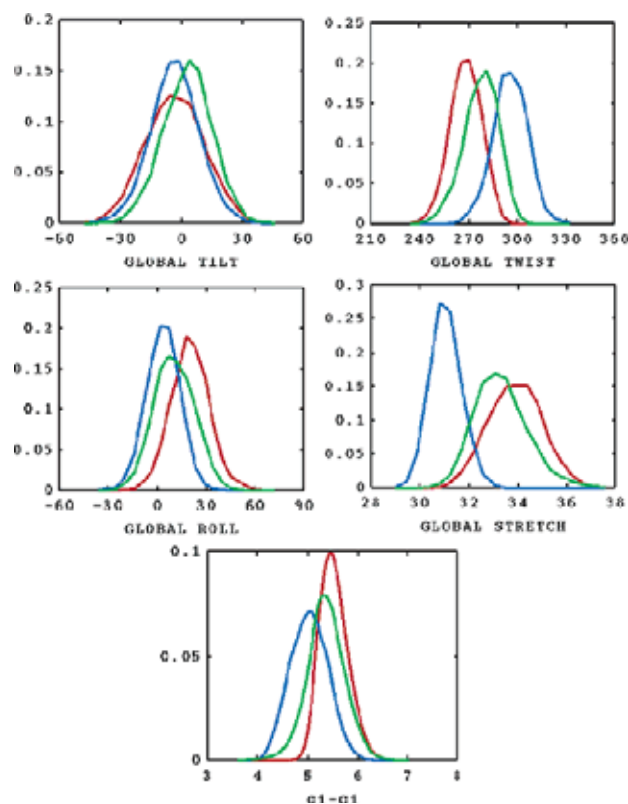


Figure 3. Distributions of selected global helical parameters in DNA₂, RNA₂, and DNA-RNA trajectories. Rotational values are in degrees, and the translational ones are in angstroms. See color code in Figure 2.

distributions found in corresponding pure duplexes (see Figure 4 for selected examples). Such a predilection is clearly reflected in the sugar pucker since all riboses are found in the North conformation (as expected for RNA), while 2'-deoxyriboses mainly populate South and South-East conformations (only 8–9% is found in North conformations) (see Figure 5). The change between S ↔ E ↔ N puckerings of 2-deoxyriboses is very fast (subnanosecond time scale), which indicates that present results cannot be ascribed to limited sampling in our MD simulations (see Figure 5). We must also notice that the amount of East conformers in the DNA strand of the hybrid is not different than that detected in normal DNA duplexes, which supports suggestions by James and co-workers^{26,27} that the anomalous sugar spectra found for the DNA-RNA hybrids are not due to a displacement of the 2'-deoxyriboses to the East conformation, but to a fast interchange between North and South puckerings.

The strand asymmetry in sugar pucker generates a unique groove distribution in the hybrid. The major groove is clearly wider than that of pure RNA and only ~2 Å narrower than that of pure DNA (see Figure 6). The minor groove of the hybrid is ~2 Å narrower than that of pure RNA and clearly wider than that of pure DNA. Very interestingly, the minor groove of the hybrid seems to be less sensitive to sequence effects (particularly to A-tracks) than does that of the DNA duplex (see Figure 6), leading to a more normal distribution in the hybrid. This difference is probably due to the geometrical restrictions imposed by the fixed conformation of the riboses in the RNA strand.

Molecular Recognition Properties. The unique groove geometry of the hybrid generates a special recognition pattern

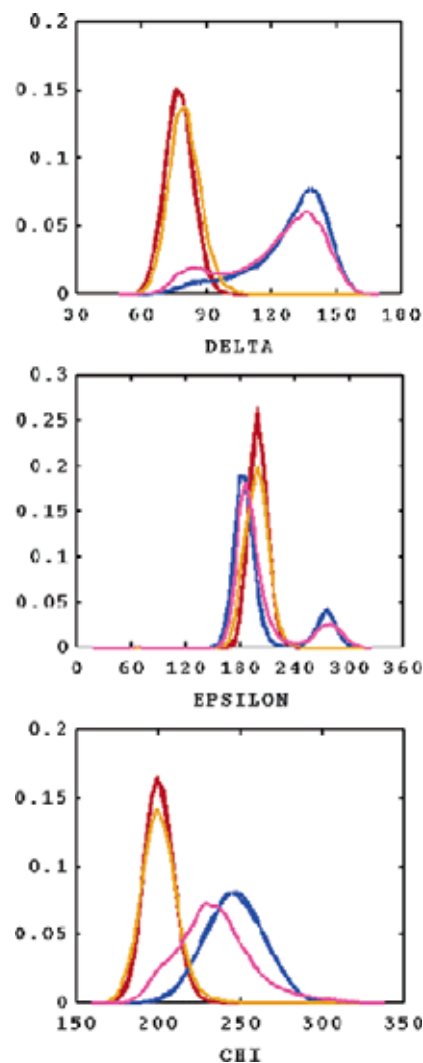


Figure 4. Distributions of selected backbone dihedrals in DNA₂, RNA₂, and DNA-RNA trajectories. Taking values for individual strands (the two DNA strand in DNA₂ (blue), the two RNA strands in RNA₂ (red), the DNA strand in the hybrid (magenta), and the RNA strand in the hybrid (orange)). Values are in degrees.

compared to DNA and RNA duplexes of the same sequence. This is noted in the cMIP distribution (Figure 7), which has characteristics of both DNA and RNA duplexes and an interesting asymmetry between strands in the major groove since the most favored region for interaction with small cations is not centered in the middle of the groove but displaced toward the RNA strand (see Figure 7). The hybrid is very well hydrated by an average number of 27.0 water molecules/nucleotide pair compared with values of 25.3 and 28.2 for DNA and RNA, respectively. In the minor groove, the amount of highly structured water is slightly larger (see Figure 7) for DNA than for RNA, the hybrid being in an intermediate situation. Due to the presence of the 2'-OH group, the backbone is better hydrated in RNA (12.8 waters/single strand) and in the RNA strand of the hybrid (12.4 waters) than in the DNA duplex (10.4 waters/single strand) or the DNA strand (10.6 waters) of the hybrid. The maximum water residence times around polar groups range from 100 ps to 1 ns, and no systematic differences are detected between DNA, RNA duplexes, and the hybrid.

Entropy Calculations. Schlitter and Andreocci-Karplus methods were used to estimate the intramolecular entropy of

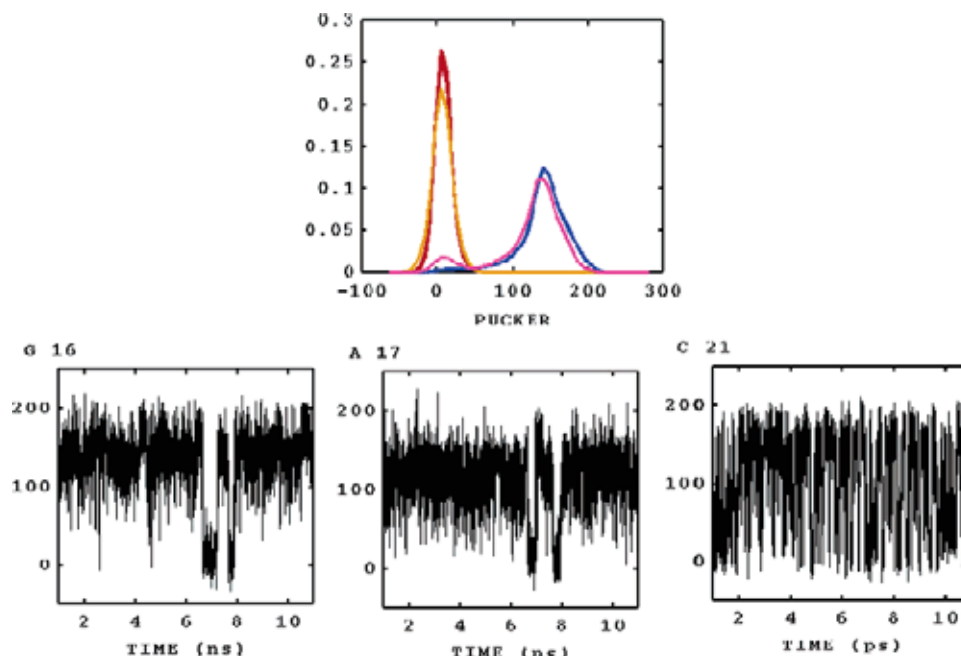


Figure 5. (Top panel) Distribution of phase angles (in degrees) in DNA₂, RNA₂, and DNA·RNA trajectories. Values are obtained by taking the two strands separately (see color codes in Figure 4). (Bottom panels) Evolution of the phase angle along the 11 ns hybrid trajectory for selected nucleotides.

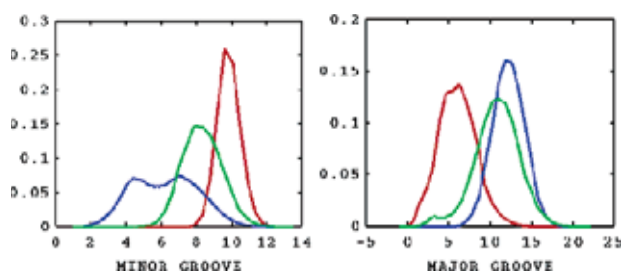


Figure 6. Distribution of minor and major groove widths in DNA₂, RNA₂, and DNA·RNA trajectories. Distances (in angstroms) are computed as the shortest P–P distance along the groove minus 5.8 Å (van der Waals radii of phosphates).

the duplexes. Both methods suggest that DNA is the most disordered structure, followed by the hybrid and finally by the RNA (Table 1). The difference in total entropy between RNA and DNA is 0.24 kcal/mol·K,⁴¹ and 0.16 kcal/mol·K between RNA and the hybrid. Therefore, while structural parameters point out that the hybrid is closer to pure RNA, the hybrid is closer to pure DNA in terms of structural disorder. As found in a previous analysis of homoduplexes,⁴¹ such an entropic difference arises from the backbone (entropy estimates obtained considering only the nucleobases are nearly identical for the three duplexes; see Table 1). Interestingly, when entropies are computed considering only the first three essential movements, DNA is the most ordered helix, while RNA and DNA·RNA show the same level of structural disorder. When the calculation includes the first 10 frequencies, the hybrid appears as the most disordered structure and RNA as the most ordered one. Clearly, entropy is not distributed uniformly in the three duplexes (see Figure 8), a fact that was already recognized in a previous study of DNA and RNA flexibility.⁴¹

The two strands of the hybrid retain the intrinsic flexibility in their respective homoduplexes. Thus, the DNA strand of the hybrid has an entropy 0.12 kcal/mol·K larger than that of the RNA strand (see Table 2). Quite impressively, the entropy estimates for the DNA and RNA strands of the hybrid nearly

Table 1. Intramolecular Entropies (in kcal/mol·K) Computed using Schlitter's (roman font) and Andreocci-Karplus's Methods (*italics*) for DNA, RNA, and HYBRID Double Strand and Extrapolated to Infinite Simulation Time (see Methods)^a

| | $S(t=\infty)$ | $S(3)$ | $S(10)$ |
|---------------|-------------------|---------------|---------------|
| DNA | 2.14(0.05) | 0.0328 | 0.0988 |
| (all atoms) | <i>1.93(0.05)</i> | <i>0.0327</i> | <i>0.0983</i> |
| RNA | 1.90(0.03) | 0.0342 | 0.0976 |
| (all atoms) | <i>1.71(0.03)</i> | <i>0.0340</i> | <i>0.0971</i> |
| HYBRID | 2.06(0.03) | 0.0341 | 0.0997 |
| (all atoms) | <i>1.85(0.03)</i> | <i>0.0339</i> | <i>0.0993</i> |
| DNA | 0.91(0.01) | 0.0286 | 0.0849 |
| (nucleobases) | <i>0.84(0.01)</i> | <i>0.0285</i> | <i>0.0845</i> |
| RNA | 0.91(0.02) | 0.0300 | 0.0860 |
| (nucleobases) | <i>0.83(0.02)</i> | <i>0.0299</i> | <i>0.0856</i> |
| HYBRID | 0.91(0.01) | 0.0299 | 0.0858 |
| (nucleobases) | <i>0.84(0.02)</i> | <i>0.0298</i> | <i>0.0853</i> |
| DNA | 1.43(0.07) | 0.0320 | 0.0962 |
| (backbone) | <i>1.34(0.10)</i> | <i>0.0319</i> | <i>0.0958</i> |
| RNA | 1.18(0.03) | 0.0333 | 0.0944 |
| (backbone) | <i>1.09(0.03)</i> | <i>0.0331</i> | <i>0.0940</i> |
| HYBRID | 1.34(0.04) | 0.0332 | 0.0973 |
| (backbone) | <i>1.24(0.04)</i> | <i>0.0331</i> | <i>0.0968</i> |

^a Partial entropies were computed considering only nucleobases or backbone atoms. In all cases, values were determined considering all frequencies, as well as only the first 3 and 10 ones.

match the average values obtained for the same strands in homopolymers of the same sequence (see Table 2). Finally, the entropy contributions associated with the first 10 essential movements of the DNA and RNA strands of the hybrid follow very closely those observed for the same strands in homopolymers (see Figure 8). In summary, the two strands of the hybrid maintain their intrinsic entropy in pure duplexes, which is not much altered by the counterpart, and the global entropy properties of the hybrid should then be understood as a combination of the local entropies of the two strands.

Essential Dynamics. As previously found in other studies,^{51,55} global twistings and bendings are the movements that explain most variances in the three duplexes. The first modes in all nucleic acids are associated with very small stiffness constants

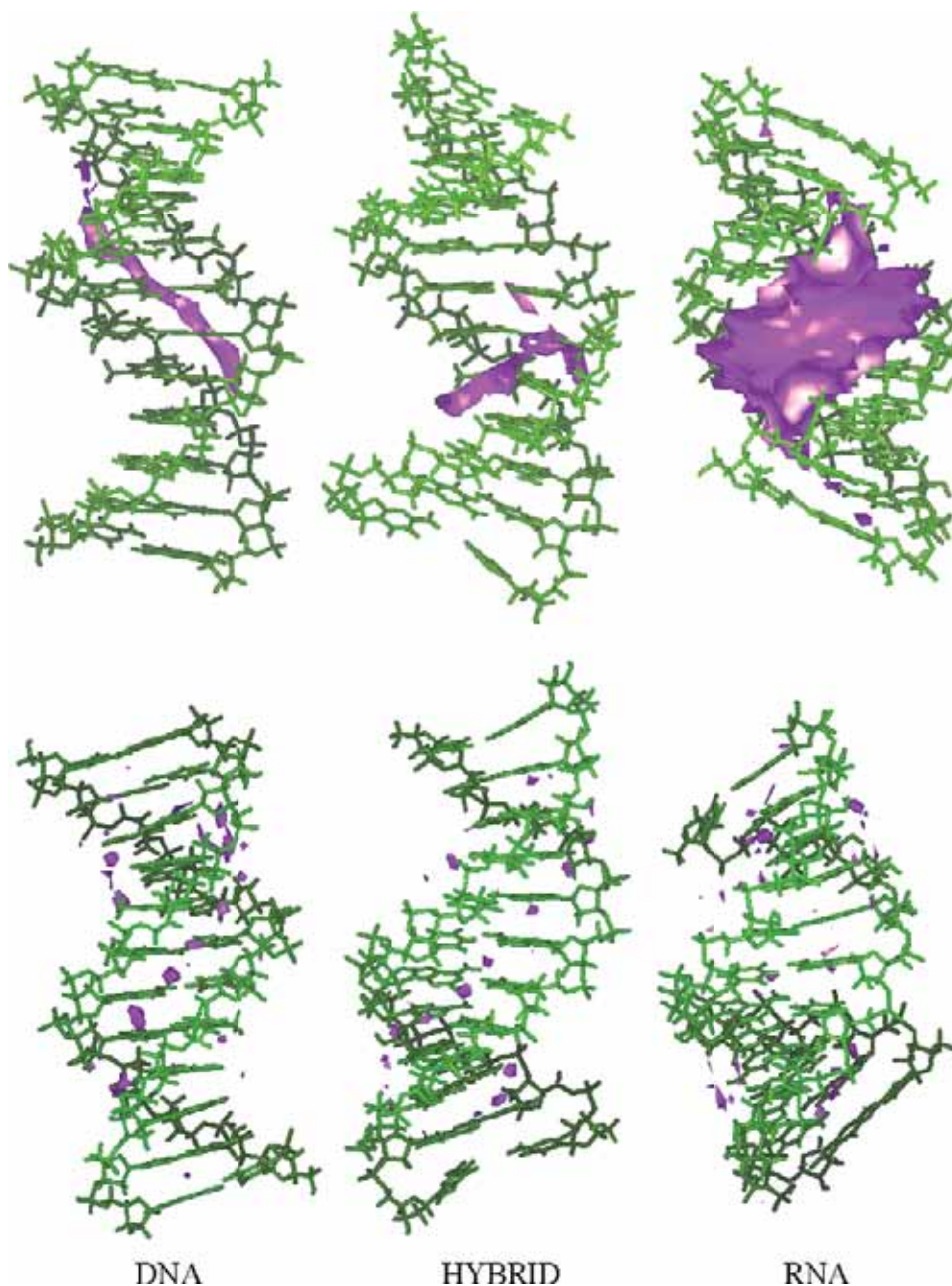


Figure 7. (Top) Classical Molecular Interaction Potential (in purple energy contour -3 kcal/mol), showing the best region for interaction of the different duplexes with a classical Na^+ probe. (Bottom) Solvation map (in purple density contour 2.5 g/cm 3) density for the three duplexes.

(below 10 cal/mol $\cdot\text{\AA}^2$; see Figure 9), reflecting the extreme plasticity of nucleic acids along their preferred deformation pathways.⁶³ As suggested by entropy plots (Figure 8), the DNA appears to be very stiff for the first deformation modes, but the situation changes for higher essential movements. The hybrid shows stiffness constants similar to those of the RNA for the first five components, while for lower modes, they approach those of the DNA homoduplex (see Figure 9).

Similarity measurements for the 10 first essential modes (those explaining more than 70% variance of all trajectories) reveal that the nature of these movements is quite common to the three duplexes, especially in movements involving nucleobases (see Table 3). The essential deformation modes of the hybrid are slightly more similar to those of the RNA than to those of the DNA (see Table 3). However, the hybrid is able to

capture characteristics of the essential dynamics of DNA which were not present in the RNA duplex and vice versa. Thus, similarity indexes between the hybrid and the two homoduplexes are larger than that obtained between pure DNA and RNA duplexes, the difference being especially clear when a large number (500) of eigenvectors are used in the comparison (see Table 3).

The essential dynamics of the RNA and DNA strands of the hybrid is significantly different, creating a unique asymmetry in the essential deformation modes of the hybrid. Thus, the relative similarities, κ , between the DNA and RNA strands of the hybrid are $\sim 76\%$ for 10 modes and 88% for 250 modes (see Table 3), while when the dynamics of the two complementary strands of pure DNA and RNA duplexes is compared, similarity indexes very close to 1.0 are obtained. The essential

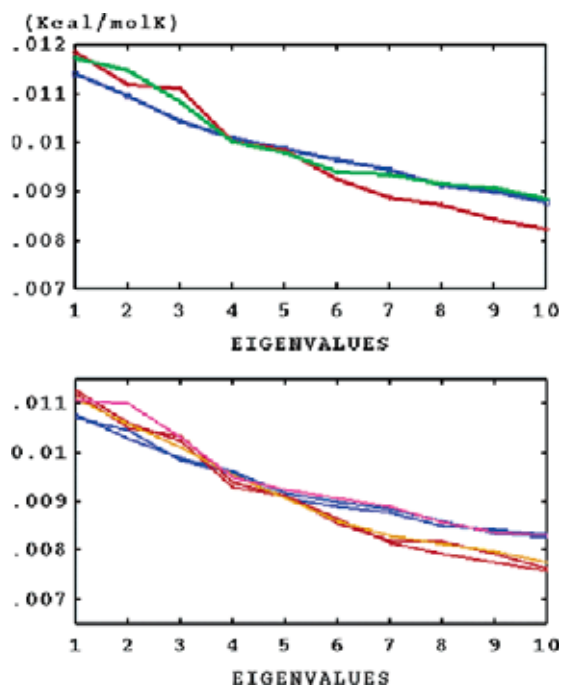


Figure 8. Entropies (in kcal/mol·K) assigned to different deformation modes of the three duplexes. (Top) Values for the two strands. (Bottom) Values for individual strands. Color code as in Figure 4.

Table 2. Same as Table 1, but Considering Each Strand Independently^a

| | $S(t = \infty)$ | $S(3)$ | $S(10)$ |
|---------------|-------------------|---------------|---------------|
| DNA | 1.11(0.05) | 0.0310 | 0.0928 |
| (all atoms) | <i>1.03(0.07)</i> | <i>0.0308</i> | <i>0.0923</i> |
| RNA | 0.99(0.02) | 0.0321 | 0.0909 |
| (all atoms) | <i>0.91(0.04)</i> | <i>0.0320</i> | <i>0.0905</i> |
| HYBRID DNA | 1.12(0.02) | 0.0324 | 0.0944 |
| (all atoms) | <i>1.04(0.03)</i> | <i>0.0323</i> | <i>0.0940</i> |
| HYBRID RNA | 1.00(0.02) | 0.0318 | 0.0912 |
| (all atoms) | <i>0.92(0.03)</i> | <i>0.0317</i> | <i>0.0908</i> |
| DNA | 0.49(0.01) | 0.0268 | 0.0786 |
| (nucleobases) | <i>0.45(0.01)</i> | <i>0.0266</i> | <i>0.0782</i> |
| RNA | 0.48(0.01) | 0.0280 | 0.0793 |
| (nucleobases) | <i>0.44(0.01)</i> | <i>0.0279</i> | <i>0.0789</i> |
| HYBRID DNA | 0.49(0.01) | 0.0281 | 0.0796 |
| (nucleobases) | <i>0.45(0.01)</i> | <i>0.0280</i> | <i>0.0792</i> |
| HYBRID RNA | 0.48(0.01) | 0.0277 | 0.0787 |
| (nucleobases) | <i>0.44(0.01)</i> | <i>0.0276</i> | <i>0.0783</i> |
| DNA | 0.75(0.09) | 0.0303 | 0.0906 |
| (backbone) | <i>0.71(0.14)</i> | <i>0.0301</i> | <i>0.0901</i> |
| RNA | 0.62(0.04) | 0.0313 | 0.0877 |
| (backbone) | <i>0.59(0.08)</i> | <i>0.0311</i> | <i>0.0873</i> |
| HYBRID DNA | 0.75(0.03) | 0.0317 | 0.0922 |
| (backbone) | <i>0.70(0.04)</i> | <i>0.0316</i> | <i>0.0918</i> |
| HYBRID RNA | 0.63(0.03) | 0.0310 | 0.0882 |
| (backbone) | <i>0.60(0.06)</i> | <i>0.0308</i> | <i>0.0877</i> |

^a For pure duplexes, values are the averages of the two strands.

movements of the RNA strand in the hybrid and in a pure RNA duplex are very similar (90% identity for 10 modes), while the maintenance of dynamics of the DNA strand is slightly worse (identity \sim 77% for 10 deformation modes (95% for 250 modes)). In summary, the two strands of the hybrid have a surprising tendency to maintain the essential dynamics of the DNA and RNA strands in pure homoduplexes. Such a predilection is especially strong for the RNA strand, whose deformability pattern in the hybrid is almost identical to that found in pure RNA duplexes.

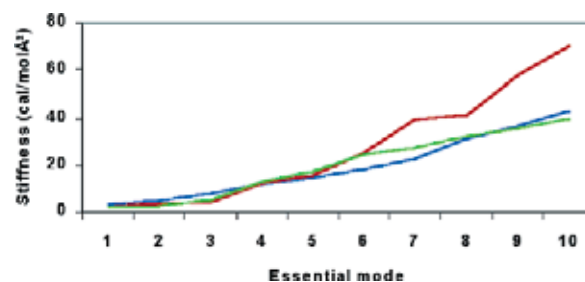


Figure 9. Force constants (in cal/mol·Å²) assigned to the most essential movements of the three duplexes. Color code as in Figure 2.

Table 3. Relative (κ) Similarity Indexes between the Essential Movements of DNA, RNA, and HYBRID at the Duplex (top) and Single Strand (bottom) Levels^a

| | all atoms | nucleobases | backbone |
|-------------------------------------|-------------|-------------|-------------|
| Duplex | | | |
| $\kappa_{\text{HYBRID/RNA}}$ | 0.788/0.911 | 0.859/0.988 | 0.805/0.935 |
| $\kappa_{\text{HYBRID/DNA}}$ | 0.722/0.921 | 0.782/0.990 | 0.728/0.919 |
| $\kappa_{\text{DNA/RNA}}$ | 0.682/0.850 | 0.701/0.970 | 0.692/0.870 |
| Single Strand | | | |
| $\kappa_{\text{HYB_RNA/RNA}}$ | 0.904/0.919 | 0.875/0.979 | 0.866/0.961 |
| $\kappa_{\text{HYB_RNA/DNA}}$ | 0.766/0.852 | 0.846/0.984 | 0.739/0.851 |
| $\kappa_{\text{HYB_DNA/RNA}}$ | 0.771/0.866 | 0.847/0.986 | 0.744/0.883 |
| $\kappa_{\text{HYB_DNA/DNA}}$ | 0.772/0.952 | 0.904/0.982 | 0.775/0.961 |
| $\kappa_{\text{HYB_RNA/HYB_DNA}}$ | 0.763/0.878 | 0.842/0.993 | 0.769/0.880 |

^a Values in roman correspond to the calculations using the first 10 modes, while values in italics are obtained considering the first 500 (duplex) or 250 (single strand) modes.

Table 4. Diagonal Elastic Force Constants for Deformations along a Reduced Set of Global Helical Parameters (angular force constants in cal/mol·deg² and displacement force constants in kcal/mol·Å²) Computed for the Central 10-mer Portion of DNA, RNA, and HYBRID Duplexes^{a,b}

| | global tilt | global roll | global twist | global stretch |
|--------|-------------|-------------|--------------|----------------|
| DNA | 4.29(0.05) | 5.92(0.07) | 5.58(0.07) | 1.51(0.039) |
| RNA | 2.98(0.05) | 5.74(0.11) | 12.89(0.21) | 0.80(0.009) |
| HYBRID | 4.56(0.06) | 4.76(0.12) | 9.52(0.23) | 0.79(0.018) |

^a Standard deviation values have been calculated by averages using different groups of snapshots taken every 5 ps. ^b The complete stiffness matrices are available upon request.

Helical Stiffness. As noted in previous papers,^{41,65} it is not clear whether DNA is more flexible than RNA in terms of global helical deformations. For the global twist, RNA is more rigid than DNA, but for global stretch and tilt, the reverse situation is found (see Table 4). Thus, DNA is more or less flexible than RNA depending on the type of global deformation. The hybrid has an intermediate behavior, though it is slightly closer to RNA than to DNA (see Table 4). Thus, the global twist of the hybrid is twice more difficult than for DNA and only 23% easier than for RNA. For the global stretch, the force constants in the hybrid and in RNA are identical and nearly one-half of that found for DNA. Finally, the hybrid is \sim 20% more flexible than both DNA and RNA in terms of global roll deformation and more rigid (50 and 20% relative to RNA and DNA, respectively) for changes in the global tilt (see Table 4). Considering a common isotropic distortion (i.e., a common distortion for all the duplexes defined to weight the same all the possible helical deformation;^{41,63} see Figure 10), there are no clear differences between the global flexibility of DNA, RNA, and DNA-RNA duplexes (Figure 10).

(65) Perez, A.; Noy, A.; Lankas, F.; Luque, F. J.; Orozco, M. *Nucleic Acids Res.* **2004**, *32*, 6144.

Table 5. Diagonal Elastic Force Constants for Deformations along Local Helical Parameters of DNA, RNA, and HYBRID^{a,b}

| | tilt ^c | roll ^c | twist ^c | shift ^d | slide ^d | rise ^d |
|--------|-------------------|-------------------|--------------------|--------------------|--------------------|-------------------|
| DNA | 31.16(0.31) | 18.50(0.15) | 14.92(0.12) | 1.22(0.01) | 1.90(0.02) | 7.20(0.01) |
| RNA | 26.48(0.22) | 15.06(0.08) | 51.72(0.21) | 1.37(0.01) | 3.19(0.02) | 6.18(0.06) |
| HYBRID | 27.04(0.17) | 12.82(0.12) | 33.12(0.17) | 1.58(0.01) | 2.42(0.01) | 7.01(0.05) |

^a Standard deviations are determined as noted in Table 4. ^b The complete stiffness matrixes are available upon request. ^c In cal/mol·deg² ^d In kcal/mol·Å²

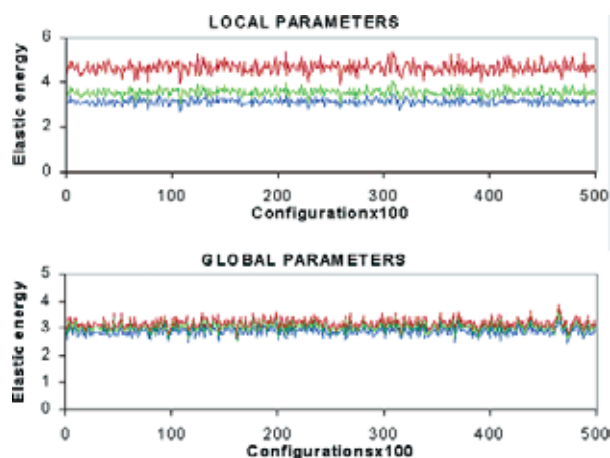


Figure 10. Elastic energy associated with helical isotropic distortions for the three nucleic acids. (Top) Perturbations in local helical parameters. (Bottom) Perturbations in global helical parameters. Perturbations along each helical variable are chosen as twice the largest standard deviation for this helical parameter in DNA₂, RNA₂, and DNA·RNA duplexes.

For most local helical parameters, RNA is stiffer than DNA, and in general, random local deformation is easier for DNA than for RNA^{41,65} (see Table 5 and Figure 10). Once again, the behavior of the hybrid is intermediate between that of DNA and RNA (see Table 5), but the local pattern of deformability of the hybrid is unique, and not just a simple scaled average of that of DNA and RNA homoduplexes (Table 5). For example, deformation in rise and tilt is equally difficult for RNA₂ and the hybrid, but it is much easier to unwind or bend (twist and roll stiffness) the hybrid than a pure RNA duplex. The isotropic deformation energy of the hybrid is slightly closer to that of DNA than to that of the RNA duplex (see Figure 10), but this situation changes if the hybrid is deformed more along a helical coordinate than along the others (anisotropic perturbation). Once again,^{41,65} the flexibility in nucleic acids emerges as a very complex concept especially for a molecule such as DNA·RNA with a very asymmetric pattern of deformability.

RNase H Susceptibility. RNase H has binding constants in the micromolar range for several oligonucleotides with A-like conformations, including RNA duplexes and DNA·RNA hybrids, but does not bind DNA duplexes^{66–71} or other B-form nucleic acids.⁷⁰ The crystal structure of a RNase H suggests that the enzyme does not recognize a pure canonical A-form and that some distortion in the helix occurs.^{72–75} It is also known that the enzyme does not show any marked sequence specific-

ity^{14,21,76,77} and is inactive against single-stranded oligonucleotides.^{69,70} Finally, in DNA·RNA duplexes, only the RNA strand of the hybrid is degraded.^{69,70}

All experimental data demonstrate that, despite the general similarity between RNA₂ and the hybrid, no appreciable amount of RNA duplex is degraded by the enzyme.^{66–70,76} The reasons for this extreme specificity have been obscure for decades since it is not easy to find structural determinants which are different for DNA·RNA and RNA₂. Thus, all crystal structures of the DNA·RNA hybrid are nearly identical to those found for pure RNA duplexes (see Introduction). This fact could explain why both RNA duplexes and DNA·RNA hybrids are recognized by the enzyme, but does not justify why RNA duplexes are inhibitors and DNA·RNA hybrids are substrates.

Crystallization conditions might bias the conformation of the DNA·RNA hybrid, and both NMR and MD simulations suggest that the hybrid adopts in physiological conditions an intermediate A/B-form. In fact, analysis of NMR or MD structures allows the determination of a few structural differences between DNA·RNA and RNA₂, such as the narrower minor groove in the hybrid compared with RNA duplex. On the basis of this finding, several authors have suggested that a minor groove with a width of ~8–9 Å is a necessary requisite for degradation by RNase H.^{16,21–24} The cMIP profiles in Figure 7 confirm that the average recognition pattern of the minor groove of the hybrid is different than that of the RNA duplex, providing an apparent support to the hypothesis that a narrow minor groove is the structural determinant for RNase H specificity. However, a more detailed analysis of the structures shows that there is large overlap in the distribution of widths of DNA·RNA and RNA duplexes, and ~20% of the time, the RNA duplex displays a minor groove with an “ideal” width of 8–9 Å (see Figure 6), which would imply some susceptibility of RNA duplexes to degradation by RNase H. Furthermore, cMIP calculations on hybrid–chimeras, which are recognized and degraded by RNase H,^{32,33} show very anomalous minor grooves, leading to cMIP distributions far from that expected for a normal DNA·RNA hybrid (compare Figures 7 and 11) and, in some cases, identical to that of a normal RNA duplex (Figures 7 and 11). In summary, equilibrium geometry can easily explain why B-type helical structures do not bind the enzyme, but it cannot explain the discriminative ability of RNase H between RNA duplex and DNA·RNA hybrid.

Discarding sequence effects and the equilibrium geometry as the unique determinants for the specificity of RNase H, we can consider that the reduced stability of the DNA·RNA hybrid compared to that of the RNA homoduplex^{78–81} can be a key

(66) Altmann, K. H.; Fabbrot, D.; Dean, N. M.; Geiger, T.; Monia, B. P.; Muller, M.; Nicklin, P. *Biochem. Soc. Trans.* **1996**, *24*, 630.

(67) Agrawal, S.; Iyer, R. P. *Pharmacol. Ther.* **1997**, *76*, 151.

(68) Crooke, S. T.; Bennett, C. F. *Annu. Rev. Pharmacol. Toxicol.* **1996**, *36*, 107.

(69) Han, G. W.; Kopka, M. L.; Cascio, D.; Grzeskowiak, K.; Dickerson, R. E. *J. Mol. Biol.* **1997**, *269*, 811.

(70) Lima, W. F.; Crooke, S. T. *Biochemistry* **1997**, *36*, 390.

(71) Stein, H.; Hausen, P. *Science* **1969**, *166*, 393.

(72) Ding, J.; Hughes, S. H.; Arnold, E. *Biopolymers* **1997**, *44*, 125.

(73) Ding, J.; Das, K.; Hsiou, Y.; Sarafianos, S. G.; Clark, A. D.; Jacobo-Molina, A.; Tantillo, C.; Hughes, S. H.; Arnold, E. *J. Mol. Biol.* **1998**, *284*, 1095.

(74) Jacobo-Molina, A.; Ding, J.; Nanni, R. G.; Clark, A. D.; Lu, X.; Tantillo, C.; Williams, R. L.; Kamer, G.; Ferris, A. L.; Clark, P. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 6320.

(75) Sarafianos, S. G.; Das, K.; Tantillo, C.; Clark, A. D.; Ding, J.; Whitcomb, J. M.; Boyer, P. L.; Hughes, S. H.; Arnold, E. *EMBO J.* **2001**, *20*, 1449.

(76) Oda, Y.; Iwai, S.; Ohtsuka, E.; Ishikawa, M.; Ikehara, M.; Nakamura, H. *Nucleic Acids Res.* **1993**, *21*, 4690.

(77) Roberts, W. R.; Crothers, D. M. *Science* **1992**, *258*, 1463.

(78) Riley, M.; Maling, B. *J. Mol. Biol.* **1966**, *20*, 359.

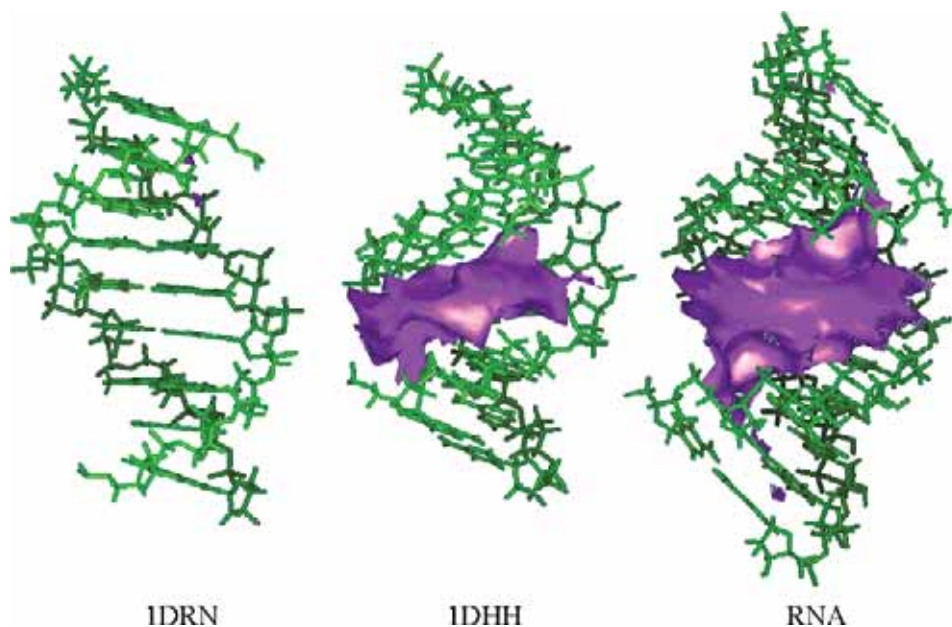


Figure 11. Classical Molecular Interaction Potential (energy contours -3 kcal/mol) for two hybrid structures known to be the substrate of RNase H (IDRN, IDHH) and a reference RNA₂ duplex which is not degraded by the enzyme.

determinant of the different susceptibility of DNA•RNA and RNA₂ to the action of RNase H. This possibility is indirectly supported by the fact that modified oligonucleotides designed to make more stable DNA•RNA hybrids^{5,67,68,82} fail to produce RNase H-susceptible hybrids. However, a few modified oligonucleotides have been generated displaying simultaneously good stability, specificity, and also RNase H susceptibility.^{30,83–86} Thus, an intrinsic instability in the duplex does not appear to be a requisite for RNase H susceptibility.

In summary, to our understanding, and without rejecting a possible role for structure and for specific interactions (for example, those involving 2'-OH groups), flexibility emerges as the major differential trend that can use RNase H to discriminate between both duplexes.^{32,33,87} Our MD simulations show that the pattern of deformability of the hybrid is quite different than that of a pure RNA duplex. Not only is the hybrid, in general, more flexible than the RNA duplex but also are there several specific local deformations which are much easier for the hybrid than for the RNA duplex, which provides more possibilities for deformability in the helix during the catalysis. Interestingly, our simulations strongly suggest that the RNA•DNA hybrid has a strong asymmetry in terms of flexibility between both strands. It is suggested that this unique asymmetry might be used by the enzyme to distinguish between the DNA and RNA strands, something that for a duplex showing a more symmetric pattern of flexibility will be very difficult.

Our results support a complex mechanism of action for RNase H. In a first step, the enzyme should bind any duplex showing a general conformation not far from the A-form and should reject B-type structures. In a second step, the enzyme should distort the duplex in a very asymmetric way, leaving the rigid RNA near the cleavage site while the DNA strand is pointed to the exterior. It is expected that this type of deformation will be too energetically costly for the rigid RNA duplex. Overall, we suggest the differential flexibility of RNA duplex and the hybrid, and especially, the asymmetry in the flexibility pattern of the latter constitutes the basis for the selective mechanism of action of RNase H. We propose that future designs of oligonucleotides for antisense purposes should explicitly consider these flexibility issues to guarantee the enzymatic susceptibility of the resulting hybrid.

Conclusions

Extended state-of-the-art MD simulations are able to reproduce with accuracy the structural properties of DNA•RNA hybrids, even when the starting conformation is far from the equilibrium conformation in solution.

The equilibrium geometry of the hybrid is closer to the A-form than to the B-form. All riboses show North puckerings, but 2'-deoxyriboses are mostly in the South and South-East regions, which lead to the definition of grooves intermediate to those typical of the A- and B-forms.

The flexibility of the DNA•RNA hybrid is unique and not a simple average of that of pure DNA and RNA hybrids. Quite surprisingly, each strand in the hybrid maintains well its essential dynamics in pure duplexes, which generates a strong asymmetry in the pattern of deformability of the helix.

Analysis of the different putative mechanisms that will allow RNase H to distinguish between different duplexes strongly suggests that while equilibrium structure can be enough for the enzyme to discard DNA₂, only the differential flexibility pattern can justify the ability of RNase H to discriminate between DNA•RNA and RNA₂.

- (79) Sugimoto, N.; Nakano, S.; Katoh, M.; Matsumura, A.; Nakamura, H.; Ohmichi, T.; Yoneyama, M.; Sasaki, M. *Biochemistry* **1995**, *34*, 11211.
 (80) Nakano, S.; Kanzaki, T.; Sugimoto, N. *J. Am. Chem. Soc.* **2004**, *126*, 1088.
 (81) Freier, S. M.; Altmann, K. H. *Nucleic Acids Res.* **1997**, *25*, 4429.
 (82) McKay, R. A.; Miraglia, L. J.; Cummins, L. L.; Owens, S. R.; Sasmor, H.; Dean, N. M. *J. Biol. Chem.* **1999**, *274*, 1715.
 (83) Moulds, C.; Lewis, J. G.; Froehler, B. C.; Grant, D.; Huang, T.; Milligan, J. F.; Matteucci, M. D.; Wagner, R. W. *Biochemistry* **1995**, *34*, 5044.
 (84) Barnes, T. W.; Turner, D. H. *J. Am. Chem. Soc.* **2001**, *123*, 4107.
 (85) Barnes, T. W.; Turner, D. H. *Biochemistry* **2001**, *40*, 12738.
 (86) Tonelli, M.; Ulyanov, N. B.; Billeci, T. M.; Karwowski, B.; Guga, P.; Stec, W. J.; James, T. L. *Biophys. J.* **2003**, *85*, 2525.
 (87) Nakamura, H.; Oda, Y.; Iwai, S.; Inoue, H.; Ohtsuka, E.; Kanaya, S.; Kimura, S.; Katsuda, C.; Katayanagi, K.; Morikawa, K.; Miyashiro, H.; Ikehara, M. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 11535.

Acknowledgment. We thank Dr. Chattopadhyaya for helpful discussions, and the Centre de Supercomputació de Catalunya for computer resources. We also thank the financial support of the Spanish Ministry of Education and Science (SAF2002-

04282; BIO2003-06848), Fundació La Caixa, and Fundación BBVA. A.N. is a fellow of the Spanish Ministry of Education and Science, and A.P. is a DURSI fellow.
JA043293V

4.3 Theoretical Analysis of Antisense Duplexes:
Determinants of the RNase H Susceptibility

Agnes Noy, Javier Luque, Modesto Orozco
Journal of American Chemical Society In press

Theoretical Analysis of Antisense Duplexes: Determinants of the RNase H Susceptibility

Agnes Noy^{1,2}, F.Javier Luque³ and Modesto Orozco^{1,2,4}*

Contribution from the Joint IRB-BSC Research Program in Computational Biology, Institut de Recerca Biomèdica Parc Científic de Barcelona, Josep Samitier 1-5, Barcelona 08028 and Barcelona Supercomputer Center, Jordi Girona 31, Edifici Torre Girona, Barcelona 08028, Spain, National Institute of Bioinformatics, Parc Científic de Barcelona, Josep Samitier 1-5, Barcelona 08028, Spain, Departament de Fisicoquímica and Institut de Biomedicina, Facultat de Farmàcia, Universitat de Barcelona, Avgda Diagonal 643, Barcelona 08028, Spain, Departament de Bioquímica, Facultat de Biologia, Universitat de Barcelona, Avgda Diagonal 647, Barcelona 08028, Spain.

modesto@mmb.pcb.ub.es

¹ Joint IRB-BSC Research Program in Computational Biology.

² National Institute of Bioinformatics.

³ Departament de Fisicoquímica and Institut de Biomedicina. Universitat de Barcelona

⁴ Departament de Bioquímica. Universitat de Barcelona.

Abstract: The structure and dynamic properties of different antisense related duplexes (DNA·RNA, 2'-O-Me-DNA·RNA, 2'-F-ANA·RNA, C5(Y)-propynyl-DNA·RNA, ANA·RNA and control duplexes DNA·DNA and RNA·RNA) have been determined by means of long molecular dynamics simulations (covering more than 0.5 μ s of fully-solvated unrestrained MD simulation). The massive analysis

presented here allows us to determine the subtle differences between the different duplexes, which in all cases pertain to the same structural family. This analysis provides information on the molecular determinants that allow RNase H to recognize and degrade some of these duplexes, while others with apparently similar conformation are not affected. Subtle structural and deformability features define the key properties used by RNase H to discriminate between duplexes.

Introduction

Antiparallel DNA·RNA hybrids are stable structures formed when complementary strands of DNA and RNA bind following recognition rules analogous to those of homopolymeric duplexes (i.e., those formed by pure DNA (DNA₂) and RNA (RNA₂) strands). A large series of crystallographic studies¹⁻⁶ suggested that in the crystal phase the DNA·RNA hybrid adopts a pure A-type structure, which should in practice be close to the homopolymeric RNA duplex. However, low resolution CD and NMR data⁷⁻⁹ and more recently high resolution NMR experiments¹⁰⁻¹⁶ showed that the structure of the DNA·RNA hybrid in solution is more complex, as it appears to have many characteristics of a canonical A-duplex, but some structural features resemble those of a B-type duplex. This view has received support by nanosecond-scale molecular dynamics (MD) simulations¹⁷, which suggest that the hybrid is closer to the A-type form than to the B-type one, but the structure does not strictly fit the canonical A-conformation (we named this conformation as A/B-form), showing an unique DNA vs RNA strand asymmetry.

The cell makes use of DNA·RNA hybrids at specific physical locations and at well defined times of its life cycle (for example, in the nuclei during replication), but in general the formation of the hybrid duplexes is associated with potential damages for the cell like viral infection¹⁸, triggering then survival responses which are mostly mediated by the enzyme RNase H¹⁹⁻²¹. This enzyme is a nuclease that degrades the RNA strand of the hybrid without affecting the complementary DNA strand^{22,23}, which can thus bind another complementary RNA strand. RNase H does not have a clear sequence-specificity^{6,9,24-26}, but exhibits an astonishing ability to discriminate between DNA·RNA hybrids and other related structures such as DNA₂, RNA₂, or single stranded DNAs or RNAs^{22,23,25, 27-30}.

The ability to degrade DNA·RNA has been exploited for biomedical and biotechnological purposes in the so called “antisense” therapy, where a single stranded DNA (i.e., antisense) is introduced in the cell to hybridize with a desired (i.e., sense) messenger RNA, which (if expressed) could trigger a given pathology. The first antisense drugs are already in the market³¹ and many other are under clinical trials to treat a wide range of pathologies, such as cancer, inflammation and viral infections^{29,32,33}. Unfortunately, the accumulated experience demonstrates that the first-generation antisense drugs face major practical problems, which has stimulated an intense research effort to design modified nucleic acids (XNA): i) resistant to interfering nucleases, and ii) able to bind with high stability and specificity to the target sense RNA^{34,35}. There are many examples of XNAs which as single strand are resistant to nucleases, can enter efficiently into the cell and form stable and specific XNA·RNA duplexes, but their usefulness is limited by the fact that the corresponding hybrid is not degraded by RNase H²⁸. Keeping in mind the chemical difficulties of developing modified nucleic acids, it would be worth to identify the molecular determinants that a XNA should satisfy in order to be recognized as substrate by RNase H prior to its synthesis. Unfortunately, despite the large amount of structural data on XNA·RNA hybrids (17 structures in PDB), the vast biochemical knowledge on the reaction^{4,9,11,14,15,22-30} and the availability of few structures of the enzyme^{36,37}, the structural reasons allowing the enzyme to distinguish between different hybrids remain to be elucidated.

In order to identify the molecular basis of the RNase H susceptibility, we report here a massive MD study for a variety of XNA·RNA duplexes, some of them degraded by the enzyme while others not. The analysis reveals the existence of clear differences in the structure and deformability pattern between duplexes susceptible or resistant to the enzyme. On the basis of these findings, we outline a clear and simple protocol to be used prior to the chemical synthesis of the XNA to evaluate its susceptibility to RNase H.

Methods

Sequence selection. As is common in the field^{38,39}, Dickerson’s dodecamer (DD⁴⁰) was selected as a typical sequence (d/r(CGCGAATTCGCG)₂) to study different types of XNA·RNA duplexes. However,

to make our conclusions more general and to escape from potential sequence-specific artifacts, we also analyzed two additional dodecamers designed by permutation of DD nucleobases subject to the following restrictions: i) all unique dinucleotide steps should be represented, ii) a similar Pyr/Pur ratio should exist in each strand to avoid the adoption of unusual structures, and iii) only palindromes were considered to avoid sequence-induced asymmetries between sense and antisense strands. With these premises, the chosen sequences were additional d/r(CATAGGCCTATG)₂ and d/r(CACAGATCTGTG)₂.

Hybrids considered. Besides the standard DNA·RNA hybrid, a variety of duplexes formed upon binding the sense RNA strand to different complementary XNAs acting as antisense candidates were considered: i) arabino-nucleic acids (ANA·RNA duplex), ii) 2'-F-arabino-nucleic acids (F-ANA·RNA duplex), iii) 2'-O-Me-DNA (OmeNA·RNA duplex), and iv) C5-propynyl-DNA (YpDNA·RNA duplex). With these XNAs, we include modifications in the sugar and in the nucleobase, which give rise to hybrids of different stability and RNase H susceptibility⁴¹⁻⁴⁴. The corresponding homopolymers (DNA₂ and RNA₂) were also included as a reference. Accordingly, a total of 7 structures were analyzed for 3 different sequences, leading to 21 different dodecamers.

Structure set-up. The generation of the starting structures was first carried out for the DD sequences, which were then used as template for the other sequences (see below). For homopolymeric duplexes (DNA₂ and RNA₂) the corresponding X-ray structures were chosen⁴⁵. For the hybrids two starting structures were generated: i) one corresponding to the pure A-form, and ii) the other to the NMR-based structure of the DNA·RNA hybrid (1EFS¹⁶) manipulated as described elsewhere¹⁷. All the structures were neutralized by Na⁺ ions and immersed in rectangular (around 60 x 60 x 70 Å³) boxes of TIP3P waters (from 4385 to 6661 molecules), so that there is at least 12 Å from any atom of the DNA duplex to the edges of the box. The systems were minimized, thermalized (T=298 K) and equilibrated using our standard equilibration protocol⁴⁶ doubling the length of the individual periods to ensure the lack of equilibration artifacts.

For DNA·RNA and YpDNA·RNA both structures converge quickly (in less than 700 ps) to samplings close to the NMR-based structure (see Figure S1 in supplementary material and Figure 1). Despite all our efforts the arabino-nucleic acids (ANA·RNA and F-ANA·RNA) failed to yield stable trajectories starting from the A-form due to the very bad interactions around the sugar when the pure A-form was imposed. Accordingly, only trajectories starting from the NMR-based structure were followed. The reverse problem was found for OmeNA·RNA duplexes, where the only stable trajectories were those starting from the pure A-form.

The starting structures for the simulations with the other two sequences ($d/r(\text{CATAGGCCTATG})_2$ and $d/r(\text{CACAGATCTGTG})_2$) were created by sequence substitution of the snapshots at the 20th ns of the trajectories obtained for the DD sequence. Because the differences in sequence only involved changes of A by G and T by C, common atoms were retained and the rest of them were added by using the LEAP module of the AMBER8.0 program⁴⁷ (those changes affected C6 and C2 positions for purines, and C5, C4 and C3 ones in pyrimidines). The set of simulations performed in this study is summarized in Table 1.

Table 1. Total simulation time (in ns) of the different trajectories considered in this study for different duplexes starting from A- and B-forms and from NMR-based structure (1EFS)^{a,b}.

| Seq 1 | DNA₂ | RNA₂ | HYB | F-ANA | ANA | YpDNA | OmeNA |
|-----------------------|------------------------|------------------------|----------------|--------------|------------|----------------|--------------|
| NMR-form | -- | -- | 50 | 50 | 50 | 50 | not stable |
| A-form | -- | 50 | 3 ^c | not stable | not stable | 3 ^c | 50 |
| B-form | 50 | -- | -- | -- | -- | -- | -- |
| Seq 2 | DNA | RNA | HYB | F-ANA | ANA | YpDNA | OmeNA |
| Equil(1) ^d | 20 | 20 | 20 | 20 | 20 | 20 | 20 |
| Seq 3 | DNA | RNA | HYB | F-ANA | ANA | YpDNA | OmeNA |
| Equil(1) ^d | 20 | 20 | 20 | 20 | 20 | 20 | 20 |

^aOnly long simulations (i.e. 50 and 20 ns) were analyzed.

^b1, 2 and 3 stand for the three sequences considered here: $d/r(\text{CGCGAATTCGCG})_2$, $d/r(\text{CATAGGCCTATG})_2$ and $d/r(\text{CACAGATCTGTG})_2$, respectively

^c After 3 ns converged to the NMR-form.

^dTrajectories starting from the snapshots at 20th ns for sequence 1.

Force-field. Canonical strands (DNA or RNA) were represented using the parmbsc0⁴⁰ revision of parm99 force-field⁴⁸, which yields reliable trajectories for a large variety of nucleic acids up to (at least) the microsecond time scale⁴⁹. Parameterization of the C5-propynyl derivative was done by using RESP/6-31G(d) charges⁵⁰ and gaff force-field⁵¹ parameters for the propynyl group. Parameters for the 2'-O-Me, arabino and 2'-F-arabino derivatives were taken from Venkateswarlu's study⁵². Charges for arabino and 2'-F-arabino derivatives were obtained from RESP/6-31G(d) calculations⁵⁰. To derive suitable torsional parameters for the modified sugars, the pseudorotation profile of thymidine nucleoside containing either arabino or 2'-F-arabino sugars was computed at the MP2/6-31G(d) level and fitted using a Monte-Carlo procedure as described elsewhere⁴⁰. Geometries were fully optimized except for the backbone dihedral angles (β , γ , ϵ , χ), which were fixed at standard hybrid values (i.e. values from NMR structure (1EFS); 177, 80, -170, -140 respectively), and the two internal dihedrals of the sugars required fixing the phase angle at a given value. The final parameters reproduce quite well the QM profiles (Figure S2 in supplementary material), and their goodness is further supported by the results obtained in test calculations (20 ns trajectories in aqueous solution) of small arabino and 2'-F-arabino hairpins⁵³, which were chosen due to the availability of NMR data (Table S1 in supplementary material). The optimized parameters (downloadable at <http://mmb.pcb.ub.es/antisense>) lead to pseudorotational profiles significantly different from those obtained for 2'-deoxyribose (Figure S3 in supplementary material).

Simulation details. The equilibrated structures of DD were subject to 50 ns of unconstrained MD simulation (100 ns for the DNA·DNA duplex⁴⁰) at constant temperature (298 K) and pressure (1 atm) using periodic boundary conditions and particle mesh Ewald⁵⁴. SHAKE⁵⁵ was used to constrain all bonds involving hydrogen atoms, which allowed us to use an integration step of 2 fs. Default AMBER8.0 parameters were used for the remaining simulation conditions⁴⁷. Based on the strong stability of the trajectories observed for DD-sequences, only 20 ns simulations were performed for the other two sequences (which in fact started from 20 ns pre-equilibrated structures; see above and Table

1). All MD simulations were performed using the PMEMD module of the AMBER8.0 computer program⁴⁷.

Analysis. Standard geometrical analysis was performed to follow the main structural features of the duplexes. The recognition properties were examined from classical molecular interaction potentials (CMIP⁵⁶) using Na⁺ as a probe. The essential dynamics of the duplexes was derived by diagonalization of the covariance matrix^{38,57}, which yields a set of eigenvectors $\{v_i\}$ describing the nature of the essential movements and the associated eigenvalues $\{\lambda_i\}$, which indicate the magnitude of the displacement expected along each mode. Note that in the harmonic limit the stiffness constants associated to each deformation mode can be derived from λ_i as shown in eq. 1

$$K_i = k_B T / \lambda_i \quad (1)$$

where k_B is Boltzman's constant and T is the absolute temperature.

To compare the similarity in the set of essential movements in two trajectories, we used absolute (eqs. 2 and 4) and relative (eqs. 3 and 5) similarity indexes previously described^{38,58}. Note that the index ξ_{AB} (eq. 4) is equivalent to γ_{AB} (eq. 2), but considers explicitly the similarity between all the associated eigenvalues included in the set of "important" eigenvectors (i.e., those needed to explain a given percentage of the total structural variance). Furthermore, note the relative indexes eliminate the noise in the absolute index arising from the limited length of simulations.

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

where n is the number of "important" eigenvectors (as usual we consider 10 eigenvectors which explains on average around 75% of variance), and v_i^X stands for the i-unitary eigenvector of trajectory X (note that γ_{AB} takes values from 0 to 1, which mean null similarity and identical essential dynamics).

$$\kappa_{AB} = 2 \frac{\gamma_{AB}}{(\gamma_{AA}^T + \gamma_{BB}^T)} \quad (3)$$

where γ_{XX}^T is the absolute self-similarity index (see eq. 2) for trajectory X obtained by comparing the first and second halves of the trajectory.

$$\xi_{AB} = \frac{2 \sum_{i=1}^{i=z} \sum_{j=1}^{j=z} \left[\left(\mathbf{v}_i^A \circ \mathbf{v}_j^B \right) \frac{\exp\left\{-\frac{(\Delta x)^2}{\lambda_i^A} - \frac{(\Delta x)^2}{\lambda_j^B}\right\}}{\sum_{i=1}^{i=z} \exp\left\{-\frac{(\Delta x)^2}{\lambda_i^A}\right\} \sum_{j=1}^{j=z} \exp\left\{-\frac{(\Delta x)^2}{\lambda_j^B}\right\}} \right]^2}{\sum_{i=1}^{i=z} \left(\frac{\exp\left\{-2\frac{(\Delta x)^2}{\lambda_i^A}\right\}}{\left(\sum_{i=1}^{i=z} \exp\left\{-\frac{(\Delta x)^2}{\lambda_i^A}\right\}\right)^2} \right)^2 + \sum_{j=1}^{j=z} \left(\frac{\exp\left\{-2\frac{(\Delta x)^2}{\lambda_j^B}\right\}}{\left(\sum_{j=1}^{j=z} \exp\left\{-\frac{(\Delta x)^2}{\lambda_j^B}\right\}\right)^2} \right)^2} \quad (4)$$

where λ_i is the eigenvalue (in \AA^2) associated to eigenvector \mathbf{v}_i , Δx is set to standard values for duplexes⁵⁸, and the sum is extended to the same “important” space used in eq. 2 ($z=10$).

$$\delta_{AB} = 2 \frac{\xi_{AB}}{(\xi_{AA}^T + \xi_{BB}^T)} \quad (5)$$

Similarity indexes between several trajectories were clustered to summarize information. To this end, we used agglomeration hierarchical clustering, which consists of building up the tree from elements by progressively merging clusters. Thus, a first cluster is defined by the two elements with shorter distance; then, values from this cluster to all elements are recalculated assuming average and shorter distance is chosen again to make new cluster or to increase one. The clustering procedure was performed using dissimilarity distances defined as $1-\delta$, an intuitive measure and fulfill statistics requirements for a distance ($d_{AB}=d_{BA}$; $d_{AA}=0$; $d_{AB} \geq 0$). When duplexes with the same sequence were compared, all common atoms including those at the nucleobases were considered (i.e. excluding 5-H/methyl/propynyl of pyrimidines and 2'-H/OH/F/O-methyl). However, when comparison involved different sequences only the backbone ending at C1' was included. In all cases the analysis was performed for the duplex as well as for the separated strands.

The ability of a given duplex to adapt its structure to that required to the productive binding to RNase H was analyzed by computing the distance in “essential space” needed to convert the “unbound” form to the “bound” conformation defined by the structure of a RNase H-bound DNA·RNA hybrid (1ZBI structure³⁷) and the associated energy. This distance in essential space was determined using

Mahalanobis metrics⁵⁹, which consists in defining Euclidean distances weighted by the variance of every degree of freedom (Eq. 6), which when computed using orthogonal vectors (as those derived from essential mode analysis) can be written as shown in Eq. 7.

$$d_M = [x^T C^{-1} x]^{1/2} \quad (6)$$

where x is the Euclidean distance vector and C is the covariance matrix.

$$d_M = \left[\sum_{i=1}^n \left(\frac{x_i}{\lambda_i^{1/2}} \right)^2 \right]^{1/2} \quad (7)$$

where x_i is the displacement along individual eigenvectors, λ_i stands for the corresponding eigenvalue (in units of distance²), and the sum extends to the space of important essential movements ($n=10$ in this paper).

The minimum Mahalanobis distance between two structures was computed by selecting iteratively small displacement along the different eigenvectors that better reduce the RMSd from original to target structure while keeping Eq. 7 to a minimum. In practice this distance can be assimilated as the easiest deformation pattern to drive a transition assuming a harmonic relationship between displacement from minimum and energy. Note that in the harmonic limit the energy associated to a displacement along normal modes can be easily determined from Mahalanobis metrics as shown in Eq. 8.

$$E = \frac{k_B T}{2} d_M^2 \quad (8)$$

In both cases transition was considered completed when the perturbed “unbound” reached a cutoff defined by eq. 9 (around 0.8 and 1.3 Å for antisense and sense strands respectively) that takes in to account RMSd value of DNA-RNA trajectory with “bound” conformation caused by thermal fluctuations.

$$RMSd(XNA, HYB_{ZI}) \leq RMSd_{avg}(HYB_{MD}, HYB_{ZBI}) - 2RMSd_{sd}(HYB_{MD}, HYB_{ZBI}) \quad (9)$$

where indexes avg and sd stand for average and standard deviation respectively.

Entropies were determined by diagonalization of the mass-weighted covariance matrix obtained during the trajectory using Schlitter model⁶⁰ for all common atoms (eq. 10). Entropy estimates at infinite simulation time were obtained using Harris's extrapolation technique⁶¹ (eq. 11).

$$S \approx 0.5k \sum_i \ln \left(1 + \frac{e^2}{\alpha_i^2} \right) \quad (10)$$

where $\alpha_i = \hbar \omega_i / k_B T$, ω denotes the eigenvalues obtained by diagonalization of the mass-weighted covariance matrix, and the sum extends to all the non-trivial vibrations (all the other symbols have the standard physical meaning).

$$S(t) = S_\infty - \frac{\alpha}{t^\beta} \quad (11)$$

where α and β are fitted parameters and t is the simulation time.

Finally, the elastic force-constants associated to helical deformation at the base-pair step level were determined by inversion of the covariance matrix in helical space, which yields a stiffness matrix (eq. 12) whose diagonal elements provide the stiffness constants associated to pure rotational (twist, roll and tilt) and translational (rise, slide and shift) deformation of the 10 unique base-pair steps. These constants were obtained for all the sequences and averaged then to provide more reliable, environment independent, stiffness parameters for the different steps. A global view of helical deformability can be obtained by further averaging of the different steps and by defining global rotational (K_{rot}) and translational (K_{trans}) deformability indexes (eqs. 13 and 14)

$$\Xi = k_B T \bullet C_h^{-1} = \begin{bmatrix} k_{twist} & k_{tw-ro} & k_{tw-ti} & k_{tw-ri} & k_{tw-sl} & k_{tw-sh} \\ k_{tw-ro} & k_{roll} & k_{ro-ti} & k_{ro-ri} & k_{ro-sl} & k_{ro-sh} \\ k_{tw-ti} & k_{ro-ti} & k_{tilt} & k_{ti-ri} & k_{ti-sl} & k_{ti-sh} \\ k_{tw-ri} & k_{ro-ri} & k_{ti-ri} & k_{rise} & k_{ri-sl} & k_{ri-sh} \\ k_{tw-sl} & k_{ro-sl} & k_{ti-sl} & k_{ri-sl} & k_{slide} & k_{sl-sh} \\ k_{tw-sh} & k_{ro-sh} & k_{ti-sh} & k_{ri-sh} & k_{sl-sh} & k_{shit} \end{bmatrix} \quad (12)$$

where C_h is the covariance matrix in helical space and Ξ is the stiffness matrix whose elements in the diagonal correspond to stiffness constants for rotations (twist, roll and tilt) and translations (rise, slide and shift) of base pairs and those out of the diagonal to coupling between helical parameters.

$$K_{tor} = k_{twist} k_{rol} k_{tilt} \quad (13)$$

$$K_{trans} = k_{rise} k_{shift} k_{slide} \quad (14)$$

The “average” helical parameters were obtained for all the steps and were then averaged to provide more reliable, environment independent values. Standard deviations associated with the averages were obtained by linear propagation of errors (eq. 15).

$$sd_{step} = \sqrt{\sum_i sd_i^2} \quad (15)$$

where *i* stand for all occurrence of a step in the three different sequences simulated.

Standard geometrical and energetic analysis was done using X3DNA⁶², PTRAJ⁴⁷ as well as in house programs. The essential dynamics was performed with the PCAZIP program⁶³ (<http://mmb.pcb.ub.es/software/pcasuite.html> & <http://www.ccpb.ac.uk/events/workshops/previous/analysis/>) and other local programs. Clustering of duplexes according to their essential deformation pattern was done using the R statistical program⁶⁴. The different trajectories are available in compressed format (95% variance threshold) at <http://mmb.pcb.ub.es/antisense/> and can be decompressed with the PCAZIP program (<http://mmb.pcb.ub.es/software/pcasuite.html>)⁶³.

Results and Discussion

General structural properties. MD simulations lead to very stable trajectories (see Figure 1), where the helical structure and the pattern of Watson-Crick hydrogen bonding are well preserved. Comparison of the DNA DD trajectory presented here with a reference 1.2 μs simulation performed in our laboratory⁴⁹ shows that current simulation times are long enough to capture most of the key structural and flexibility features of the equilibrium structure of 12-mer duplexes like those considered here.

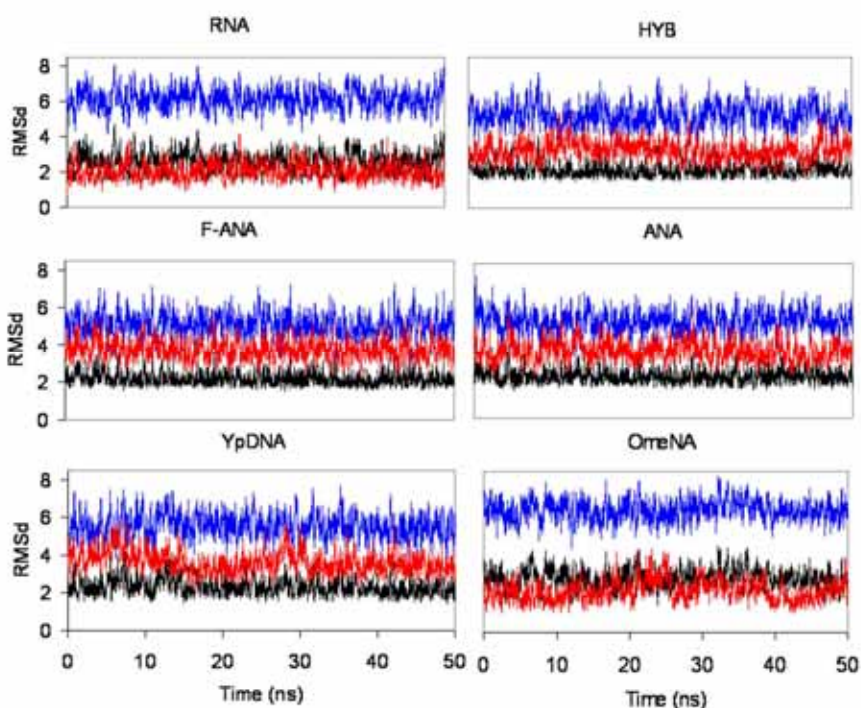


Figure 1. RMSd (for the central 10-mer equivalent (see Methods) backbone atoms; in Å) with respect to DD sequence canonical A (red), B (blue) and NMR-based (1EFS; black) structures of the different duplexes with the DD sequence containing at least 1 RNA strand. Equivalent plots for the other two sequences (Figs. S4 and S5), as well as for the DNA duplexes (Fig. S6) are displayed in supplementary material.

Inspection of the different trajectories (see Figure 1 and S4-S6 in supplementary material) suggest the existence of three structural groups: i) the DNA duplex, which samples B-like conformations close to those experimentally characterized in crystal and solution phases, ii) the DNA·RNA, YpDNA·RNA, ANA·RNA and F-ANA·RNA hybrids, whose conformation resembles the NMR-based structure for DNA·RNA duplexes (i.e., it pertains to the A-family, but with some B-like features; A/B-conformation), and iii) RNA₂ and OrreNA·RNA duplexes, which display a pure A-conformation. The analysis of the helical parameters (Table 2) shows that all the duplexes but DNA₂ pertain to the A-family with average twist angles around 30 degrees, but also reveals differences in roll and rise between pure A-duplexes and A/B-structures. The sequence-dependent variability is very strong in DNA₂, small for pure A-forms and moderate for A/B-structures. It is worth noting that sequence-dependence rules valid for one type of

duplexes are not for the others, showing that not only nucleobase-nucleobase interactions (mostly stacking) but also the nature of the backbone modulate the local geometry of the different sequences.

Table 2. Rotational^a and translational^b helical parameters for the ten unique dinucleotide steps for the different duplexes studied here^{c,d}.

| Step | DNA ₂ | RNA ₂ | DNA | F-ANA | ANA | YpDNA | OmeNA |
|-------|------------------|------------------|----------|----------|----------|----------|----------|
| GC·GC | 34.7±9.6 | 30.2±6.0 | 31.8±7.1 | 29.6±6.6 | 30.8±7.6 | 33.7±6.3 | 30.0±5.8 |
| | -0.5±9.6 | 4.7±8.5 | 3.4±8.7 | 1.8±8.1 | 3.2±8.7 | 0.9±7.7 | 5.4±8.5 |
| | 0.4±7.5 | -0.1±7.1 | 0.9±7.2 | 2.7±7.0 | 3.8±7.4 | 2.3±6.8 | 0.4±7.0 |
| | 3.3±0.5 | 3.3±0.4 | 3.3±0.5 | 3.3±0.4 | 3.4±0.5 | 3.4±0.4 | 3.3±0.4 |
| | -0.3±1.0 | -1.6±0.9 | -1.1±0.9 | -1.3±0.9 | -1.3±1.0 | -1.6±0.9 | -1.8±0.8 |
| | -0.1±1.1 | 0.0±0.9 | 0.1±1.0 | -0.2±0.9 | 0.2±1.0 | 0.2±0.8 | 0.1±0.8 |
| GG·CC | 32.0±7.5 | 29.6±4.8 | 29.2±5.3 | 29.6±4.9 | 27.8±5.5 | 28.9±4.8 | 29.4±4.5 |
| | 4.0±7.4 | 8.3±7.1 | 6.6±7.1 | 5.4±7.3 | 6.8±7.1 | 5.1±6.6 | 8.8±7.0 |
| | 0.1±6.2 | 0.0±5.8 | 2.9±5.9 | 2.9±5.7 | 3.3±5.8 | 2.5±5.5 | 0.3±5.7 |
| | 3.5±0.5 | 3.3±0.4 | 3.4±0.4 | 3.4±0.4 | 3.4±0.4 | 3.3±0.4 | 3.3±0.4 |
| | -1.1±1.0 | -1.9±0.5 | -1.8±0.6 | -1.8±0.5 | -1.7±0.6 | -1.8±0.5 | -2.1±0.4 |
| | 0.0±1.0 | 0.0±0.8 | 0.1±0.8 | 0.3±0.8 | 0.2±0.7 | 0.1±0.7 | 0.0±0.7 |
| GT·AC | 30.6±8.8 | 29.0±5.1 | 29.9±5.7 | 29.4±6.7 | 29.4±5.4 | 30.2±5.5 | 29.8±4.6 |
| | 0.6±8.1 | 8.7±8.4 | 5.1±7.9 | 2.3±8.1 | 3.8±7.4 | 4.5±7.8 | 7.2±7.0 |
| | 0.2±6.0 | 0.1±6.5 | 0.0±6.0 | 4.0±9.9 | 2.8±6.2 | 1.5±6.1 | -0.7±5.8 |
| | 3.3±0.4 | 3.3±0.4 | 3.2±0.4 | 3.3±0.5 | 3.4±0.4 | 3.2±0.4 | 3.3±0.4 |
| | -0.5±0.7 | -1.3±0.7 | -1.0±0.7 | -1.2±1.0 | -1.4±0.8 | -1.1±0.7 | -1.5±0.7 |
| | -0.1±1.1 | 0.0±1.0 | 0.0±1.0 | -0.3±1.2 | -0.2±1.0 | 0.2±1.1 | 0.2±0.8 |
| GA·TC | 36.9±11.7 | 30.3±6.7 | 29.7±8.1 | 30.9±7.3 | 30.0±9.1 | 30.3±7.2 | 29.8±6.4 |
| | 1.8±11.2 | 8.3±12.8 | 7.9±12.7 | 5.8±12.2 | 5.9±12.0 | 5.2±11.2 | 9.2±12.0 |
| | -0.3±9.0 | 0.0±9.6 | 2.8±9.7 | 2.6±9.3 | 2.6±9.5 | 1.7±8.8 | 1.4±8.8 |
| | 3.5±0.6 | 3.2±0.7 | 3.4±0.7 | 3.4±0.7 | 3.4±0.7 | 3.3±0.7 | 3.3±0.6 |
| | -0.3±1.4 | -1.6±0.9 | -1.3±1.1 | -1.5±1.0 | -1.6±1.0 | -1.5±0.9 | -1.9±0.8 |
| | 0.0±1.4 | 0.0±1.4 | 0.1±1.3 | 0.2±1.3 | 0.2±1.4 | 0.1±1.2 | -0.1±1.2 |
| AA·TT | 35.0±6.4 | 28.3±5.1 | 28.7±5.3 | 28.9±5.1 | 28.2±4.9 | 28.0±5.1 | 28.6±4.4 |
| | 2.1±7.0 | 9.0±9.1 | 7.1±8.5 | 5.0±8.4 | 4.8±8.4 | 6.0±8.6 | 8.0±8.3 |
| | -0.1±5.7 | 0.2±7.6 | 1.0±7.3 | 1.5±6.9 | 1.8±7.1 | 2.9±6.7 | -0.3±6.8 |
| | 3.4±0.4 | 3.3±0.5 | 3.3±0.5 | 3.3±0.5 | 3.3±0.5 | 3.3±0.5 | 3.2±0.5 |
| | -0.6±0.8 | -1.6±0.6 | -1.3±0.7 | -1.4±0.7 | -1.6±0.7 | -1.6±0.6 | -1.8±0.6 |
| | -0.1±0.8 | 0.0±1.0 | 0.1±0.9 | 0.2±0.9 | 0.1±0.8 | 0.1±0.9 | -0.1±0.8 |
| AG·CT | 30.2±12.7 | 28.3±6.7 | 27.5±7.8 | 27.7±8.5 | 28.1±7.6 | 27.5±7.2 | 27.8±6.6 |
| | 2.7±11.6 | 10.6±12.0 | 6.9±11.6 | 4.8±11.6 | 2.0±11.4 | 6.8±11.0 | 9.0±10.8 |
| | 0.3±8.7 | 0.1±9.2 | 1.6±9.3 | 3.2±9.6 | 0.5±9.7 | 1.8±9.0 | -1.2±8.7 |
| | 3.3±0.7 | 3.4±0.6 | 3.4±0.7 | 3.4±0.7 | 3.5±0.7 | 3.4±0.6 | 3.3±0.6 |
| | -0.7±1.4 | -1.7±0.8 | -1.6±1.0 | -1.6±0.9 | -1.8±0.9 | -1.6±0.8 | -1.9±0.7 |
| | 0.2±1.5 | 0.0±1.4 | -0.1±1.4 | 0.0±1.6 | -0.2±1.3 | 0.1±1.4 | -0.1±1.4 |
| AT·AT | 30.9±7.8 | 27.6±7.1 | 28.3±7.5 | 26.8±7.0 | 27.5±6.4 | 27.5±6.5 | 27.8±6.8 |
| | -0.1±9.6 | 8.3±12.6 | 4.3±11.4 | 1.6±10.3 | 2.3±8.7 | 2.3±9.5 | 5.8±10.2 |
| | 0.1±7.7 | 0.0±10.3 | 0.7±9.4 | 3.1±9.1 | 2.2±7.8 | 2.2±7.8 | -0.4±8.8 |
| | 3.3±0.6 | 3.2±0.6 | 3.2±0.6 | 3.3±0.6 | 3.3±0.5 | 3.3±0.5 | 3.3±0.5 |
| | -0.9±0.9 | -1.4±1.0 | -1.2±1.0 | -1.4±1.0 | -1.7±1.0 | -1.7±0.9 | -1.9±1.0 |

| | | | | | | | |
|-------|---|--|---|---|--|---|--|
| | <i>0.1±1.3</i> | <i>0.0±1.6</i> | <i>-0.1±1.5</i> | <i>-0.4±1.4</i> | <i>-0.3±1.3</i> | <i>-0.3±1.3</i> | <i>0.2±1.3</i> |
| CG·CG | 26.0±12.6 <i>7.3±8.4</i> <i>0.7±7.4</i> <i>3.1±0.5</i> <i>-0.2±0.8</i> <i>0.1±1.1</i> | 31.0±4.8 <i>14.9±9.7</i> <i>-0.1±6.8</i> <i>3.6±0.6</i> <i>-1.6±0.6</i> <i>0.0±1.0</i> | 31.8±6.9 <i>7.9±9.8</i> <i>-1.1±8.0</i> <i>3.8±0.6</i> <i>-1.4±0.8</i> <i>-0.3±1.0</i> | 31.0±5.4 <i>9.8±9.3</i> <i>2.9±6.8</i> <i>3.5±0.6</i> <i>-1.4±0.7</i> <i>0.4±0.9</i> | 30.1±6.5 <i>8.7±9.6</i> <i>2.7±7.1</i> <i>3.4±0.6</i> <i>-1.2±0.8</i> <i>0.1±1.1</i> | 26.3±8.0 <i>11.2±10.2</i> <i>4.4±7.3</i> <i>3.8±0.7</i> <i>-1.8±1.0</i> <i>0.1±1.2</i> | 31.0±4.7 <i>15.4±9.2</i> <i>-1.1±6.7</i> <i>3.6±0.6</i> <i>-1.8±0.6</i> <i>-0.1±1.0</i> |
| CA·TG | 30.1±12.6 <i>8.5±9.8</i> <i>-0.4±7.3</i> <i>3.3±0.5</i> <i>-0.4±1.0</i> <i>-0.1±1.1</i> | 31.5±4.7 <i>16.7±9.9</i> <i>0.1±6.5</i> <i>3.6±0.5</i> <i>-1.5±0.5</i> <i>0.0±0.9</i> | 29.6±6.4 <i>12.9±10.5</i> <i>3.0±7.4</i> <i>3.4±0.7</i> <i>-1.0±0.9</i> <i>0.2±1.0</i> | 30.6±4.9 <i>10.6±9.0</i> <i>1.2±6.5</i> <i>3.4±0.5</i> <i>-1.2±0.6</i> <i>0.4±0.9</i> | 30.5±5.4 <i>10.2±9.0</i> <i>1.3±6.8</i> <i>3.4±0.5</i> <i>-1.3±0.6</i> <i>0.3±0.9</i> | 29.5±5.3 <i>11.5±9.5</i> <i>2.4±6.5</i> <i>3.5±0.6</i> <i>-1.2±0.6</i> <i>-0.1±1.0</i> | 31.4±4.4 <i>16.1±9.1</i> <i>-0.3±6.3</i> <i>3.5±0.5</i> <i>-1.6±0.5</i> <i>-0.1±0.8</i> |
| TA·TA | 32.1±10.7 <i>8.1±11.6</i> <i>-0.5±7.3</i> <i>3.3±0.5</i> <i>-0.3±1.1</i> <i>-0.2±1.3</i> | 31.3±4.8 <i>17.5±10.2</i> <i>-0.1±7.1</i> <i>3.4±0.5</i> <i>-1.4±0.5</i> <i>0.0±0.8</i> | 29.4±5.1 <i>16.5±9.8</i> <i>1.6±7.3</i> <i>3.4±0.5</i> <i>-1.2±0.5</i> <i>0.3±0.8</i> | 31.3±5.1 <i>11.9±9.1</i> <i>-0.5±7.5</i> <i>3.3±0.5</i> <i>-1.2±0.6</i> <i>0.4±0.7</i> | 30.4±5.3 <i>10.5±9.9</i> <i>0.2±7.4</i> <i>3.2±0.5</i> <i>-1.2±0.6</i> <i>0.3±0.8</i> | 29.8±5.4 <i>14.3±10.1</i> <i>2.0±7.8</i> <i>3.5±0.7</i> <i>-1.2±0.6</i> <i>0.2±0.9</i> | 31.8±4.3 <i>16.4±9.8</i> <i>0.1±6.6</i> <i>3.4±0.5</i> <i>-1.5±0.5</i> <i>0.0±0.7</i> |
| all | 32.1±7.1 <i>2.9±6.6</i> <i>0.0±4.7</i> <i>3.3±0.3</i> <i>-0.5±0.7</i> <i>0.0±0.8</i> | 29.5±3.7 <i>10.1±7.3</i> <i>0.0±4.8</i> <i>3.3±0.4</i> <i>-1.6±0.5</i> <i>0.0±0.7</i> | 29.4±4.4 <i>7.4±7.4</i> <i>1.5±5.2</i> <i>3.4±0.4</i> <i>-1.3±0.6</i> <i>0.0±0.7</i> | 29.3±4.3 <i>5.3±7.0</i> <i>2.5±5.1</i> <i>3.3±0.3</i> <i>-1.4±0.6</i> <i>0.1±0.7</i> | 29.2±4.3 <i>5.3±6.7</i> <i>2.1±5.0</i> <i>3.4±0.4</i> <i>-1.5±0.6</i> <i>0.1±0.7</i> | 29.6±5.1 <i>6.6±7.2</i> <i>2.3±4.7</i> <i>3.4±0.4</i> <i>-1.5±0.5</i> <i>0.1±0.7</i> | 29.5±3.5 <i>9.7±7.0</i> <i>-0.3±4.5</i> <i>3.3±0.3</i> <i>-1.8±0.4</i> <i>0.0±0.6</i> |

^a Twist, roll and tilt in roman and degrees.

^b Rise, slide and shift in italics and Å.

^c For non-homopolymeric duplexes it is labelled by the antisense strand.

^d Standard deviations were obtained by linear propagation of deviations in samplings and structures (see Methods).

The width of the minor groove reflects very clearly the different nature of the three conformational groups mentioned above. The DNA duplex shows a spread distribution of minor groove widths reflecting the intrinsic flexibility of the duplex and the dependence of conformation on sequence (see Figure 2). Averaging data for all steps and sequences yields a minor groove width equal to 6.2 Å. The pure A-form duplexes (RNA₂ and OmeNA·RNA) display wide (11.2 Å), sequence-independent rigid minor grooves. Finally, the A/B-hybrids show intermediate minor grooves ranging from 9.0 Å (ANA·RNA) to 9.9 Å (DNA·RNA). The rigidity of these minor grooves is also intermediate compared to pure A and B-forms.

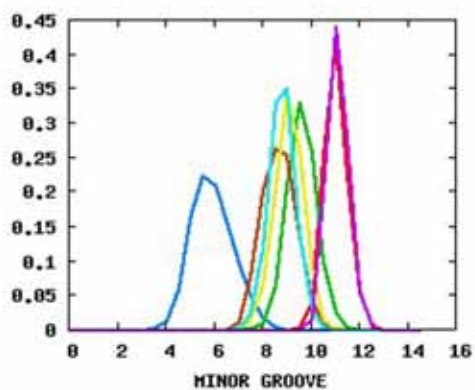


Figure 2. Population (in fractions of 1) of minor groove widths (in Å) for the different duplexes considered here. Dark blue: DNA₂, red: RNA₂, green: DNA·RNA, brown: F-ANA·RNA, light-blue: ANA·RNA, yellow: YpDNA·RNA, magenta: OmeNA·RNA. Values derived by pooling the results obtained for the three sequences.

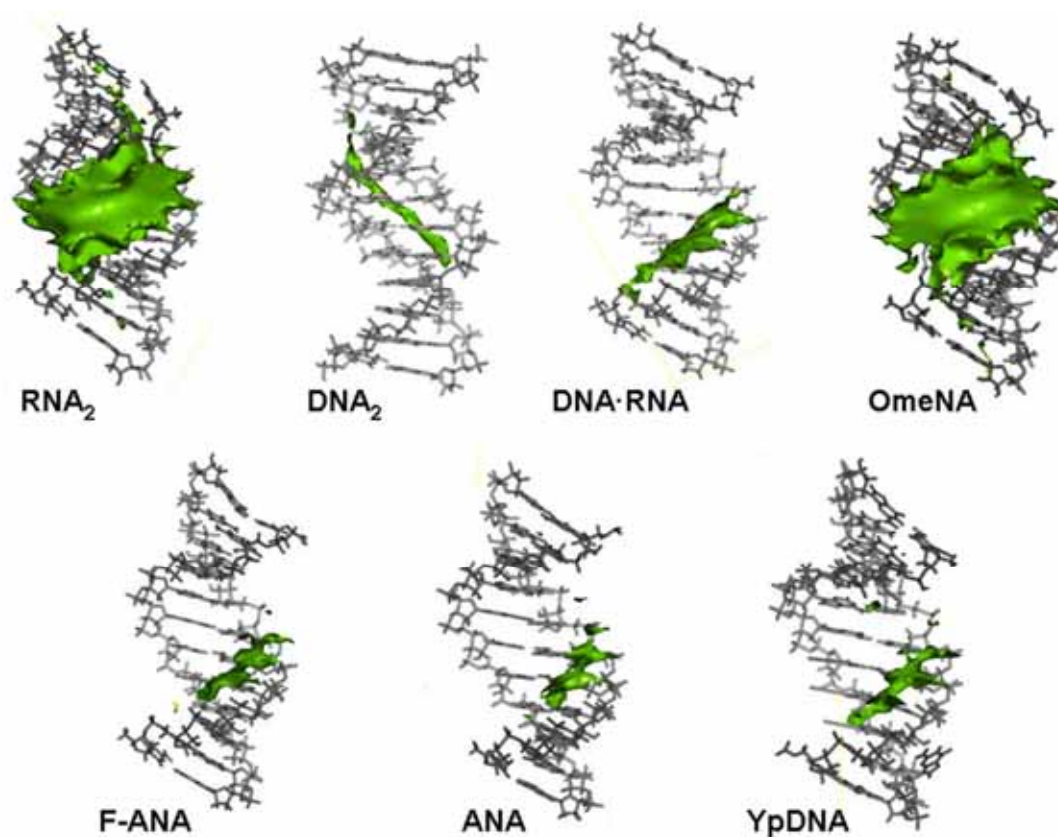


Figure 3. cMIP isocontours (-3 kcal/mol) for the interaction with a Na⁺ probe for the different average duplexes with the DD sequence.

As expected, the different geometry of the grooves changes drastically the interaction profile of the duplexes (see Figure 3; for the sake of simplicity this analysis is shown only for the DD sequence). Thus, pure A-form duplexes mainly interact with cationic groups along the major groove, while DNA₂ displays the classical profile with strong propensity for binding cations along the bottom of the narrow minor groove. All the A/B-duplexes display a common pattern with well defined regions of favorable interaction located asymmetrically in the major groove around the phosphate groups of the sense strand. Note then that despite the greater similarity with the A-form, the A/B-like hybrids show a quite distinct interaction pattern in the grooves, which also differs from that of pure B-DNA duplexes.

Global flexibility and strand asymmetry. Intra-molecular entropy analysis for the duplexes with the DD sequence (those for which longer trajectories are available) demonstrate that DNA₂ is more flexible than any of the other duplexes studied here, while pure A-forms, particularly the OmeNA·RNA duplex, are the most rigid ones (see Figure 4) All the A/B-hybrids display entropy values intermediate between those of DNA₂ and RNA₂, with the arabino derivatives closer to RNA₂ and the DNA·RNA and YpDNA·RNA duplexes closer to DNA₂.

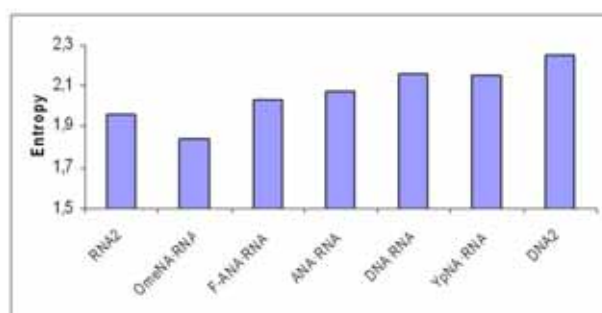


Figure 4. Entropies (in kcal/mol K) extrapolated at infinite simulation time for trajectories performed for duplexes with the DD sequence (see methods for details).

The higher global flexibility of DNA₂ relative to hybrids and these relative to pure A-duplexes is not homogeneous for all the essential deformations, as noted by the force constants associated to the essential deformation modes. Thus, for the very first modes pure A-duplexes can be very soft, but the situation is reversed after c.a. the fifth mode (Figure 5), thus confirming⁶⁵ that A-duplexes are guided by a simple dynamics where only a very small number of modes contribute to the flexibility, while A/B-

hybrids and specially DNA₂ show a more complex deformability pattern, which involves a larger number of essential movements. Interestingly, the two strands of A/B-hybrids show a quite remarkable difference in the deformability scheme, as can be seen in the stiffness constants associated to the first deformation modes of the two independent strands (Figure 6). Thus, sense strands behave close to those of pure A-forms (RNA₂ and OmeNA·RNA), while the antisense strands (of A/B-hybrids) are closer to those of DNA₂ (Figure 6). This makes clear that the duplex structure does not make uniform the two strands in terms of flexibility.

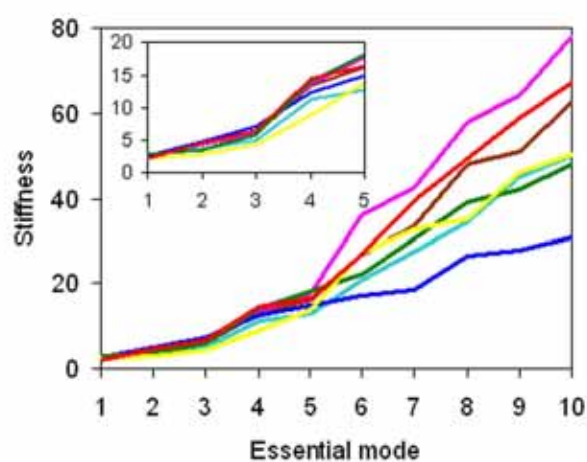


Figure 5. Force constants (in cal/mol Å²) assigned to the first 10 essential movements (five in the insert) of the duplexes with the DD sequence (very similar plots were obtained for the other two sequences). Dark blue: DNA₂, red: RNA₂, green: DNA·RNA, brown: F-ANA·RNA, light-blue: ANA·RNA, yellow: YpDNA·RNA, magenta: OmeNA·RNA.

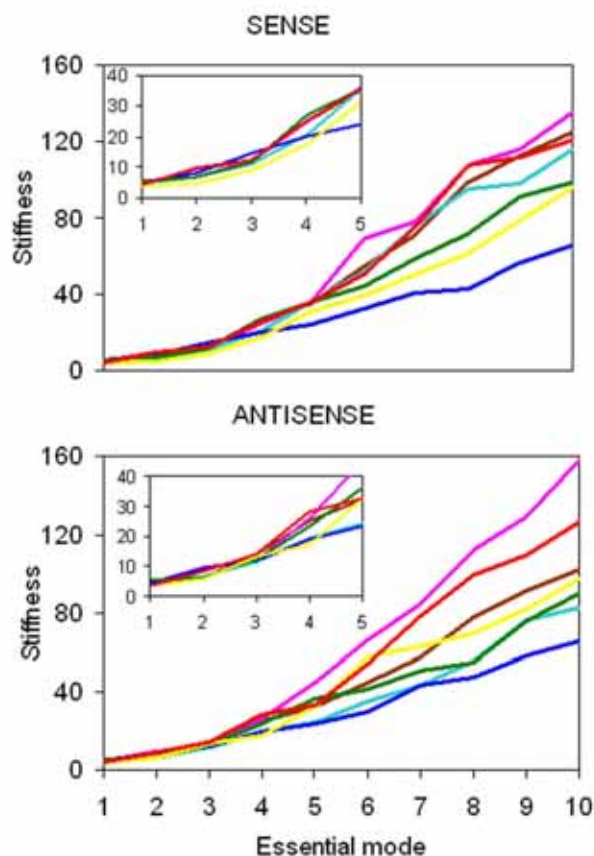


Figure 6. Force constants (in cal/mol \AA^2) of the first 10 essential movements (five in the insert) of the sense (RNA) and antisense (XNA) strands of duplexes with the DD sequence (very similar plots were obtained for the other two sequences). Dark blue: DNA₂, red: RNA₂, green: DNA·RNA, brown: F-ANA·RNA, light-blue: ANA·RNA, yellow: YpDNA·RNA, magenta: OmeNA·RNA.

Comparison of the nature of the essential deformation movements (see Methods) shows an overall good similarity between the different duplexes (Table S2). However, a careful clustering of similarity data shows the existence of two main groups: one corresponds to the three DNA₂ considered here, while the other contains all the other duplexes (Figure 7). This latter cluster contains two other well differentiated families: i) the pure A-duplexes and ii) the A/B-hybrids. Note that the differences between the essential dynamics related to sequences might be sizable, but do not alter the assignment of a duplex into a given cluster, thus confirming the consistency of the deformation behavior of the duplexes and the possibility to classify duplexes based on the deformability pattern. When the similarity analysis is performed for the individual strands, the asymmetry between them becomes clear (see Table S2 and

Figure 7). Three clusters appear: i) DNA, ii) all the antisense strands of A/B hybrids and iii) RNA₂, OmeNA·RNA and all the sense strands. This demonstrates that irrespective of the nature of the hybrid all sense strands maintain an internal deformability pattern that resembles that of a RNA strand in an A-duplex, which differs from the deformability pattern of the antisense strands.

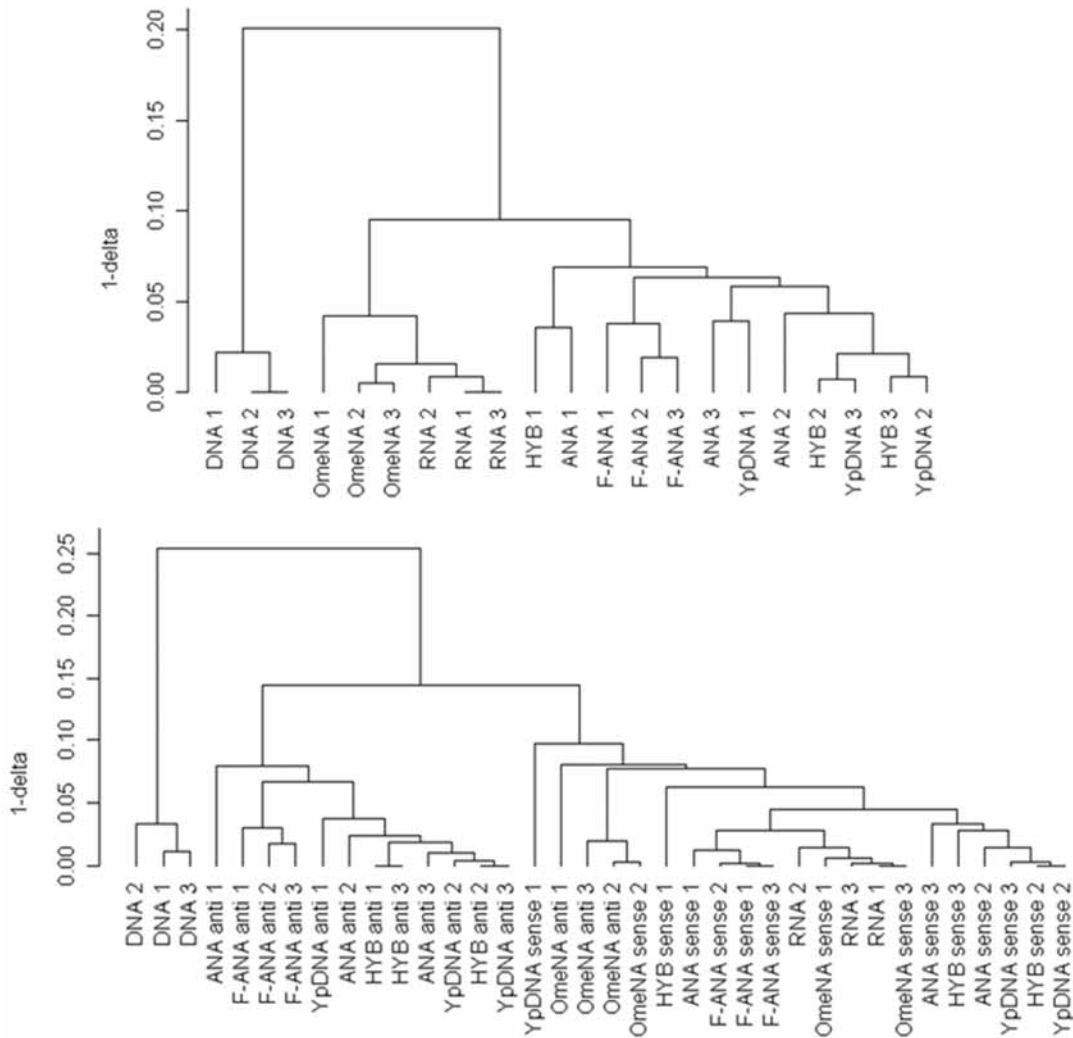


Figure 7. Clustering of the different molecules (and sequences) based on similarity measures (results correspond to $1-\delta$ distances (eq. 5), though similar values were obtained using the κ index). TOP: considering the simultaneous dynamics of both strands. BOTTOM: considering the dynamics of each strand separately. Similarities were determined using the 10 central steps. 1, 2 and 3 stand for the three sequences considered here: $d/r(\text{CGCGAATTCGCG})_2$, $d/r(\text{CATAGGCCTATG})_2$ and $d/r(\text{CACAGATCTGTG})_2$.

Stiffness analysis associated to helical deformations provides information on the flexibility of the different duplexes at the base-pair step level. The results show that DNA₂ is more flexible than all the other duplexes both in terms of rotations and translations of base-pair steps (Table 3), which is mainly due to twist and shift deformations, since no relevant stiffness differences are found for the other helical deformations. The OmeNA·RNA duplex is the stiffest one at the base-pair step level, followed by RNA₂, whose rigidity at this level is rather similar to that obtained for A/B-hybrids. The DNA₂ has the largest sequence-dependent variability of helical stiffness, and pure A-duplexes (RNA₂ and OmeNA·RNA) display the smallest, while the A/B-hybrids are placed in between, but generally closer to the pure A-family than to the DNA₂ (Table 3).

Table 3. Rotational^a and translational^b helical force constants of the ten unique dinucleotide steps for the different duplexes^c.

| Step | DNA ₂ | RNA ₂ | DNA | F-ANA | ANA | YpNA | OmeNA |
|-------|------------------|------------------|------|-------|------|------|-------|
| GC·GC | 31.1 | 59.3 | 47.3 | 55.1 | 47.1 | 57.5 | 66.4 |
| | 27.1 | 27.1 | 26.8 | 29.2 | 26.0 | 36.0 | 27.0 |
| | 37.9 | 36.7 | 36.0 | 38.8 | 35.5 | 43.6 | 39.3 |
| | 9.7 | 12.7 | 11.1 | 11.5 | 10.0 | 11.6 | 13.0 |
| | 3.1 | 3.2 | 3.0 | 3.4 | 3.0 | 3.6 | 3.8 |
| | 1.7 | 2.3 | 2.1 | 2.3 | 2.0 | 3.0 | 2.8 |
| GG·CC | 32.3 | 61.8 | 58.0 | 61.8 | 52.3 | 62.8 | 70.8 |
| | 22.8 | 25.4 | 24.8 | 23.7 | 25.2 | 30.9 | 25.3 |
| | 43.4 | 49.7 | 45.1 | 48.0 | 46.9 | 51.5 | 50.7 |
| | 7.8 | 9.5 | 8.9 | 8.5 | 9.0 | 8.8 | 9.9 |
| | 1.7 | 5.4 | 4.6 | 5.3 | 4.9 | 5.2 | 7.0 |
| | 1.5 | 2.3 | 2.2 | 2.3 | 2.3 | 2.8 | 2.8 |
| GT·AC | 21.4 | 50.9 | 41.2 | 63.2 | 53.0 | 44.9 | 64.5 |
| | 23.1 | 20.0 | 24.6 | 28.5 | 24.3 | 22.9 | 27.8 |
| | 36.6 | 29.2 | 36.9 | 41.4 | 34.5 | 34.5 | 42.2 |
| | 8.1 | 10.4 | 8.4 | 9.6 | 9.5 | 10.6 | 10.0 |
| | 2.5 | 3.1 | 3.3 | 3.7 | 2.8 | 3.4 | 3.5 |
| | 1.1 | 1.1 | 1.2 | 1.5 | 1.2 | 1.0 | 1.9 |
| GA·TC | 33.8 | 63.9 | 50.9 | 57.7 | 47.3 | 59.7 | 67.2 |
| | 21.8 | 19.8 | 19.4 | 20.6 | 20.8 | 24.9 | 19.9 |
| | 41.9 | 40.2 | 37.0 | 39.9 | 39.1 | 42.2 | 43.3 |
| | 8.6 | 8.9 | 8.1 | 8.0 | 8.0 | 8.1 | 9.2 |
| | 2.1 | 4.0 | 2.9 | 3.6 | 3.5 | 4.0 | 5.1 |
| | 1.6 | 1.5 | 1.5 | 1.8 | 1.6 | 1.9 | 1.8 |
| AA·TT | 49.6 | 57.4 | 54.0 | 61.1 | 65.2 | 60.2 | 70.0 |
| | 25.4 | 19.2 | 21.8 | 22.6 | 21.5 | 22.1 | 21.4 |
| | 43.5 | 30.1 | 32.9 | 35.7 | 33.1 | 36.1 | 38.4 |

| | | | | | | | |
|--|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| | <i>8.1</i> | <i>8.4</i> | <i>7.9</i> | <i>7.8</i> | <i>7.0</i> | <i>8.1</i> | <i>8.8</i> |
| | <i>3.2</i> | <i>3.8</i> | <i>2.9</i> | <i>3.6</i> | <i>3.6</i> | <i>3.9</i> | <i>4.6</i> |
| | <i>2.2</i> | <i>1.3</i> | <i>1.7</i> | <i>1.8</i> | <i>2.0</i> | <i>1.9</i> | <i>2.0</i> |
| AG·CT | 24.9 | 62.9 | 56.3 | 63.3 | 61.6 | 62.1 | 69.7 |
| | 20.2 | 19.6 | 21.1 | 21.7 | 21.6 | 23.5 | 22.8 |
| | 39.9 | 34.5 | 36.6 | 37.9 | 35.7 | 38.1 | 41.6 |
| | <i>7.4</i> | <i>7.8</i> | <i>6.9</i> | <i>6.9</i> | <i>6.5</i> | <i>7.4</i> | <i>8.1</i> |
| | <i>1.7</i> | <i>4.5</i> | <i>3.7</i> | <i>4.5</i> | <i>4.8</i> | <i>4.6</i> | <i>6.0</i> |
| | <i>1.5</i> | <i>1.3</i> | <i>1.5</i> | <i>1.7</i> | <i>1.8</i> | <i>1.6</i> | <i>1.6</i> |
| AT·AT | 47.8 | 54.9 | 51.8 | 61.6 | 58.4 | 52.6 | 62.3 |
| | 29.2 | 19.5 | 22.8 | 25.8 | 26.5 | 24.9 | 26.5 |
| | 40.4 | 22.9 | 31.7 | 34.3 | 33.8 | 34.2 | 34.8 |
| | <i>8.5</i> | <i>10.3</i> | <i>9.5</i> | <i>10.0</i> | <i>9.1</i> | <i>9.3</i> | <i>10.8</i> |
| | <i>3.9</i> | <i>3.3</i> | <i>3.4</i> | <i>3.9</i> | <i>3.0</i> | <i>3.4</i> | <i>3.5</i> |
| | <i>1.6</i> | <i>1.0</i> | <i>1.1</i> | <i>1.3</i> | <i>1.2</i> | <i>1.2</i> | <i>1.5</i> |
| CG·CG | 11.8 | 59.1 | 45.7 | 55.9 | 42.4 | 38.4 | 62.4 |
| | 19.5 | 14.5 | 14.6 | 14.7 | 15.1 | 15.3 | 16.1 |
| | 27.7 | 29.3 | 25.6 | 29.8 | 26.4 | 29.5 | 31.0 |
| | <i>7.1</i> | <i>3.9</i> | <i>3.7</i> | <i>3.7</i> | <i>3.9</i> | <i>3.4</i> | <i>4.2</i> |
| | <i>2.4</i> | <i>3.6</i> | <i>3.3</i> | <i>3.6</i> | <i>2.9</i> | <i>2.8</i> | <i>4.2</i> |
| | <i>1.2</i> | <i>1.4</i> | <i>1.5</i> | <i>1.7</i> | <i>1.4</i> | <i>1.4</i> | <i>1.4</i> |
| CA·TG | 13.5 | 59.0 | 39.8 | 56.0 | 51.2 | 48.3 | 66.7 |
| | 15.9 | 15.3 | 14.4 | 17.5 | 16.8 | 16.8 | 17.4 |
| | 25.1 | 28.9 | 24.7 | 29.0 | 27.1 | 29.3 | 30.8 |
| | <i>5.5</i> | <i>4.7</i> | <i>3.7</i> | <i>4.8</i> | <i>4.6</i> | <i>4.5</i> | <i>5.2</i> |
| | <i>1.3</i> | <i>4.7</i> | <i>2.3</i> | <i>3.3</i> | <i>3.7</i> | <i>3.8</i> | <i>5.1</i> |
| | <i>1.1</i> | <i>1.6</i> | <i>1.4</i> | <i>1.9</i> | <i>1.8</i> | <i>1.6</i> | <i>2.0</i> |
| TA·TA | 18.9 | 55.2 | 50.5 | 53.8 | 50.1 | 46.2 | 67.0 |
| | 15.2 | 13.6 | 15.5 | 15.9 | 14.8 | 14.2 | 15.3 |
| | 23.3 | 24.4 | 24.3 | 25.7 | 24.0 | 22.6 | 28.6 |
| | <i>6.4</i> | <i>5.1</i> | <i>5.3</i> | <i>5.2</i> | <i>5.2</i> | <i>3.7</i> | <i>5.5</i> |
| | <i>1.5</i> | <i>4.9</i> | <i>4.3</i> | <i>4.5</i> | <i>4.1</i> | <i>4.0</i> | <i>5.7</i> |
| | <i>0.9</i> | <i>2.1</i> | <i>2.3</i> | <i>2.7</i> | <i>2.3</i> | <i>1.8</i> | <i>2.7</i> |
| Average | 28.5 | 58.4 | 49.6 | 59.0 | 52.9 | 52.3 | 66.7 |
| | 22.0 | 19.4 | 20.6 | 22.0 | 21.3 | 23.2 | 22.0 |
| | 36.0 | 32.6 | 33.1 | 36.1 | 33.6 | 36.2 | 38.1 |
| | <i>7.7</i> | <i>8.2</i> | <i>7.4</i> | <i>7.6</i> | <i>7.3</i> | <i>7.6</i> | <i>8.5</i> |
| | <i>2.3</i> | <i>4.1</i> | <i>3.4</i> | <i>3.9</i> | <i>3.6</i> | <i>3.9</i> | <i>4.9</i> |
| | <i>1.4</i> | <i>1.6</i> | <i>1.7</i> | <i>1.9</i> | <i>1.8</i> | <i>1.8</i> | <i>2.1</i> |
| $\Xi_{\text{rot}}(\text{av})^{\text{d}}$ | 23 10 ³ | 37 10 ³ | 34 10 ³ | 46 10 ³ | 38 10 ³ | 45 10 ³ | 58 10 ³ |
| $\Xi_{\text{trans}}(\text{av})$ | 26 | 53 | 41 | 57 | 47 | 53 | 84 |

^a Twist, roll and tilt in roman and cal/mol·deg²

^b Rise, slide and shift in italics and kcal/mol·Å²

^c For non-homopolymeric duplexes it is labelled by the antisense strand

^d Values in the last rows correspond to average global translational and rotational constants (in kcal³/mol³·Å⁶ and cal³/mol³·deg⁶).

Interestingly, steps which are very stiff for one type of helical deformation might be soft for others duplexes (for example the CG step which is the easiest point for unwinding in DNA₂ is quite stiff for the same deformation in RNA₂), suggesting that concepts such as “step deformability or rigidity” might be used with caution, and that in contrast to with general belief, alterations in the backbone can affect in different ways distinct base steps, changing the sequence-rules for stiffness (Table 3). These findings, in conjunction with the geometrical data for steps (see above and Table 2), demonstrate that there is a subtle and complex coupling between sequence-dependent helical geometry/deformability and the backbone, which must not be ignored.

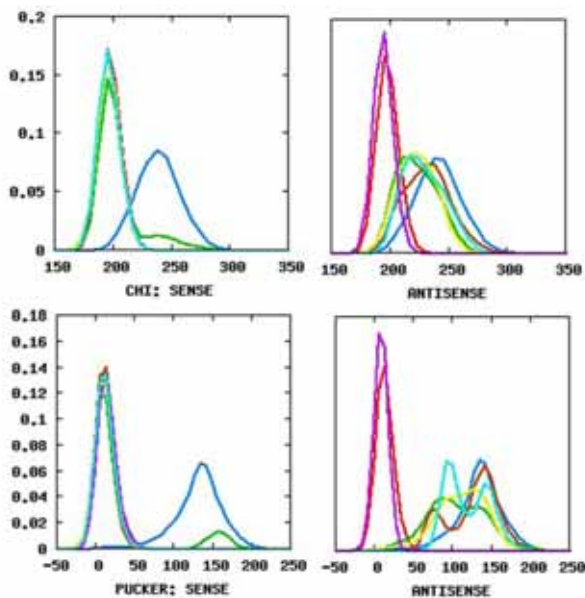


Figure 8. Population (in fractions of 1) of glycosidic torsion (χ in degrees) and phase angle (pucker in degrees) for the sense and antisense strands of the different duplexes considered here. Blue: DNA₂, red: RNA₂, green: DNA·RNA, brown: F-ANA·RNA, light-blue: ANA·RNA, yellow: YpDNA·RNA, magenta: OmeNA·RNA. Values shown are obtained by pooling results from the three sequences.

The flexibility of the backbone is mostly dominated at the microscopic level by: i) rotations around glycosidic angle (χ), ii) sugar puckering, iii) the concerted α/γ rotation and iv) ϵ/ζ coupled rotations. DNA₂ shows the largest flexibility in terms of χ torsions (Figure 8), which correlates with the largest flexibility in terms of twist. The pure A-duplexes show a very sharp χ -distribution (Figure 8), while the

situation for the A/B-hybrids depends on the strand (see below) with the sense strand identical to pure RNA and the antisense one closer to DNA. Sugar puckering widely oscillates in the South to South-East region for DNA₂, while it is fixed in the North conformation for A-duplexes. Again the two strands of A/B-duplexes show clear distinct trends: sense sugars fixed in the North conformation (a residual population of South puckering is found in DNA·RNA hybrids), while wide distributions of phase angles in the South-South East regions occur in the antisense strand.

The use of the new parmbsc0 force-field allowed us to evaluate the concerted α/γ and ϵ/ζ coupled rotations, which are crucial to understand backbone flexibility. As previously reported⁴⁰, DNA₂ displays reversible α/γ rotations in the multi-ns time scale, with a minor population (< 1 %) of all α/γ pairs in non-canonical regions (Table 4). The population of non-canonical conformations becomes nearly zero for pure A-duplexes, but increases significantly for all the A/B-hybrids due to the tendency to display unusual α/γ conformers in the antisense strand, which seems to agree with the fact that unusual α/γ conformers are found in the DNA·RNA·RNase H complex in hybrid regions close to the enzyme active site³⁷. The largest flexibility with respect to ϵ/ζ rotation is found for DNA₂, due to the well known (B_I/B_{II}) transition (around 20% of all ϵ/ζ torsions are in the non-canonical conformation). Such transitions do not exist for pure A-forms and are less prevalent for the hybrids, where the transitions are located in the antisense strand.

Table 4. Percentage of non-canonical α/γ and ϵ/ζ torsions in the different duplexes^a.

| | | | | | | | |
|------------------|------------------|------------------|------|-------|------|-------|-------|
| α/γ | DNA ₂ | RNA ₂ | HYB | F-ANA | ANA | YpDNA | OmeNA |
| Duplex | 0.61 | 0.06 | 2.14 | 7.26 | 1.31 | 6.24 | 0.02 |
| Sense | -- | -- | 0.25 | 0.64 | 0.94 | 12.39 | 0.04 |
| Antisense | -- | -- | 4.04 | 13.87 | 1.67 | 0.1 | 0.01 |
| ϵ/ζ | DNA | RNA | HYB | F-ANA | ANA | YpDNA | OmeNA |
| Duplex | 20.14 | 0.01 | 2.27 | 0.07 | 4.33 | 0.98 | 0.00 |
| Sense | -- | -- | 1.72 | 0.04 | 0.03 | 0.88 | 0.00 |
| Antisense | -- | -- | 2.82 | 0.10 | 8.63 | 1.08 | 0.00 |

^a Due to the slow convergence of α/γ torsion, values are shown only for the duplexes with the DD sequence.

In summary, our analysis shows that the flexibility of A/B-hybrids is larger than that of pure A-form duplexes and smaller than that of DNA₂. The backbone of all A/B-hybrids has a unique deformation pattern, with a strong asymmetry between sense and antisense strands and a good memory of each strand of its behavior in the corresponding homopolymer.

Biologically relevant deformations. One of the most distinctive features of the A/B-hybrids is the unique structure of the grooves. However, considering the flexibility of nucleic acids, this structural feature alone is not expected to be enough as to discriminate between RNase H substrates and non-substrates^{14,15}. Therefore, the enzyme should exploit additional discriminative properties. Our results suggest that deformability can be a key differential feature, since hybrids susceptible to RNase H have a deformability pattern clearly different to that of non-substrate hybrids. Thus, the higher flexibility of A/B-hybrids relative to A-forms should facilitate the binding to the enzyme in a suitable conformation and the strong strand asymmetry should help the enzyme recognizes and cleaves the sense strand, keeping intact the antisense one.

The deformation of pure A-type hybrids or RNA₂ implies a pathway in the essential space associated with high energy requirements (Table 5), due mostly to the difficulty in deforming the very rigid antisense (and sense) strand (see Figure 9). As a result, a productive binding of A-type duplexes to the enzyme is energetically very demanding. On the contrary, deformation of A/B-hybrids is achieved without a significant energy cost. Finally, it is worth noting that the productive binding of DNA₂ is mostly handicapped by the energy cost of deforming the sense strand from a pure B to a pure A-conformation, in other words the cost of changing its general shape to the general A-form expected by the enzyme. It is worth noting that these findings are not dependent on the sequence and should be considered “universal” for a given duplex type.

Table 5. Minimum Mahalanobis distances in essential deformation space between the structure of the antisense (roman), sense (italics) strands in relaxed forms and in the RNase H-bound conformations (see Methods)^{a,b,c}.

| Sequence | Seq. 1 | Seq. 2 | Seq. 3 | Av distance | Energy |
|----------------|--------|--------|--------|-------------|--------|
| DNA(a-sense) | 3.41 | 1.96 | 2.54 | 2.69 | 2.17 |
| DNA(sense) | 7.59 | 6.04 | 7.26 | 6.96 | 14.53 |
| RNA(a-sense) | 7.82 | 8.60 | 8.34 | 8.25 | 20.3 |
| RNA(sense) | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| HYB(a-sense) | 1.92 | 1.35 | 1.49 | 1.59 | 0.75 |
| HYB(sense) | 2.20 | 1.83 | 1.12 | 1.72 | 0.89 |
| F-ANA(a-sense) | 1.68 | 2.16 | 1.96 | 1.90 | 1.1 |
| F-ANA(sense) | 2.09 | 2.25 | 1.93 | 2.09 | 1.31 |
| ANA(a-sense) | 1.80 | 1.34 | 1.62 | 1.59 | 0.75 |
| ANA(sense) | 2.53 | 3.80 | 2.20 | 2.84 | 2.42 |
| YpDNA(a-sense) | 1.55 | 1.85 | 1.62 | 1.67 | 0.83 |
| YpDNA(sense) | 2.03 | 2.17 | 1.75 | 1.98 | 1.18 |
| OmeNA(a-sense) | 8.74 | 9.84 | 7.23 | 8.60 | 22.1 |
| OmeNA(sense) | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

^a The last two columns correspond to the average distance values and the associated harmonic deformation energy (in kcal/mol)

^b All the values were determined considering the central 6-mer portion (i.e., the region in contact with the protein)

^c 1, 2 and 3 stand for the three sequences considered here: $d/r(\text{CGCGAATTCGCG})_2$, $d/r(\text{CATAGGCCTATG})_2$ and $d/r(\text{CACAGATCTGTG})_2$, respectively.

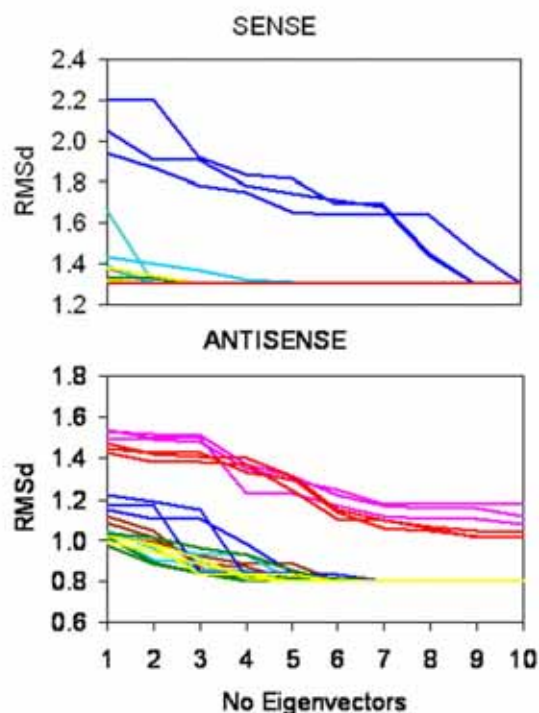


Figure 9. RMSd (in Å) of the sense and antisense strands to the “RNase H-bound” structure reached when movements are allowed along an increasing (from 1 to 10) number of eigenvectors (i.e. selecting iteratively small distances over eigenvectors to reduce as much as possible RMSd while keeping

Mahalanobis distance and associated energy to a minimum; see Eqs 7 and 8; see Methods). Using values of RMSd produced by thermal fluctuations for DNA·RNA trajectory as reference, a transition is considered done when RMSd(sense) $<$ 1.3 Å and RMSd(antisense) $<$ 0.8 Å (see Methods). The different lines with the same colour correspond to the 3 sequences considered. Blue: DNA₂, red: RNA₂, green: DNA·RNA, brown: F-ANA·RNA, light-blue: ANA·RNA, yellow: YpDNA·RNA, magenta: OmeNA·RNA.

In summary, our results point towards a double sieve: the general shape, which precludes the binding of B-like structure and the flexibility pattern which avoids the productive binding of rigid pure A-forms, whose general structure is not so far from that expected by the enzyme. Our results not only rationalize the apparently paradoxical behaviour of the enzyme, but also outline a protocol to recognize RNase H substrates. This implies 20-50 ns MD simulations from which geometry and flexibility will be determined. Based on this any hybrid can be grouped in those susceptible to the enzyme and those which can not be recognized in a productive way, opening then the possibility for more efficient design of anti-sense drugs.

Acknowledgments. We are indebted to Prof. C.A.Laughton for suggestions and comments. This work has been supported by Spanish Ministry of Science BIO2006-01602) and the National Institute of Bioinformatics (Structural Bioinformatics Node). Calculations were performed at the *MareNostrum* supercomputer at the Barcelona Supercomputer Center.

Supporting Information Available. Complete ref. 47, figures S1-S6 and table S1, S2. This material is available free of charge via the Internet at <http://pubs.acs.org>

- (1) Wang, A. H.; Fujii, S.; van Boom, J. H.; van der Marel, G. A.; van Boeckel, S. A.; Rich, A. *Nature*. **1982**, *299*, 601.
- (2) Katahira, M.; Lee, S. J.; Kobayashi, Y.; Sujeta, H.; Kyogoku, Y.; Iwai, S.; Ohtsuka, E.; Benevides, J. M.; Thomas, G. J. *J. Am. Chem. Soc.* **1990**, *112*, 4508.
- (3) (a) Egli, M.; Usman, N.; Zhang, S. G.; Rich, A. *Proc. Natl. Acad. Sci. USA*. **1992**, *89*, 534. (b) Egli, M.; Usman, N.; Rich, A. *Biochemistry*. **1993**, *32*, 3221

- (4) Horton, N. C.; Finzel, B. C. *J. Mol. Biol.* **1996**, *264*, 521.
- (5) Conn, G. L.; Brown, T.; Leonard, G. A. *Nucl. Acids Res.* **1999**, *27*, 555.
- (6) Xiong, Y.; Sundaralingam, M. *Nucl. Acids Res.* **2000**, *28*, 2171.
- (7) (a) Chou, S. H.; Flynn, P.; Reid, B. *Biochemistry.* **1989**, *28*, 2422. (b) Chou, S. H.; Flynn, P.; Reid, B. *Biochemistry.* **1989**, *28*, 2435.
- (8) Hall, K. B.; McLaughlin, L. W. *Biochemistry.* **1991**, *30*, 10606.
- (9) Fedoroff, O. Y.; Salazar, M.; Reid, B. R. *J. Mol. Biol.* **1993**, *233*, 509.
- (10) (a) Lane, A. N.; Ebel, S.; Brown, T. *Eur. J. Biochem.* **1993**, *215*, 297. (b) Gyi, J. I.; Conn, G. L.; Lane, A. N.; Brown, T. *Biochemistry.* **1996**, *35*, 12538. (c) Gyi, J. I.; Lane, A. N.; Conn, G. L.; Brown, T. *Biochemistry.* **1998**, *37*, 73. (d) Gyi, J. I.; Gao, D.; Conn, G. L.; Trent, J. O.; Brown, T.; Lane, A. N. *Nucl. Acids Res.* **2003**, *31*, 2683.
- (11) (a) Salazar, M.; Fedoroff, O. Y.; Miller, J. M.; Ribeiro, N. S.; Reid, B. R. *Biochemistry.* **1993**, *32*, 4297. (b) Fedoroff, O. Y.; Ge, Y.; Reid, B. R. *J. Mol. Biol.* **1997**, *269*, 225.
- (12) (a) Gao, X.; Jeffs, P. W. *J. Biomol. NMR.* **1994**, *4*, 367. (b) Cross, C. W.; Rice, J. S.; Gao, X. *Biochemistry.* **1997**, *36*, 4096.
- (13) (a) Gonzalez, C.; Stec, W.; Kobylanska, A.; Hogrefe, R. I.; Reynolds, M.; James, T. L. *Biochemistry.* **1994**, *33*, 1062. (b) Gonzalez, C.; Stec, W.; Reynolds, M.; James, T. L. *Biochemistry.* **1995**, *34*, 4969.
- (14) Nishizaki, T.; Iwai, S.; Ohkubo, T.; Kojima, C.; Nakamura, H.; Kyogoku, Y.; Ohtsuka, E. *Biochemistry.* **1996**, *35*, 4016.
- (15) Szyperski, T.; Gotte, M.; Billeter, M.; Perola, E.; Cellai, L.; Heumann, H.; Wuthrich, K. *J. Biomol. NMR.* **1999**, *13*, 343.
- (16) Hantz, E.; Larue, V.; Ladam, P.; Le Moyec, L.; Gouyette, C.; Huynh Dinh, T. *Int. J. Biol. Macromol.* **2001**, *28*, 273.
- (17) Noy, A.; Pérez, A.; Márquez, M.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **2005**, *127*, 4910
- (18) Alberts, B.; Bray, D.; Lewis, J.; Raff, M.; Roberts, K.; Watson, J. D. *Molecular Biology of the Cell.* 3rd Edition. Garland Publishing Inc. NY, 1994.
- (19) Peliska, J. A.; Benkovic, S. J. *Science* **1992**, *258*, 1112
- (20) Rumbaugh, J. A.; Murante, R. S.; Shi, S.; Bambara, R. A. *J. Biol. Chem.* **1997**, *272*, 22591
- (21) Crouch, R. J.; Toulmé, J. J. (Eds.). INSERM: Paris **1998**
- (22) Han, G. W.; Kopka, M. L.; Cascio, D.; Grzeskowiak, K.; Dickerson, R. E. *J. Mol. Biol.* **1997**, *269*, 811.
- (23) Lima, W. F.; Crooke, S. T. *Biochemistry.* **1997**, *36*, 390.
- (24) Roberts, W. R., Crothers, D. M. *Science.* **1992**, *258*, 1463.
- (25) Oda, Y.; Iwai, S.; Ohtsuka, E.; Ishikawa, M.; Ikehara, M.; Nakamura, H. *Nucl. Acids Res.* **1993**, *21*, 4690.

- (26) Ho, S. P.; Britton, D. H. O.; Stone, B. A.; Behrens, D. L.; Leffet, L. M.; Hobbs, F. W.; Miller, J. A.; Trainor G. L. *Nucl. Acids Res.* **1993**, *21*, 4690.
- (27) Stein, H.; Hausen, P. *Science* **1969**, *166*, 393.
- (28) Altmann, K. H.; Fabbrot, D.; Dean, N. M.; Geiger, T.; Monia, B. P.; Muller, M.; Nicklin, P. *Biochem. Soc. Trans.* **1996** *24*, 630.
- (29) Crooke, S. T. *Annu. Rev. Med.* **2004**, *55*, 61.
- (30) Agrawal, S.; Iyer, R. P. *Pharmacol. Ther.* **1997**, *76*, 151.
- (31) The Vitravene Study Group. *Am. J. Ophtalmol* **2002**, *133*, 467.
- (32) Kurreck, J. *Eur. J. Biochem.* **2003**, *270*, 1628.
- (33) Gleave, M. E.; Monia, B.P. *Nat. Rev. Cancer* **2005**, *5*, 468.
- (34) Mangos, M. M.; Damha, M. J. *Curr. Top. Med. Chem.* **2002**, *2*, 1147.
- (35) Chan, J. H. P.; Lim, S.; Wong, W. S. F. *Clin. Exp. Pharmacol. Physiol.* **2006**, *33*, 533.
- (36) Sarafianos, S. G.; Das, K.; Tantillo, C.; Clark, A. D.; Ding, J.; Whitcomb, J. M.; Boyer, P. L.; Hughes, S. H.; Arnold, E. *Embo J.* **2001**, *20*, 1449.
- (37) Nowotny, M.; Gaidamakov S. A.; Crouch, J.; Yang, W. *Cell* **2005**, *121*, 1005.
- (38) Orozco, M.; Perez, A.; Noy, A.; Luque, F. J. *Chem. Soc. Rev.* **2003**, *32*, 350.
- (39) Cheatham, T.E.^{3rd}. *Curr. Opin. Struct. Biol.* **2004**, *14*, 360.
- (40) Pérez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E. 3rd; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817.
- (41) Wagner, R. W.; Matteucci, M. D.; Grant, D.; Huang, T.; Froehler, B. C. *Nat. Biotechnol.* **1996**, *14*, 840.
- (42) Nishizaki, T.; Iwai, S.; Ohtsuka, E.; Nakamura, H. *Biochemistry* **1997**, *36*, 2577.
- (43) (a) Wilds, C. J.; Damha, M. J. *Nucl. Acids Res.* **2000**, *28*, 3625. (b) Noronha, A. M.; Wilds, C. J.; Lok, C. N.; Viazovkina, K.; Arion, D.; Parniak, M. A.; Damha, M. J. *Biochemistry* **2000**, *39*, 7050. (c) Damha, M. J.; Noronha, A. M.; Wilds, C. J.; Trempe, J. F.; Denisov, A.; Pon, R. T.; Gehring, K. *Nucleosides Nucleotides Nucleic Acids* **2001**, *20*, 429.
- (44) (a) Barnes, T. W.; Turner, D. H. *Biochemistry* **2001**, *40*, 12738. (b) Barnes, T. W.; Turner, D. H. *J. Am. Chem. Soc.* **2001**, *123*, 4107.
- (45) Dickerson, R. E.; Drew, H. R.; Conner, B. N.; Wing, R. M.; Fratini, A. V. *Science* **1982**, *216*, 475.
- (46) Shields, G. C.; Laughton, C. A.; Orozco, M. *J. Am. Chem. Soc.* **1999**, *119*, 7463.
- (47) Case, D.A. *et al.* *AMBER8*, San Francisco, 2004.
- (48) (a) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179. (b) Cheatham, T. E. 3rd; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845.

- (49) Pérez, A.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **2007**, *129*, 14739-45.
- (50) Bayly, C. E.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phy. Chem.* **1993**, *97*, 10269.
- (51) Wang, J., Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comp. Chem.* **2004**, *25*, 1157.
- (52) (a) Venkateswarlu, D.; Lind, K. E.; Mohan, V.; Manoharan, M.; Ferguson, D. M. *Nucl. Acids Res.* **1999**, *27*, 2189. (b) Venkateswarlu, D.; Ferguson, D. M. *J. Am. Chem. Soc.* **1999**, *121*, 5509.
- (53) (a) Trempe, J.; Wilds, C. J.; Denisov, A. Y.; Pon, R. T.; Damha, M. J.; Gehring, K. *J. Am. Chem. Soc.* **2001**, *123*, 4896. (b) Denisov, A. Y.; Noronha, A. M.; Wilds, C. J.; Trempe, J.; Pon, R. T.; Gehring, K.; Damha, M. J. *Nucl. Acids Res.* **2001**, *29*, 4284.
- (54) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
- (55) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comp. Phys.* **1977**, *23*, 327.
- (56) Gelpi, J. L.; Kalko, S. G.; Barril, X.; Cirera, J.; de La Cruz, X.; Luque, F. J.; Orozco, M. *Proteins* **2000**, *45*, 428.
- (57) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. *Proteins* **1993**, *17*, 412.
- (58) (a) Cubero, E.; Abrescia, N. G. A.; Subirana, J. A.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **2003**, *125*, 14603. (b) Rueda, M.; Kalko, S. G.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **2003**, *125*, 8007. (c) Pérez, A.; Blas, J. R.; Rueda, M.; Lopez-Bes, J. M.; de la Cruz, X.; Orozco, M. *J. Chem. Theor. Comp.* **2005**, *1*, (5), 790-800.
- (59) (a) Noy, A.; Pérez, A.; Laughton, C. A.; Orozco, M. *Nucl. Acids Res.* **2007**, *35*, 3330. (b) Mahalanobis, P. C. *Proc. Nat. Inst. Sci. India* **1936**, *2*, 49.
- (60) Schlitter, J. *Chem. Phys. Letters.* **1993**, *215*, 617-21
- (61) Harris, S. A.; Gavathiotis, E.; Searle, M. S.; Orozco, M.; Laughton, C. A. *J. Am. Chem. Soc.* **2001**, *123*, 12658.
- (62) Lu, X. J.; Shakked, Z.; Olson, W. K. *J Mol Biol.* **2000**, *300*, 819.
- (63) Meyer, T.; Ferrer-Costa, C.; Pérez, A.; Rueda, M.; Bidon-Chanal, A.; Luque, F. J.; Laughton, C. A.; Orozco M. *J. Chem. Theor. Comp.* **2006**, *2*, 251.
- (64) Ihaka, R.; Gentleman, R. *J. Comp. Graph. Stat.* **1996**, *5*, 299.
- (65) Noy, A.; Perez, A.; Lankas, F.; Luque, F. J.; Orozco, M. *J. Mol. Biol.* **2004**, *343*, 627.

Supporting Information

Theoretical Analysis of Antisense Duplexes: Determinants of the RNase H Susceptibility

*Agnes Noy^{1,2}, F.Javier Luque³ and Modesto Orozco^{1,2,4}**

modesto@mmb.pcb.ub.es

¹ Joint IRB-BSC Research Program in Computational Biology.

² National Institute of Bioinformatics.

³ Departament de Fisicoquímica and Institut de Biomedicina. Universitat de Barcelona

⁴ Departament de Bioquímica. Universitat de Barcelona.

Complete reference 47.

(47) D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, K.M. Merz, B. Wang, D.A. Pearlman, M. Crowley, S. Brozell, V. Tsui, H. Gohlke, J. Mongan, V. Hornak, G. Cui, P. Beroza, C. Schafmeister, J.W. Caldwell, W.S. Ross, and P.A. Kollman (2004), AMBER 8, University of California, San Francisco.

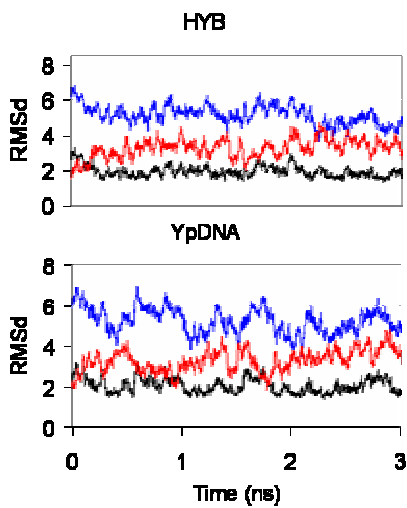


Figure S1. RMSd (for the central 10-mer equivalent (see Methods) backbone atoms; in Å) with respect to DD sequence canonical A (red), B (blue) and NMR-based (1EFS; black) structures of the DNA·RNA and YpDNA·RNA duplexes with the DD sequence starting from A-form.

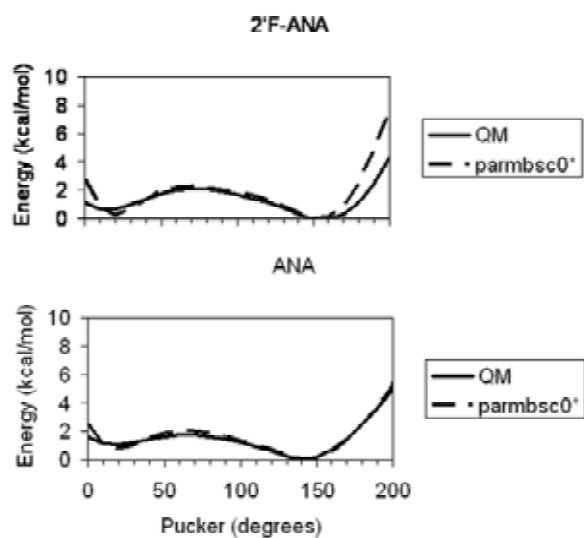


Figure S2. Puckering profile (phase angle vs energy) of arabino and 2'F-arabino sugars in the corresponding thymidine nucleoside from MP2 and parmbsc0 with optimized parameters calculations. To give more detail of the crucial N-S-E region the large maximum at W region is omitted.

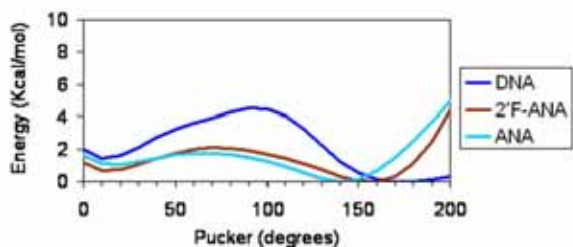


Figure S3. Comparison of the puckering profiles (phase angle vs energy) of 2'deoxyribose, arabino and 2'F-arabino sugars in the corresponding thymidine nucleoside from parmbsc0 with optimized parameters calculations. To give more detail of the crucial N-S-E region the large maximum at W region is omitted.

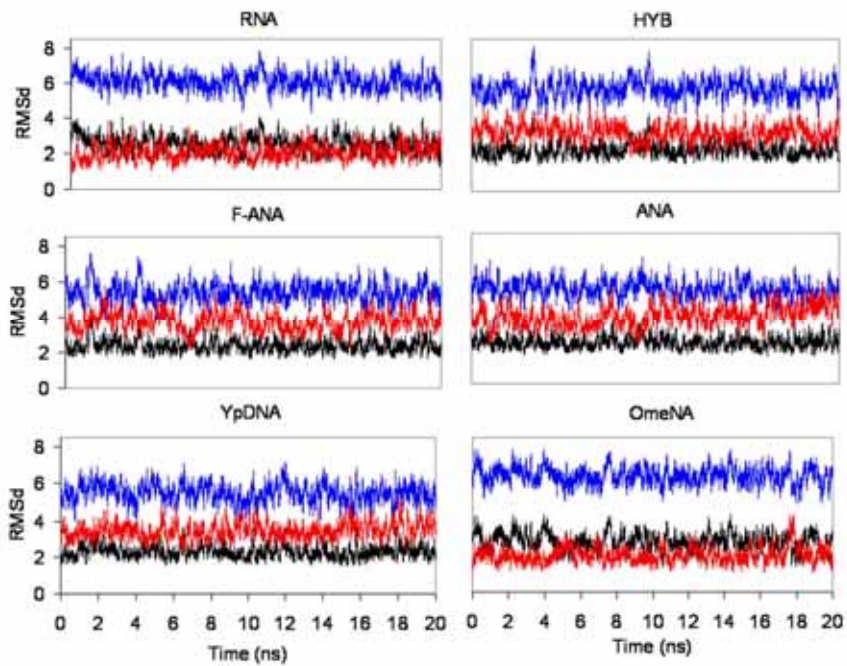


Figure S4. RMSd (for the central 10-mer equivalent (see Methods) backbone atoms in Å) with respect to DD sequence canonical A (red), B (blue) forms and the NMR structure 1EFS (black) of the different seq. 2 duplexes (containing at least 1 RNA strand).

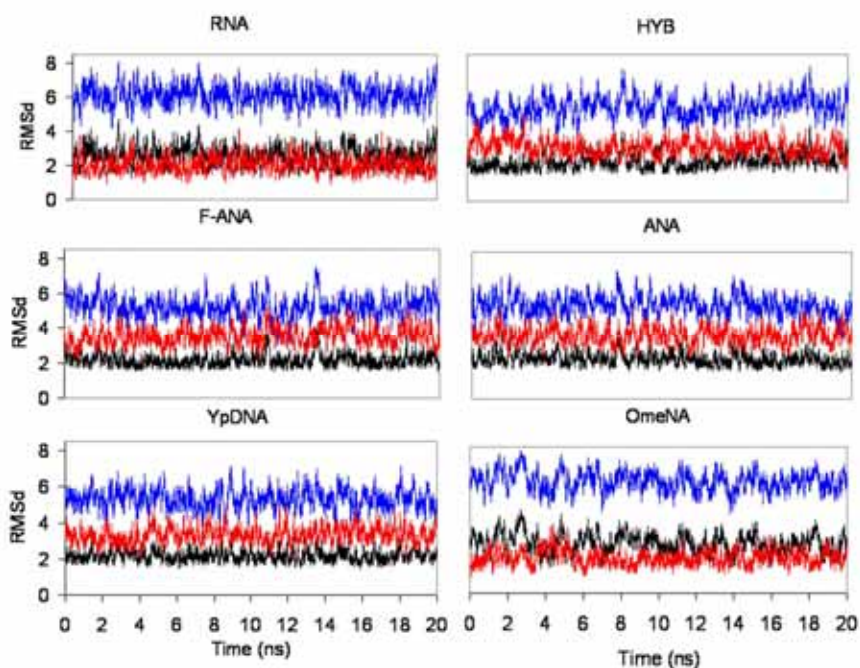


Figure S5. RMSd (for the central 10-mer equivalent (see Methods) backbone atoms in Å) with respect to DD sequence canonical A (red), B (blue) forms and the NMR structure 1EFS (black) of the different seq. 3 duplexes (containing at least 1 RNA strand).

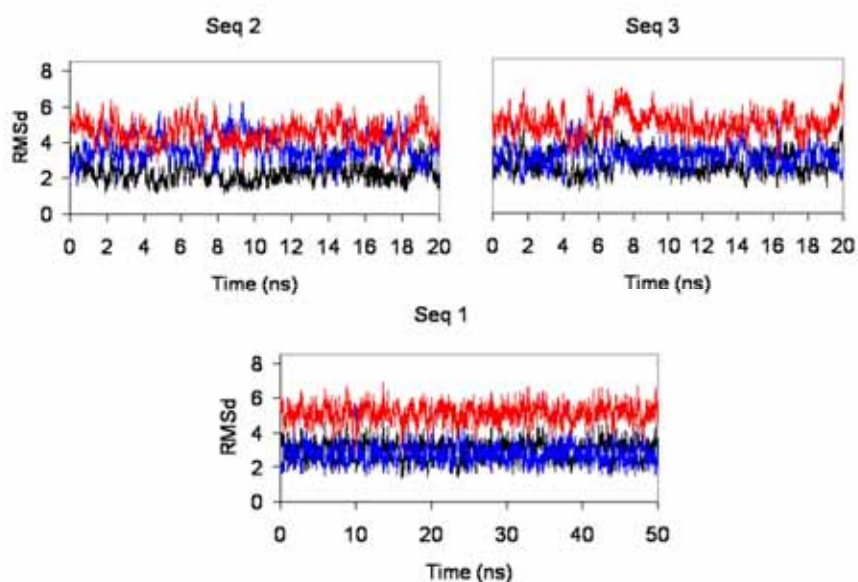


Figure S6. RMSd (for the central 10-mer equivalent (see Methods) backbone atoms in Å) with respect to DD sequence canonical A (red), B (blue) forms and the NMR structure 1EFS (black) of the different DNA₂ duplexes considered here.

Table S1. Percentage of South ($P < 90$) puckerings in hairpins 5'-r(GGAC) d(TTCG) ara(GUCC)-3' and 5'-r(GGAC) d(TTCG) 2'-F-ara(GTCC)-3' detected in the last 10 ns of 20 ns MD simulations (carried out with parmbsc0 and arabino parameters for sugars) compared to the values obtained for the same hairpins from NMR experiments^a.

| Nucleotide | % South (MD) | | % South (NMR) | |
|------------|--------------|----|---------------|--------|
| G9 / FG9 | 99 | 75 | 82-100 | 80-100 |
| U10 / FT10 | 79 | 52 | 80-100 | 72-100 |
| C11 / FC11 | 90 | 80 | 87-100 | 86-100 |
| C12 / FC12 | 86 | 60 | 46-64 | 42-53 |

^a First and second halves of each cell correspond to values for arabino (PDB entry 1HO6) and 2'-F-arabino (PDB entry 1FC8) derivatives.

Table S2A. Relative similarity indexes^a between the eigenvectors associated to the central 10-mer backbone of the different duplexes with the 3 sequences considered here.

| | | DNA | | | RNA | | | HYB | | | F-ANA | | | ANA | | | YpDNA | | | OmeNA | | |
|-------|---|------|------|------|------|------|------|------|------|------|-------|------|------|------|------|------|-------|------|------|-------|------|------|
| | | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| DNA | 1 | | 0.97 | 0.99 | 0.73 | 0.73 | 0.72 | 0.82 | 0.77 | 0.80 | 0.79 | 0.79 | 0.80 | 0.82 | 0.80 | 0.79 | 0.77 | 0.76 | 0.79 | 0.69 | 0.68 | 0.68 |
| | 2 | 0.96 | | 1.02 | 0.81 | 0.81 | 0.81 | 0.90 | 0.87 | 0.85 | 0.85 | 0.86 | 0.87 | 0.84 | 0.85 | 0.89 | 0.86 | 0.83 | 0.87 | 0.77 | 0.76 | 0.76 |
| | 3 | 0.96 | 1.00 | | 0.75 | 0.76 | 0.75 | 0.85 | 0.84 | 0.87 | 0.81 | 0.84 | 0.83 | 0.80 | 0.84 | 0.82 | 0.81 | 0.82 | 0.83 | 0.72 | 0.71 | 0.71 |
| RNA | 1 | 0.66 | 0.78 | 0.70 | | 0.99 | 1.01 | 0.93 | 0.95 | 0.91 | 0.92 | 0.90 | 0.91 | 0.90 | 0.90 | 0.92 | 0.95 | 0.92 | 0.96 | 0.97 | 0.98 | 0.98 |
| | 2 | 0.67 | 0.75 | 0.71 | 0.94 | | 1.00 | 0.93 | 0.96 | 0.91 | 0.92 | 0.91 | 0.91 | 0.89 | 0.92 | 0.93 | 0.92 | 0.93 | 0.96 | 0.94 | 0.98 | 0.98 |
| | 3 | 0.66 | 0.79 | 0.71 | 1.08 | 0.98 | | 0.93 | 0.93 | 0.90 | 0.92 | 0.90 | 0.91 | 0.89 | 0.90 | 0.93 | 0.93 | 0.91 | 0.95 | 0.96 | 0.99 | 0.99 |
| HYB | 1 | 0.77 | 0.86 | 0.81 | 0.88 | 0.88 | 0.90 | | 0.95 | 0.91 | 0.96 | 0.92 | 0.93 | 0.96 | 0.92 | 0.93 | 0.95 | 0.92 | 0.96 | 0.90 | 0.88 | 0.89 |
| | 2 | 0.69 | 0.81 | 0.79 | 0.89 | 0.90 | 0.88 | 0.92 | | 0.98 | 0.92 | 0.94 | 0.93 | 0.92 | 0.95 | 0.94 | 0.95 | 0.98 | 0.99 | 0.88 | 0.91 | 0.90 |
| | 3 | 0.75 | 0.80 | 0.85 | 0.89 | 0.88 | 0.87 | 0.86 | 0.95 | | 0.93 | 0.93 | 0.94 | 0.91 | 0.97 | 0.94 | 0.93 | 0.99 | 0.97 | 0.85 | 0.88 | 0.87 |
| F-ANA | 1 | 0.71 | 0.77 | 0.74 | 0.84 | 0.86 | 0.85 | 0.92 | 0.84 | 0.85 | | 0.96 | 0.97 | 0.95 | 0.94 | 0.94 | 0.95 | 0.92 | 0.96 | 0.90 | 0.90 | 0.89 |
| | 2 | 0.72 | 0.82 | 0.81 | 0.81 | 0.83 | 0.81 | 0.87 | 0.91 | 0.88 | 0.91 | | 0.98 | 0.92 | 0.95 | 0.94 | 0.92 | 0.92 | 0.94 | 0.86 | 0.88 | 0.88 |
| | 3 | 0.75 | 0.85 | 0.77 | 0.85 | 0.85 | 0.86 | 0.89 | 0.86 | 0.89 | 0.91 | 0.94 | | 0.93 | 0.93 | 0.93 | 0.92 | 0.92 | 0.96 | 0.87 | 0.89 | 0.89 |
| ANA | 1 | 0.76 | 0.78 | 0.75 | 0.81 | 0.82 | 0.84 | 0.95 | 0.86 | 0.84 | 0.92 | 0.87 | 0.89 | | 0.92 | 0.94 | 0.93 | 0.90 | 0.94 | 0.87 | 0.86 | 0.86 |
| | 2 | 0.78 | 0.82 | 0.82 | 0.85 | 0.87 | 0.87 | 0.89 | 0.92 | 0.96 | 0.87 | 0.91 | 0.85 | 0.89 | | 0.94 | 0.91 | 0.96 | 0.94 | 0.86 | 0.88 | 0.88 |
| | 3 | 0.71 | 0.87 | 0.80 | 0.83 | 0.86 | 0.84 | 0.87 | 0.88 | 0.91 | 0.85 | 0.90 | 0.85 | 0.88 | 0.91 | | 0.96 | 0.93 | 0.97 | 0.90 | 0.91 | 0.91 |
| YpDNA | 1 | 0.72 | 0.84 | 0.79 | 0.93 | 0.85 | 0.88 | 0.90 | 0.91 | 0.88 | 0.89 | 0.86 | 0.82 | 0.87 | 0.86 | 0.93 | | 0.93 | 0.96 | 0.92 | 0.91 | 0.91 |
| | 2 | 0.71 | 0.78 | 0.77 | 0.91 | 0.95 | 0.91 | 0.88 | 0.93 | 0.99 | 0.89 | 0.86 | 0.87 | 0.87 | 0.95 | 0.90 | 0.90 | | 0.99 | 0.87 | 0.89 | 0.88 |
| | 3 | 0.72 | 0.83 | 0.79 | 0.92 | 0.97 | 0.94 | 0.94 | 0.94 | 0.96 | 0.92 | 0.89 | 0.93 | 0.91 | 0.89 | 0.97 | 0.94 | 1.00 | | 0.92 | 0.93 | 0.92 |
| OmeNA | 1 | 0.60 | 0.70 | 0.67 | 0.99 | 0.85 | 0.97 | 0.81 | 0.78 | 0.80 | 0.79 | 0.75 | 0.75 | 0.75 | 0.78 | 0.79 | 0.84 | 0.81 | 0.84 | | 0.95 | 0.96 |
| | 2 | 0.63 | 0.73 | 0.68 | 0.95 | 0.97 | 0.95 | 0.80 | 0.85 | 0.88 | 0.81 | 0.80 | 0.80 | 0.76 | 0.83 | 0.84 | 0.85 | 0.89 | 0.94 | 0.87 | | 1.00 |
| | 3 | 0.65 | 0.75 | 0.71 | 0.95 | 0.91 | 0.97 | 0.84 | 0.83 | 0.86 | 0.77 | 0.81 | 0.81 | 0.76 | 0.84 | 0.86 | 0.84 | 0.85 | 0.90 | 0.89 | 0.92 | |

^a δ above the diagonal in Roman font and κ below in *Italics*

Table S2B. Relative similarity indexes^a between the eigenvectors associated to the central 10-mer backbone of the different strands for the DD sequence^b.

| | | DNA | RNA | HYB | | F-ANA | | ANA | | YpDNA | | OmeNA | |
|-------|-------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------|
| | | | | anti | sense | anti | sense | anti | sense | anti | sense | anti | sense |
| DNA | | | 0.63 | 0.86 | 0.69 | 0.77 | 0.66 | 0.86 | 0.68 | 0.73 | 0.73 | 0.64 | 0.61 |
| RNA | | <i>0.61</i> | | 0.83 | 0.93 | 0.84 | 0.99 | 0.78 | 0.97 | 0.90 | 0.90 | 0.95 | 1.00 |
| HYB | anti | <i>0.80</i> | <i>0.75</i> | | 0.89 | 0.95 | 0.85 | 1.00 | 0.87 | 0.94 | 0.89 | 0.79 | 0.80 |
| | sense | <i>0.67</i> | <i>0.85</i> | <i>0.82</i> | | 0.90 | 0.96 | 0.83 | 0.96 | 0.93 | 0.92 | 0.88 | 0.92 |
| F-ANA | anti | <i>0.68</i> | <i>0.69</i> | <i>0.82</i> | <i>0.74</i> | | 0.86 | 0.92 | 0.87 | 0.95 | 0.87 | 0.80 | 0.82 |
| | sense | <i>0.64</i> | <i>0.94</i> | <i>0.78</i> | <i>0.89</i> | <i>0.70</i> | | 0.79 | 1.00 | 0.91 | 0.91 | 0.93 | 0.97 |
| ANA | anti | <i>0.85</i> | <i>0.71</i> | <i>0.95</i> | <i>0.78</i> | <i>0.86</i> | <i>0.74</i> | | 0.81 | 0.91 | 0.84 | 0.75 | 0.76 |
| | sense | <i>0.63</i> | <i>0.96</i> | <i>0.77</i> | <i>0.88</i> | <i>0.71</i> | <i>1.00</i> | <i>0.73</i> | | 0.91 | 0.95 | 0.92 | 0.97 |
| YpDNA | anti | <i>0.73</i> | <i>0.75</i> | <i>0.90</i> | <i>0.79</i> | <i>0.78</i> | <i>0.77</i> | <i>0.89</i> | <i>0.76</i> | | 0.91 | 0.85 | 0.90 |
| | sense | <i>0.60</i> | <i>0.77</i> | <i>0.73</i> | <i>0.77</i> | <i>0.67</i> | <i>0.80</i> | <i>0.69</i> | <i>0.81</i> | <i>0.72</i> | | 0.91 | 0.90 |
| OmeNA | anti | <i>0.62</i> | <i>0.90</i> | <i>0.73</i> | <i>0.83</i> | <i>0.66</i> | <i>0.87</i> | <i>0.72</i> | <i>0.86</i> | <i>0.72</i> | <i>0.87</i> | | 0.95 |
| | sense | <i>0.61</i> | <i>1.00</i> | <i>0.75</i> | <i>0.86</i> | <i>0.69</i> | <i>0.96</i> | <i>0.71</i> | <i>0.95</i> | <i>0.75</i> | <i>0.81</i> | <i>0.96</i> | |

^aδ above the diagonal in Roman font and κ below in *Italics*

^b Equivalent results for the other two sequences are available upon request

4.4 Theoretical Study of Large Conformational Transitions in DNA:
the B \leftrightarrow A Conformational Change in Water and Ethanol/water

Agnes Noy, Alberto Pérez, Charles A. Laughton, Modesto Orozco

Nucleic Acids Research (2007) 35, 3330-8

Theoretical study of large conformational transitions in DNA: the B \leftrightarrow A conformational change in water and ethanol/water

Agnes Noy¹, Alberto Pérez¹, Charles A. Laughton² and Modesto Orozco^{1,3,4,*}

¹Molecular Modeling and Bioinformatics Unit, Institut de Recerca Biomèdica & Instituto Nacional de Bioinformática, Parc Científic de Barcelona, Josep Samitier 1-5, Barcelona 08028, Spain, ²School of Pharmacy and Centre for Biomolecular Sciences, University of Nottingham, Nottingham NG7 2RD, UK, ³Departament de Bioquímica i Biologia Molecular. Facultat de Biologia. Universitat de Barcelona. Avgda Diagonal 645, Barcelona 08028, Spain and ⁴Computational Biology Program, Barcelona Supercomputer Centre, Jordi Girona 31, Edifici Torre Girona, Barcelona 08028, Spain

ABSTRACT

We explore here the possibility of determining theoretically the free energy change associated with large conformational transitions in DNA, like the solvent-induced B \leftrightarrow A conformational change. We find that a combination of targeted molecular dynamics (tMD) and the weighted histogram analysis method (WHAM) can be used to trace this transition in both water and ethanol/water mixture. The pathway of the transition in the A \rightarrow B direction mirrors the B \rightarrow A pathway, and is dominated by two processes that occur somewhat independently: local changes in sugar puckering and global rearrangements (particularly twist and roll) in the structure. The B \rightarrow A transition is found to be a quasi-harmonic process, which follows closely the first spontaneous deformation mode of B-DNA, showing that a physiologically-relevant deformation is in coded in the flexibility pattern of DNA.

INTRODUCTION

DNA is a flexible and polymorphic structure, whose conformation changes as a response to the presence of ligands, variations in the temperature or modifications in the solvent (1–3). Such flexibility is implicitly coded in the structure of the DNA (4–8) and is crucial to its biological functionality, since, for example, binding of DNA-regulatory proteins can require dramatic changes in the structure of nucleic acid. Understanding DNA's flexibility and conformational transition is thus a necessary step in transforming structural data into biologically important information.

In general, the determination of DNA structure is no longer the challenge for experimental techniques had it used to be, both NMR and X-ray techniques are now able to provide accurate structures for most sequences

(more than 2500 (experimentally-solved) DNA structures are deposited in the Protein Data Bank in June 2006). Unfortunately, these experimental techniques are less effective at discovering flexibility or tracing conformational transitions, and so for this we currently make major use of simulation techniques like molecular dynamics (MD) to study these transitions at the atomic level. However, many conformational transitions in DNA occur on the micro or millisecond time scale, while timescales currently accessible by atomistic MD are in the 5–50 ns range. To overcome this problem, it is common to use biasing techniques to move simulations along a reaction coordinate at an 'artificially' high rate, but many of the transitions we wish to study involve large conformational changes, where many internal degrees of freedom move in a coordinated way. In this situation, it is difficult to apply standard biasing techniques based on the regular change of internal degrees of freedom along the reaction coordinate.

In summary, the analysis of large conformational changes is still a major challenge for both experimental and theoretical techniques. A clear example is the B \leftrightarrow A transition of DNA duplexes (9–12). It has been known since 1953 that physiological DNA is mostly B-form, while physiological RNA is always in the A-conformation (1,2,9–12). However, the binding of some proteins to DNA can induce local B \rightarrow A changes, which are needed to form some protein-DNA complexes crucial for the control of gene functionality (13–15). The B \rightarrow A transition can also be brought about by other stress conditions, like crystal lattice restrictions (16 and references therein), or the addition of a large proportion of ethanol, which is believed to displace water molecules hydrating the helix, leading to a pseudo-anhydrous environment where the most compact A-form is more stable (1,12,17,18). The B \leftrightarrow A conformational transition has been also the subject of numerous theoretical studies, aimed at understanding the mechanism of the transition and the atomic reasons for the A/B preference in water and

*To whom Correspondence should be addressed: Tel: +34 934037155; Fax: +34 934037157; Email: modesto@mmb.pcb.ub.es

other solvents. The latest generations of both CHARMM and AMBER force-fields (the ones most used in simulations of nucleic acids) recognize the A-form as the only stable conformation of duplex RNA, and the B-form as the most important conformation of DNA duplex in aqueous solution (18–28). In fact, MD trajectories of DNA duplexes started in the A-form convert very rapidly to the B-conformation (23), showing that MD simulations are able to drive the DNA from an incorrect conformation to a correct one. Unfortunately, MD simulations are unable to detect reversible transitions due to the limited length of current trajectories, which makes very unlikely to sample unstable regions of the conformational space. This lack of reversibility in the transition precludes the determination of the free energy associated with the conformational change and the determination of the atomic mechanism of the transition. For this reason, these simulations need to be biased to force them to sample reversibly the B↔A transition.

In this article, we re-visit the B↔A transition for a duplex DNA dodecamer using the techniques of essential dynamics, unbiased molecular dynamics simulations and a combination of targeted MD (tMD (29–31)) and the weighted histogram analysis method (WHAM (32)) in both water and a mixture of 85:15 ethanol/water. Using these techniques, we were able to trace the B↔A transition in a smooth reversible way providing estimates of the associated free energy and of the molecular mechanism of this conformational change.

METHODS

Model selection

We felt that in order to obtain reliable conclusions on B/A preference, a DNA duplex containing at least one helix turn should be used as the model system (i.e. at least 10–12 base pairs); smaller duplexes might not provide a good model of interactions along the major and minor grooves and would maximize the effect of artifactual end-effects. Unfortunately, the increase in the size of the duplex leads to a parallel increase in the computational difficulty of the simulation. Thus, as a compromise, we decided to use Dickerson's dodecamer (dGCGCAATTGCGC)₂ a B-type 12-mer oligo (33) for which several experimental structures are available (see <http://ndbserver.rutgers.edu>) and for which dozens of simulations have been published (4,34–37).

Systems setup for aqueous simulations

Input coordinates for trajectories started in the B-form were generated by taking Dickerson's crystal geometries (33). The experimental structure was neutralized by adding Na⁺ in the best regions according to classical molecular interaction potential (CMIP) calculations (38) and immersed in a rectangular box of water (around 4454 TIP3P molecules). As described elsewhere (38), MIP neutralization protocol allows the definition of a reasonable ionic atmosphere around DNA based on Poisson–Boltzman potentials, reducing then the time needed for counterion equilibration. The solvated systems were then

optimized, thermalized and pre-equilibrated using our standard protocol (39–41). The final conformations were then re-equilibrated (constant pressure and temperature: 1 atm, 298 K) for 1 ns more to ensure the stability of the trajectories. Simulations starting in the A-conformation were generated by first building a standard A-type DNA duplex of the desired sequence, which was then neutralized and hydrated as below. Optimization, thermalization and pre-equilibration was carried out following the same procedure than for B-DNA but adding an harmonic restraint ($K=160$ kcal/mol.rad²) that restrained all the δ angles to 80°. The same restraints were used in equilibration (4 ns) and in production runs; when they are removed the structure changes quickly to the B-form.

Ten different structures obtained during the equilibration of the A-form (with δ angles restrained to 80°) were used to perform 10 parallel un-restrained 3 ns MD simulations. Within this short simulation time the A→B transition was completed in all replicas, which allowed us to obtain reasonable statistics on the microscopic mechanism of the 'unbiased' transition.

Systems setup for ethanol/water simulations

Equilibrated boxes containing 85% ethanol: 15% water were generated from Monte Carlo simulations (300 million configurations) at constant pressure (1 atm) and temperature (300 K). These boxes were then used to solvate neutralized B and A-DNAs (starting conformations of B and A-DNAs were those obtained at the end of the respective MD equilibrations in water (see above)). The solvated systems were thermalized ($T=298$ K) and pre-equilibrated using 2 ns of MD where only the solvent was free to move. Finally, all the systems were equilibrated for 4 ns in both A and B-conformations. The δ angles were restrained to 80° in A-simulations and 120° in B-simulations. Fifteen structures obtained during the last 1 ns of the restrained B-type simulations were used to start 3 ns unbiased MD simulations intended to capture spontaneous B→A transitions in (85:15) ethanol/water. Mirror calculations with unbiased trajectories started in the A-form were used to confirm that the A-form is a stable minimum in the ethanol/water mixture considered here.

Targeted MD simulations

These simulations drive a transition by introducing a harmonic penalty that forces the sampled structure to be at a given RMSd from a reference(s) structure(s). By changing smoothly the target RMSd, we can then approach or separate the molecule from reference geometries. Several formalisms can be introduced to introduce the harmonic restraints (25), but in our hands smoother and more reliable A↔B transitions were obtained by using eq. (1) (see results). As described below, a selected group, rather than all the atoms (25) were used to compute the RMSd.

$$E_{\text{rest}}^{\lambda} = K_{\text{rest}}(\text{RMSd}(X, X_{\text{final}}) - \text{RMSd}(\lambda))^2 \quad (1)$$

where $\text{RMSd}(\lambda)$ is the desired RMSd to final structure at point λ of the transition path.

After several preliminary tests, simulations were performing using windows of 0.25 Å with $K_{\text{rest}} = 0.1$ kcal/mol Å² × atom for all windows except the terminal ones, where we used spacing 0.1 Å and $K_{\text{rest}} = 8$ (water) or 2 (ethanol/water) kcal/mol Å² × atom. Unless otherwise noted, each window was simulated for 1.2 ns, the first 0.2 being considered as equilibration. To make transitions as smooth as possible, the end point of each window was used in general as starting point for the next. Using these conditions smooth transitions and good overlap between windows was obtained.

The biased samplings obtained were used to derive potentials of mean force (PMF) for the transition using the WHAM method (32). Free energies were recovered by integrating the PMF using suitable boundaries (see below).

Essential dynamics

Unbiased B-DNA and A-RNA 10 ns trajectories were used to derive the essential dynamics of relaxed B and A-forms (4,7,8,42,43). The first eigenvector (that describing the most important deformation mode) of the relaxed B-DNA (or A-RNA) trajectory was compared with the B→A transition vector (see eq. (2)) to obtain an estimate of the overlap between normal B-DNA fluctuations and B→A transitions.

$$\gamma = \frac{v_{\text{relax}}}{|v_{\text{relax}}|} \bullet \frac{v_{\text{trans}}}{|v_{\text{trans}}|} \quad (2)$$

where v_{relax} stands for the first eigenvector describing the essential dynamics in A or B-forms and v_{trans} represents the transition vector A→B or B→A.

As described in Results, the first eigenvector of the essential dynamics of B-DNA correlates well with the B→A transition vector. We then explored the harmonic energy needed to animate this eigenvector (from the B-form) to reach conformations close to the A-form. For this purpose, we add vibrational energy to the mode obtaining the associated displacement through eq. (3) from which perturbed geometries were determined. Pursuing this approach, we performed calculations of the dimension-less Mahalanobis distance (d_M ; see eq. (4)) between B and A forms. The Mahalanobis distance is a unit-less metric, directly related to the deformation energy (see eq. (5)) which defines the minimum pathway in essential space between two conformations. Thus, using d_M calculations, we can trace the B→A transition by activating many different essential modes of the B-equilibrium trajectories computing then the energy needed to reach structures closer (to a certain threshold) to the A-form.

$$\Delta x = \pm \left(\frac{2E\lambda}{k_b T} \right)^{1/2} \quad (3)$$

where E is the vibrational energy, k_b is Boltzman's constant, T is absolute temperature, λ is the eigenvalue (in distance² units).

$$d_M = \left[\sum_{i=1}^n \left(\frac{x_i}{\lambda_i^{1/2}} \right)^2 \right]^{1/2} \quad (4)$$

where x_i is the displacement along individual eigenvectors and λ_i stands for the corresponding eigenvalue (in distance² units). The sum extends of n -important essential movements, in this article we consider the first 10th ones, which accounts for more than 85% of the DNA variance.

$$E = \frac{k_B T}{2} d_M^2 \quad (5)$$

Technical details

All simulations were carried in the isothermal–isobaric ensemble ($P = 1$ atm, $T = 300$ K). Monte Carlo simulations (for preparing the hybrid solvent box) were performed using the BOSS3.4 computer program (44) allowing internal rotations in the solvent molecules, but no other internal changes. A non-bonded cutoff of 12 Å was used in conjunction with periodic boundary conditions to reduce the number of non-bonded interactions in these Monte Carlo simulations. Molecular dynamics (MD) calculations were carried out using the AMBER8 computer programs (45) and long-range electrostatic effects were taken into account by means of the Particle Mesh Ewald method (46). PARM99 was used to represent DNA interactions (47,48), while TIP3P (49) and all-atoms OPLS (50) parameters were used for the solvents. The use of SHAKE (51) to maintain all the bonds at equilibrium distance allowed us to use 2 fs as the integration step size.

Analysis of trajectories was carried out using PTRAJ module included with the AMBER8 release as well as software developed in house. All simulations were performed using the *Mare Nostrum* supercomputer at the Barcelona Supercomputer Center.

RESULTS AND DISCUSSION

Unbiased A→B trajectories

As reported by others (19–24), MD simulations of DNA using the PARM99 force-field produces a spontaneous A→B conformational change in water in short simulation times, which makes possible the study of the atomic mechanism of such an unforced transition. Ten different trajectories starting from well-equilibrated A-form conformations move to the B-conformation within the 3 ns simulation times. Extension of five of these trajectories to 10 ns did not shown any back-transition to the A-form (data not shown, but available upon request). The analysis of the RMSd plots from canonical A and B-forms show that the transition starts very quickly, and after around 500 ps, the lines of RMSd(versus A) and RMSd(versus B) cross for the first time (see Figure 1). After this period, the structures fluctuate around

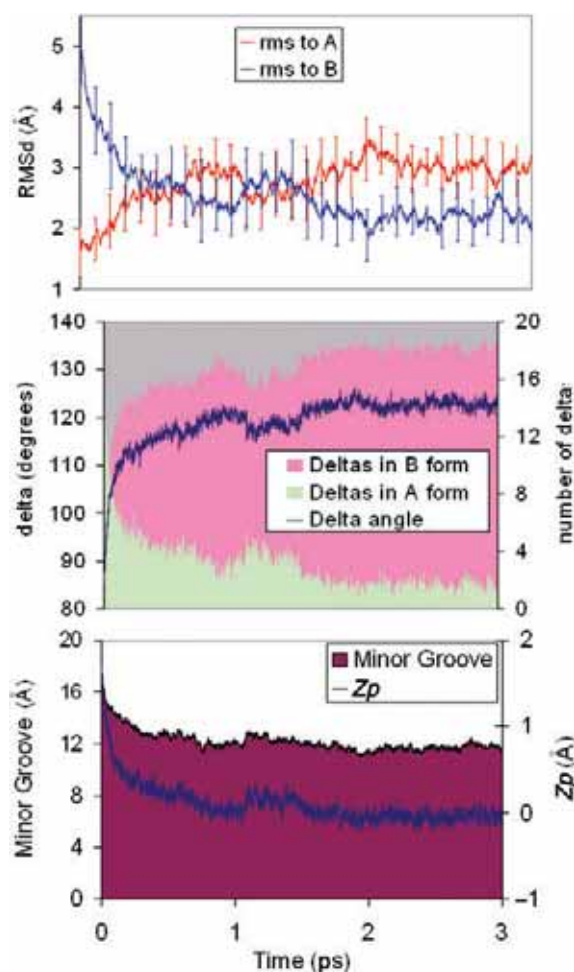


Figure 1. TOP: Average variation with time of the RMSd (from A and B canonical structures) of 10 unbiased trajectories starting from A-form in water. Standard deviations are shown. MIDDLE: average variations of δ and number of sugars in North conformation for the same trajectories. BOTTOM: Variations in the average minor groove width (*mGw*) and *Zp*.

a conformation with B-like properties (see Figure 1; individual plots for trajectories are shown at <http://mmb.pcb.ub.es/A-B>), but still not far from the A-form in terms of RMSd. For all 10 trajectories, the transition in RMSd only happens after major changes in sugar pucker, supporting the idea that sugar re-pucker is the driving force for A \rightarrow B transition (24). Once delta change, helical parameters like roll and twist are adjusted to the standard B values. Thus, in just 100 ps half of the sugars which were originally in the N-conformation (δ below 95°) have moved into the E and S regions (δ greater than 95°) and after 500 ps only four sugars are (in average) in the N-region. Slow re-pucker of these residual North sugars occur along the 1–3 ns simulation period. Analysis of individual sugars shows that in general, sugars in purine nucleotides change faster than those in pyrimidines. Within purines G shows faster transitions than A and within pyrimidines C is that with the slower rate of N \rightarrow S transitions.

Table 1. Dot products between the unit vectors representing the first essential movement of different trajectories. A value of 1 (or -1) indicates a perfect match between essential deformation movements. Note that the first essential movement obtained in the unbiased A \rightarrow B simulation or in the tMD B \rightarrow A one should correspond to the B \rightarrow A transition vector. The non-optimum overlap between them (around 90%) gives an indication of the uncertainties in the definition of this transition vector

| | DNA | RNA | B \rightarrow A (Unbiased) | B \rightarrow A (tMD) |
|------------------------------|------|-------|------------------------------|-------------------------|
| DNA | 1.00 | -0.81 | 0.94 | 0.87 |
| RNA | | 1.00 | -0.81 | -0.90 |
| B \rightarrow A (unbiased) | | | 1.00 | 0.86 |

Very interestingly, the transition vector correlates very well (see Table 1) with the first deformation mode of both relaxed B and A-forms (taken from ref. (7)). This indicates that the deformation pattern needed to perform the biologically relevant B \leftrightarrow A transition is implicitly coded in the polymeric structure of these nucleic acids and that in a rough approximation, the A-form can be interpreted as a vibrational-activated state of B-DNA. This hypothesis is supported by analyzing the structures obtained by moving the B-DNA along the first essential mode (see Figure 2). Thus, a simple deformation along the first mode yields structures with $\text{RMSd(A)} = \text{RMSd(B)}$ for vibrational energies around 4 kcal/mol, and for deformation energy around 12 kcal/mol to structures that are at only 2 Å from the A-form (see Figure 2). The activation of additional deformation modes (up to 10) using the Mahalanobis metric allowed us to slightly reduce the RMSd from target A-form (to 1.7 Å), but at the expense of a very large vibrational energy (19 kcal/mol). Clearly, B \rightarrow A transitions follows closely the first mode, and local rearrangement from a distorted to the real A-conformation requires some local, probably non-harmonic deformations.

Targeted MD transitions in water

Unbiased simulations provide an intuitive picture of the A \rightarrow B transition in water, but they are not able to provide any estimate of the thermodynamics, or kinetics of the process. Essential dynamics allows us to obtain only a rough qualitative picture of the B \rightarrow A transition, making necessary the use of more accurate non-harmonic techniques like the tMD/WHAM one.

Our previous work (24) and the unbiased simulations discussed above suggested that sugar pucker could be a good variable to discriminate between B and A-form. Accordingly, we undertook a study of the B \rightarrow A transition by means of restrained MD simulations, slowly changing the δ angles of all sugars from 140° to 70° . As shown in Figure 3, and in data from equilibrium simulations in Figure 4, the B \rightarrow A transition happens for δ values around 90° . Note that such a transition is clear not only in the RMSd from A and B-conformations, but also in the width of the minor groove (*mGw*) and in Olson's (52) *Zp* parameter (see Figure 3). Thus, a structure can be assigned to the A-family if $\text{RMSd(A)} < \text{RMSd(B)}$,

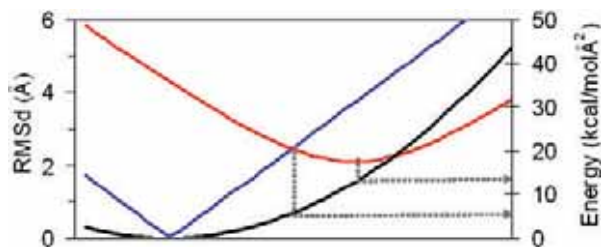


Figure 2. Representation of the energy (black), and RMSd with respect to A (in red) and B (in blue), when a relaxed DNA structure is perturbed along its first essential mode. RMSds were computed with reference to the MD-averaged A and B-conformations. Blue line is not smooth because perturbation along first essential mode is made using MD-averaged B-conformation.

$\langle Zp \rangle > 1.0$; $\langle \delta \rangle < 95^\circ$ and $\langle mGw \rangle > 14 \text{ \AA}$. When only the RMSd criteria is fulfilled, the structure should be labeled as distorted B-form. These threshold need to be consider to validate when a tMD simulation is really driving the structure to the target conformation and not to a distorted geometry, which might have small RMSd to target conformation but very poor internal geometry.

While the use of restraints on delta permitted the required transition, and was extremely useful to define clear threshold values to classify structures as A or B, it does not allow us to easily recover the free energy associated to the conformational change. Thus, we decided to follow the $A \leftrightarrow B$ transition by defining the RMSd function considering only heavy atoms in the sugar ring (Reference set 1, see Figure 5). Additional calculations were performed considering an extended set of restraint atoms which include all ring atoms plus O5' (Reference set 2) and an alternative set of restraint atoms formed by the phosphate group and O4' (Reference set 3). Using any of these three RMSd definitions and the restraint function shown in eq. (1), we were able to reproduce correct $B \rightarrow A$ transitions (see Figure 6 and Figure S1 in Supplementary material), where both starting and end structures fits the canonical B and A-forms. It is worth noting that such smooth transitions cannot be so easily obtained with other choices of the RMSd function. Caution is then necessary against a un-supervised pure-force use of the targeted MD technique to follow complex conformational transitions.

It is worth to note that the transition pathway closely follows the direction of the first essential mode of relaxed DNAs and RNAs (see Table 1), confirming the results obtained for unbiased $A \rightarrow B$ transitions. This finding demonstrates that the ability to change between the B and A-conformers is implicitly coded in the structure of nucleic acid duplexes. Interestingly, irrespective of the set of atoms used to define the RMSd with respect to target structures the transition is similar: for $\text{RMSd}(\lambda) \approx 2.0 \text{ \AA}$, the $\text{RMSd}(A)$ and $\text{RMSd}(B)$ lines cross, but until very late in the reaction coordinate ($\text{RMSd}(\lambda) \approx 1.0 \text{ \AA}$) full transition is not achieved (see Figure 6 and Supplementary Figure S1).

The PMF associated to the $B \rightarrow A$ transitions in water are well converged and seems robust to the selection of the

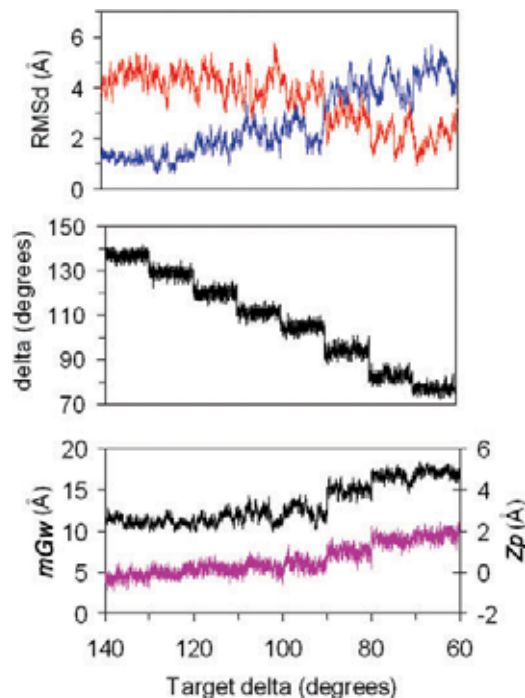


Figure 3. TOP: Profile of RMSds (from A in red and B in blue) obtained from MD trajectories where the target δ (δ_t) was varied from 140 to 70° (see text). Profiles corresponding the average δ are shown in the MIDDLE panel and those for the average mGw (black) and Zp (pink) are shown BOTTOM.

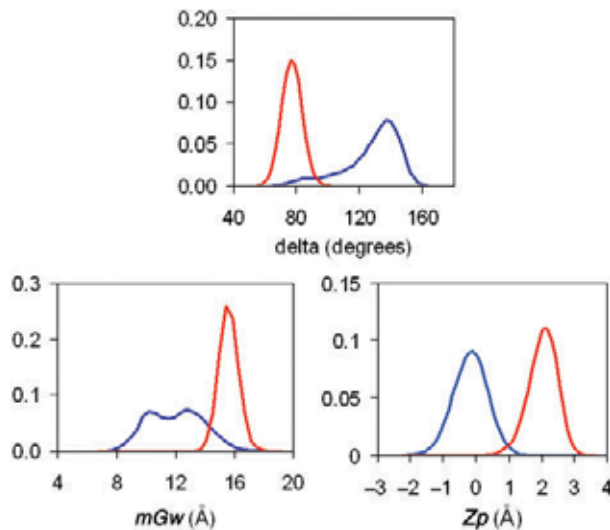


Figure 4. Distribution of TOP: δ angle, BOTTOM-left: mGw and BOTTOM-right: Zp for equilibrium trajectories of B-DNA blue and A-RNA (red). Trajectories were taken from ref.(7).

set of restrained atoms and to changes in the 'forward' or 'reverse' direction of the transition (see Figure 7). As predicted, the $A \rightarrow B$ PMF takes the form of a downhill landscape, where the A-form is not defined as a minimum. Integration of the PMF profiles provides a direct estimate of the free energy associated with conformational transitions. Inspection of Figure 6 (see text) allowed us to

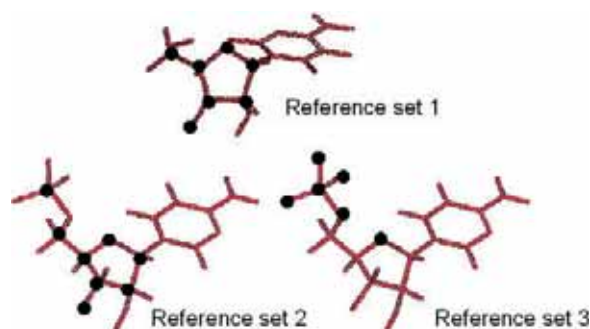


Figure 5. Representation of different sets of atoms used to restrain the system in tMD simulations of the B↔A transition in water and 85:15 ethanol/water mixture.

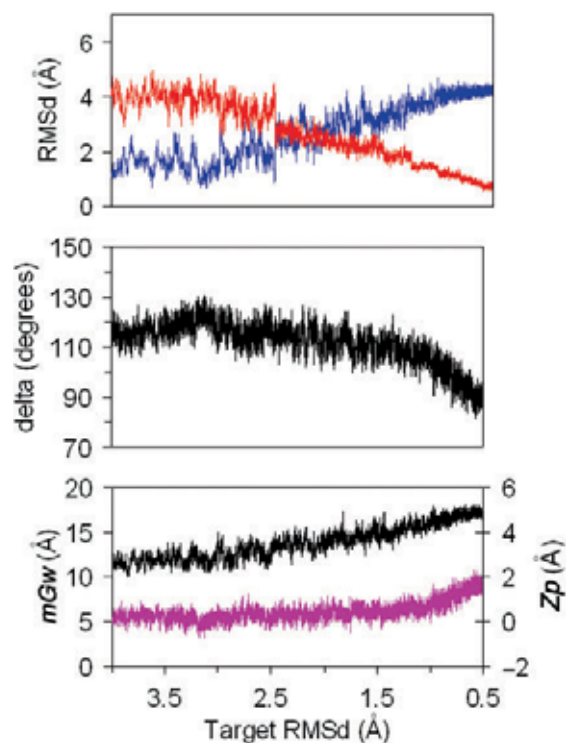


Figure 6. Variation of TOP: RMSds (from A in red and B in blue), MIDDLE: δ angle; BOTTOM: Z_p (pink) and mGw (black), obtained in targeted MD trajectories in pure water using Reference set 1.

classify the PMF into three regions: (i) pure A-form ($\text{RMSd}(\lambda) \leq 1.0 \text{ \AA}$), pure B-form ($\text{RMSd}(\lambda) \geq 2.0 \text{ \AA}$), and distorted B-forms ($2.0 \text{ \AA} > \text{RMSd}(\lambda) > 1.0 \text{ \AA}$). Using these values, the deformation of B-DNA to achieve structures with $\text{RMSd}(A) \sim \text{RMSd}(B)$ requires around 3 kcal/mol (see Table 2), while a full B→A transition is associated with a free energy penalty of 11.4 kcal/mol. After sending the first version of this manuscript for publication, a colleague addressed us to the work by Zhurkin's group (53) who reported empirical scales for the free energy of B→A transition in water. Applying Zhurkin's data in our duplex, we obtain an 'empirical' estimate of 11 kcal/mol for the B→A transition, a value that matches perfectly our estimate. It is also worth noting

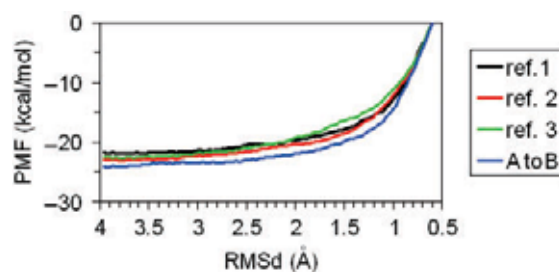


Figure 7. Potential of mean force (PMF) obtained from tMD-WHAM calculations of the B→A transition in water. Profiles were obtained using different sets of atoms to define the target RMSd in each window (see insert and text) and using both forward and reverse directions.

Table 2. Free energy (in kcal/mol) difference for B→A conformational transition in water and 85:15 ethanol/water obtained by integration of the PMF curves. For definition of the states, see text

| Solvent | B→B(dist) | B(dist)→A | B→A(all) | A→A(dist) |
|--------------|-----------|-----------|----------|-----------|
| Water | 3.2 | 8.1 | 11.4 | – |
| (EtOH/water) | – | – | –0.8 | 0.4 |

the very good agreement between rigorous tMD/WHAM estimates and rough values obtained by essential dynamics in Figure 2, something that confirms that the B→A transition is a pure 'up-hill' process which can be simplified as a vibrational activation of the first deformation mode of B-DNA. A small amount of energy is enough to distort the DNA to conformations that are not far in terms of RMSd from the A-form, however, a full transition is very unlikely and requires major changes in the environment, such as the introduction of proteins or co-solvents.

In summary, all our simulations suggest that in water the A→B transition is a downhill process, and that the A-form is not a stable conformation of Dickerson's dodecamer in water. This result is in full agreement with all previous MD simulations irrespective of the force-field, sequence or simulation conditions used (18–22,24,25), which argues against force-field or simulation protocols. Special agreement is found with data by Banavali and Roux (25), who using other force-field a shorter oligo, and a different functional for the restrain energy obtained quite similar results, demonstrating that the method, when applied with common sense is robust. However, these theoretical results have been severely criticized by Jose and Porschke (54,55), who found that experimentally the transition time for the A↔B change (around 70:30 ethanol/water) is in the range 10^0 – $10^1 \mu\text{s}$, which should indicate the existence of a sizeable transition barrier and the presence of two well-characterized canonical forms. We will show below that the criticism is not justified since the shape of the free-energy curve in pure water or ethanol-rich solutions is completely different, and conclusions obtained in rich ethanol mixtures cannot be simply extrapolated to water solution (see next sections).

Study of B→A transitions in 85:15 ethanol/water by unbiased MD simulations

It is experimentally known that in the presence of large concentration of ethanol, the DNA changes from a B to an A-type conformation (1,12,17,18,54,55). Up to 15 MD simulations of Dickerson's dodecamer starting from the A-conformation remain in this region, as previously found by McConnell and Beveridge (21). Extension of one of this simulation to 10 ns does not show any dramatic change from the shorter ones (see Supplementary Figure S2), suggesting that the force-field recognize the A-form as a stable conformation of DNA in ethanol/water solution. However, 10 unbiased 3 ns trajectories starting in the B-conformation remains within this conformational family (see Figure 8), confirming again previous results by other authors (20,21,55). Overall, unbiased MD simulations suggest that in the presence of high concentration of ethanol, the free-energy landscape is different to that in pure water, since at least two minima (close to the B and A-forms) exist, separated by a sizeable energy barrier.

Targeted MD transitions in ethanol/water (85:15)

Targeted MD simulations (using restraint set 1) are able to drive a smooth and complete A→B transition (see Figure 9) in the presence of a high concentration of ethanol. Analysis of RMSd and internal geometry changes along the transition pathway demonstrate that the A→B transition in 85:15 ethanol/water starts with a fast change ($\text{RMSd}(\lambda) \approx 3.3 \text{ \AA}$) of sugar pucker from N to S which leads to an immediate change in the Z_p parameter, and later ($\text{RMSd}(\lambda) \approx 2.0 \text{ \AA}$) to the crossing of the RMSd(A/B) lines and to the reduction of the minor groove width to canonical B-values. In summary, we see in ethanol/water a mirror copy of the transition found in pure water. Combining results in ethanol/water and pure water, we can obtain a quite complete picture of the nature of the A↔B conformational change. Accordingly, when the DNA moves from the A to the B-form, the driving force is a change in the sugar pucker that is soon followed by a global arrangement of the structure. On the contrary, for the B→A transition a global change of shape and reduction of the width of the minor groove happens first, and only when the global structure is close to the A-form do the sugars adopt a North pucker, defining at that point a true A-type structure.

As suggested by unbiased simulations, PMF profile for the A→B transition in ethanol/water shows the existence of two minima regions separated by wide region of higher free energy (see Figure 10). Current tMD simulations cannot be used to determine precisely the free energy barrier, but a rough estimate of $\sim 2 \text{ kcal/mol}$ (A→B) and $\sim 1 \text{ kcal/mol}$ (B→A) can be derived from the PMF curve in Figure 10. Clearly, the shape of the free-energy profile associated with the A↔B transition is strongly dependent on the solvent, the barrier-less A→B transition found in water is changed to a barrier-limited B→A transition in 85:15 ethanol/water. The apparent discrepancy between MD simulations (in water) and experimental studies

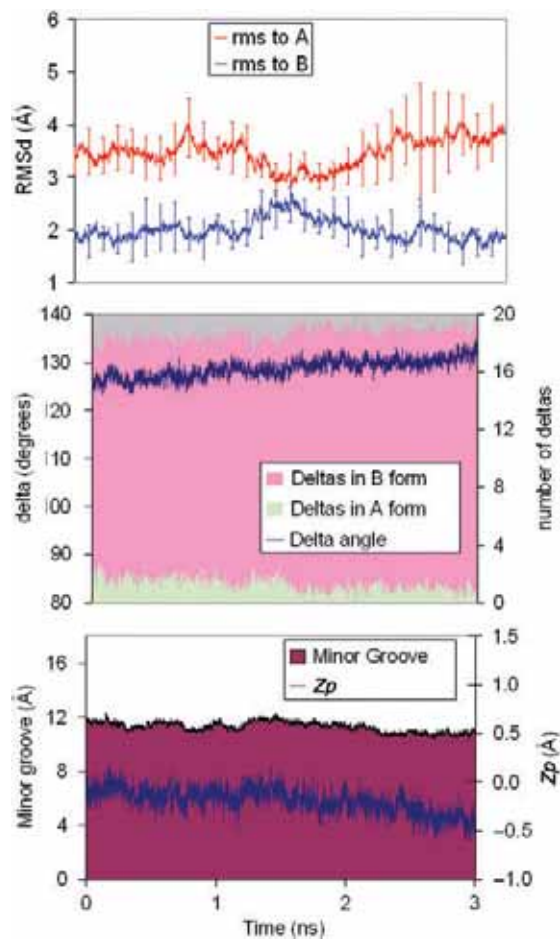


Figure 8. TOP: Average variation along the time of the RMSds (from A and B canonical structures, in Å) of 10 unbiased trajectories starting from B-form in 85:15 ethanol/water mixture. Standard deviations are shown. MIDDLE: average variations in δ and number of sugars in North conformation for the same trajectories. BOTTOM: variations of the average minor groove width (mG_w) and Z_p .

(in high ethanol concentration) of the B↔A transition can then be easily understood, and illustrates that caution is always necessary when MD simulation results are interpreted in the light of experimental measures performed under different conditions.

Analysis of geometrical variation along the tMD simulation pathway allows us to define boundaries for integration of the PMF curve. Thus, the pure A-form is obtained just in a narrow valley ($4 \text{ \AA} \geq \text{RMSd}(\lambda) \geq 3.3 \text{ \AA}$), an intermediate distorted A-conformation is obtained in the region: $3.3 \text{ \AA} \geq \text{RMSd}(\lambda) \geq 2.0 \text{ \AA}$. Finally, a canonical B-form is defined in the region: $2.0 \text{ \AA} \geq \text{RMSd}(\lambda)$. Note that these partitions agree well with the positioning of maxima and minima in the PMF curve (see Figure 10). Using these definitions, we conclude that in 85:15 ethanol/water the A-form (considering both canonical and distorted species) is 0.8 kcal/mol more stable than the B-one, while the distortion of the canonical A-form is disfavored by only 0.4 kcal/mol .

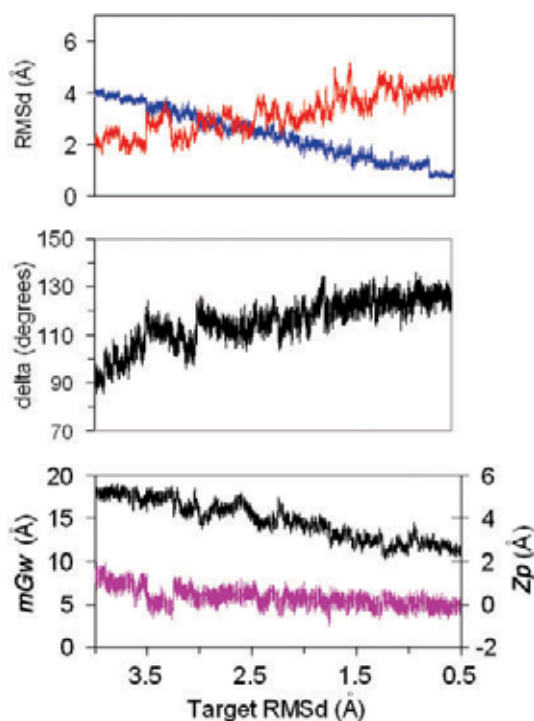


Figure 9. Variation of TOP: RMSds (from A in red and B in blue); MIDDLE: δ angle; BOTTOM: Z_p (pink) and mGw (black); obtained in targeted MD trajectories in 85:15 ethanol/water mixture using Reference set 1.

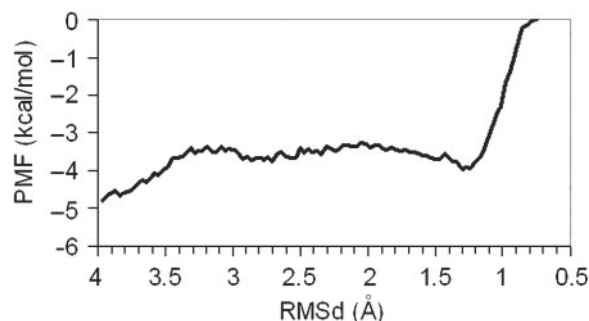


Figure 10. Potential of mean force (PMF) obtained from tMD-WHAM calculations of the A \rightarrow B transition in a mixture of ethanol (85%) and water (15%).

In summary, tMD simulations show that the presence of large amounts of ethanol produces a dramatic change in the A/B conformational equilibrium, in good agreement with experimental data (see Introduction). The A-form, which is negligibly populated (ratio $B/A=10^9$) in water, is the favored conformer in 85% ethanol (ratio $B/A=0.2$). Clearly, as discussed elsewhere, the differential screening of water and ethanol/water explains this dramatic solvent effect (19–21,56). However, the important point to emphasize here is that despite the different and numerous sources of errors, force-field simulations coupled to tMD and WHAM are able to reproduce a dramatic solvent-induced conformational transition.

ACKNOWLEDGEMENT

This work was supported by the Spanish Ministry of Education and Science (BIO2006-01602), Fundación La Caixa and the Barcelona Supercomputer Center (Computational Biology Program). AN and AP are fellows of the Spanish and Catalan Ministries of Science. Funding to pay the Open Access publication charge was provided by Institut de Recerca Biomèdica.

Conflict of interest statement. None declared.

REFERENCES

1. Saenger, W. (1984) *Principles of nucleic acid structure*. Springer-Verlag, New York, p.556.
2. Bloomfield, V.A., Crothers, D.M. and Tinoco, I. (2000) *Nucleic acids: structures, properties, and functions*. University Science Books, Sausalito, Calif., p. 794, [15] of plates.
3. Ghosh, A. and Bansal, M. (2003) A glossary of DNA structures from A to Z. *Acta Crystallogr. D Biol. Crystallogr.*, **59**, 620–626.
4. Orozco, M., Perez, A., Noy, A. and Luque, F.J. (2003) Theoretical methods for the simulation of nucleic acids. *Chem. Soc. Rev.*, **32**, 350–364.
5. Ohshima, T. (2005) *DNA conformation and transcription*. Landes Bioscience, Springer Science Business Media, Georgetown, Tex, New York, NY, p.211.
6. Olson, W.K., Gorin, A.A., Lu, X.J., Hock, L.M. and Zhurkin, V.B. (1998) DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 11163–11168.
7. Noy, A., Perez, A., Lankas, F., Luque, F.J. and Orozco, M. (2004) Relative flexibility of DNA and RNA: a molecular dynamics study. *J. Mol. Biol.*, **343**, 627–638.
8. Perez, A., Noy, A., Lankas, F., Luque, F.J. and Orozco, M. (2004) The relative flexibility of B-DNA and A-RNA duplexes: database analysis. *Nucleic Acids Res.*, **32**, 6144–6151.
9. Watson, J.D. and Crick, F.H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, **171**, 737–738.
10. Wilkins, M.H., Stokes, A.R. and Wilson, H.R. (1953) Molecular structure of deoxypentose nucleic acids. *Nature*, **171**, 738–740.
11. Klug, A. (2004) The discovery of the DNA double helix. *J. Mol. Biol.*, **335**, 3–26.
12. Franklin, R.E. and Gosling, R.G. (1953) Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate. *Nature*, **172**, 156–157.
13. Jones, S., van Heyningen, P., Berman, H.M. and Thornton, J.M. (1999) Protein-DNA interactions: a structural analysis. *J. Mol. Biol.*, **287**, 877–896.
14. Lu, X.J., Shakked, Z. and Olson, W.K. (2000) A-form conformational motifs in ligand-bound DNA structures. *J. Mol. Biol.*, **300**, 819–840.
15. Cheatham, G.M. and Steitz, T.A. (1999) Structure of a transcribing T7 RNA polymerase initiation complex. *Science*, **286**, 2305–2309.
16. Ng, H.L., Kopka, M.L. and Dickerson, R.E. (2000) The structure of a stable intermediate in the A \leftrightarrow B DNA helix transition. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 2035–2039.
17. Ivanov, V.I., Minchenkova, L.E., Minyat, E.E., Frank-Kamenetskii, M.D. and Schyolkina, A.K. (1974) The B to A transition of DNA in solution. *J. Mol. Biol.*, **87**, 817–833.
18. Basham, B., Schroth, G.P. and Ho, P.S. (1995) An A-DNA triplet code: thermodynamic rules for predicting A- and B-DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **92**, 6464–6468.
19. Srinivasan, J., Cheatham, T.E., Cieplak, P., Kollman, P.A. and Case, D.A. (1998) Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. *J. Am. Chem. Soc.*, **120**, 9401–9409.
20. Cheatham, T.E., Crowley, M.F., Fox, T. and Kollman, P.A. (1997) A molecular level picture of the stabilization of A-DNA in mixed ethanol-water solutions. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 9626–9630.

21. McConnell, K.J. and Beveridge, D.L. (2000) DNA structure: what's in charge? *J. Mol. Biol.*, **304**, 803–820.
22. Sprous, D., Young, M.A. and Beveridge, D.L. (1998) Molecular dynamics studies of the conformational preferences of a DNA double helix in water and an ethanol/water mixture: theoretical considerations of the A double left right arrow B transition. *J. Phys. Chem. B.*, **102**, 4658–4667.
23. Cheatham, T.E. and Kollman, P.A. (1996) Observation of the A-DNA to B-DNA transition during unrestrained molecular dynamics in aqueous solution. *J. Mol. Biol.*, **259**, 434–444.
24. Soliva, R., Luque, F.J., Alhambra, C. and Orozco, M. (1999) Role of sugar re-puckering in the transition of A and B forms of DNA in solution. A molecular dynamics study. *J. Biomol. Struct. Dyn.*, **17**, 89–99.
25. Banavali, N.K. and Roux, B. (2005) Free energy landscape of A-DNA to B-DNA conversion in aqueous solution. *J. Am. Chem. Soc.*, **127**, 6866–6876.
26. Foloppe, N. and MacKerell, A.D. (2000) All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Com. Chem.*, **21**, 86–104.
27. MacKerell, A.D. and Banavali, N.K. (2000) All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *J. Com. Chem.*, **21**, 105–120.
28. Pastor, N. (2005) The B- to A-DNA transition and the reorganization of solvent at the DNA surface. *Biophys. J.*, **88**, 3262–3275.
29. Schlitter, J., Engels, M., Kruger, P., Jacoby, E. and Wollmer, A. (1993) Targeted molecular-dynamics simulation of conformational change—application to the T[−]R transition in insulin. *Mol. Sim.*, **10**, 291–297.
30. Schlitter, J., Engels, M. and Kruger, P. (1990) Targeted molecular dynamics: a new approach for searching pathways of conformational transitions. *J. Mol. Graph.*, **12**, 84–90.
31. Ma, J. and Karplus, M. (1997) Molecular switch in signal transduction: reaction paths of the conformational changes in ras p21. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 11905–11910.
32. Kumar, S., Rosenberg, J.M., Bouzida, D., Swendsen, R.H. and Kollman, P.A. (1995) Multidimensional free-energy calculations using the weighted histogram analysis method. *J. Com. Chem.*, **16**, 1339–1350.
33. Dickerson, R.E. and Ng, H.L. (2001) DNA structure from A to B. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 6986–6988.
34. Beveridge, D.L. and Ravishanker, G. (1994) Molecular-dynamics studies of DNA. *Curr. Opin. Struct. Biol.*, **4**, 246–255.
35. Auffinger, P. and Westhof, E. (1998) Simulations of the molecular dynamics of nucleic acids. *Curr. Opin. Struct. Biol.*, **8**, 227–236.
36. Cheatham, T.E. and Kollman, P.A. (2000) Molecular dynamics simulation of nucleic acids. *Annu. Rev. Phys. Chem.*, **51**, 435–471.
37. Cheatham, T.E. and Young, M.A. (2000) Molecular dynamics simulation of nucleic acids: successes, limitations, and promise. *Biopolymers*, **56**, 232–256.
38. Gelpi, J.L., Kalko, S.G., Barril, X., Cirera, J., de la Cruz, X., Luque, F.J. and Orozco, M. (2001) Classical molecular interaction potentials: improved setup procedure in molecular dynamics simulations of proteins. *Proteins*, **45**, 428–437.
39. Shields, G.C., Laughton, C.A. and Orozco, M. (1997) Molecular dynamics simulations of the d(T center dot A center dot T) triple helix. *J. Am. Chem. Soc.*, **119**, 7463–7469.
40. Soliva, R., Laughton, C.A., Luque, F.J. and Orozco, M. (1998) Molecular dynamics simulations in aqueous solution of triple helices containing d(G center dot C center dot C) trios. *J. Am. Chem. Soc.*, **120**, 11226–11233.
41. Shields, G.C., Laughton, C.A. and Orozco, M. (1998) Molecular dynamics simulation of a PNA center dot DNA center dot PNA triple helix in aqueous solution. *J. Am. Chem. Soc.*, **120**, 5895–5904.
42. Perez, A., Blas, J.R., Rueda, M., Lopez-Bes, J.M., de la Cruz, X. and Orozco, M. (2005) Exploring the essential dynamics of B-DNA. *J. Chem. Theor. Comp.*, **1**, 790–800.
43. Sherer, E.C., Harris, S.A., Soliva, R., Orozco, M. and Laughton, C.A. (1999) Molecular dynamics studies of DNA A-tract structure and flexibility. *J. Am. Chem. Soc.*, **121**, 5981–5991.
44. Jorgensen, W.L. (2000) BOSS 3.4 computer program.
45. Case, D.A., Darden, T.E., Cheatham, T.E., Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Merz, K.M., Wang, B. *et al.* (2004) *AMBER8*. University of California, San Francisco.
46. Darden, T., York, D. and Pedersen, L. (1993) Particle Mesh Ewald—an N·Log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, **98**, 10089–10092.
47. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A. (1995) A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.
48. Cheatham, T.E., Cieplak, P. and Kollman, P.A. (1999) A modified version of the Cornell *et al.* force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dynam.*, **16**, 845–862.
49. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.
50. Kaminski, G.A., Friesner, R.A., Tirado-Rives, J. and Jorgensen, W.L. (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B.*, **105**, 6474–6487.
51. Ryckaert, J.P., Ciccotti, G. and Berendsen, H.J.C. (1977) Numerical-integration of Cartesian equations of motion of a system with constraints—molecular-dynamics of N-alkanes. *J. Comp. Phys.*, **23**, 327–341.
52. Lu, X.J. and 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.
53. Tolstorukov, M.Y., Ivanov, V.I., Malenkov, G.G., Jernigan, R. and Zhurkin, V.B. (2001) Sequence dependent B \leftrightarrow A transitions in DNA evaluated with dimeric and trimeric scales. *Biophys. J.*, **81**, 3409–3421.
54. Jose, D. and Porschke, D. (2004) Dynamics of the B-A transition of DNA double helices. *Nucleic Acids Res.*, **32**, 2251–2258.
55. Jose, D. and Porschke, D. (2005) The dynamics of the B-A transition of natural DNA double helices. *J. Am. Chem. Soc.*, **127**, 16120–16128.
56. Jayaram, B., Sprous, D., Young, M.A. and Beveridge, D.L. (1998) Free energy analysis of the conformational preferences of A and B forms of DNA in solution. *J. Am. Chem. Soc.*, **120**, 10629–10633.

SUPPLEMENTARY MATERIAL

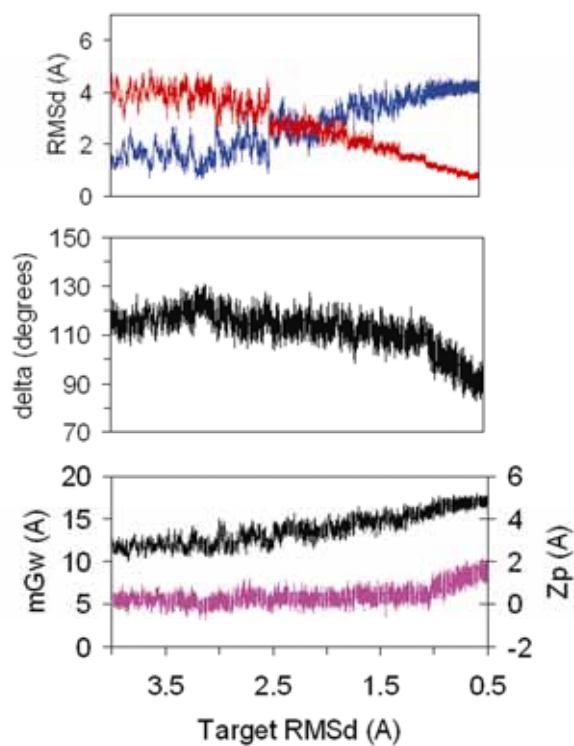


Figure S1A. Variation of TOP: RMSDs (from A in red and B in blue); BOTTOM-left: δ angle; BOTTOM-right: Zp (pink) and mGw (black); obtained in targeted MD trajectories in pure water using reference set 2.

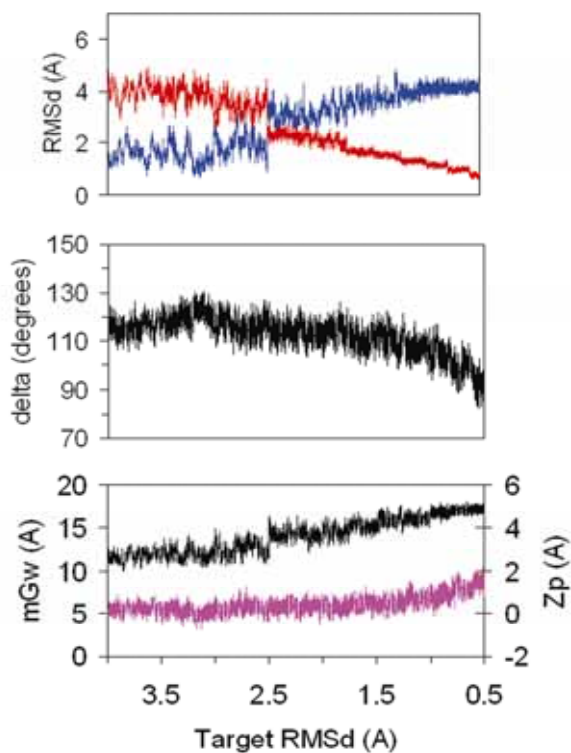


Figure S1B. Variation of TOP: RMSDs (from A in red and B in blue); BOTTOM-left: δ angle; BOTTOM-right: Zp (pink) and mGw (black); obtained in targeted MD trajectories in pure water using reference set 3.

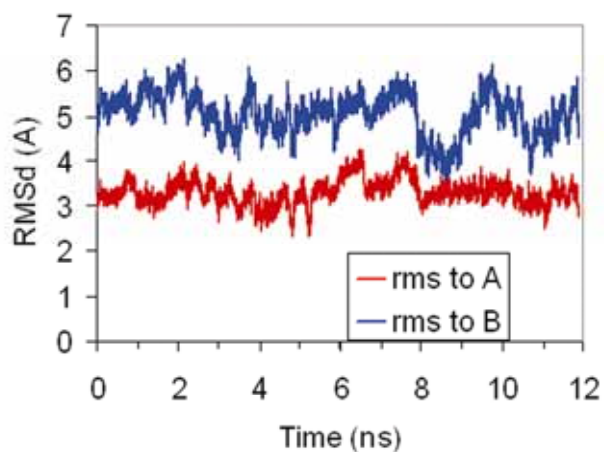


Figure S2. Variation along the time of the RMSDs (from A and B canonical structures, in Å) of a unbiased trajectory starting from A-form in 85:15 ethanol/water mixture.

5. DISCUSIÓN

5.1 Estudio Comparativo de la Flexibilidad en las Moléculas de ADN y ARN

5.1.1 Motivación

Las moléculas de ADN y ARN son fundamentales para la vida celular, si bien presentan funciones claramente diferenciadas. Mientras el ADN es responsable de almacenar y transmitir la información genética, el ARN desempeña un amplio abanico de funciones, como ser partícipe del proceso de expresión de un gen (ARNt y ARNr) o la regulación de la expresión génica (ARNi y ribozimas), aunque posiblemente existen más funciones por descubrir (189).

A pesar de su diferente papel en la célula, las diferencias químicas entre ADN y ARN son realmente pequeñas, afectando únicamente a la presencia de un grupo metilo y un grupo hidroxilo. Estas modificaciones provocan una estructura secundaria diferente cuando se trata de una doble hélice canónica (forma B para ADN y A para ARN), que permite entender como la célula puede distinguir ambas moléculas. No obstante, la estructura no es suficiente *per se* para explicar la especificidad tan elevada que presentan algunas enzimas para discernir entre ambos tipos de ácidos nucleicos, especialmente considerando que algunas enzimas reconocen el ADN forzando un cambio conformacional hacia la forma A. Por lo tanto, es razonable pensar que mecanismos discriminatorios basados en la estructura se vean complementados con otros relacionados con posibles diferencias en la deformabilidad entre ambos ácidos nucleicos. Estas reflexiones nos han llevado a estudiar la flexibilidad relativa del ADN y ARN, para determinar hasta qué punto la célula puede utilizar este mecanismo para diferenciarlas.

En principio, la comunidad científica acepta que el ADN es más flexible que el ARN, pero esta afirmación general se basa en datos experimentales indirectos. Así, experimentos electroforéticos donde se midió la *persistence length* dieron un valor de 50 nm para ADN y de 80 nm para ARN (11,12), aunque estos datos tienen un rango de error similar a esta diferencia. Por otra parte, la difracción de fibras ha demostrado que la familia B presenta varios tipos de estructuras (B, C, D y T) según las condiciones y secuencia, mientras que la

familia A es mucho más homogénea (182). Asimismo, experimentos de RMN donde se siguen los acoplamientos J (190,191) sugieren que el ADN es más flexible que el ARN, aunque debe indicarse que este tipo de medidas solamente son sensibles a distorsiones locales. Finalmente, la difracción de cristales ha puesto de relieve la capacidad del ADN de cristalizar en distintas unidades asimétricas (192), siendo así más sensible que el ARN a distorsiones creadas por el entorno cristalino (184,193), aunque experimentos más recientes han mostrado una igual variabilidad para el ARN (194,195). Además, un análisis reciente de datos cristalográficos del *Protein Data Bank* (PDB) arroja dudas sobre tal afirmación, pues no existen diferencias significativas entre los factores térmicos y el número de grupos espaciales entre ambos ácidos nucleicos (196).

El estudio de la flexibilidad del ADN y ARN ha sido el objetivo de investigaciones teóricas previas. Se indicó que el ADN es más flexible que el ARN debido a un mayor número de transiciones diédricas en el esqueleto azúcar-fosfato, argumentando que podría ser debido al diferente número de aguas estructurales (192). Así, el ARN tiene más aguas asociadas, que además forman puentes entre los átomos de los distintos residuos intracatenarios, lo que reduciría la flexibilidad global de la molécula (192,197). No obstante, tras analizar la facilidad de ambas moléculas para realizar movimientos de abertura de pares de bases, se observó que esta transición se daba con mucha más frecuencia en ARN (198), lo cual parece apuntar en la dirección opuesta y por lo tanto que el ARN es más flexible que el ADN.

En resumen, a pesar de la falta de evidencias, se asume que el ADN es más flexible que el ARN. No obstante, el término flexibilidad resulta ambiguo cuando nos referimos a moléculas tan complejas como los ácidos nucleicos, pues tienen una multitud de grados de libertad siendo sistemas altamente anisotrópicos.

5.1.2 Plan de trabajo

En este trabajo (apartado 4.1) se ha planteado caracterizar y describir las dos moléculas en términos de flexibilidad realizando dos simulaciones de DM de 10 ns a partir de las estructuras canónicas con la secuencia del dodecámero de Dickerson (DD) y con el campo de

fuerzas AMBER-parm99. En concreto, se han analizado diferentes descriptores globales capaces de discernir entre las estructuras canónicas. Además, se ha estimado la flexibilidad total mediante cálculos de entropía configuracional, juntamente con análisis de dinámica esencial, que permiten determinar si la variabilidad deriva de movimientos locales o bien de movimientos globales. Para obtener una visión intuitiva del patrón de flexibilidad de ambas moléculas, se ha estudiado la deformabilidad asociada a los parámetros helicoidales. Finalmente, se ha investigado el papel de las aguas estructurales y de los iones sobre la diferente elasticidad de los dos ácidos nucleicos.

5.1.3 Discusión de los resultados principales

El análisis de los descriptores globales no da una respuesta clara sobre la flexibilidad relativa de ambas moléculas. El RMSd promedio a lo largo de la trayectoria y su desviación estándar en relación a la estructura promedio da un resultado equivalente en los dos casos (tabla 1). Si miramos la distancia C1'-C1', el ARN presenta mayor variabilidad temporal, aunque es más independiente de secuencia, mientras que el ADN presenta mayor estabilidad en relación a su valor promedio a lo largo de la simulación, pero también mayor variación en relación a los distintos pasos (figura 1). Finalmente, la anchura del surco menor presenta mayor fluctuación para el ADN, mientras que la anchura del surco mayor fluctúa más para el ARN (figura 2).

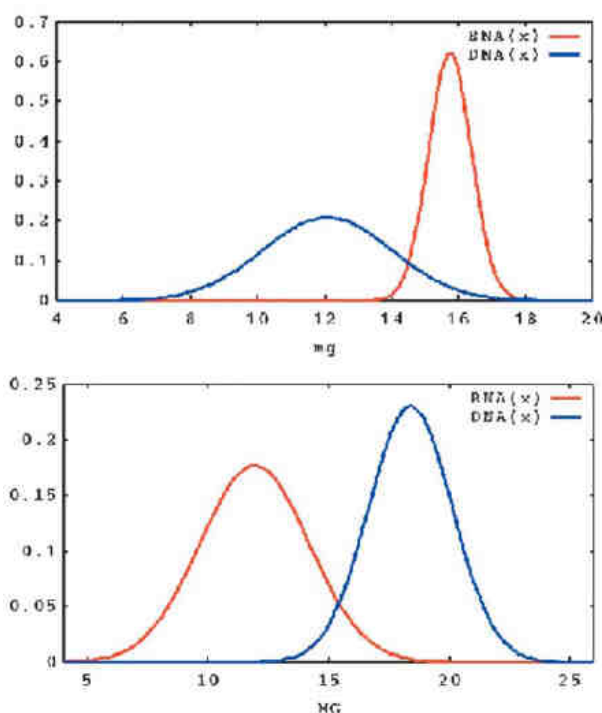


Figura 2. Distribuciones (ajustadas a gaussianas normalizadas) de la anchura (en Å) del surco menor (arriba) y mayor (abajo) de las trayectorias de DM de ADN y ARN (apartado 4.1)

El valor de entropía (extrapolado a tiempo infinito, S_{∞}) es un 11% mayor para ADN que para ARN. No obstante, si consideramos sólo los 10 primeros componentes principales, la diferencia entre ambas moléculas se reduce al 1%, y considerando sólo los 3 primeros el ARN es un 4% más flexible que el ADN. Por tanto, se puede concluir que el ADN es más flexible gracias a una multitud de movimientos deformables, en tanto que la flexibilidad del ARN recae en un pequeño número de movimientos muy deformables (tabla 2). Al analizar la contribución a la entropía del esqueleto azúcar-fosfato y de las bases nitrogenadas, se observa que es en los primeros donde se dan las diferencias entre los dos ácidos nucleicos. Además, la adición de valores de las partes se aproxima notablemente al valor global de la molécula, confirmándose así la relativa independencia de los movimientos de bases y esqueleto. De esta manera, podemos decir que el ADN es más elástico que el ARN debido a la mayor cantidad de movimientos locales del esqueleto.

| | $S(t=9 \text{ ns})$ | $S(t=\infty)$ | $S(3)$ | $S(10)$ |
|-------------------------|---------------------|---------------|--------|---------|
| <i>Todos los átomos</i> | | | | |
| ADN | 2.0067 | 2.14(0.05) | 0.0328 | 0.0988 |
| ARN | 1.8228 | 1.90(0.03) | 0.0342 | 0.0976 |
| <i>Nucleobases</i> | | | | |
| ADN | 0.8946 | 0.91(0.01) | 0.0286 | 0.0849 |
| ARN | 0.8893 | 0.91(0.02) | 0.0300 | 0.0860 |
| <i>Esqueleto</i> | | | | |
| ADN | 1.3424 | 1.43(0.07) | 0.0320 | 0.0962 |
| ARN | 1.1429 | 1.18(0.03) | 0.0333 | 0.0944 |

Tabla 2. Entropías intramoleculares (kcal/molK) calculadas con el método de Schlitter. Se han considerado todos los componentes principales, así como los tres y diez primeros y la extrapolación a tiempo infinito (apartado 4.1).

En el análisis de dinámica esencial se han calculado los índices de similitud usando tanto los 10 como los 500 primeros modos esenciales. Si tenemos en cuenta sólo los 10 primeros, se incluyen los subespacios creados por aquellos movimientos más globales. En cambio, al

considerar los 500 primeros, los espacios son mucho más amplios e incluyen un gran número de pequeñas transiciones locales. Los valores de autosimilitud con 10 modos del ADN son más bajos que los del ARN, mientras que son similares para los 500 modos, lo cual indica que se necesita un mayor número de modos para describir propiamente la flexibilidad del ADN, reflejando así la mayor complejidad dinámica de éste comparada con la del ARN (tabla 3).

| <i>Dúplexes</i> | $\gamma_{\text{ARN/ADN}}$ | $\gamma_{\text{ARN/ARN}}$ | $\gamma_{\text{ADN/ADN}}$ | $\kappa_{\text{ARN/ADN}}$ |
|-----------------|---------------------------|---------------------------|---------------------------|---------------------------|
| 10 | 0.564 | 0.860 | 0.794 | 0.682 |
| 500 | 0.744 | 0.873 | 0.878 | 0.850 |

Tabla 3. Índices de similitud absolutos (γ) y relativos (κ) para los movimientos esenciales del ADN y ARN (apartado 4.1).

El cálculo de las constantes de elasticidad para los movimientos esenciales confirma la mayor complejidad dinámica del ADN. En la figura 4 se observa que el ARN presenta las constantes más bajas para los 3 primeros modos, mientras que en el 4º modo las líneas se cruzan, siendo el ADN más deformable a partir de este punto. En consecuencia, se puede afirmar que el ARN presenta más diferencia entre sus modos esenciales, mientras que el ADN contiene un patrón de elasticidad más degenerado, donde la dinámica aportada por movimientos locales no se puede despreciarse.

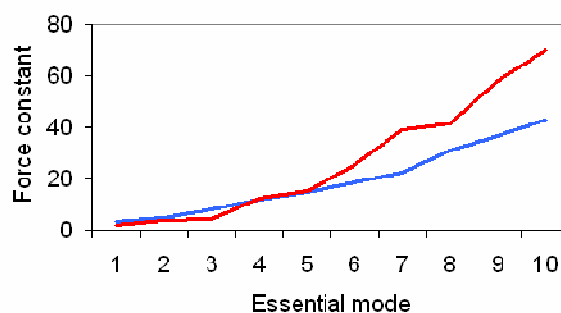


Figura 4. Constantes de fuerza (en cal/molK) asociadas a los movimientos esenciales del ADN (azul) y ARN (rojo) (Notar que la figura 4 del apartado 4.1 contiene un error tipográfico: el eje y se tiene que multiplicar por un factor de 10 (apartado 4.2)).

Desde nuestro conocimiento, la dinámica esencial se había utilizado para eliminar la redundancia de los grados de libertad que perfilan el espacio conformacional de las

biomoléculas. De esta manera, se define el subespacio esencial donde unos pocos modos explican la mayor parte de la variabilidad (33,76-80). No obstante, en este estudio dicho análisis se aprovecha como herramienta para describir el patrón de flexibilidad, permitiendo así describir la partición de la variabilidad total del sistema en forma bien degenerada o bien jerárquica. Paralelamente, en otro trabajo de nuestro grupo, estos análisis se han aplicado al estudio de las estructuras experimentales depositadas en el PDB (196), donde los resultados apuntan en la misma dirección que los mostrados aquí.

Para obtener una visión intuitiva del patrón de flexibilidad de ambas moléculas, se han calculado las constantes elásticas relativas a las deformaciones respecto a los parámetros helicoidales globales y locales. Los resultados muestran la alta anisotropía de ambas moléculas y, en consecuencia, el complejo escenario de su deformabilidad. Las constantes de elasticidad globales indican que el *twist* en ADN es más flexible que en ARN, el *roll* es similar en ambas moléculas, y que el ARN es más flexible que el ADN en cuanto a *tilt* y *stretch* (tabla 4). En el caso de las constantes de elasticidad locales, el ARN es sobre 190%, 75%, 35% y 12% más rígido que el ADN para *twist*, *slide*, *shift* y *rise*, respectivamente. Por el contrario, el ADN es más rígido que el ARN en términos de *roll* y *tilt* en 10% y 20% respectivamente. Cuando se tienen en cuenta los términos cruzados, los valores apenas cambian (tablas 5 y 6), pues la matriz de covarianzas está significativamente dominada por la diagonal.

| <i>Global</i> | <i>Tilt</i> | <i>Roll</i> | <i>Twist</i> | <i>Stretch</i> | | |
|---------------|-------------|-------------|--------------|----------------|--------------|-------------|
| ADN | 0.0043 | 0.0059 | 0.0056 | 1.50 | | |
| ARN | 0.0030 | 0.0058 | 0.0129 | 0.80 | | |
| <i>Local</i> | <i>Tilt</i> | <i>Roll</i> | <i>Twist</i> | <i>Shift</i> | <i>Slide</i> | <i>Rise</i> |
| ADN | 0.031 | 0.016 | 0.016 | 1.094 | 1.564 | 5.174 |
| ARN | 0.028 | 0.013 | 0.046 | 1.484 | 2.709 | 5.814 |

Tabla 4 y 5. Constantes elásticas respecto los parámetros helicoidales globales y locales (constantes angulares en kcal/mol grado² y constantes traslacionales en kcal/mol Å²) (apartado 4.1).

En términos generales, el ARN tiende a ser un poco más elástico en los parámetros globales, mientras que el ADN lo es más para los parámetros locales. Para confirmar este resultado, se perturbaron al azar los parámetros helicoidales dentro del margen [promedio ±

2* desviación estándar] con la intención de evaluar la elasticidad molecular. La energía elástica local de las conformaciones generadas es claramente más elevada para el ARN, en tanto que no hay diferencias claras en la energía global (figura 5). No obstante, queda claro que el concepto general de flexibilidad tiene poco sentido y que hay que considerar en cualquier caso el parámetro helicoidal en cuestión. Así, por ejemplo, el ADN es más deformable según el *twist* aunque éste sea global, mientras que el ARN es más deformable según el *roll*, aunque éste sea local.

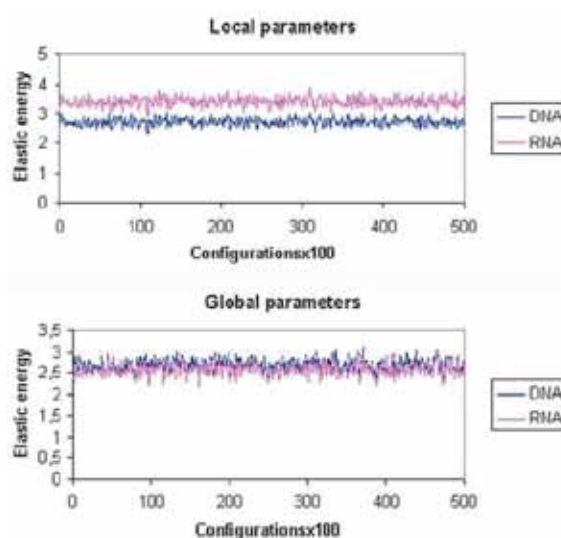


Figura 5. Energía elástica (en kcal/mol) asociada a las distorsiones de los parámetros helicoidales globales y locales. Con el fin de reducir el ruido, se representan valores promedio de cada 100 configuraciones (apartado 4.1).

Las diferentes propiedades de hidratación e interacción de iones en ambas moléculas ha sido la razón dada por diferentes autores para explicar los diferentes patrones de flexibilidad (192,197). Además, nuestros resultados indican que las diferencias recaen en el esqueleto azúcar-fosfato y que casi no hay acoplamiento entre los movimientos del esqueleto y de las bases. Esto sugiere que el *stacking* es una interacción no discriminante en términos de flexibilidad, mientras que las interacciones electrostáticas deben tener un papel destacado. Por lo tanto, el papel apantallante del solvente debe ser clave para perfilar sus propiedades de elasticidad.

Los mapas de solvatación son distintos cuando comparamos las dos moléculas (figura 6). En el surco menor del ADN existe un mayor número de aguas estructurales, comprobando así la espina de hidratación descrita en experimentos de difracción de rayos X (116). El número de aguas asociadas a la primera capa de hidratación del ADN es 10.4 por nucleótido, mientras que en el ARN es 12.8, pudiendo atribuirse dicha diferencia al grupo hidroxilo presente en el esqueleto azúcar-fosfato del ARN (tabla 7). Cabe decir que dichos valores coinciden con los obtenidos a partir de datos cristalográficos (116) y en estudios teóricos previos (192,197). Al analizar los tiempos de residencia máximos de las aguas, en el surco menor los más largos corresponden al ADN, mientras que en el surco mayor son para el ARN, si bien no se observan diferencias significativas entre ambos ácidos nucleicos en un análisis detallado a nivel de tipo de átomo (tabla 8).

| | Esqueleto | Surco menor | Surco mayor | Surcos (ambos) | Total |
|-----|------------------|--------------------|--------------------|-----------------------|--------------|
| ADN | 10.4 | 6.5 | 8.4 | 14.9 | 25.3 |
| ARN | 12.8 | 6.1 | 9.3 | 15.4 | 28.2 |

Tabla 7. Número de aguas promedio localizadas en cada par de bases de ADN y ARN. Se considera contacto cuando el oxígeno del agua está a menos de 3.5 Å del heteroátomo en cuestión (apartado 4.1)

| <i>Esqueleto</i> | O2' | O4' | O3' | O1P | O2P | O5' |
|--------------------|-------------|-------------|-------------|---------------|-------------|-------------|
| ADN | - | 617 | 270 | 518/668 | 410/298 | 266 |
| ARN | 244 | 171 | 113 | 234/723 | 510/600 | 388 |
| <i>Surco menor</i> | C O2 | G N3 | G N2 | T/U O2 | A N3 | |
| ADN | 600/353 | 438 | 534 | 521 | 571 | |
| ARN | 200/273 | 145 | 244 | 245 | 175 | |
| <i>Surco mayor</i> | C N4 | G O6 | G N7 | T/U O4 | A N6 | A N7 |
| ADN | 297 | 537 | 352/121 | 402/351 | 563 | 558/415 |
| ARN | 332 | 54 | 465/655 | 374/449 | 442 | 500/2006 |

Tabla 8. Tiempos de residencia máximos (en ps) para moléculas de agua y cationes (itálica) unidos a grupos polares de ADN y ARN (apartado 4.1).

Nuestra conclusión es que no hay una correlación obvia entre el número de aguas estructurales y sus tiempos de residencia con el grado de flexibilidad. De hecho, un análisis similar para el caso de los cationes conduce a la misma conclusión. De hecho, tanto en ADN

como en ARN, precisamente el número de aguas/cationes y sus tiempos de residencia son siempre mayores para el surco más flexible: el menor en ADN y el mayor en ARN (tabla 7, 8 y figura 2). Aunque sería necesario realizar simulaciones más largas, los resultados sugieren que ambas moléculas tienen una capacidad limitada de coordinar Na^+ y aguas. De esta manera, el solvente se puede reorganizar rápidamente delante de las transiciones conformacionales, ya que éstas son suficientemente lentas. Por ello, se propone que la presencia del solvente con cierto grado de organización no provoca mayor rigidez.

5.1.4 Conclusión

Los resultados indican que el patrón de flexibilidad de los dúplexes de ADN y ARN es realmente complejo debido al elevado número de grados de libertad, que introducen una alta anisotropía, si bien existe una deformabilidad diferencial dependiendo de la variable en cuestión. El dúplex de ADN es globalmente más flexible que el de ARN debido a la multitud de movimientos notablemente deformables del esqueleto. No obstante, en los movimientos globales, donde se abarca toda la molécula, el ARN presenta mayor flexibilidad. Asimismo, en los parámetros helicoidales locales, el ADN es más deformable para el *twist*, *slide* y *shift*, ligeramente más deformable en el *rise*, y ligeramente más rígido en el *roll* y *tilt*. Por otro lado, en los parámetros globales, el ARN es más elástico en el *tilt* y *stretch*, pero más rígido en el *twist*. Una conclusión importante es, pues, que se precisa cierta cautela a la hora de utilizar la palabra flexibilidad en términos absolutos para sistemas tan complejos como los ácidos nucleicos.

5.2 Estructura, Propiedades de Reconocimiento y Flexibilidad del Híbrido ADN·ARN

5.2.1 Motivación

El dúplex heterogéneo ADN·ARN se forma en procesos cruciales de la vida celular como la replicación, la transcripción y la retrotranscripción. A veces aparece en forma de quimera, donde el ARN y ADN coexisten en la misma cadena, como es el caso de los fragmentos de Okazaki. Además, es el sustrato de la RNasa H, enzima crítico en la terapia antisentido, que es inespecífico de secuencia pero que no puede actuar sobre los dúplex puros de ARN y ADN.

Las técnicas de difracción, tanto de fibras como de cristales, han encontrado en el dúplex híbrido una forma A, con todos los azúcares en conformación Norte (199-204). Sin embargo, las técnicas de CD y RMN sugieren una estructura intermedia A/B, con muchas características tipo A, pero con alteraciones en los surcos que la hacen cercana a la forma B debido al *puckering* Sur de la cadena ADN (152-154,205-209). Estas características han sido confirmadas por diversos estudios teóricos (163,165,192), así como por una estructura cristalina del complejo RNasa H:ADN·ARN (158). Concretamente, la anchura intermedia del surco menor en el dúplex híbrido ha sido sugerida como clave para explicar la especificidad de la RNasa H (152,154,158), aunque diversas estructuras quimera susceptibles a RNasa H han reflejado propiedades dispares en este surco (135,136) sugiriendo que existen otros factores para explicar su actividad.

Este estudio (apartado 4.2), pues, persigue identificar las propiedades de reconocimiento y flexibilidad del dúplex ADN·ARN, comparándolas con las de los homopolímeros de ADN y ARN, con el fin de averiguar qué factores, aparte de la estructura promedio, explican la especificidad de la RNasa H.

5.2.2 Plan de trabajo

Se han simulado trayectorias con la secuencia del DD, usando como estructuras inicial i) la estructura de difracción de cristal 1FIX (202), y ii) la estructura de difracción de fibras (210). Se ha estudiado la convergencia de ambas trayectorias, y posteriormente se ha analizado la distribución de los parámetros helicoidales y de los ángulos diedros, así como las propiedades de reconocimiento molecular. También se ha investigado el patrón de flexibilidad del dúplex híbrido mediante dinámica esencial, así como la elasticidad relacionada con los parámetros helicoidales. Dichos resultados han sido interpretados con el fin de obtener una explicación mecánica acerca de la especificidad de la RNasa H.

5.2.3 Discusión de los resultados principales

Ambas trayectorias empiezan con una conformación tipo A y en menos de 1 ns se aprecia una transición hacia una estructura intermedia A/B, más cerca de A que de B, donde se explora la región configuracional cercana a la estructura de RMN 1EFS (209) (figura 1). Así, pues, las simulaciones están principalmente de acuerdo con las estructuras de RMN, apartándose de los resultados cristalográficos.

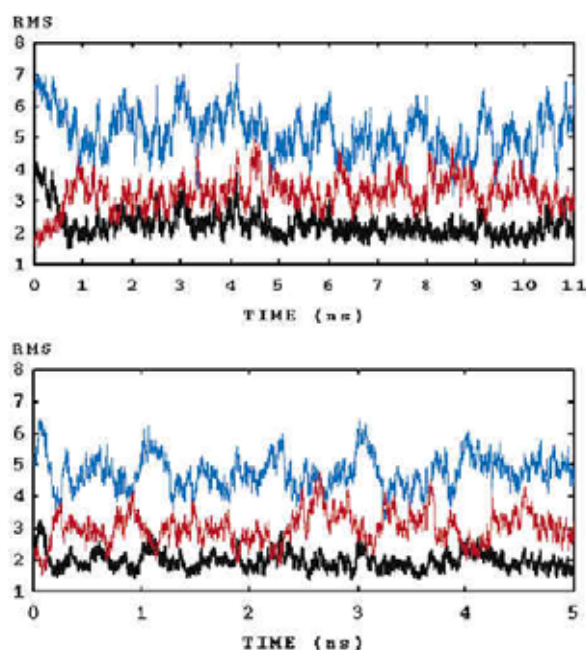


Figura 1. RMSd (en Å) de ambas trayectorias del híbrido, con 1FIX (arriba) y difracción de fibras (abajo) como estructuras iniciales, respecto a la forma A (rojo), a la B (azul) y a la de tipo RMN (1EFS, negro) (apartado 4.2)

El carácter intermedio A/B del híbrido queda patente en los parámetros helicoidales globales (figura 3). Respecto a los parámetros locales, existe una distribución intermedia en el caso del *roll*, mientras que para el *slide* y *twist* es más parecido al ARN, sin observarse diferencias significativas en los otros parámetros (figura 2).

Respecto a los ángulos diedros del esqueleto, se observa una asimetría entre las cadenas del dúplex, manteniendo cada una su naturaleza intrínseca (figura 4). Así, el *puckering* de la cadena de ADN del híbrido muestra un comportamiento muy parecido a las cadenas individuales del dúplex de ADN (figura 5), predominando la conformación Sur, aunque se detecta una población residual de conformémeros Norte.

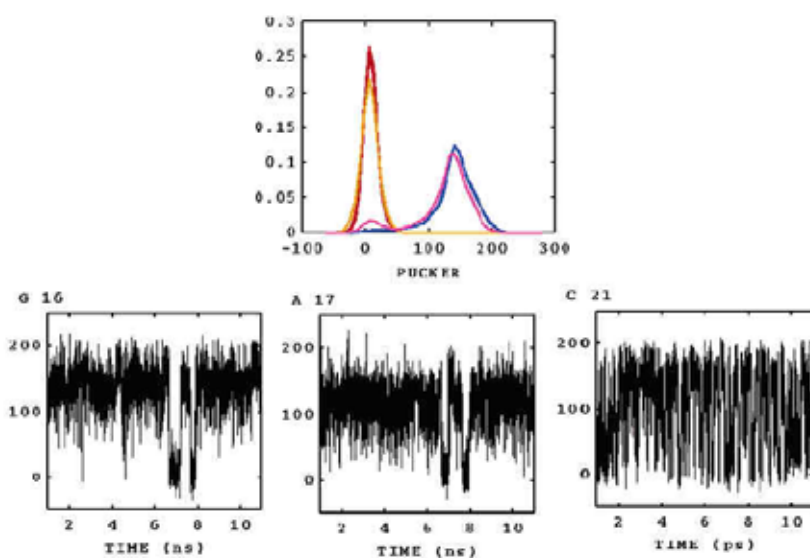


Figura 5. Arriba: distribución del *puckering* (en grados) considerando las cadenas por separado de las trayectorias de ADNds (azul), ARNds (rojo) e híbrido (violeta la cadena ADN y naranja la cadena ARN). Abajo: evolución del *puckering* a lo largo de la simulación del híbrido de algunos nucleótidos de la cadena ADN (apartado 4.2).

Las diferentes configuraciones en el esqueleto de las dos cadenas del dúplex provocan una anchura intermedia de los surcos menor y mayor (figura 6), afectando asimismo a las propiedades de reconocimiento. En el caso de Na^+ , la unión preferente se encuentra distribuida de manera asimétrica cerca de la cadena de ARN en el surco mayor. Los rasgos de hidratación en el surco menor son intermedios, habiendo cierta espina de hidratación, pero no tan importante como en el ADN (figura 7). El número de aguas estructurales de cada cadena

del híbrido se mantiene fiel a las cadenas correspondientes de los homopolímeros, 12.6 en las de ARN y 10.5 en las de ADN.

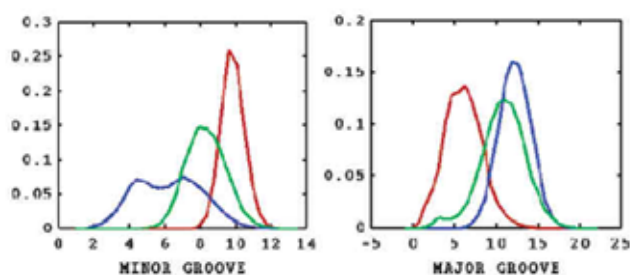


Figura 6. Distribuciones de las anchuras del surco menor y mayor (en Å) para las trayectorias ADNds (azul), ARNds (rojo) e híbrido (verde) (apartado 4.2).

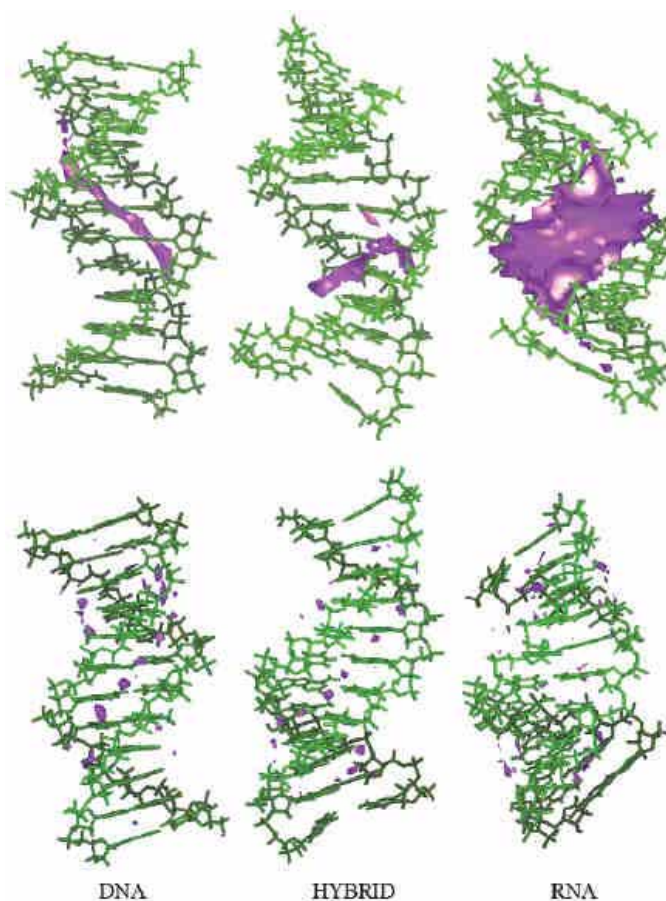
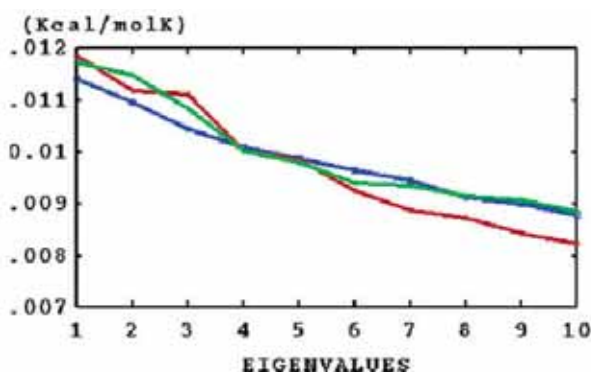


Figura 7. Arriba: potencial de interacción clásico molecular para Na^+ (en violeta la zona de menos de -3 kcal/mol). Abajo: mapa de solvatación (en violeta los puntos con más densidad de 2.5 g/cm^3) (apartado 4.2).

La entropía del dúplex ADN·ARN es intermedia entre la de ADN y de ARN (tabla 1), aunque los valores de los principales modos esenciales muestran un patrón de flexibilidad único para el dúplex híbrido. En los tres primeros modos los dúplexes de ARN y ADN·ARN son más deformables, existiendo un cruzamiento a partir de ese punto, y de hecho los dúplexes de ADN y ADN·ARN son más deformables hasta el décimo modo (figura 8). Cuando se consideran las cadenas por separado, se observa que cada cadena mantiene su deformabilidad intrínseca. Así, las cadenas de ADN y ARN del híbrido tienen un valor de 1.11 y 1.00 kcal/molK, respectivamente, que coinciden con los de las mismas cadenas en los respectivos homopolímeros (tabla 2). La memoria conformacional es especialmente fuerte en la cadena de ARN, mientras que la cadena de ADN parece ligeramente más influenciada, concretamente en los tres primeros modos, donde adopta valores entrópicos similares a la cadena complementaria (figura 8).

| <i>Dúplexes</i> | ADNs | ARNs | Híbrido | |
|------------------|-------------|-------------|--------------------|--------------------|
| Todos los átomos | 2.14(0.05) | 1.90(0.03) | 2.06(0.03) | |
| Nucleobases | 0.91(0.01) | 0.91(0.02) | 0.91(0.01) | |
| Esqueleto | 1.43(0.07) | 1.18(0.03) | 1.34(0.04) | |
| <i>Cadenas</i> | ADNs | ARNs | ADN híbrido | ARN híbrido |
| Todos los átomos | 1.11(0.05) | 0.99(0.02) | 1.12(0.02) | 1.00(0.02) |
| Nucleobases | 0.49(0.01) | 0.48(0.01) | 0.49(0.01) | 0.48(0.01) |
| Esqueleto | 0.75(0.09) | 0.62(0.04) | 0.75(0.03) | 0.63(0.03) |

Tabla 1 y 2. Entropías intramoleculares extrapoladas a tiempo infinito (kcal/molK) calculadas con el método de Schlitter (apartado 4.2).



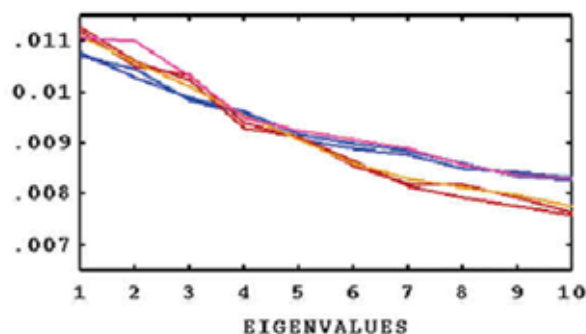


Figura 8. Arriba: entropías intramoleculares (kcal/molK) calculadas con el método de Schlitter para los movimientos esenciales de los dúplexes. Abajo: considerando las cadenas por separado. ADN en azul, ARN en rojo e híbrido en verde (la cadena ADN en violeta y la cadena ARN en naranja) (apartado 4.2).

Asimismo, los índices de similitud indican que la molécula híbrida tiene un carácter intermedio, y que las máximas diferencias se hallan entre los homopolímeros. Respecto a las cadenas individuales, las del híbrido se parecen más a las correspondientes de los dúplexes que entre ellas mismas, quedando patente la asimetría de la molécula en su comportamiento dinámico. Para la cadena ARN del híbrido, este efecto se revela considerando tanto los 10 como los 250 primeros nodos (tabla 3). Sin embargo, para la cadena ADN, no se observa una clara diferencia en los 10 primeros modos respecto a las de los dúplexes de ADN, ARN e híbrido, pero sí es más parecida a las cadenas del dúplex ADN al considerar los 250 primeros modos. Este resultado confirma la mayor adaptabilidad de la cadena ADN del híbrido a su pareja en los modos más globales. Así, pues, las propiedades intermedias dinámicas reflejan las dos distintas naturalezas de las hebras que mantienen totalmente (ARN) o parcialmente (ADN) las características de las correspondientes cadenas en los homodúplexes.

| <i>Dúplexes</i> | $\kappa_{\text{HYBRID/RNA}}$ | $\kappa_{\text{HYBRID/DNA}}$ | $\kappa_{\text{DNA/RNA}}$ | | |
|-----------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|-------------------------------------|
| 10 | 0.788 | 0.722 | 0.682 | | |
| 500 | 0.911 | 0.921 | 0.850 | | |
| <i>Cadenas</i> | $\kappa_{\text{HYB_RNA/RNA}}$ | $\kappa_{\text{HYB_RNA/DNA}}$ | $\kappa_{\text{HYB_DNA/RNA}}$ | $\kappa_{\text{HYB_DNA/DNA}}$ | $\kappa_{\text{HYB_RNA/HYB_DNA}}$ |
| 10 | 0.904 | 0.766 | 0.771 | 0.772 | 0.763 |
| 250 | 0.919 | 0.852 | 0.866 | 0.952 | 0.878 |

Tabla 3. Índices de similitud relativos (κ) considerando los 10 y 500 primeros vectores propios en los dúplexes y los 10 y 250 en las cadenas (apartado 4.2).

El patrón de elasticidad asociado a los parámetros helicoidales del híbrido es único y no simplemente el promedio entre homopolímeros. Así, el híbrido es especialmente rígido en el *tilt*, intermedio en el *twist*, igual de flexible que ARN en el *stretch* y especialmente flexible en el *roll* (tabla 4). Una fluctuación al azar de los valores helicoidales globales dentro del margen [promedio \pm 2* desviación estándar] muestra una misma deformabilidad (figura 10). En cambio si perturbamos al azar los parámetros locales, el perfil energético nos discrimina perfectamente los tres dúplexes.

5.2.4 Conclusión

Se ha sugerido previamente (152,154,158) que la anchura intermedia del surco menor observada en estructuras experimentales subyace en la especificidad del enzima. Según nuestros resultados, existe un solapamiento significativo entre ADN·ARN y el dúplex de ARN, lo que arroja dudas sobre esta hipótesis. Además, si observamos la estructura de RMN de dos quimeras sustratos de la enzima, se observan diferencias obvias en los correspondientes surcos menores (135,136) (figura 11). Así, la estructura promedio explica porqué la enzima no puede unir el dúplex de ADN, pero no es suficiente para entender porqué no actúa sobre el dúplex de ARN.

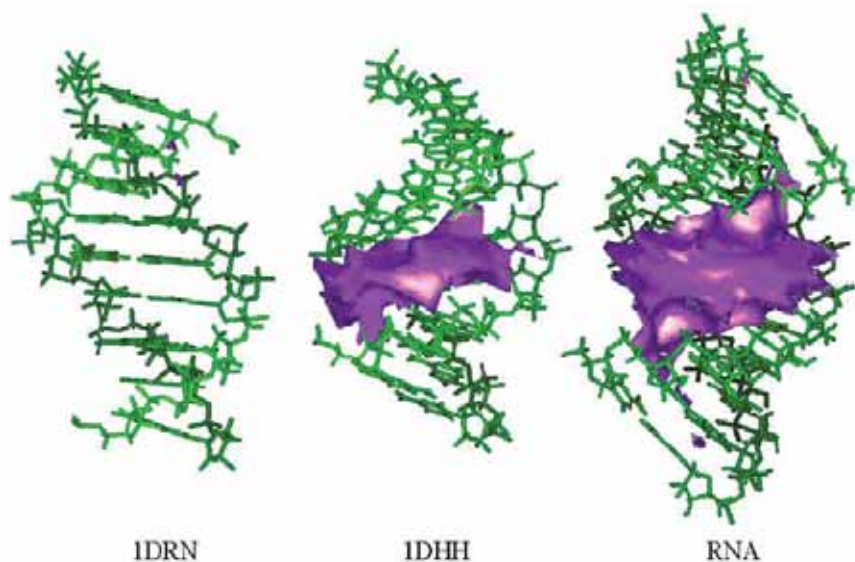


Figura 11. Potencial de interacción clásico molecular para Na^+ (en violeta la zona de menos de -3 kcal/mol) de dos estructuras quimeras susceptibles a la RNAsa H (135,136) (apartado 4.2).

Nuestra propuesta es que la principal diferencia entre ADN·ARN y el dúplex de ARN reside en la flexibilidad, especialmente a nivel local. Gracias a una estructura global de tipo A, RNasa H podría unir dichos dúplexes, pero no el de ADN. Para la catálisis, sería necesario una serie de distorsiones locales posibles para la cadena de ADN del híbrido, pero muy costosas para una cadena de ARN, pudiendo diferenciar entre el dúplex de ARN y ADN·ARN. En consecuencia, el mecanismo de especificidad hipotético otorga a la asimetría en el perfil de flexibilidad un papel clave para entender la susceptibilidad del híbrido ADN·ARN a la RNasa H.

5.3 Análisis Teórico de Dúplexes Antisentido: Determinantes de la Susceptibilidad de la RNasa H

5.3.1 Motivación

La terapia antisentido consiste en inyectar en la célula un oligonucleótido capaz de unirse específicamente a un ARNm con el fin de bloquear su expresión. Desde que se ideó hace 25 años (141) hasta la actualidad, se ha llevado a cabo una intensa investigación para mejorar las propiedades de los oligonucleótidos antisentido (AXN), dando lugar a los primeros fármacos basados en esta tecnología (142) y a otros muchos en fase de ensayos clínicos para tratar patologías como cáncer e infecciones virales (143,147).

Los AXN de primera generación persiguieron conjugar la estabilidad de unión al ARNm y la resistencia a nucleasas, aunque la mayoría no mantenía la susceptibilidad hacia la RNasa H. Ésta, al poder degradar la cadena de ARN en un dúplex híbrido (AXN·ARN), hace más eficiente la terapia antisentido, ya que permite el reciclaje del AXN. Así, los AXN de segunda generación han sido diseñados para mantener la actividad frente a RNasa H. Sin embargo, el mecanismo de especificidad de esta enzima aun está sin resolver, lo que hace complejo el diseño de futuros fármacos antisentido (13,152-159).

Nuestra hipótesis (ver apartado anterior) es que la flexibilidad asimétrica del dúplex era crucial para la especificidad de la enzima. En consecuencia, este estudio (apartado 4.3) persigue comprobar hasta qué punto esta afirmación explica la susceptibilidad de un conjunto de dúplexes AXN·ARN creados en el contexto de la estrategia antisentido.

5.3.2 Plan de trabajo

En este capítulo hemos considerado:

2'-O-Metil-ADN (AOmeN·ARN).

Es un representante de los AXN de primera generación, siendo no susceptible a la RNasa H. Según una estructura de RMN, la conformación global del dúplex AOmeN·ARN es de

tipo A, presentando todos los azúcares en *puckering* Norte (211). Dichos resultados han sido confirmados por simulaciones de DM (163).

Ácidos arabinonucleicos (AAN·ARN y F-AAN·ARN).

Son AXNs de segunda generación, donde se ha invertido la quiralidad del C2' manteniendo el grupo hidroxilo (AAN) o sustituyéndolo por un fluor (F-AAN) para conseguir más estabilidad. Estudios de CD (187,212) y de RMN (213,214) indican estructuras intermedias A/B parecidas al dúplex ADN·ARN, distinguiendo dos configuraciones de *puckering* para los arabinoazúcares, una Este y otra Sur. Un estudio previo de DM sugiere una conformación Sur para la cadena AAN, debido a un puente de hidrógeno entre C2'-OH y C5'-O, y una conformación Este para la cadena F-AAN (215). Ante la complejidad del espacio conformacional de la arabinosa, se ha parametrizado su *puckering* a nivel MP2/6-31G(d) y después se han ajustado los parámetros de torsión con el método de Monte-Carlo (38) (figura S2 del material suplementario). Los valores para estos azúcares son ligeramente distintos a los típicos de la desoxirribosa, desplazándose el mínimo hacia una región Este y disminuyendo la barrera entre ésta y la configuración Norte (figura S3). Los nuevos parámetros han sido testeados con simulaciones de 20 ns a partir de las estructuras de RMN (213,214), obteniendo unas distribuciones de *puckering* muy similares a los valores experimentales (tabla S1).

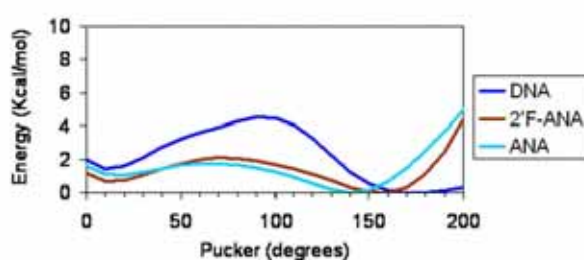


Figura S3. Comparación de los perfiles de *puckering* de la desoxirribosa (DNA), arabinosa (ANA) y 2'F-arabinosa (2'F-ANA) correspondientes a un nucleósido de timina utilizando el parm-bsc0 con los parámetros optimizados. Para dar más detalle de la zona clave Norte-Este-Sur, la zona Oeste, la cual contiene un máximo, ha sido eliminada (apartado 4.3)

C5-propinil-Y-ADN (YpADN·ARN).

Es un AXN de segunda generación con bases modificadas, donde la resolución de una estructura de RMN indica una conformación de tipo A/B sin dejar claro el *puckering* del

azúcar (165). Los parámetros para el grupo propinil se han obtenido del campo de fuerza GAFF ajustando cargas con el procedimiento RESP 6-31G(d).

Se han simulado 4 dúplexes antisentido (AXN·ARN) con la secuencia del DD y con el campo de fuerza AMBER-parm-bsc0, empezando a partir de dos estructuras iniciales distintas con el fin de estudiar sus propiedades de convergencia: i) la forma A, y ii) la estructura basada en un dúplex ADN·ARN de RMN (1EFS). Para el dúplex AOMeN·ARN, ha sido imposible obtener una simulación estable empezando con la estructura tipo RMN, mientras que para las moléculas AAN·ARN y F-AAN·ARN no ha sido posible empezar con la forma A. En estos casos, la inestabilidad siempre ha sido causada por la imposibilidad de realizar la transición conformacional en los azúcares, debido al volumen de los grupos sustituidos en el C2'. Las trayectorias de ADN·ARN y YpADN·ARN empezadas en la forma A han convergido a una conformación de tipo RMN (figura S1). En todos los casos se han confirmado los resultados experimentales. Las simulaciones estables han sido alargadas hasta obtener trayectorias de 50 ns. Además, se han realizado cálculos de DM de los tres dúplexes control (ADNds, ARNds, ADN·ARN) con la misma longitud y secuencia.

Con el fin de garantizar las conclusiones, se han simulado dos secuencias más a partir de la configuración obtenida en el vigésimo nanosegundo de la trayectoria de 50 ns, por permutación de la secuencia, a fin de garantizar que todos los dímeros de pares de bases queden representados. Estas trayectorias tienen 20 ns de longitud. Además, se han realizado simulaciones de los tres dúplexes control (ADNds, ARNds, ADN·ARN). Por lo tanto, hemos estudiado siete dúplexes distintos en tres secuencias diferentes (CGCGAATTCGCG, CATAGGCCTATG, CACAGATCTGTG), lo que hemos sintetizado en la tabla 1:

| | ADNds | ARNds | HYB | F-AAN | AAN | YpADN | AOMeN |
|--------------|-------|-------|-----|------------|------------|-------|------------|
| <i>Seq 1</i> | | | | | | | |
| RMN | -- | -- | 50 | 50 | 50 | 50 | No estable |
| Forma A | -- | 50 | 3 | No estable | No estable | 3 | 50 |
| Forma B | 50 | -- | -- | -- | -- | -- | -- |
| <i>Seq 2</i> | | | | | | | |

| | | | | | | | |
|--------------|----|----|----|----|----|----|----|
| Equil(1) | 20 | 20 | 20 | 20 | 20 | 20 | 20 |
| <i>Seq 3</i> | | | | | | | |
| Equil(1) | 20 | 20 | 20 | 20 | 20 | 20 | 20 |

Tabla 1. Longitud de las diferentes trayectorias consideradas en este estudio junto a su estructura inicial. Equil denota la estructura equilibrada de la seq 1 en el vigésimo ns (apartado 4.3)

A fin de examinar las características estructurales y dinámicas que utiliza la RNasa H, se han estudiado las propiedades de un conjunto de dúplexes antisentido, algunos susceptibles a la enzima, mientras que otros no. Se han analizado las propiedades de reconocimiento, flexibilidad y rasgos globales de los dúplexes, así como de sus cadenas por separado. Es remarcable comentar que posteriormente a la publicación del anterior artículo, apareció la estructura cristal del ADN·ARN unido a la proteína RNasa H (158), lo que aquí nos ha permitido incorporar el cálculo de la energía necesaria para deformar cada híbrido hasta esta conformación.

5.3.3 Discusión de los principales resultados

Las gráficas de RMSd de las trayectorias en relación con la forma A, B y RMN del híbrido (1EFS) (figuras 1 y S4-S6) permiten clasificar los dúplexes en tres grupos estructurales de manera independiente a la secuencia:

- Tipo B: Muestran el espacio conformacional muy cercano a la forma B y lejos de la A, siendo adoptado por los dúplexes ADN, que no pueden unirse a la RNasa H.
- Tipo A: Próximos a la conformación A, no muy lejos de la estructura del híbrido determinado por RMN (1EFS), pero lejos de la B; los dúplexes ARNs y AOMeN·ARN, que son capaces de unirse a la enzima pero no son degradados, pertenecen a este grupo.
- Tipo A/B: Cercanos a la estructura tipo 1EFS, no muy alejada de la A y lejos de la B; la adoptan todos los dúplexes susceptibles a RNasa H: AAN·ARN, F-AAN·ARN, YpADN·ARN y ADN·ARN.

A excepción del ADNds, los parámetros helicoidales muestran que el resto de dúplexes tienen un comportamiento parecido al ARNds, sobretodo en términos de *twist* y *slide* y con excepción del *roll* (tabla 1). El efecto de estos parámetros en relación a los distintos pasos depende del tipo de dúplex. Así, los dúplexes tipo B y A presentan alta y baja variabilidad, respectivamente, siendo intermedia las de tipo A/B. Por otro lado, los dúplexes que mantienen la actividad de RNasa H poseen una anchura y flexibilidad intermedia en el surco menor, y una región óptima de interacción con Na^+ asimétrica en el surco mayor cerca de la cadena ARN (figura 2 y 3), mientras que el dúplex AOMeN·ARN manifiesta un comportamiento parecido al ARNds. Así, la estructura del surco menor podría ser un elemento importante para la especificidad de RNasa H.

En base a los cálculos de entropía (figura 4) y a las constantes elásticas sobre los parámetros helicoidales (tabla 2), se observa que el dúplex AOMeN·ARN es el más rígido, seguido de ARNds, los dúplexes tipo A/B y finalmente ADNds. Este último es más flexible en términos de *twist*, *slide* y *rise*, mientras que no hay diferencias significativas en otros términos (*roll*, *tilt* y *shift*). Los pasos YpR son los más flexibles en todos los dúplexes, mientras que depende del tipo de molécula para los pasos RpY y RpR. Por ejemplo, para el paso GpC el ADNds presenta una constante de *twist* rígida en relación a su promedio, mientras que los dúplexes A/B muestran constantes relativamente flexibles. En consecuencia, la deformabilidad característica de un paso determinado no sólo depende de las interacciones de *stacking*, sino que la configuración del esqueleto también tiene relevancia.

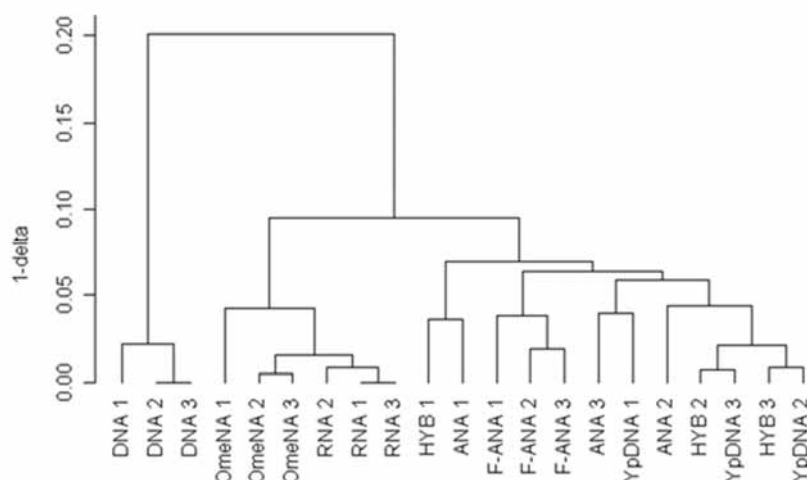


Figura 7A. Análisis de *cluster* de las diferentes dúplexes y secuencias basado en los índices de similitud ponderados (distancias correspondientes a $1-\delta$, ver apartado 4.3).

El cálculo de las constantes elásticas relativas a los modos esenciales muestra un perfil jerarquizado para los dúplexes tipo A, siendo más degenerado para el resto, especialmente en ADNds (figura 5). El análisis *cluster* de los índices de similitud relativos confirma los tres grupos estructurales independientes a las diferentes secuencias examinadas (figura 7). En concreto, primero se distinguen dos grupos generales, uno con los dúplexes tipo B y otro con el resto, y posteriormente, el segundo grupo se divide en dúplexes tipo A y A/B. Así, los dúplexes diana a RNasa H presentan un patrón de flexibilidad característico que les permite agruparse mediante los índices de similitud relativos. Asimismo, presentan una entropía global intermedia y un patrón de deformabilidad bastante degenerado.

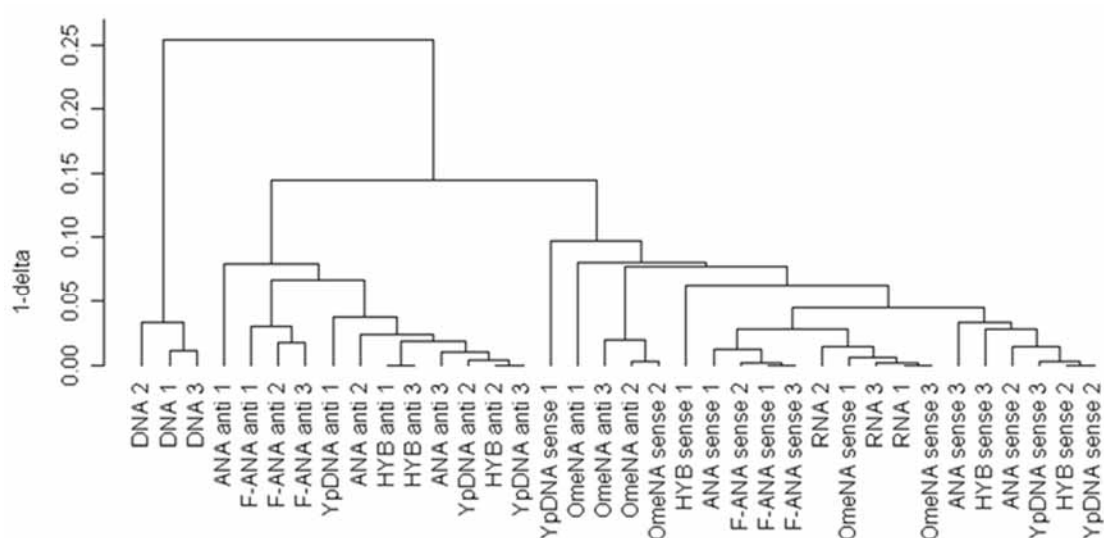


Figura 7B. Análisis de *cluster* de las diferentes cadenas y secuencias basado en los índices de similitud ponderados (distancias correspondientes a $1-\delta$, ver apartado 4.3)

Las constantes de deformabilidad de los modos esenciales de las cadenas sentido muestran un comportamiento parecido en todos ellos (de tipo A), excepto lógicamente en la cadena ADN (figura 6). En cambio, para las cadenas antisentido, el patrón de los dúplexes diana es degenerado, mientras que el de los dúplexes no diana es jerarquizado, especialmente en AOMeN·ARN. Estas propiedades dinámicas se reflejan en el *clustering*, que por un lado agrupa las cadenas antisentido de los dúplexes diana, y por otro todas las cadenas sentido y

antisentido de los dúplexes no diana. Las cadenas de los dúplexes ADN quedan concentradas en un tercer grupo (figura 7). Ello indica que los dúplexes susceptibles a RNasa H presentan una asimetría dinámica característica, lo cual se confirma con las distribuciones de los ángulos diedros del azúcar y glucosídico (figura 8). Así, los arabinoazúcares presentan dos valores preferidos, Este y Sur, tal y como apuntaban las estructuras RMN. Además, mientras que los dúplexes tipo A no suelen presentar transiciones relativas a los ángulos diedros α/γ y ϵ/ζ , existe un número considerable en los dúplexes A/B, especialmente en las cadenas antisentido (tabla 3).

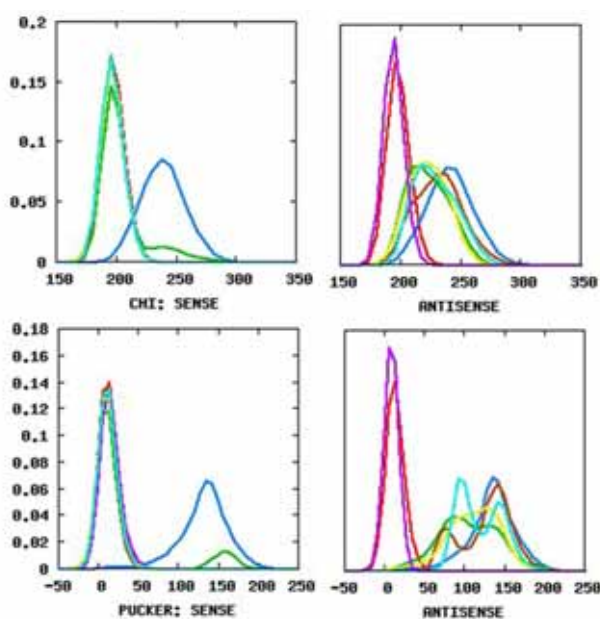


Figura 8. Distribuciones del *puckering* y la torsión glucosídica (χ) (en grados) considerando las cadenas sentido y antisentido por separado para las trayectorias ADNds (azul), ARNds (rojo), ADN·ARN (verde), F-AAN·ARN (marrón), AAN·ARN (azul claro), YpADN·ARN (amarillo) y AOMeN·ARN (violeta) (apartado 4.3).

Finalmente, utilizando el subespacio esencial construido con los 10 primeros modos se ha intentado alcanzar la conformación de la molécula híbrida unida a RNasa H, calculando la mínima distancia de Mahalanobis con su correspondiente energía elástica (tabla 4 y figura 9). Los dúplexes tipo A adquieren la estructura activa de la cadena sentido, pero no la antisentido, ya que su rigidez hace muy costosa la deformación local de esta cadena para obtener una unión productiva con RNasa H. Para los dúplexes tipo B se observa el resultado opuesto, mientras que aquéllos A/B alcanzan la conformación activa en ambas cadenas con valores bajos de energía de deformación.

| <i>Distancia de Mahalanobis</i> | | Seq. 1 | Seq. 2 | Seq. 3 | Av distance | Energy |
|---------------------------------|--------------------|--------|--------|--------|-------------|--------|
| ADN | <i>antisentido</i> | 3.41 | 1.96 | 2.54 | 2.69 | 2.17 |
| | <i>sentido</i> | 7.59 | 6.04 | 7.26 | 6.96 | 14.53 |
| ARN | <i>antisentido</i> | 7.82 | 8.60 | 8.34 | 8.25 | 20.3 |
| | <i>sentido</i> | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| HYB | <i>antisentido</i> | 1.92 | 1.35 | 1.49 | 1.59 | 0.75 |
| | <i>sentido</i> | 2.20 | 1.83 | 1.12 | 1.72 | 0.89 |
| F-AAN | <i>antisentido</i> | 1.68 | 2.16 | 1.96 | 1.90 | 1.1 |
| | <i>sentido</i> | 2.09 | 2.25 | 1.93 | 2.09 | 1.31 |
| AAN | <i>antisentido</i> | 1.80 | 1.34 | 1.62 | 1.59 | 0.75 |
| | <i>sentido</i> | 2.53 | 3.80 | 2.20 | 2.84 | 2.42 |
| YpADN | <i>antisentido</i> | 1.55 | 1.85 | 1.62 | 1.67 | 0.83 |
| | <i>sentido</i> | 2.03 | 2.17 | 1.75 | 1.98 | 1.18 |
| AOmeN | <i>antisentido</i> | 8.74 | 9.84 | 7.23 | 8.60 | 22.1 |
| | <i>sentido</i> | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

Tabla 4. Distancias de Mahalanobis (en Å) de las cadenas *antisentido* y *sentido* (itálica) entre las estructuras relajadas y las unidas a la RNasa H dentro del subespacio esencial construido con los 10 primeros modos (ver apartado 4.3).

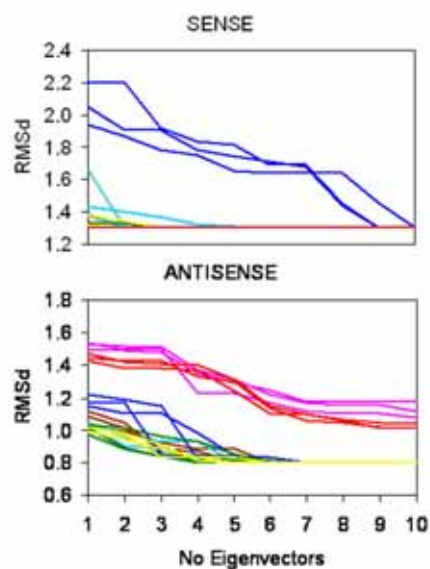


Figura 9. RMSd (en Å) de las cadenas de ADNds (azul), ARNds (rojo), ADN·ARN (verde), F-AAN·ARN (marrón), AAN·ARN (azul claro), YpADN·ARN (amarillo) y AOmeN·ARN (violeta) respecto la estructura unida a RNasa H, alcanzada a medida que se incrementan el número de vectores propios (del 1 al 10). El *cutoff*

se define según las fluctuaciones térmicas de las trayectorias ADN·ARN. Las diversas líneas en el mismo color corresponden a las 3 secuencias consideradas (ver apartado 4.3).

5.3.4 Conclusión

Los resultados corroboran la intuición concebida en el anterior artículo, donde la asimetría entre cadenas es clave para diferenciar dúplexes susceptibles a RNasa H de aquéllos que no lo son. Así, la RNasa H puede unir dúplexes tipo A, pero no puede actuar debido a que precisa un conjunto de transiciones locales altamente costosas en la cadena *antisentido*.

5.4 Estudio Teórico de Grandes Transiciones Conformacionales en el ADN: El Cambio B \leftrightarrow A en Agua y Etanol/Agua

5.4.1 Motivación

El ADN es una molécula flexible y polimórfica. Aunque su estructura más estable es la forma B, presenta transiciones conformacionales hacia estructuras canónicas, como la forma A o Z, y no canónicas, como hélices paralelas o formas cruciformes. Esta deformabilidad es necesaria para su función, ya que le permite unirse a proteínas que pueden forzar un cambio estructural dramático (118,169) o empaquetarse en nucleosomas (216). Por ello, la flexibilidad es clave para transformar la información estructural en biológica.

Aunque el número de estructuras experimentales han aumentado notablemente en los últimos años, ello no permite obtener información directa sobre las transiciones dinámicas de biomoléculas. Además, si bien la RMN proporciona cierta información dinámica, está limitada a transiciones locales, lo que hace ineludible el uso de algoritmos teóricos como la DM, que permite observar procesos en la escala del sub-microsegundo. No obstante, las grandes transiciones conformacionales requieren una escala temporal más elevada, siendo necesario cálculos más extensos o el uso de métodos de sesgo. Además, no se pueden caracterizar con una simple coordenada geométrica distinguida. Por ello, el análisis de los grandes cambios conformacionales es un reto para los estudios teóricos.

Mientras la forma B es la más estable en un entorno acuoso, la forma A se adopta típicamente en condiciones de baja hidratación (18,127,128)) o por la unión a ciertas proteínas (17), además de existir en el dúplex de ARN. El cambio conformacional B \leftrightarrow A ha sido objeto de interés en numerosos estudios tanto teóricos como experimentales. Así, parece que la principal contribución a la transición es la repulsión electrostática de los fosfatos (180). Asimismo, la restricción de todos los azúcares en bloque de la conformación Sur a la Norte, modificando la estructura de los surcos, fue suficiente para observar la transición B \leftrightarrow A (179). Por otro lado, se observó que la reducción del número de moléculas de agua en una simulación de ADN permitía alcanzar la forma A, viendo cómo se producía un colapso de los iones en el surco mayor (217). Estos resultados corroboran que la estructura de los surcos, y

por lo tanto de las cargas negativas, es clave para determinar la conformación canónica más estable, y confirman que la deshidratación (inducida por la adición de etanol) desencadena la transición de B a A.

Cheatham y Kollman observaron que la DM en un entorno acuoso conducía un dúplex de ADN de la forma A a la B en la escala del nanosegundo en dos simulaciones independientes (64), indicando que la forma A no es una estructura estable en este solvente (218). También se comprobó que una solución etanol/agua al 85% en volumen la forma A era estable, mientras que cuando se empezaban las simulaciones en la forma B tampoco se apreciaban transiciones hacia la forma A, lo que sugería la presencia de una barrera de activación (219,220). Estos resultados fueron confirmados mediante CD en 70% etanol/agua, donde se observó la transición en el rango del microsegundo (221).

Los estudios anteriores sugieren que la alteración del solvente modifica el paisaje de energía libre del ADNds. En este artículo se han utilizado los métodos de sesgo de *Umbrella Sampling* (US) (14) y *Weighted Histogram Analysis Methods* (WHAM) (15) acoplados por primera vez, acorde con nuestro conocimiento, a técnicas de *Targeted Molecular Dynamics* (TMD) (16), con el fin de describir los caminos mayoritarios de alteración de la estructura y calcular la diferencia de energía libre.

5.4.2 Plan de trabajo

Se han realizado simulaciones de ADN en agua, que comprenden 10 trayectorias no sesgadas A→B, así como cálculos US aplicada a todos los azúcares B→A y simulaciones TMD para la transición B→A. Por otro lado, se han realizado simulaciones en solución etanol/agua al 85% en volumen, incluyendo 15 trayectorias no sesgadas para el cambio A→B, 10 trayectorias no sesgadas para la transición B→A, y finalmente cálculos TMD para el cambio A→B.

5.4.3 Discusión de los principales resultados

Como indicaban estudios anteriores (64), la transición A→B en las simulaciones no sesgadas en agua se obtiene en menos de 3 ns (figura 1). Se aprecia un cambio de *puckering* de Norte a Sur en los primeros 500 ps y posteriormente un ajuste conformacional global. Cuando se compara el primer modo esencial de la transición con los primeros modos de las simulaciones de ADN y ARN, se observa una alta correlación (tabla 1), lo cual sugiere que la transición B→A aprovecharía el modo de deformación más importante del ADN. En consecuencia, la forma A se podría interpretar como una configuración activada de un estado vibracional natural del dúplex de ADN.

| | ADN | ARN | B→A (no sesgada) | B→A (TMD) |
|------------------|------|-------|------------------|-----------|
| ADN | 1.00 | -0.81 | 0.94 | 0.87 |
| ARN | | 1.00 | -0.81 | -0.90 |
| B→A (no sesgada) | | | 1.00 | 0.86 |

Tabla 1. Productos escalares entre los vectores propios unidad del primer movimiento esencial de las distintas trayectorias (apartado 4.4).

Para confirmar esta hipótesis, se obtuvieron diversas estructuras a lo largo del primer modo esencial derivado de la trayectoria de ADN en equilibrio, que se acercaban a una forma A hasta un RMSd de 2 Å (figura 2). Asumiendo una aproximación cuasi-harmónica, el punto de cruce de los valores de RMSd respecto a las formas A y B representaría un valor energético de 4 kcal/mol, que alcanzaría una energía de 12 kcal/mol para un RMSd de 2 Å. Considerando los 10 primeros nodos, se obtiene un RMSd de 1.7 Å, pero a expensas de una energía vibracional más alta (19 kcal/mol). Ello sugiere que la transición viene guiada por el primer modo de deformación del ADN, aunque requiere un conjunto de transiciones locales o de naturaleza no harmónica para alcanzar una forma A real.

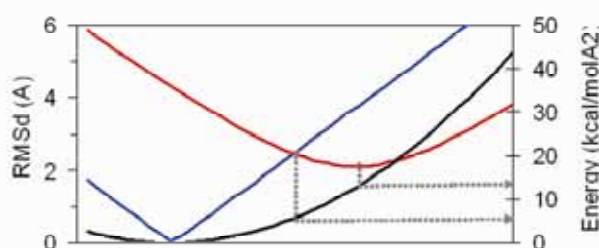


Figura 2. Representación de la energía (negro) y el RMSd respecto la forma A (rojo) y la B (azul) cuando la estructura en equilibrio del ADN se perturba a lo largo de su primer movimiento esencial (apartado 4.4).

La conformación del azúcar es un buen parámetro discriminante entre las formas A y B (179). Por esta razón, hemos forzado una transición B→A aplicando US a todos los ángulos δ (representativo del *puckering* del azúcar). La transición ocurre en valores de δ alrededor de 90° , junto con un cambio en la anchura del surco menor y en el parámetro Z_p , los cuales también son discriminantes ($\langle Z_p \rangle > 1.0$; $\langle \delta \rangle < 95^\circ$ y $mGw > 14 \text{ \AA}$; figuras 3 y 4). En cambio, para el punto de cruce de RMSd respecto las formas A y B, no se observa una estructura A correcta, sino una B distorsionada, lo que señala el riesgo de emplear el RMSd como índice de transición.

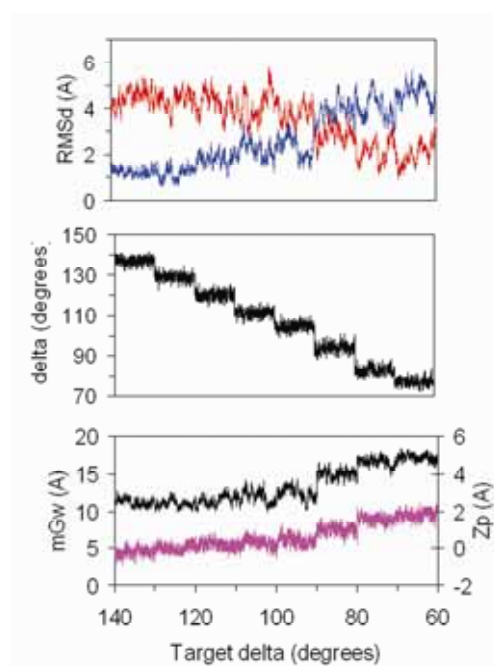


Figura 3. Arriba: RMSd (en \AA) respecto a la forma A (en rojo) y a la B (en azul) de las trayectorias obtenidas aplicando US donde el δ_{target} varía de 140° a 70° . Medio: ángulo δ (en grados). Abajo: anchura del surco menor (mGw, en \AA , negro) y Z_p (en \AA , violeta) (apartado 4.4).

Debido a que la simulación precedente incluye 20 restricciones (todos los ángulos δ), no es posible calcular el potencial de fuerza media (PMF), para el que se necesita tener una sola coordenada de reacción que conduzca la transición de manera reversible. Por ello, hemos aplicado TMD usando el RMSd respecto a un conjunto reducido de átomos. En concreto, se han elegido tres combinaciones de átomos del esqueleto (figura 5), que producen transiciones parecidas y perfiles de PMF semejantes y reversibles (figuras 6, 7 y S1). Cuando el RMSd hacia la forma A se restringe a un valor de 2 \AA , el punto de cruce corresponde a una forma B

distorsionada, donde el surco menor adopta una estructura típica A. Sin embargo, no es hasta un valor de 1 Å que la transición es completa, detectándose el cambio en la conformación de los azúcares (figuras 6, 7 y S1). Si integramos las curvas de PMF usando estos límites, la forma B es más favorable que la B distorsionada por 3.2 kcal/mol y más favorable que la A por 11.4 kcal/mol (tabla 2), que concuerdan con los obtenidos por el análisis del primer modo esencial. Asimismo, existe una alta correlación entre el primer modo de la transición reversible B→A con los primeros modos de las dinámicas no restringidas y la transición irreversible A→B (tabla 1). Finalmente, también cabe destacar la concordancia con la estimación obtenida a partir de datos experimentales (11 kcal/mol; 222).

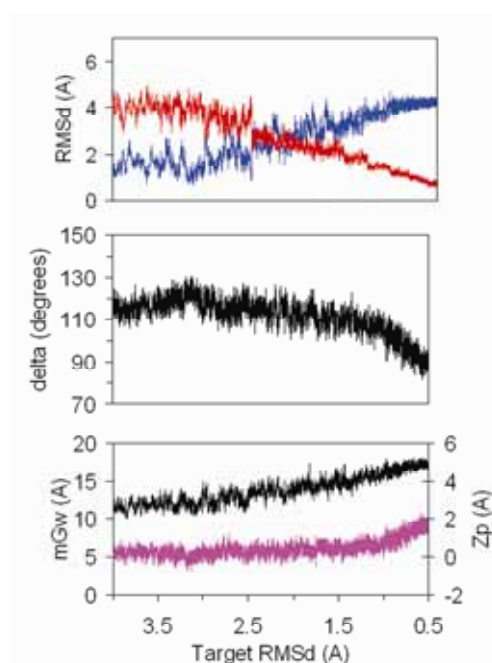


Figura 6. Arriba: RMSd (en Å) respecto a la forma A (en rojo) y a la B (en azul) de las trayectorias obtenidas aplicando TMD B→A en agua. Medio: ángulo δ (en grados). Abajo: anchura del surco menor (mGw, en Å, negro) y Z_p (en Å, violeta) (apartado 4.4).

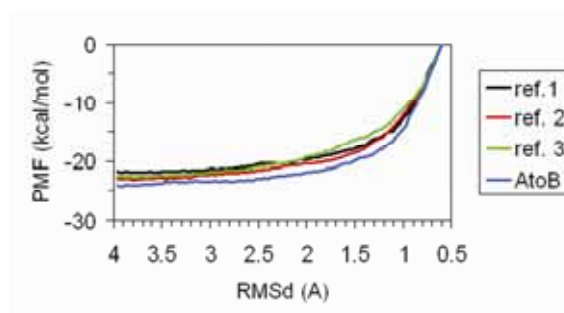


Figura 7. PMF de la transición B→A en agua calculado con la técnica TMD-WHAM. Las distintas curvas se obtienen de los 3 conjuntos de átomos utilizados para definir el RMSd diana y también de considerar el sentido reverso (apartado 4.4).

| Energía | B→B(dist) | B(dist)→A | B→A | A→A(dist) |
|-------------|-----------|-----------|------|-----------|
| Agua | 3.2 | 8.1 | 11.4 | --- |
| Etanol/agua | --- | --- | -0.8 | 0.4 |

Tabla 2. Diferencias de energía libre (en kcal/mol) para la transición conformacional B↔A en agua y etanol/agua 85:15 obtenidas de integrar las curvas de PMF. Dist. corresponde a distorsionada (apartado 4.4).

Las simulaciones no sesgadas en 85% etanol/agua empezadas tanto en la forma A como en la forma B no presentan ninguna transición, como se había observado teórica (219,220) y experimentalmente (18,221) (figuras 8 y S2). En base al PMF obtenido a partir de simulaciones TMD para la transición B→A (figura 10), la forma A es la más estable, siendo la B un mínimo local, con una barrera de activación cercana a 2 kcal/mol respecto a A y 1 kcal/mol respecto a B. Se observa en el mecanismo de transición A→B un reflejo del propio de la transición B→A (figura 6 y 9). Cuando se restringe el RMSd respecto a la forma B en 3.3 Å, se advierte un *repuckering* de Norte a Sur alcanzando así una estructura A distorsionada. La transición no es completa hasta alcanzar el valor de RMSd de 2 Å, percibiendo el cambio global en el surco menor característico de la forma B. Cabe, pues, identificar la estructura A distorsionada como la propia del estado de transición, con un *puckering* Sur aunque un surco menor ancho (figura 9). Si se integra la curva con los límites establecidos, la forma A aparece más estable que la A distorsionada por 0.4 kcal/mol y que la B por 0.8 kcal/mol.

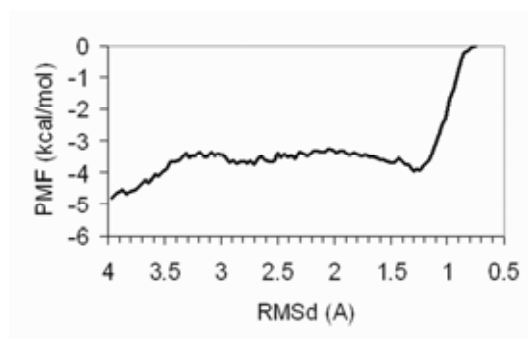


Figura 10. PMF de la transición A→B en etanol/agua 85:15 calculado con la técnica TMD-WHAM (apartado 4.4).

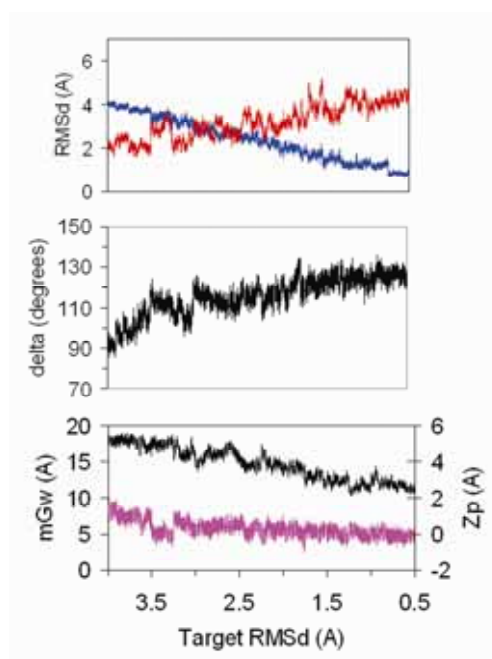


Figura 9. Arriba: RMSd (en Å) respecto a la forma A (en rojo) y a la B (en azul) de las trayectorias obtenidas aplicando TMD A→B en etanol/agua 85:15. Medio: ángulo δ (en grados). Abajo: anchura del surco menor (mGw, en Å, negro) y Z_p (en Å, violeta) (apartado 4.4).

Combinando los resultados de ambas transiciones reversibles, podemos obtener una imagen completa del mecanismo del cambio $B \leftrightarrow A$. El motor en la dirección $B \rightarrow A$ es un reajuste en la geometría del surco menor, teniendo lugar después el cambio conformacional de Sur a Norte en los azúcares. De esta manera, aparece una estructura intermedia (estructura B distorsionada) que presenta el surco menor ancho y los azúcares en forma Sur. En cambio, en la dirección $A \rightarrow B$, primero sucede la transición Norte→Sur de los azúcares y posteriormente el estrechamiento del surco menor, generando una estructura A distorsionada muy similar a la B distorsionada. En consecuencia, se puede concluir que la transición conformacional $B \leftrightarrow A$ está guiada por dos procesos elementales: el cambio en la conformación de los azúcares y la modificación del surco menor.

5.4.4 Conclusión

Las simulaciones de DM sesgadas muestran que en agua la forma B es mayoritaria ($B/A = 10^9$), mientras que en 85% etanol/agua la forma A es la más abundante ($B/A = 0.2$), en concordancia con los resultados experimentales. En solución acuosa, la dificultad de la reacción $B \rightarrow A$ recae en la necesidad de cambiar la conformación de los azúcares. Es

interesante observar que en trayectorias en el rango del nanosegundo no es extraño observar un azúcar individual en una conformación Norte. Sin embargo, debido a que la vida media del cambio es en la escala del picosegundo, la probabilidad de encontrar todos los azúcares en este *puckering* es muy baja. Cuando el solvente es modificado a 85% etanol/agua, la conformación Norte se estabiliza en relación a la Sur, favoreciendo la conformación general tipo A (219).

6. CONCLUSIONES

El patrón de flexibilidad diferencial entre las moléculas de ADN y ARN es realmente complejo y depende del grado de libertad considerado. En general, el ADN es más flexible gracias a la gran cantidad de movimientos locales propios del esqueleto azúcar-fosfato, mientras que el ARN presenta movimientos globales más deformables.

Las simulaciones de DM reproducen las estructuras experimentales de RMN del híbrido ADN·ARN, así como las propias del conjunto de dúplexes antisentido estudiados, independientemente de la estructura inicial. Las moléculas susceptibles a RNasa H presentan una forma intermedia, más cerca de la forma A que de la B, con propiedades únicas en los surcos.

Mientras que la estructura promedio puede justificar la incapacidad de la RNasa H de unir el dúplex de ADN, no es suficiente para discernir entre ADN·ARN y el dúplex de ARN, así como entre dúplexes antisentido susceptibles y no susceptibles. La principal diferencia es la asimetría dinámica entre cadenas, lo que provoca que la deformación local de la cadena *antisentido* en los dúplexes tipo A sea demasiado costosa para conseguir una unión productiva a RNasa H.

Aunque la TMD puede ser fuente de numerosos errores cuando se usa sin rigor, permite identificar el cambio en el paisaje de energía libre del ADN delante de la alteración del solvente. Así, en agua la forma B es mayoritaria y la forma A puede describirse como una configuración vibracional activada del primer modo esencial. En cambio, en 85% etanol/agua, la forma A es la más estable, aunque la forma B constituye un mínimo local.

La transición conformacional $B \leftrightarrow A$ está constituida por dos procesos principales: el cambio Sur \leftrightarrow Norte de los azúcares y el de la anchura del surco menor. Esto genera una estructura intermedia con un surco menor ancho típico A y un *puckering* Sur típico B.

7. BIBLIOGRAFÍA

1. Schlick T (2006) *Molecular Modeling and Simulation*. Ed. Springer
2. Sanbonmatsu KY, Joseph S, Tung CS (2007) *Proc Natl Acad Sci USA* **102**,15854-9
3. Alder BJ, Wainwright TE (1957) *J Chem Phys* **27**, 1208-9
4. Pérez A, Luque FJ, Orozco M (2007) *J Am Chem Soc* **129**, 14739-45
5. Watson JD, Crick FH (1953) *Nature* **171**, 964-7
6. Wing R, Drew H, Takano T, Broka C, Tanaka S, Itakura K, Dickerson RE. (1980) *Nature* **287**, 755-8
7. Olson WK, Gorin AA, Lu X, Hock LM, Zhurkin VB (1998) *Proc Natl Acad Sci USA* **95**, 11163-8
8. Jacob F, Monod J (1961) *J Mol Biol* **3**, 318-56
9. Alberts B, Bray D, Johnson A, Lewis J, Raff M, Roberts K, Walter P (1999) *Biología Molecular de la Célula*. 3rd ed, Ediciones Omega, Barcelona
10. Orozco M, Pérez A, Noy A, Luque FJ (2003) *Chem Soc Rev* **32**, 350-64
11. Hagerman PJ (1988) *Annu Rev Biophys Biomol Struct* **17**, 265-86
12. Hagerman PJ (1997) *Annu Rev Biophys Biomol Struct* **26**, 139-56
13. Lima WF, Crooke ST (1997) *Biochemistry* **36**, 390-8
14. Torrie GM, Valleau, JP (1977) *J Comput Phys* **23**, 187-99
15. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM (1992) *J Comput Chem* **13**, 1011-21
16. Schlitter J, Engels M, Krüger P, Jacoby E, Wollmer A (1993) *Mol Simul* **10**, 291-308
17. Lu XJ, Shakked Z, Olson WK (2000) *J Mol Biol* **300**, 819-40
18. Ivanov VI, Minchenkova LE, Minyat EE, Frank-Kamenetskii MD, Schyolkina AK (1974) *J Mol Biol* **87**, 817-33
19. Born M, Oppenheimer R (1927) *Annalen der Physik* **389**, 457-84
20. Szabo A, Ostlund NS (1982) *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. McGraw-Hill, New York.
21. Leach AR (2001) *Molecular Modelling. Principles and Applications*, 2nd Ed. Essex, Pearson Education (Prentice Hall)
22. Cramer CJC (2002) *Essentials of Computational Chemistry: Theories and Models*. Chichester, Wiley
23. Roothaan CC (1951) *Rev Mod Phys* **23**, 69-89
24. Slater JC (1930) *Phys Rev* **36**, 57-64

25. Hehre WJ, Stewart RF, Pople JA (1969) *J Chem Phys* **51**, 2657-64
26. Bates DR (1961) *Quantum Theory*. Academic Pres, New York
27. Møller C, Plesset MS (1934) *Phys Rev* **46**, 618-22
28. Weiner SJ, Kollman PA, Nguyen DT, Case DA (1986) *J Comput Chem* **7**, 230-52
29. Karlus M (1983) *J Comput Chem* **4**, 187-217
30. Cheatham TE 3rd (2004) *Curr Opin Struc Biol* **14**, 360-7
31. Beveridge DL, Dixit SB, Barreiro G, Thayer KM. (2004) *Biopolymers* **73**, 380-403
32. van Gunsteren WF, Bakowies D, Baron R, Chandrasekhar I, Christen M, Daura X, Gee P, Geerke DP, Glattli A, Hunenberger PH, Kastenholz MA, Oostenbrink C, Schenk M, Trzesniak D, van der Vegt NF, Yu HB (2006) *Angew Chem Int Ed* **45**, 4064-92
33. Cubero E, Luque FJ, Orozco M. (2001) *J Am Chem Soc* **123**, 12018-25
34. Cubero E, Abrescia NG, Subirana JA, Luque FJ, Orozco M. (2003) *J Am Chem Soc* **125**, 14603-12
35. Cubero E, Luque FJ, Orozco M. (2006) *Biophys J* **90**, 1000-8
36. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. (1995) *J Am Chem Soc* **117**, 5179-97
37. Wang J, Cieplak P Kollman PA (2000) *J Comput Chem* **21**, 1049-74
38. Pérez A, Marchán I, Svozil D, Sponer J, Cheatham TE 3rd, Laughton CA, Orozco M (2007) *Biophys J* **92**, 3817-29
39. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) *J Comp Chem* **25**, 1157-68.
40. Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S, Weiner P (1984) *J Am Chem Soc* **106**, 765-84
41. Bayly CI, Cieplak P, Cornell WD, Kollman PA (1993) *J Phys Chem* **97**, 10269-80
42. Cieplak P, Caldwell J, Kollman PA (2001) *J Comput Chem* **22**, 1048-57
43. Applequist J, Carl J, Fung K (1972) *J Am Chem Soc* **94**, 2952-60
44. Babin V, Baucom J, Darden TA, Sagui C (2006) *J Phys Chem B* **110**, 11571-81
45. Chen AA, Pappu RV (2007) *J Phys Chem B* **111**, 11884-7
46. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) *J Chem Phys* **79**, 926-35

47. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J (1981) *Intermolecular Forces*. B. Pullman (ed). Reidel. Dordrecht. 331-5
48. Warshel A (1976) *Nature* **260**, 679-83
49. McCammon JA (1977) *Nature* **267**, 585-8
50. Tidor B (1983) *J Biomol Struct Dyn* **1**, 231-40
51. Hockney RW (1970) *Meth Comput Phys* **9**, 135-211
52. Verlet L (1967) *Phys Rev* **159**, 98-103
53. Ryckaert JP, Cicotti G, Berendsen HJC (1977) *J Comput Phys* **23**, 327-41
54. Cheatham TE 3rd, Young MA (2001) *Biopolymers* **56**, 232-56
55. Ewald PP (1921) *Ann Phys* **64**, 253-7
56. Darden TA, York DM, Pedersen LG (1993) *J Chem Phys* **98**, 10089-92
57. Kastenzholz MK, Hünenberger PH (2004) *J Phys Chem B* **108**, 774-88
58. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) *J Chem Phys* **81**, 3684-90
59. Hobza P, Kabelac M, Sponer J, Mejzlik P, Vondrasek J (1997) *J Comput Chem* **18**, 1136-50
60. Cheatham TE 3rd, Cieplak P, Kollman PA (1999) *J Biomol Struct Dyn* **16**, 845-55
61. Foloppe N, Mackerell AD (2000) *J Comput Chem* **21**, 86-104
62. Mackerell AD, Banavali NK (2000) *J Comput Chem* **21**, 105-20
63. Langley DR (1998) *J Biomol Struct Dynam* **16**, 487-97
64. Cheatham TE 3rd, Kollman PA (1996) *J Mol Biol* **259**, 434-44
65. Shields GC, Laughton CA, Orozco M (1997) *J Am Chem Soc* **119**, 7463-69
66. Arhanari H, McConnell KJ, Beger R, Young MA, Beveridge DL, Bolton PH (2003) *Biopolymers* **68**, 3-15
67. Dixit SB, Beveridge DL, Case DA, Cheatham TE 3rd, Giudice E, Lankas F, Lavery R, Maddocks JH, Osman R, Sklenar H, Thayer KM, Varnai P (2005) *Biophys J* **89**, 3147-57
68. Roux B (1995) *Comput Phys Commun* **91**, 275-82
69. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM (1995) *J Comput Chem* **16**, 1339-50
70. Apostolakis J, Ferrara P, Caflisch A (1999) *J Chem Phys* **110**, 2099-108

71. Frenkel D, Smith B (1996) *Understanding Molecular Simulations: from Algorithms to Applications*. Academic Press, Inc., San Diego
72. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) *J Chem Phys* **21**, 1087-92
73. van Gusteren WF, Mark AE (1998) *J Chem Phys* **108**, 6109-16
74. Gelpí JL, Kalko SG, Barril X, Cirera J, de la Cruz X, Luque FJ, Orozco M (2001) *Proteins* **45**, 428-37
75. Orozco M, Luque FJ (2000) *Chem Rev* **100**, 4187-225
76. Kitao A, Go N (1999) *Curr Opin Struct Biol* **9**, 164-9
77. Go N, Scheraga HA (1976) *Macromolecules* **9**, 535-40
78. Sherer EC, Harris SA, Soliva R, Orozco M, Laghton CA (1999) *J Am Chem Soc* **121**, 5981-91
79. Amadei A, Linssen ABM, Berendsen HJC (1993) *Proteins* **14**, 412-24
80. Rueda M, Kalko SG, Luque FJ, Orozco M (2003) *J Am Chem Soc* **125**, 8007-14
81. Hess B (2000) *Physical Review E* **62**, 8438-48
82. Mahalanobis PC (1936) *Proc Nat Inst Sci India* **2**, 49-55
83. Karplus M, Kushick JN (1981) *Macromolecules* **14**, 325-32
84. Pérez A, Blas JR, Rueda M, Lopez-Bes JM, de la Cruz, X, Orozco M (2005) *J Chem Theor Comp* 2005 **1**, 790-800
85. Schlitter J (1993) *Chem Phys Let* **215**, 617-21
86. Andricioaei I, Karplus, M (2001) *J Chem Phys* **115**, 6289-92
87. Harris S, Gavathiotis E, Searle MS, Orozco M, Laughton CA (2001) *J Am Chem Soc* **123**, 12658-63
88. Saenger W (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York
89. Bloomfield VA, Crothers DM, Tinoco I (2000) *Nucleic Acids. Structures, Properties and Functions*, University Science Books, Sausalito
90. Blackburn GM, Gait MJ (1990) *Nucleic Acids in Chemistry and Biology*, IRL Press, New York
91. Neidle S (1995) *DNA Structure and Reactivity*. Oxford University Press, London
92. Altona C, Sundaralingam M (1972) *J Am Chem Soc* **94**, 8205-10
93. de Leeuw HPM, Haasnoot CAG, Altona C (1980) *Isr J Chem* **20**, 108-20
94. Sundaralingam M (1976) *Annu NY Acad Sci* **255**, 3-10

-
95. Hruska FE, Smith AA, Dalton JG (1971) *J Am Chem Soc* **93**, 4334-45
 96. Varnai P, Djuranovic D, Lavery R, Hartmann B (2002) *Nuc Acids Res* **30**, 5398-406
 97. Bertrand HO, Ha-Duong T, Femandjian S, Hartmann B (1998) *Nuc Acids Res* **26**, 1261-7
 98. Renugopalakrishnan V, Lakshminarayanan AV, Sasisekharan V (1971) *Biopolymers* **10**, 1159-67
 99. Pullman B, Pullman A (1971) *Adv Heterocyc Chem* **13**, 77-85
 100. Pauling L (1978) *The Nature of the Chemical Bond*. Cornell University Press, Ithaca, New York
 101. Nowak MJ, Szcsepaniak K, Barski A, Shugar D (1978) *Z Naturforsch C* **33**, 876-85
 102. Topal MD, Fresco JR (1976) *Nature* **263**, 285-8
 103. Hoogsteen K (1963) *Acta Crystallogr* **16**, 907-15
 104. Crick FH (1966) *J Mol Biol* **19**, 548-58
 105. Hanlon S (1966) *Biochem Biophys Res Commun* **23**, 861-70
 106. Luo R, Gilson HSR, Potter MJ, Gilson MK (2001) *Biophys J* **80**, 140-50
 107. Ts'o POP (1974) En *Basic Principles in Nucleic Acid Chemistry*. Volume 1. Academic, New York
 108. Solie TN, Shellman JA (1968) *J Mol Biol* **33**, 61-70
 109. Tribolet R, Sigel H (1987) *Biophys Chem* **27**, 119-30
 110. Guckian KM, Schweitzer BA, Ren RXF, Sheils CJ, Paris PL, Tahmassebi DC, Kool ET (1996) *J Am Chem Soc* **118**, 8182-90
 111. Cieplak P, Kollman PA (1988) *J Am Chem Soc* **119**, 3734-45
 112. Alhambra C, Luque FJ, Gago F, Orozco M (1997) *J Phys Chem B* **101**, 3846-54
 113. Hobza P, Sponer J (1999) *Chem Rev* **99**, 3247-55
 114. Nelson JC, Saven JG, Moore JS, Wolynes PG (1997) *Science* **277**, 1793-5
 115. Orozco M, Cubero E, Barril X, Colominas C, Luque FJ (1999) *Computational Molecular Biology. Theoretical Computational Chemistry*. Volume 8. Elsevier Science, Amsterdam
 116. Drew HR, Dickerson RE (1981) *J Mol Biol* **151**, 535-43
 117. McFail-Isom L, Shui X, Williams LD (1998) *Biochemistry* **37**, 17105-14
 118. Dickerson RE (1998) *Nuc Acids Res* **26**, 1906-26

119. Dickerson RE, Bansal M, Calladine CR, Diekmann S, Hunter WN, O Kennard, von Kitzing E, Lavery R, Nelson HCM, Olson WK, Saenger W, Shakked Z, Sklenar H, Soumpasis DM, Tung CS, Wang AHJ, Zhurkin VB (1989) *EMBO J* **8**, 1–4
120. El Hassan MA, Calladine CR (1995) *J Mol Biol* **251**, 648-64
121. Lu X, Olson WK (2003) *Nuc Acids Res* **31**, 5108-21
122. Lu X, El Hassan MA, Hunter CA (1997) *J Mol Biol* **273**, 668-80
123. Lankas F, Sponer J, Hobza P, Langowski J (2000) *J Mol Biol* **299**, 695-709
124. Kratky O, Porod G (1949) *Rec Trav Chim Pays-Bas* **68**, 1106-23
125. Arnott S, Hukins DWL (1972) *Biochem Biophys Res Commun* **47**, 1504-13
126. Drew HR, Wing RM, Takano T, Broka C, Tanaka S, Itakura K, Dickerson RE (1981) *Proc Nat Acad Sci USA* **78**, 2179-84
127. Fuller W, Wilkins MHF, Wilson HR, Hamilton LD, Arnott S (1965) *J Mol Biol* **12**, 60-70
128. Ivanov VI, Minchenkova LE, Schyolkina AK, Poletayev AI (1973) *Biopolymers* **12**, 98-110
129. Wahl MC, Sundaralingam M (1999) *Oxford Handbook of Nucleic Acid Structure*, ed Neidle (Oxford, Sci) 117-44
130. Arnott S, Chandrasekaran R, Birdsall DL, Leslie AGW (1980) *Nature* **283**, 743-5
131. Wittig B, Dorbic T, Rich A (1991) *Proc Natl Acad Sci USA* **88**, 2259-62
132. Chastain M, Tinoco Jr (1991) *Prog Nuc Acid Res Mol Biol* **41**, 131-77
133. Fry M (2007) *Front Biosci* **12**, 4336-51
134. Crick FH (1958) *Symp Soc Exp Biol* **XII**, 139-63
135. Nishizaki T, Shigenori I, Ohkubo T, Kojima C, Nakamura H, Kyogoku Y, Ohtsuka E (1996) *Biochemistry* **35**, 4016-25
136. Syperski T, Gotte M, Billeter M, Perola E, Cellai L, Heumann H, Wüthrich K (1999) *J Biomol NMR* **13**, 343-55.
137. Sancar A (1996) *Annu Rev Biochem* **65**, 43-81
138. Lebars I, Yoshizawa S, Stenholm AR, Guittet E, Douthwaite S, Fourmy D (2003) *EMBO J* **22**, 183–92
139. Borovinskaya MA, Shoji S, Holton JM, Fredrick K, Cate JH (2007) *Acs Chem Biol* **2**, 545-52

-
140. Telesnitsky A, Goff SP (1997) *Reverse transcriptase and the generation of retroviral DNA*. In Coffin JM, Hughes SH, Varmus HE. (eds), *Retroviruses*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
 141. Zamecnik PC, Stephenson ML (1978) *Proc Natl Acad Sci USA*, **75**, 280-5
 142. The Vitravene Study Group (2002) *Am J Ophthalmol* **133**, 467-70
 143. Gleave ME, Monia BP (2005) *Nat Rev Cancer* **5**, 468-79
 144. Kurreck J (2003) *Eur J Biochem* **270**, 1628-37
 145. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998) *Nature* **391**, 806-12
 146. Lavorgna G, Dahary D, Lehner B, Sorek R, Sanderson CM, Casar G. (2004) *Trends Biochem Sci* **29**, 88-100
 147. Sullenger BA, Millich L, Jones III JP *Gene and Cell Therapy, Therapeutic mechanisms and Strategies 2nd Edition* (2004) Templeton NS Ed; Marcel Dekker, Inc New York, 333-45
 148. Doherty EA, Doudna JA. (2001) *Annu Rev Biophys Biomol Struct* **30**, 475-98
 149. Moelling K, Matskevich A, Jung JS (2006) *Cold Spring Har Symp Quant Biol* **71**, 365-8
 150. Cerutti H, Casas-Mollano JA (2006) *Curr Gen* **50**, 81-99
 151. Worrall JAR, Luisi BF (2007) *Curr Opin Struct Biol* **17**, 128-37
 152. Fedoroff OY, Salazar M, Reid BR (1993) *J Mol Biol* **233**, 509-23
 153. Fedoroff OY, Ge Y, Reid BR (1997) *J Mol Biol* **269**, 225-35
 154. Gyi JI, Lane AN, Conn GL, Brown T (1998) *Biochemistry* **37**, 73-82.
 155. Croke ST (2004) *Annu Rev Med* **55**, 61-95
 156. Stein H, Hausen P (1969) *Science* **166**, 393-7
 157. Sarafianos SG, Das K, Tantillo C, Clark AD, Ding J, Whitcomb JM, Boyer PL, Hughes SH, Arnold E (2001) *EMBO J* **20**, 1449-58
 158. Nowotny M, Gaidamakov SA, Crouch J, Yang W (2005) *Cell* **121**, 1005-16
 159. Mangos MM, Damha MJ (2002) *Curr Top Med Chem* **2**, 1147-71
 160. Crouch RJ, Dirksen ML (1982) *Nucleases* Linn SM, Roberts RJ, eds. Cold Spring Harbor Laboratory Press, Plainview, NY. 211-41
 161. Ceritelli SM, Frolova EG, Feng C, Grinberg A, Love PE, Crouch RJ (2003) *Mol Cell* **11**, 807-15

162. Aboul-Fadl T (2005) *Curr Med Chem* **12**, 2193-21
163. Venkateswarlu D, Lind KE, Mohan V, Manoharan M, Ferguson DM (1999) *Nuc Acids Res* **27**, 2189-95
164. Damha MJ, Noronha AM, Wilds CJ, Trempe JF, Denisov A, Pon RT, Gehring K (2001) *Nucleosides Nucleotides* **20**, 429-40
165. Gyi JI, Gao D, Conn GL, Trent JO, Brown T, Lane AN (2003) *Nuc Acids Res* **31**, 2683-93
166. Znosko BM, Barnes TW 3rd, Krugh TR, Turner DH (2002) *J Am Chem Soc* **125**, 6090-7
167. Hannon GJ (2002) *Nature* **418**, 244-251
168. Chatterton JE, Hu X, Wong-Staal F (2004) *Drug Disc Tod Targ* **3**, 10-7
169. Luscombe NM, Austin SE, Berman HM, Thornton JM (2000) *Genome Biol* **1**, 1-37
170. Lankas F, Sponer J, Langowski J, Cheatham TE 3rd (2003) *Biophys J* **85**, 2872-83
171. El Hassan MA, Calladine CR (1997) *Phil Trans R Soc Lond A* **355**, 43-100
172. Calladine CR (1982) *J Mol Biol* **161**, 343-52
173. Hunter CA (1993) *J Mol Biol* **230**, 1025-54
174. Hunter CA, Lu X (1997) *J Mol Biol* **265**, 603-19
175. Calladine CR, Drew HR (1992) *Undertanding DNA*. Academic Press
176. Hud NV, Plavec J (2003) *Biopolymers* **69**, 144-59
177. Calladine CR, Drew HR (1984) *J Mol Biol* **178**, 773-82
178. Ng H, Kopka ML, Dickerson RE (2000) *Proc Natl Acad Sci USA* **97**, 2035-39
179. Soliva R, Luque FJ, Alhambra C, Orozco M (1999) *J Biomol Struct Dyn* **17**, 89-99
180. Jayaram B, Sproun D, Young MA, Beveridge DL (1998) *J Am Chem Soc* **120**, 10629-33
181. Franklin RE, Gosling RG (1953) *Nature* **172**, 156-7
182. Arnott S, Chandrasekharan R, Puigjaner IH, Birdsall DL, Ratliff RL (1983) *Nuc Acids Res* **11**, 1457-74
183. Arnott S (2006) *Trends Biochem Sci* **16**, 349-54.
184. Dickerson RE, Goodsell DS, Neidle S (1994) *Proc Natl Acad Sci USA* **91**, 3579-83
185. Davies DB (1978) *Prog NMR Spectrosc* **12**, 135-22
186. Holme DJ, Peck H (1998) *Analytical Biochemsitry 3 ed* Pearson Education Limited
187. Wilds CJ, Damha MJ (2000) *Nuc Acids Res* **28**, 3625-35

-
188. Marini JC, Levene SD, Crothers DM, Englund PT (1982) *Proc Natl Acad Sci USA* **79**, 7664-68
 189. Kannanganattu VP, Spector DL (2007) *Cold Spring Har Symp Quant Biol* **21**, 11-42
 190. Fujiwara T, Shindo H (1985) *Biochemistry* **24**, 896-902
 191. Shindo H, Fujiwara T, Akutsu H, Matsumoto U, Kyogoku Y (1985) *Biochemistry* **24**, 887-95
 192. Cheatham TE, Kollman PA (1997) *J Am Chem Soc* **119**, 4805-25
 193. Lipanov A, Kopka ML, Kaczor-Grzeskowiak M, Quintana J, Dickerson RE (1993) *Biochemistry* **32**, 1373-89
 194. Portmann S, Usman N, Egli M (1995) *Biochemistry* **34**, 7569-77
 195. Shakked Z, Guerstein-Guzikevich G, Eisenstein M, Frolow F, Rabinovich D (1989) *Nature* **342**, 456-60
 196. Pérez A, Noy A, Lankas F, Luque FJ, Orozco M (2004) *Nuc Acids Res* **32**, 6144-51
 197. Auffinger P, Westhof E (2001) *J Mol Biol* **305**, 1057-72
 198. Pan Y, MacKerell AD (2003) *Nuc Acids Res* **31**, 7131-40
 199. Milman G, Langridge R, Chamberlin MJ (1967) *Proc Natl Acad Sci USA* **57**, 1804-8
 200. O'Brien EJ, MacEwan AW (1970) *J Mol Biol* **48**, 243-41
 201. Katahira M, Lee SJ, Kobayashi Y, Sujeta H, Kyogoku Y, Iwai S, Ohtsuka E, Benevides JM, Thomas GJ (1990) *J Am Chem Soc* **112**, 4508-14
 202. Horton NC, Finzel BC (1996) *J Mol Biol* **264**, 521-30
 203. Conn GL, Brown T, Leonard GA (1999) *Nucl. Acids Res.* **27**, 555-64
 204. Xiong Y, Sundaralingam M (2000) *Nucl Acids Res* **28**, 2171-80
 205. Chou SH, Flynn P, Reid B (1989) *Biochemistry* **28**, 2422-34
 206. Hall KB, McLaughlin LW (1991) *Biochemistry* **30**, 10606-12
 207. Cross CW, Rice JS, Gao X (1997) *Biochemistry* **36**, 4096-104
 208. Gonzalez C, Stec W, Reynolds M, James TL (1995) *Biochemistry* **34**, 4969-78
 209. Hantz E, Larue V, Ladam P, Le Moyec L, Gouyette C, Huynh Dinh T (2001) *Int J Biol Macromol* **28**, 273-80
 210. Arnott S, Bond PJ, Selmsung E, Smith PJ (1976) *Nuc Acids Res* **3**, 2459-67
 211. Nishizaki T, Iwai S, Ohtsuka E, Nakamura H (1997) **36**, 2577-85
 212. Noronha AM, Wilds CJ, Lok CN, Viazovkina K, Arion D, Parniak MA, Damha MJ (2000) *Biochemistry* **39**, 7050-62

213. Trempe J, Wilds CJ, Denisov AY, Pon RT, Damha MJ, Gehring K (2001) *J Am Chem Soc* **123**, 4896-903
214. Denisov AY, Noronha AM, Wilds CJ, Trempe J, Pon RT, Gehring K, Damha MJ (2001) *Nuc Acids Res* **29**, 4284-93
215. Venkateswarlu D, Ferguson DM (1999) *J Am Chem Soc* **121**, 5809-10
216. Segal E, Fondufe-Mittendorf Y, Chen L, Thastrom A, Field Y, Moore YK, Wanq JP, Widom J (2006) *Nature* **442**, 772-8
217. Mazur AK (2003) *J Am Chem Soc* **125**, 7849-59
218. Banavali NK, Roux B (2005) *J Am Chem Soc* **127** 6866-76
219. Cheatham TE, Crowley MF, Fox T, Kollman PA (1997) *Proc Natl Acad Sci USA* **94**, 9626-30
220. Sprous D, Young MA, Beveridge DL (1998) *J Phys Chem B* **102**, 4658-67
221. Jose D, Porschke D (2005) *J Am Chem Soc* **127**, 16120-8
222. Tolstorukov MY, Ivanov VI, Malenkov GG, Jernigan RL, Zhurkin VB (2001) *Biophys J* **81**, 3409-21

8. APÉNDICES

A. Abreviaturas

| | |
|--------|---|
| A | Adenina |
| AAN | Ácido arabinonucleico |
| ACP | Análisis de Componentes Principales |
| ADN | Ácido desoxorribonucleico |
| ADNds | <i>ADN double strand</i> |
| AMBER | <i>Assisted Model Building with Energy Refinement</i> |
| ARN | Ácido ribonucleico |
| ARNds | <i>ARN double strand</i> |
| ARNi | ARN de interferencia |
| ARNm | ARN mensajero |
| ARNr | ARN ribosomal |
| ARNsi | <i>ARN small interference</i> |
| ARNt | ARN de transferencia |
| B3LYP | <i>Becke 3 parameter hybrids & Lee-Yang-Parr correlation</i> |
| BO | Born-Oppenheimer |
| C | Citosina |
| CD | <i>Circular Dichroism</i> |
| CEHS | <i>Cambridge University Engineering Department Helix computation Scheme</i> |
| CHARMM | <i>Chemistry Harvard Molecular Mechanics</i> |
| CMIP | <i>Classical Molecular Interaction Potencial</i> |
| DD | Dodecámero de Dickerson |
| DM | Dinámica Molecular |
| ESP | <i>Electrostatic Potencial</i> |
| F-AAN | Ácido 2'F-arabinonucleico |
| FDP | Función de Distribución Promedio |
| FFT | <i>Fast Fourier Transform</i> |
| fs | femtosegundos |
| G | Guanina |
| GAFF | <i>General Amber Force Field</i> |
| GTO | <i>Guassian-Type Orbital</i> |

| | |
|-----------|--|
| HF | Hartree-Fock |
| IR | <i>Infrared</i> |
| ISA | Índice de Similitud Absoluto |
| ISR | Índice de Similitud Relativo |
| LCAO | <i>Linear Combination of Atomic Orbitals</i> |
| MBT | <i>Mid-Base Triad</i> |
| MC | Monte Carlo |
| MP | Møller-Plesset |
| MST | <i>Mid-Step Triad</i> |
| NOE | <i>Nuclear Overhauser Effect</i> |
| NOESY | <i>Nuclear Overhauser Enhancement Spectroscopy</i> |
| ns | nanosegundo |
| parm94 | parámetros del 1994 |
| parm99 | parámetros del 1999 |
| parm-bsc0 | parámetros <i>barcelona super computer 0</i> |
| PBC | <i>Periodic Boundary Conditions</i> |
| PDB | <i>Protein Data Bank</i> |
| PME | <i>Particle Mesh Ewald</i> |
| PMF | <i>Potencial of Mean Force</i> |
| ps | picosegundo |
| R | purinas |
| RESP | <i>Restrained Electrostatic Potencial</i> |
| RISC | <i>RNA-induced silencing complex</i> |
| RMN | Resonancia Magnética Nuclear |
| RMSd | <i>Root Mean Squared desviation</i> |
| RNasa | Ribonucleasa |
| SEP | Superficie de Energía Potencial |
| SIDA | Síndrome de inmunodeficiencia adquirida |
| SPC/E | <i>Extended Simple Point Charge</i> |
| STO | <i>Slater-Type Orbital</i> |
| T | Timina |
| TIP3P | <i>Transferable Intermolecular Potencial 3 Point</i> |

| | |
|----------------|--|
| T _m | <i>melting Temperature</i> |
| TMD | <i>Targeted Molecular Dynamics</i> |
| U | Uracil |
| US | <i>Umbrella Sampling</i> |
| UV | Ultravioleta |
| vdW | van der Waals |
| VIH | Virus de la Inmunodeficiencia Humana |
| WC | Watson-Crick |
| WLC | <i>Worm-like Chain</i> |
| WHAM | <i>Weighted Histogram Analysis Methods</i> |
| Y | Pirimidinas |

B. Otras publicaciones

- 1 Orozco M, Pérez A, Noy A, Luque FJ
“Theoretical methods for the simulation of nucleic acids.”
Chemical Society Reviews 2003, 32 (6): 350-364
- 2 Pérez A, Noy A, Lankas F, Luque FJ, Orozco M
“The relative flexibility of B-DNA and A-RNA duplexes: database analysis.”
Nucleic Acids Research 2004, 32(20): 6144-51
- 3 Noy A, Meyer T, Rueda M, Ferrer C, Valencia A, Pérez A, de la Cruz X, Lopez-Bes JM, Pouplana R, Fernandez-Recio J, Luque FJ, Orozco M
“Data mining of molecular dynamics trajectories of nucleic acids.”
Journal of Biomolecular Structure & Dynamics 2006, 23(4): 447-56
4. Orozco M, Noy A, Meyer T, Rueda M, Ferrer C, Valenca A, Pérez A, Carrillo O, Fernandez-Recio J, de la Cruz J, Luque, F J
“Data mining of molecular dynamic trajectories of nucleic acids.”
Computational Studies of RNA and DNA. 5 121-145. Ed. Springer 2006
- 5 Orozco M, Noy A, Pérez A
“Recent advances in the study of nucleic acid flexibility by molecular dynamics.”
Current Opinion in Structural Biology 2008 In Press

C. Agradecimientos

Al Modesto i al Javi per confiar amb mi, acollir-me al grup i ensenyar-me la seva visió de la ciència així com els seus petits secrets. A tot mmb. Thanks to Sarah, Ramin and Charlie for encourage me in my work and for the excepcional discussions.

Al meu nen que és casa, als meus pares pels dinars terapèutics, a la meva família, als desayunos, QTCeros, bionenes, bioquímics, al supergrupo, the people I met in Sheffield, thanks Sarah and Geoff for welcomed me warmly to their house, als qui hem compartit aquesta última etapa d'escriptura i a tothom qui m'ha escoltat i animat.

A tots vosaltes, moltes gràcies