

**TESIS DE DOCTORADO EN BIOLOGÍA  
SUBAREA: BIOFÍSICA  
PEDECIBA**

**“Desarrollo de Métodos para Acelerar  
Simulaciones de Dinámica Molecular”**

Lic. Leonardo Darré

Director: Dr. Sergio Pantano  
Co-Director: Dr. Fernando Herrera

Octubre 2012

# Contenidos

## Resumen

### Capítulo I. Introducción a las Simulaciones Moleculares Coarse Grain

Simulaciones Moleculares	I-1
Modelos Coarse Grain	I-9
Simulaciones Multiescala	I-16

### Capítulo II. Desarrollo de un Nuevo Modelo CG para Solvatación Acuosa

Introducción	II-1
Métodos	II-3
Resultados	II-13
Discusión y Conclusiones	II-29
Bibliografía	II-31

### Capítulo III. Curvatura en el ADN inducida por el entorno iónico: aplicando los modelos CG desarrollados

Introducción	III-1
Métodos	III-17
Resultados y Discusión	III-20
Conclusiones	III-37
Bibliografía	III-38

### Capítulo IV. Desarrollo de un nuevo modelo de solvatación multiescala

Introducción	IV-1
Métodos	IV-2
Resultados	IV-7
Conclusiones	IV-30
Bibliografía	IV-32

## Glosario de Símbolos

$\rho^{eq}(i)$	densidad de probabilidad de encontrar un sistema en la configuración "i" en condiciones de equilibrio
m	masa
$O(x^n)$	término de orden $x^n$ o menor
$k_B$	constante de Boltzmann ( $1.3806488(13) \times 10^{-23} \text{ J K}^{-1}$ )
$\langle v_x(t')v_x(t'') \rangle$	función de autocorrelación de velocidades (para la componente en la coordenada x)
$r_b (=r_{eq}=d_{eq})$	distancia de equilibrio para interacciones ligadas de dos cuerpos
$k_b (=k_{bond})$	constante de fuerza para interacciones ligadas de dos cuerpos
$\theta (= \theta_{eq})$	ángulo de equilibrio para interacciones ligadas de tres cuerpos
$k_\theta$	constante de fuerza para interacciones ligadas de tres cuerpos
$n_k$	multiplicidad (interacciones ligadas de cuatro cuerpos)
$\gamma_k^{eq}$	fase (interacciones ligadas de cuatro cuerpos)
$V_k$	constante de fuerza para interacciones ligadas de tres cuerpos
$\epsilon_{lm} (= \epsilon)$	energía en el mínimo de la función de Lennard-Jones
$\sigma_{lm} (= \sigma)$	distancia de equilibrio de la función de Lennard-Jones
$q_i$	Carga de la partícula "i"
$\epsilon_0$	permitividad del vacío ( $8.85 \times 10^{-12} \text{ C}^2 \text{ J}^{-1} \text{ m}^{-1}$ )
$\epsilon_r$	permitividad relativa
$L_z$	longitud de la caja en la dirección z
$P_{ii}$	componente ii del tensor de presión (p.e.: $P_{xx}$ =componente xx)
V	volumen
$\rho$	concentración molar
F	constante de Faraday ( $96485.3399 \text{ C mol}^{-1}$ )
$\beta$	$1/k_b T$
R	constante de los gases ideales ( $8.314472 \text{ J mol}^{-1} \text{ K}^{-1}$ )
$\Delta G$	diferencia de energía libre de Gibbs
h	constante de Planck ( $6.626068 \times 10^{-34} \text{ m}^2 \text{ kg}^{-1} \text{ s}$ )

## Glosario de Siglas y Nombres

AA	"All Atoms"
FG	"Fine Grain"
UA	"United Atoms"
CG	"Coarse Grain"
WT4	"Wat Four"
NaW	representación "coarse grain" del ion Na <sup>+</sup>
KW	representación "coarse grain" del ion K <sup>+</sup>
ClW	representación "coarse grain" del ion Cl <sup>-</sup>
SPC	"Single Point Charge"
TIP3P	"Transferable Interaction Potential of 3 Points"
RMN	Resonancia Magnética Nuclear
DM	Dinámica Molecular
MMC	Metropolis Monte Carlo
NVT	"ensemble" canónico (Numero de partículas, Volumen y Temperatura constantes)
NPT	"ensemble" isobárico-isotérmico (Numero de partículas, Presión y Temperatura constantes)
$\mu$ V	"ensemble" micro-canónico (Potencial Químico, Volumen y Temperatura constantes)
BPTI	"Bovine Pancreatic Trypsin Inhibitor"
RDF (=g(r))	"Radial Distribution Function"
vdW	van der Waals
SSD	"Soft Sticky Dipole"
MS-CG	"Multi-Scale Coarse Graining"
FG-CG	"Fine Grain" – "Coarse Grain"
O <sub>WT4</sub>	superátomo de WT4 cargado negativamente
H <sub>WT4</sub>	superátomo de WT4 cargado positivamente
DNX	representación "coarse grain" de los nucleótidos del ADN (N=A,C,G,T)
BRIM	"Bilayer with Restricted Ionic Motion"
RMSD	"Root Mean Square Deviation"
TBP	"TATA Binding Protein"
CAP	"Catabolite Activator Protein"
IHF	"Integration Host Factor"
HPV	"Human PapillomaVirus"
TSE	"Transition State Ensemble"
PME	"Particle Mesh Ewald"
PC	"Principal Component"
VAB	Vector Asociado al "Bending"

IS	Indice de Similitud
SM	Surco Menor
SNARE	"Soluble Nsf Attachment protein REceptor"
SB (=sb)	Synaptobrevina
SX (=sx)	Syntaxina
SN1, SN2	cadena de SNAP-25 modeladas
RMSF	"Root Mean Square Fluctuations"
DTM	Dominio Trans-Membrana
POPC	1-palmitoil-2oleoil-fosfatidilcolina
POPS	1-palmitoil-2oleoil-fosfatidilserina
MFA	Microscopía de Fuerza Atómica

## Resumen

Las simulaciones de dinámica molecular son actualmente una alternativa confiable para obtener detalles atómicos, propiedades estructurales y dinámicas de sistemas moleculares para los cuales resulta difícil o imposible obtener datos experimentales. Sin embargo, debido al alto costo computacional, las escalas espacio-temporales accesibles mediante este método son insuficientes para el estudio de muchos procesos de interés biológico. Una opción para encarar este problema es el uso de representaciones moleculares simplificadas (abreviada CG, del inglés “Coarse Grain”), donde grupos de átomos son reemplazados por sitios de interacción efectivos.

Por otra parte, la precisión en la simulaciones requiere de la correcta consideración de los efectos de solvatación. En este sentido, el modelado del entorno acuoso a nivel CG es al día de hoy un área de intenso trabajo. Entre los principales desafíos se destaca el tratamiento de las interacciones electrostáticas y la capacidad de autogenerar un entorno dieléctrico correcto en un modelo basado en potenciales de interacción estándar (Lennard-Jones y Coulomb). El estado del arte sobre esta rama de la ciencia se detalla en el Capítulo I.

En el Capítulo II de la presente Tesis de Doctorado se desarrolla un modelo CG de agua basado en la formación de arreglos tetraédricos transitorios presentes en el agua pura. Este modelo de solvente acuoso simplificado, llamado Wat Four (WT4), condensa 11 moléculas de agua en cuatro sitios de interacción interconectados de manera tetraédrica. Cada sitio de interacción transporta una carga parcial permitiendo considerar explícitamente las interacciones electrostáticas de largo alcance, generando su propia permitividad dieléctrica y una correcta representación de las propiedades del agua líquida para un rango de temperaturas de relevancia biológica (278-328 K). La aplicación conjunta de este esquema de simulación, junto a la modelización de electrolitos simples ( $\text{Na}^+$ ,  $\text{K}^+$  y  $\text{Cl}^-$ ) posibilitó reproducir varias características de la hidratación específica del ADN en la escala de los microsegundos. Este estudio de la hidratación del ADN se profundizó en el Capítulo III poniendo énfasis en los cambios estructurales inducidos en el ADN por la unión de cationes al surco menor, en un sistema de relevancia biológica como es la interacción de la proteína E2 de papiloma-virus con sus sitios de unión al ADN. Estos resultados permiten proponer un papel activo del entorno iónico en el reconocimiento proteína-ADN.

Finalmente, se logró extender el uso del modelo WT4 a simulaciones “híbridas”, combinándolo con modelos de agua a resolución atómica (Capítulo IV). La aplicación exitosa de este método al estudio de un sistema molecular que incluye un complejo proteico de gran tamaño interaccionando con dos bicapas lipídicas en un medio acuoso con concentración iónica fisiológica avalan la utilidad del método.

Los resultados de los Capítulos I, II y IV han sido publicados en revistas internacionales.<sup>1-3</sup> Los datos presentados en el Capítulo III son parte de un artículo actualmente en preparación (“Active role of ions on the indirect readout mechanism”). A la vez que la extensión del método híbrido para otros modelos de agua atomística presentada en el Capítulo IV forma parte de un segundo artículo en preparación (“Extension of the WT4-based multiscale solvation approach to SPC/E and TIP3P water models”). Cabe destacar que los resultados presentados en el Capítulo IV fueron obtenidos en colaboración con Alex Tek y el Dr. Marc Baaden del Institut de Biologie et Physico-Chimie de París.

1. Darré, L., Machado, M. R. & Pantano, S. Coarse-grained models of water. *Wiley Interdisciplinary Reviews: Computational Molecular Science* (2012).doi:10.1002/wcms.1097
2. Darré, L., Machado, M. R., Dans, P. D., Herrera, F. E. & Pantano, S. Another Coarse Grain Model for Aqueous Solvation: WAT FOUR? *Journal of Chemical Theory and Computation* **6**, 3793–3807 (2010).
3. Darré, L., Tek, A., Baaden, M. & Pantano, S. Mixing Atomistic and Coarse Grain Solvation Models for MD Simulations: Let WT4 Handle the Bulk. *Journal of Chemical Theory and Computation* (2012).doi:10.1021/ct3001816

# Capítulo I. Introducción a las Simulaciones Moleculares Coarse Grain

## Simulaciones Moleculares

Las simulaciones computacionales de sistemas moleculares comenzaron como una herramienta para explotar las máquinas de cómputo electrónico desarrolladas durante y posteriormente a la Segunda Guerra Mundial. Estas máquinas fueron desarrolladas para llevar a cabo cálculos de alto costo computacional vinculadas con el desarrollo de armas nucleares y criptografía. A principios de los años 50 los computadores electrónicos se volvieron relativamente accesibles para uso no militar dando comienzo a la disciplina de las simulaciones computacionales. Probablemente uno de los principales aportes del uso de computadores en el estudio de los fenómenos naturales ha sido la posibilidad de resolver numéricamente ecuaciones que no pueden resolverse de manera exacta, excepto para algunos casos muy especiales. Por ejemplo, si se desea estudiar el movimiento de más de dos cuerpos que interactúan entre sí, incluso la relativamente sencilla mecánica newtoniana hace que el problema se vuelva esencialmente irresoluble analíticamente. Sin embargo, el uso del computador permite su resolución al grado deseado de precisión.

Previo a la aparición de las simulaciones computacionales, la predicción de propiedades de un material recaía sobre el uso de teorías que solo pueden proveer una descripción aproximada. Muy pocos son los sistemas para los que se pueden predecir propiedades teóricas de manera exacta (gas ideal, cristal armónico, modelo de Ising). Por lo tanto, modificaciones de dichas teorías que incorporan diversas aproximaciones (ecuación de van der Waals, teoría de Debye-Hückel para electrolitos, ecuación de Boltzmann para transporte de gases diluidos), han sido desarrolladas para la predicción de propiedades de materiales reales. Estas teorías, para dar buenos resultados, además de ser en sí mismas correctas, necesitan contar con el conocimiento de las interacciones intermoleculares, el cual es de por sí muy limitado. Esto constituye un doble problema a la hora de la comparación con los datos experimentales: si no hay concordancia puede ser porque la teoría es equivocada, porque la aproximación a las interacciones intermoleculares es mala, o por ambas.

Las simulaciones surgieron como una herramienta intermedia en la comparación entre teoría y experimento, además de convertirse en sí mismas en una forma de predicción teórica de propiedades. En primer lugar son una herramienta que depende únicamente de la estimación de las interacciones intermoleculares y no de una teoría aproximada como en los casos anteriores. Esto permite dos cosas de gran valor: primero, la comparación con

resultados experimentales sugiere directamente la validez de las estimaciones de las interacciones intermoleculares y por ende permite su refinamiento; y segundo, la comparación con las predicciones hechas con una teoría aproximada provee información directa sobre la validez de la teoría, pues la validez de las interacciones habría sido evaluada en la comparación entre la simulación y el experimento. Es así como cobra sentido la denominación de experimentos computacionales para las simulaciones.

El uso más común de las simulaciones computacionales de sistemas moleculares es la predicción de propiedades de materiales de interés. Claramente es preferible obtener dichas propiedades a partir de medidas experimentales, y en muchos casos es más fácil de esa manera que recurrir a las simulaciones. Sin embargo, son muchos los casos en que no se cuenta con las posibilidades técnicas para obtener una medida de una propiedad de interés y es allí donde las simulaciones cobran valor. Además es posible también utilizar las simulaciones para estudiar materiales nuevos, que ni siquiera han sido sintetizados aún. Por último las simulaciones son de gran utilidad para el análisis de datos como es la técnica de obtención de información estructural de macromoléculas a partir de datos de Resonancia Magnética Nuclear (RMN). En este caso se alimenta al algoritmo de simulación con datos obtenidos por RMN y se busca la estructura que sea tanto energéticamente favorable como compatible con dichos datos experimentales.

Cabe mencionar que las simulaciones pueden ser utilizadas como herramientas netamente exploratorias, avanzando más allá de lo que se conoce de un sistema dado y prediciendo nuevos fenómenos. Es de destacar en tal sentido el debate hacia 1950 sobre la formación de cristales en un sistema de partículas esféricas con repulsiones fuertes de corta distancia, sin atracción mutua. En aquel momento las simulaciones de Alder y Wainwright y las de Wood y Jacobson dieron respuesta afirmativa a dicha interrogante. A pesar del escepticismo, 30 años de trabajo posterior han demostrado dicha predicción.

Las dos técnicas de simulación de sistemas moleculares más ampliamente utilizadas han sido: simulaciones Metropolis Monte Carlo (MMC) y simulaciones de Dinámica Molecular (DM). La versión del método Monte Carlo denominada Metropolis Monte Carlo fue introducida por un grupo de físicos del Laboratorio Nacional de Los Alamos en la década de 1950 mientras trabajaban en el desarrollo de armas nucleares. Se trata de un método estocástico para muestrear puntos de un espacio multidimensional de acuerdo a una predeterminada distribución de probabilidad definida sobre tal espacio. Dicho espacio multidimensional es el espacio configuracional generado por las coordenadas de las moléculas constituyentes del sistema a lo cual se suman algunas variables macroscópicas extensivas a las que se les permite fluctuar. La densidad de probabilidad se



define de acuerdo con el tipo de conjunto estadístico o “ensemble”. En el algoritmo original,<sup>1</sup> cada paso de simulación consiste de dos etapas: primero se intenta la transición desde el estado actual  $i$  a un nuevo estado  $j$  con probabilidad  $\alpha(i \rightarrow j)$ , donde la matriz estocástica de probabilidad de intentos es simétrica por lo que:

$$\alpha(i \rightarrow j) = \alpha(j \rightarrow i)$$

En segundo lugar se acepta la transición con la probabilidad definida por:

$$P_{\text{aceptar}}(i \rightarrow j) = \min \left[ 1, \frac{\rho^{\text{eq}}(j)}{\rho^{\text{eq}}(i)} \right]$$

Si se rechaza la transición, se mantiene el estado  $i$  como el nuevo estado. Con estas definiciones, las probabilidades de transición  $P(i \rightarrow j) = \alpha(i \rightarrow j)P_{\text{aceptar}}(i \rightarrow j)$ , desde el estado  $i$  al  $j$ , satisfacen la condición de reversibilidad microscópica. Por lo tanto, la cadena de estados o conformaciones generada muestrea asintóticamente la distribución de probabilidad elegida. Es decir que en la etapa de la simulación MMC donde todas las propiedades fluctúan alrededor de valores correspondientes al equilibrio termodinámico, cada estado  $i$  es muestreado con una frecuencia proporcional a su densidad de probabilidad de equilibrio. Este tipo de muestreo se denomina “importance sampling”.

MMC fue originalmente desarrollado para muestrear el “ensemble” canónico (NVT), sin embargo, ha sido extendido para aplicarse a los ensembles isotérmico-isobárico (NPT), y gran canónico ( $\mu$ VT).

Una de las principales carencias de MMC es que no da información sobre la relación temporal entre los estados muestreados. Sin embargo presenta al menos tres características que lo mantienen con gran vigencia: a- puede adaptarse rápidamente a nuevos ensembles; b- puede alcanzar el estado de equilibrio mucho más eficientemente que las simulaciones de DM; c- permite el estudio de sistemas con hipersuperficies de potencial muy accidentadas.

En sus comienzos fue instrumentado para probar las teorías basadas en la mecánica estadística, pero actualmente se lo considera como una herramienta predictiva y de diseño en sí misma. Es posible, utilizando MMC, realizar predicciones cuanti- o semicuantitativas sobre la composición, estructura, propiedades

termodinámicas, diagramas de fase y energía libre de diversos sistemas. Entre algunas aplicaciones más recientes puede destacarse su uso en el marco de la teoría del estado de transición (Kinetic Monte Carlo), que permite estimar constantes de velocidad de eventos poco frecuentes; así como también su papel en el modelado “multiscale”<sup>1</sup> que permite reducir el “gap” entre las interacciones a nivel molecular y las propiedades macroscópicas.

Siendo uno de los grandes logros intelectuales del siglo 20, las simulaciones Metropolis Monte Carlo han encontrado una amplia gama de aplicaciones no solo en el área las ciencias naturales y exactas, sino también en la ingeniería y la economía.

Por su lado, las primeras simulaciones de DM fueron reportadas hacia fines de los años 50 y principios de los 60, con los trabajos de: Alder y Wainwright sobre interacción de esferas rígidas (1956); Vineyard quien simuló el daño radiactivo en Cu cristalino (1959); y Rahman y Argonne que realizaron la primera simulación por DM de argón líquido (1964). La primer simulación por DM de agua líquida fue realizada por Rahman y Stillinger en 1971 y la primer simulación de una proteína (el inhibidor de la tripsina pancreática bovina, BPTI) fue publicada en 1977 por McCammon, Gelin y Karplus. Esta última ayudó a cambiar la concepción relativamente rígida de las macromoléculas biológicas, pasando a una visión dinámica, en la que los movimientos internos juegan un papel funcional.

En los últimos 60 años una amplia variedad de fenómenos han sido estudiados utilizando simulaciones de DM, incluyendo ácidos nucleicos, proteínas, bicapas lipídicas, diversos polímeros, azúcares, polielectrolitos, y una gran variedad de sistemas complejos. En todos los casos incrementando progresivamente la ventana espacio-temporal simulada, y refinando la descripción de las interacciones interatómicas.

Tres tipos de aplicaciones de los métodos de simulación se destacan: a-muestreo del espacio configuracional (p.e. refinamiento de estructuras obtenidas experimentalmente), b-obtención de descripción del sistema en equilibrio, incluyendo propiedades estructurales y de movimiento, así como el valor de parámetros termodinámicos, c-examinar la dinámica del sistema siguiendo el curso temporal. Para las primeras dos aplicaciones ambas técnicas mencionadas, MMC y MD, son útiles, por el contrario para la tercera, solo las simulaciones de DM son adecuadas en tanto proveen la secuencia temporal de los cambios estructurales.

El fundamento de las simulaciones de DM se detallan en la siguiente sección.

---

1 Construcción de modelos simplificados a partir de modelos atomísticos.

**Dinámica Molecular.** La simulación de DM es una técnica para el cálculo de propiedades de equilibrio y transporte de un sistema clásico<sup>2</sup> constituido por un cierto número de partículas que interactúan entre sí. Se basa en la resolución de la ecuación de movimiento de Newton para dicho sistema, estando éste sometido a un determinado campo de fuerza<sup>3</sup>. Esta es una buena aproximación para una amplia gama de materiales, ya que cuando solo se consideran los movimientos traslacionales o rotacionales de átomos o moléculas livianas (He, H<sub>2</sub>, D<sub>2</sub>) o movimientos vibracionales con frecuencia  $\nu$  tal que  $h\nu > k_B T$ , los efectos cuánticos dejan de ser despreciables.

Las simulaciones de DM son similares en varios sentidos a los experimentos reales. Cuando se realiza un experimento real se siguen una serie de pasos: preparación de la muestra en condiciones que permitan controlar varias propiedades utilizando diversos instrumentos de medida (termómetro, manómetro, viscosímetros, etc), y finalmente se miden las propiedades de interés durante un período de tiempo suficientemente largo, teniendo en cuenta el ruido estadístico. Por su lado, en las simulaciones de DM se sigue el mismo procedimiento: en primer lugar se debe preparar la muestra, es decir se selecciona un sistema de estudio de  $N$  partículas, se lo acopla a instrumentos de medida (termostato y/o barostato, etc) y se resuelven las ecuaciones de movimiento de Newton hasta que las propiedades del sistema se mantienen constantes (equilibración del sistema). Luego de preparada la muestra se toman las medidas durante la fase de producción, teniendo también en consideración el ruido estadístico para determinar el tiempo de medición adecuado.

El siguiente algoritmo ejemplifica el procedimiento general detrás de una simulación de DM:<sup>1</sup>

1. Lectura de los parámetros que especifican las condiciones de simulación (p.e.: temperatura inicial, número de partículas, densidad, paso de tiempo utilizado para la integración de las ecuaciones de movimiento).
2. Inicialización del sistema: asignación de posiciones y velocidades iniciales (de acuerdo con la temperatura inicial asignada, ver más abajo).
3. Cálculo de fuerzas sobre todas las partículas.
4. Integración de las ecuaciones de movimiento de Newton.

---

<sup>2</sup> Obedece las leyes clásicas de movimiento.

<sup>3</sup> El concepto general de “campo de fuerzas” permite describir la interacción mutua entre las partículas de un sistema. En lugar de hablar de la acción de una partícula sobre otra, se puede decir que una partícula genera un campo entorno de sí, de forma que una determinada fuerza actúa sobre cada una de las demás partículas situadas en dicho campo [1]. En el caso particular de las simulaciones clásicas dicho campo de fuerzas está determinado por la forma funcional que se adopte para describir la energía potencial y por la parametrización de la misma de acuerdo a cada tipo de átomo en el sistema.

5. Se repiten los pasos 3 y 4 hasta que se obtiene la evolución temporal del sistema por el intervalo de tiempo deseado.
6. Una vez completado el paso 5, se calculan los promedios de las propiedades medidas.

### Calculo de las fuerzas

Este representa el paso más costoso del punto de vista computacional ya que para un sistema de  $N$  partículas con interacciones “pairwise” (de a pares) aditivas, al menos  $N(N-1)/2$  distancias deben evaluarse, dando un algoritmo de orden  $N^2$ . Sin embargo, comúnmente se aplican técnicas eficientes para acelerar la evaluación de las fuerzas de forma tal que el orden del algoritmo sea  $N\log(N)$ .

Para todas parejas de partículas  $(i,j)$  se calcula la distancia “ $r$ ” entre ellas y a partir de ésta, la fuerza entre dichas partículas como el gradiente de la energía potencial “ $u$ ”. Por ejemplo el cálculo del componente  $x$  de la fuerza viene dado por:

$$f_x(r) = -\frac{\partial u(r)}{\partial x} = -\left(\frac{x}{r}\right)\left(\frac{\partial u(r)}{\partial r}\right)$$

### Integración de las ecuaciones de movimiento de Newton

Una vez calculadas las fuerzas entre todas los pares de partículas se pueden integrar las ecuaciones de movimiento de Newton. Existen varios algoritmos para esta tarea, y a continuación se expone uno de ellos, el denominado algoritmo Verlet. Su derivación parte de considerar la expansión de Taylor de la coordenada de una partícula en el entorno de un tiempo  $t$ :

$$r(t+\Delta t) = r(t) + v(t)\Delta t + \frac{f(t)}{2m}\Delta t^2 + \ddot{r}\frac{\Delta t^3}{3!} + O(\Delta t^4)$$

$$r(t-\Delta t) = r(t) - v(t)\Delta t + \frac{f(t)}{2m}\Delta t^2 - \ddot{r}\frac{\Delta t^3}{3!} + O(\Delta t^4)$$

Sumando estas dos ecuaciones término a término se obtiene:

$$r(t + \Delta t) + r(t - \Delta t) = 2r(t) + \frac{f(t)}{m} \Delta t^2 + O(\Delta t^4)$$

$$\downarrow$$

$$r(t + \Delta t) \approx 2r(t) - r(t - \Delta t) + \frac{f(t)}{m} \Delta t^2$$

De donde se observa que conociendo la fuerza a tiempo  $t$  y la posición a  $t - \Delta t$ , puede estimarse la posición a tiempo  $t + \Delta t$ , con un error asociado del orden de  $\Delta t^4$ , donde  $\Delta t$  es el paso de tiempo de la simulación de DM. Una vez conocida la nueva posición, pueden estimarse las nuevas velocidad en la misma a partir de:

$$r(t + \Delta t) - r(t - \Delta t) = 2v(t) \Delta t + O(\Delta t^3)$$

$$\downarrow$$

$$v(t) = \frac{r(t + \Delta t) - r(t - \Delta t)}{2 \Delta t} + O(\Delta t^2)$$

Como se menciona más arriba, existen diversos algoritmos para llevar a cabo simulaciones de DM y caben destacar algunos criterios que ayudan a identificar aquellos más adecuados según el caso de estudio: errores introducidos al truncar la expansión de Taylor, velocidad de procesamiento, requerimientos de memoria, máxima longitud del paso de tiempo sin pérdida de precisión, conservación de la energía a cortos y largos tiempos de simulación, reversibilidad, preservación del área del sistema, entre otras.<sup>1</sup>

**Medición de propiedades.** El comportamiento de un sistema en el transcurso de una simulación de DM, posterior a la fase de equilibración, corresponde al estado de equilibrio del mismo, y por ende las mediciones que se obtienen no son sino propiedades en el estado de equilibrio. En tal sentido pueden medirse dos tipos de propiedades: estáticas y dinámicas. La manera en que se obtienen dichas medidas es expresando el observable deseado en términos de las posiciones y momentos de las  $N$  partículas del sistema.

Las propiedades estáticas son aquellas que no dependen, del punto de vista macroscópico, de la evolución temporal del sistema y pueden obtenerse por ende, tanto de simulaciones de DM como MMC. Este tipo de propiedades pueden clasificarse en dos clases: propiedades termodinámicas y propiedades estructurales. Entre

las propiedades termodinámicas de interés, se encuentran dos categorías, aquellas expresables como promedio de alguna función dependiente de las coordenadas y momento de las partículas del sistema (temperatura, presión, capacidad calorífica. etc), y aquellas que no (entropía, energía libre de Helmholtz, energía libre de Gibbs, etc). Para estas últimas son necesarias técnicas de simulación específicas, mientras que para las primeras basta una simulación de DM estándar. A modo de ejemplo, la medición de la temperatura, definida de acuerdo con la equipartición de la energía sobre los grados de libertad, se obtiene de la siguiente manera:<sup>1</sup>

$$\sum_{i=1}^N \frac{m_i v(t)_i^2}{2 N_f} = \left\langle \frac{1}{2} m v_\alpha^2 \right\rangle = \frac{1}{2} k_B T$$

$$\downarrow$$

$$T(t) = \sum_{i=1}^N \frac{m_i v(t)_i^2}{k_B N_f}$$

siendo  $N_f$  el número de grados de libertad del sistema ( $3N-3$  para momento total fijo),  $m_i$  la masa de la partícula “i”,  $v(t)_i$  la velocidad a tiempo “t” de dicha partícula y  $k_B$  la constante de Boltzmann. De aquí se puede observar que la temperatura del sistema, a tiempo t, se calcula en función de la velocidad de cada partícula a dicho tiempo.

Entre las propiedades estructurales caben mencionar las funciones que caracterizan la estructura local de los fluidos, siendo probablemente la más destacable la función de distribución radial (RDF, por su nombre en inglés: Radial Distribution Function). La importancia de la misma resulta de interés principalmente por dos razones: i)- métodos experimentales como difracción de rayos-X y difracción de neutrones en fluidos simples, así como experimentos de difracción de luz en suspensiones coloidales, proveen información sobre la RDF; ii)-la RDF juega un papel central en las teorías del estado líquido. A partir de una simulación, puede medirse la RDF como la relación entre la densidad del número de partículas promedio a una distancia r de un átomo dado, y la densidad a distancia r desde un átomo en un gas ideal con la misma densidad global.

Las propiedades dinámicas, a diferencia de las estáticas, sí son dependientes del tiempo y por ende no pueden estudiarse utilizando métodos MMC, siendo entonces las simulaciones de DM el método de elección. Debe notarse que el estudio de propiedades dependientes del tiempo en el marco de un estado de equilibrio no lleva a contradicción siempre que la perturbaciones que sufra el sistema sean leves. Es decir, el curso temporal de un

sistema perturbado levemente está completamente descrito por las características del estado del equilibrio. En este marco, algunas propiedades que caben mencionar son aquellas relacionadas con procesos de transporte y que se obtienen a partir de integrar funciones de correlación temporal, a saber: coeficiente de difusión ( $D$ ), viscosidad ( $\eta$ ), conductividad térmica ( $\lambda_T$ ) y conductividad eléctrica ( $\sigma_e$ ). A modo de ejemplo, para el cálculo del coeficiente de difusión a partir de las velocidades obtenidas en una simulación de DM se aplica la siguiente ecuación:<sup>1</sup>

$$D = \int_0^{\infty} d(\tau) \langle v_x(\tau) v_x(0) \rangle$$

( $\tau = t - t'$ )

## Modelos Coarse Grain

Las simulaciones computacionales de sistemas biológicos están en continua y pujante expansión motivadas por el constante incremento del poder de cómputo, permitiendo así el estudio de sistemas de mayor complejidad y escalas de tiempo cuya intersección con las biológicamente relevantes es cada vez mayor.<sup>1</sup> Dicha capacidad computacional depende, además del poder de cómputo propiamente dicho, de la complejidad y nivel de detalle que se desea obtener de las simulaciones. La complejidad está determinada por la forma funcional de la energía y el detalle en la descripción de los átomos del sistema. El tratamiento de la energía depende del tipo de descripción física necesaria (cuántica o clásica). Por su lado el nivel de detalle atómico está determinado por el modo en que los átomos y las moléculas que conforman el sistema de estudio sean representados. Según el problema, pueden utilizarse diversos niveles para el tratamiento de las interacciones, cambiando el efecto concreto que ejerce una partícula (ya sea un electrón, un átomo o un conjunto de átomos), sobre las otras partículas, por un efecto medio. Es así que puede adoptarse desde una descripción explícita de núcleo y electrones (métodos cuánticos MC), tratar implícitamente los electrones (métodos clásicos “All Atoms”, AA) o los electrones e hidrógenos no polares (métodos clásicos “United Atoms”, UA), agrupar varios átomos en “superátomos” tratando también los electrones implícitamente (métodos clásicos “Coarse-Grain”, CG), hasta tratar una especie molecular completa de manera implícita (métodos de continuo). Esta última aproximación es comúnmente utilizada para representar el solvente que rodea un soluto dado bajo estudio. A su vez, estos métodos pueden combinarse en una misma simulación para describir con mayor precisión una porción del

sistema mientras que el resto se trata de manera más aproximada, simulando de forma detallada solo las regiones de mayor interés.

La necesidad de lograr descripciones a nivel estructural y dinámico de sistemas de mayor envergadura ha empujado a la comunidad científica a realizar esfuerzos en el desarrollo de tales modelos simplificados de los sistemas moleculares. En este contexto los modelos CG han tomado particular relevancia en tanto permiten capturar las propiedades físicas fundamentales del fenómeno bajo estudio, utilizando una representación molecular que reduce fuertemente las demandas computacionales.<sup>3,4</sup> De esta manera colocan la frontera de simulación en la escala de los micro-milisegundos y las decenas de nanómetros, permitiendo por ejemplo simular la dinámica de una cápside viral.<sup>5</sup>

Comenzando por los modelos pioneros utilizados para describir el plegamiento de proteínas<sup>6,7</sup> una enorme cantidad de aplicaciones exitosas se han publicado, cubriendo un amplio rango de aplicaciones de interés biológico y nanotecnológico.<sup>5,8-19</sup>

En este contexto, el tratamiento de los efectos del solvente acuoso reviste gran interés debido a las propiedades fisico-químicas únicas del agua que le asignan un rol central en los procesos biológicos. El efecto hidrofóbico es un buen ejemplo de la importancia del agua, actuando como motor de procesos de autoensamblaje como la formación de bicapas lipídicas o el plegamiento de proteínas.<sup>20,21</sup>

Consecuentemente una inmensa cantidad de trabajo se ha dedicado a caracterizar el agua y sus propiedades como solvente biológico, utilizando la más amplia gama de métodos disponibles, tanto experimentales como teóricos, incluyendo los métodos de simulación molecular mencionados en los párrafos anteriores. Sin embargo la capacidad computacional disponible actualmente limita las escalas temporal y espacial al orden de los microsegundos y decenas de nanómetros, respectivamente. Además, la abundancia de agua en los sistemas biológicos típicos, implica una dedicación de ~80% del tiempo computacional en el cálculo de las interacciones agua-agua, que no son, en general el motivo de estudio. En este contexto, simplificaciones respecto de una descripción a nivel atomístico o Fine Grain (FG, por contraposición a CG) resultan muy atractivos permitiendo extender los límites espacio temporales, a la vez de mantener la esencia de las propiedades fisico-químicas del fenómeno de interés.<sup>17</sup> Para abordar esta problemática, dos aproximaciones pueden destacarse, por un lado el uso de métodos de solvatación implícita, en la cual las moléculas de agua no están presentes en el sistema a simular sino que sus características se incorporan via interacciones efectivas entre las partículas del soluto.<sup>22-28</sup> Sin embargo, tal aproximación puede resultar un muchos casos demasiado extrema, en particular cuando la presencia explícita de moléculas de agua es importante en el fenómeno que se pretende estudiar. En tales casos



la segunda aproximación que se ha propuesto es el uso de modelos CG de solvente acuoso, existiendo actualmente varias alternativas que se comentan en los siguientes párrafos. El desarrollo de una de éstas será presentado en el Capítulo II del presente trabajo. Adicionalmente, una extensión del uso de modelos CG es el desarrollo de métodos multiescala, que combinan representaciones FG y CG para mantener el detalle atómico en regiones de interés, mientras que el resto de sistema de estudio se trata de manera simplificada. Estos métodos serán revisados al final del presente capítulo, y el desarrollo de un modelo de ese tipo será el tema central del Capítulo IV de la presente Tesis.

Previo a introducir los modelos CG de agua presentes en la literatura, cabe comentar brevemente sobre algunos modelos de agua FG comúnmente utilizados en la comunidad de las simulaciones, de manera de tener un marco de referencia. Un importante número de tales modelos han sido publicados, utilizando diferentes estrategias de ajuste y número de partículas por molécula. Los modelos más simples y comúnmente utilizados en las simulaciones FG son el TIP3P (“Transferable Interaction Potential 3 Points”) y el SPC (“Simple Point Charge”).<sup>29,30</sup> En ambos casos, una molécula de agua es representada por medio de un átomo de oxígeno rígidamente unido a dos átomos de hidrógeno, con cargas parciales sobre cada átomo. Las interacciones de Lennard-Jones se definen únicamente para el átomo de oxígeno. Esquemas más elaborados incluyendo partículas adicionales para representar momentos dipolares o pares solitarios de electrones también han sido propuestos.<sup>31</sup> Adicionalmente, la inclusión de flexibilidad en los enlaces y ángulos permite incorporar efectos de polarizabilidad en el marco de simulaciones clásicas. Para una exhaustiva revisión de los modelos existentes en la literatura, así como de las propiedades que son capaces de reproducir se sugiere ver el sitio web: <http://www.lsbu.ac.uk/water/models.html>. Sin embargo, debe tenerse en cuenta que el aumento en la complejidad del modelo no asegura una mejor reproducción de las propiedades experimentales.<sup>32</sup> De hecho, el buen balance entre eficiencia computacional y precisión en la reproducción de propiedades del agua en el “bulk” han posicionado a los modelos SPC y TIP3P como las opciones más comunes en la gran mayoría de las simulaciones de DM.

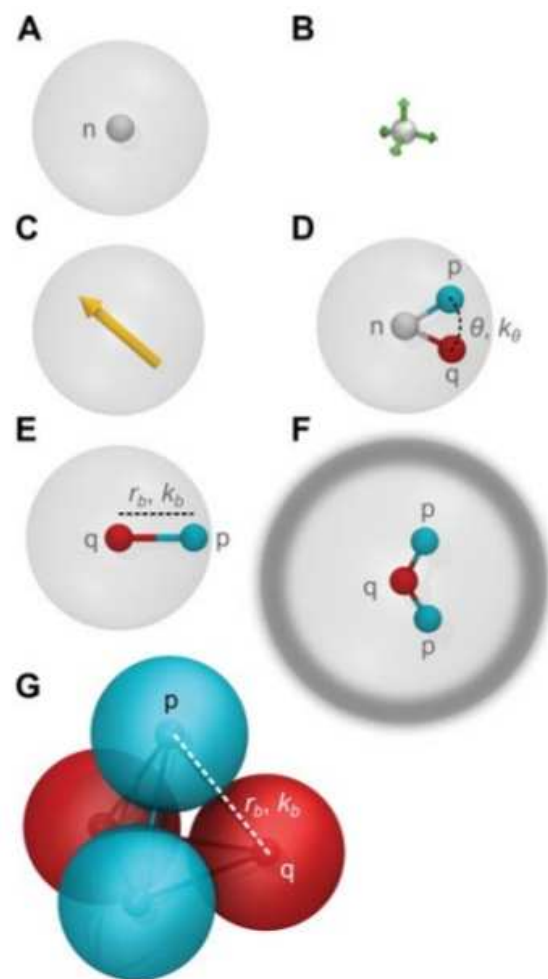
Varios modelos CG de agua han sido propuestos en la literatura con el objetivo de reducir el costo computacional relacionado con la consideración explícita del entorno acuoso. En general, tres características principales determinan la capacidad de un modelo simplificado de agua de reproducir las propiedades requeridas: el nivel de simplificación o de CG, es decir el número de aguas reales representadas por una partícula CG; la forma funcional del potencial de interacción; y la distancia máxima utilizada para calcular las interacciones (“cutoff”). En tal sentido es importante tener presente que las partículas CG no son moléculas reales, por lo cual, no existe un único procedimiento general para derivar los parámetros de interacción.

La representación más sencilla corresponde a un líquido de van der Waals (vdW) con un único sitio de interacción por partícula (Figura 1-1.A). En general, tales modelos condensan desde una a cinco moléculas de agua por partícula CG,<sup>33-35</sup> permitiendo una mayor aceleración en la simulación, a medida que se aumenta el número de moléculas de agua por partícula CG. Sin embargo, también aumenta la dificultad de reproducir o incluso de relacionar las propiedades calculadas a partir del modelo con aquellas del agua real.<sup>36</sup> Por otro lado las interacciones intermoleculares son generalmente calculadas de manera "pair-wise" (de a pares) utilizando potenciales empíricos con parámetros derivados de datos experimentales. Alternativamente, la parametrización puede realizarse de manera sistemática a partir de simulaciones FG utilizando procedimientos como Force Matching, Iterative Boltzmann Inversion o Inverse Monte Carlo.<sup>37</sup> Independientemente del método utilizado, es imposible para cualquier potencial aditivo "pair-wise" reproducir todas las propiedades de una sustancia dada.<sup>38</sup> Esto implica que, si se intenta ajustar el modelo para reproducir una cierta propiedad de interés, la precisión en otras propiedades se verá comprometida, dificultando la transferibilidad del modelo. Un buen ejemplo de este tipo de dificultades puede encontrarse en el trabajo de He et al,<sup>39</sup> en el que varios modelos de agua de tipo líquido de vdW se estudian analizando diferentes potenciales de interacción, distancia de "cutoff" de las interacciones no enlazantes y niveles de CG entre una y cuatro moléculas de agua por partícula CG. Los autores demuestran que la densidad, tensión superficial y compresibilidad son propiedades muy difíciles de reproducir simultáneamente, lográndose únicamente con niveles de CG de una o dos moléculas de agua por sitio CG. Niveles de CG mayores implican que alguna de dichas propiedades no será reproducida correctamente.<sup>39</sup> Adicionalmente los autores muestran que el uso de distancias de "cutoff" muy cortas o funciones de potenciales de interacción con gradientes muy elevados, pueden llevar a aumentos en la temperatura de fusión del líquido, con serias consecuencias en sistemas biológicos, pues al agua podría comportarse en estado sólido a temperatura ambiente. Esto ocurre de hecho con el modelo implementado en el campo de fuerza MARTINI<sup>35</sup> (nivel de CG: 4-1 y "cutoff" : 1.2nm). En este caso los autores proponen una estrategia práctica para evitar tal problema que es introducir una pequeña cantidad (~0.1M) de partículas CG más voluminosas que actúan como "anticongelantes", evitando el empaquetamiento.<sup>35</sup>

Un problema importante de los modelos de tipo líquido de vdW es que las interacciones electrostáticas no pueden incorporarse utilizando un único sitio de interacción. Una posible mejora es incluir anisotropía en las interacciones de vdW, como es el caso del modelo propuesto por Molinero y Moore.<sup>40</sup> Este modelo, basado en un nivel de CG 1-1, reproduce los arreglos tetraédricos observables en el agua líquida, así como también la estructura de solvatación de sales monovalentes.<sup>41</sup> Sin embargo, dicha aproximación no incluye el

apantallamientos electrostático inherente al agua real. Una corrección a tal carencia es incluir el efecto de

**Figura 1-1. Modelos CG de agua explícita.** A- Modelo de agua del campo de fuerza CG MARTINI. Este es un ejemplo típico de modelo de tipo van der Waals como los descritos por He et al.<sup>42</sup> B- Modelo tridimensional Mercedes Benz.<sup>43</sup> Las flechas verdes representan los vectores que imitan las interacciones de enlace de hidrógeno. C- Modelo CG de solvente polarizable.<sup>44</sup> Las flechas amarillas representan el momento dipolar. D- Modelo MARTINI p.<sup>45</sup> E- Modelo GROMOS.<sup>46</sup> F- Modelo Big Multipole.<sup>47</sup> G- Modelo WT4.<sup>48</sup> Los colores de las esferas se refieren a la carga en cada partícula. Azul, rojo y gris corresponden a cargas positiva (p), negativa (q) y neutra (n), respectivamente. El tamaño de las esferas semitransparentes corresponde con el radio de Lennard-Jones de cada modelo. Los parámetros para las interacciones de enlaces y ángulos se indican en cada caso ( $r_b$ ,  $k_b$ ,  $\theta$ ,  $k_\theta$ ).



apantallamiento del agua asumiendo que el sistema entero se encuentra embebido en un medio dieléctrico continuo. Esta aproximación es la utilizada en el campo de fuerza MARTINI, en el que el sistema es embebido en un medio con constante dieléctrica de 15. Sin embargo, dado que los sistemas biomoleculares son altamente heterogéneos, la combinación de interacción de Lennard-Jones con apantallamientos implícitos no es suficiente para reproducir los efectos electrostáticos.

Entre los trabajos pioneros en la inclusión de los efectos electrostáticos se puede citar el trabajo de Warshel y Levitt.<sup>49</sup> Estos autores modelaron las moléculas de agua como una grilla de dipolos alrededor de un soluto de interés. Esta aproximación permitió una determinación empírica de los efectos electrostáticos sobre una proteína debido al entorno acuoso. Desde este trabajo hasta ahora, varias aproximaciones se han propuesto para

incluir las interacciones electrostáticas de manera más realista.

Un ejemplo simple de solvente CG es el modelo bidimensional llamado “Mercedes-Benz”, en el cual las moléculas de agua son representadas como círculos de Lennard-Jones con tres vectores para representar la formación de enlaces de hidrógeno, como forma de incorporar características propias de las interacciones electrostáticas. Este esquema tan sencillo fue suficiente para reproducir varias propiedades del agua.<sup>50</sup> Una extensión de tal modelo para tres dimensiones fue presentada recientemente (Figura 1-1.B). Sin embargo, la incorporación de la direccionalidad de los enlaces de hidrógeno podría ser de difícil implementación en los programas de simulación comúnmente utilizados.

Otro modelo que cabe mencionar es el denominado Soft Sticky Dipole (SSD).<sup>51</sup> Este presenta un nivel de CG 1-1, con un centro de Lennard-Jones, un dipolo y un término de interacción tetraédrico que es responsable de la formación de enlaces de hidrógeno. Este modelo muestra una buena reproducción de propiedades estructurales, termodinámicas y dieléctricas del agua. Sin embargo, el nivel modesto de CG utilizado no generara una notable ganancia del punto de vista computacional respecto de los modelos FG.

Por otra parte, efectos de polarizabilidad han sido incorporados en modelos CG de agua como es el ejemplo del modelo reportado por Borgis et al,<sup>52</sup> denominado PCGS (Figura 1-1.C). En este, un dipolo se localiza en el centro de una partícula con nivel CG 3-1. Las partículas PCGS son sensibles al campo eléctrico generado por el soluto pero no al generado por ellas mismas. En cambio, las interacciones solvente-solvente son tratadas a corta distancia vía la combinación de potenciales de Lennard-Jones y Gaussiano como modo de representar las fuerzas repulsivas y atractivas, respectivamente. A pesar que la aproximación Gaussiana (que varía como  $e^{-r^2}$ ) solo reproduce de manera muy aproximada las componentes atractivas de las interacciones de Coulomb ( $r^{-1}$ ) y vdW ( $r^{-6}$ ), esta aproximación provee buenas estimaciones de las energías libres de solvatación y las fuerzas derivadas del efecto hidrofóbico. Sin embargo, la implementación de esta aproximación en los programas de simulación comúnmente utilizados y la eficiencia computacional podrían verse comprometidos por el uso de los mencionados potenciales.

Más recientemente se han propuesto modelos con la característica común de que las interacciones electrostáticas son calculadas de modo "pair-wise" usando términos de Coulomb estándar entre sitios que transportan carga parcial. En primer lugar se destaca una extensión del modelo del campo de fuerza MARTINI, denominado MARTINI<sub>p</sub>, en el que se agrega a la partícula estándar que solo presenta interacciones de Lennard-Jones, dos partículas adicionales que poseen cargas parciales opuestas. Estas se mantienen a una distancia fija de la partícula de Lennard-Jones central por medio del uso de vínculos (“constraints”), a la vez que un potencial

angular armónico con un ángulo de equilibrio de cero grados se impone entre las tres partículas, como modo de modular la distribución del momento dipolar (Figura 1-1.D). Las dos partículas cargadas interactúan con las partículas análogas en otras moléculas únicamente vía interacciones de Coulomb. A pesar del uso de interacciones electrostáticas explícitas, el uso de un medio dieléctrico continuo de  $\epsilon=2.5$  es necesario para reproducir la permitividad del agua real. Esta nueva versión del agua CG del campo de fuerza MARTINI presenta varias mejoras entre las que se destaca, además de la mejor reproducción de las propiedades dieléctricas del agua, una disminución del punto de fusión (282 K). Además, el modelo es compatible con el resto de las moléculas para las cuales puede utilizarse el campo de fuerzas MARTINI. Una contraindicación de este nuevo modelo radica en el aumento del número de partículas para representar el agua respecto de la versión original, aumentando el costo computacional en un factor de tres.<sup>45</sup>

Otro modelo CG de agua fue propuesto recientemente por Riniker y van Gunsteren<sup>46</sup> (Figura 1-1.E), en el cual el desarrollo de los parámetros se basa en el análisis sistemático de diferentes propiedades de pequeños “clusters” de agua extraídas a partir de simulaciones FG. Este modelo, denominado GROMOS, presenta un nivel de CG 5-1 con dos sitios de carga puntual opuesta conectados por una distancia variable, lo cual le permite reproducir fenómenos de polarizabilidad. Las interacciones electrostáticas son consideradas explícitamente entre las partículas CG utilizando una distancia de “cutoff” para las interacciones de corta distancia mientras que un campo de reacción (“reaction field”) se emplea como corrección para las interacciones de largo alcance. Solo la partícula central incluye interacciones de Lennard-Jones estándar (12-6). Estas características permiten la transferibilidad a otros programas de simulaciones, además del GROMOS, en el cual se hizo el desarrollo.

Por último, cabe mencionar el modelo propuesto por Wu et al<sup>47</sup> denominado Big Multipole Water (BMW, Figura 1-1.F). Este presenta tres sitios de carga parcial unidos por “constraints” entre sí, para dar lugar a una partícula CG que representa cuatro moléculas de agua. En consecuencia, cada partícula CG lleva consigo un dipolo y un cuadrupolo. Las interacciones electrostáticas de largo alcance son incorporadas utilizando el método Particle Mesh Ewald (PME).<sup>53,54</sup> Al igual que en el caso del modelo MARTINI, una constante dieléctrica continua de  $\epsilon=1.3$  es necesaria para mejorar el ajuste a la permitividad del agua real. Adicionalmente, las interacciones no electrostáticas son calculadas utilizando un potencial “suave” entre los sitios centrales de las diferentes moléculas.

En base a la revisión presentada en los párrafos anteriores, actualmente el desarrollo de modelos CG de agua es un área de intenso trabajo en el que la reproducción de las propiedades del agua líquida, en el bulk y como solvente para biomoléculas, junto con una elevada eficiencia computacional, son requisitos cada vez más

exigentes. En tal sentido, la descripción de las interacciones electrostáticas de corto y largo alcance, la capacidad de autogenerar un entorno dieléctrico, el uso de potenciales estándar que faciliten tanto la transferibilidad entre los programas de simulación más comúnmente utilizados así como la eficiencia computacional, son objetivos necesarios para nuevos modelos.

En este marco es que se plantea como primer objetivo de la presente Tesis, el desarrollo de un modelo CG de agua que cumpla con los objetivos recién mencionados. Concretamente, que se base en el uso combinado de los potenciales de interacción no enlazante Lennard-Jones y Coulomb, que localice cargas parciales en la partículas CG, las cuales deberán ajustarse junto con la topología del modelo para reproducir lo mejor posible la permitividad dieléctrica del agua real. Adicionalmente, el modelo deberá basar el tratamiento de las interacciones electrostáticas en un método como el PME, que garantice el tratamiento correcto de las interacciones del largo alcance. Por otro lado, debe ser un modelo que permita llevar la escala de las simulaciones de DM a los micro-milisegundos y ser funcional como solvente acuoso para biomoléculas de interés, entre las cuales se destaca el modelo CG de ADN publicado recientemente en nuestro grupo,<sup>55</sup> así como otros desarrollo que sigan la misma filosofía. Un objetivo corolario, pero no menos importante es el desarrollo de iones CG compatibles con el modelo de agua, que permitan reproducir las propiedades estructurales y termodinámicas de una solución acuosa de electrolitos en la cual embeber solutos biomoleculares de interés.

Los esfuerzos realizados para llevar a cabo este objetivo se presentan en el Capítulo II de la presente Tesis. En la Figura 1-1.G se muestra un esquema del modelo que desarrolló.

## Simulaciones Multiescala

La descripción de los procesos biológicos puede a menudo requerir el estudio de múltiples escalas espacio-temporales acopladas. En este sentido, aproximaciones donde regiones de interés son representadas a nivel FG, mientras que el resto del sistema es tratado a nivel CG (representación multiescala) son muy atractivas. Particularmente importante, en este sentido, es el rol que juega la naturaleza discreta de las moléculas de agua en la formación y ruptura de enlaces de hidrógeno, fuerzas resultantes de interacciones hidrofóbicas, apantallamiento electrostático, etc., que son difíciles de reproducir con una aproximación puramente CG. Adicionalmente, dado que el agua constituye en promedio 80% de los sistemas biológicos, una representación multiescala donde el soluto y las aguas vecinas son tratadas a nivel atómico mientras que las aguas más alejadas que constituyen el “bulk”, son representadas a nivel CG, representa una herramienta muy útil que

permitiría reducir el número de partículas del sistema y consecuentemente reducir el costo computacional, sin comprometer la precisión de los resultados.

Algunos ejemplos reportados en la literatura sugieren que estas técnicas multiescala constituyen una alternativa práctica para alcanzar varias de las escalas espaciales y/o temporales involucradas en sistemas biomoleculares complejos, manteniendo el detalle atómico en el soluto.<sup>9,56-59</sup>

En términos generales pueden distinguirse dos líneas principales de desarrollo de métodos multiescala: 1- métodos de resolución adaptativa y 2- métodos de resolución fija. Mientras la primera estrategia permite el cambio en el grado de representación durante la simulación, en la segunda, la representación de cada partícula, sea FG o CG, es fijada al comienzo de la simulación y se mantiene invariada durante la misma. A continuación se detallan las principales características de tales métodos.

**Métodos de Resolución Adaptativa.** Usualmente estos métodos dividen el sistema molecular en tres regiones: FG, híbrida FG-CG, y CG. Las partículas pueden cruzar de una región a otra cambiando su resolución en el curso de su trayectoria dentro de la caja de simulación, es decir una partícula CG puede convertirse en una FG y viceversa. La transición es gobernada por una función de escalado que interpola las fuerzas o potenciales de interacción entre las regiones FG y CG, asignando la identidad de una partícula dada de acuerdo a su posición. El uso de una región de interfase es necesaria para “suavizar” la energía potencial y cinética del sistema cuando un grupo cambia de resolución al moverse por la caja de simulación.<sup>60</sup> Distintas alternativas se han propuesto en la literatura. A modo de ejemplo las estrategias propuestas por Ensing et al<sup>61</sup> y Praprotnik et al<sup>62</sup> utilizan diferentes campos de fuerza en cada región e integran las ecuaciones de movimiento de Newton para propagar la dinámica. Por el contrario, Heyden y Threlkoff<sup>63</sup> definen un único Lagrangiano para el sistema completo integrando las ecuaciones de movimiento de Euler-Lagrange. La importancia de esta distinción radica en que, cuando los grados de libertad de un grupo de átomos cambia dependiendo de las coordenadas, el funcional de energía potencial del sistema es dependiente a la vez de la velocidad y las coordenadas. En este marco, el tercer método ofrece la ventaja de conservar la energía y los momentos lineal y angular cuando se pasa de una región del sistema a otra. A pesar de problemas menores (pérdida de orden y pequeñas oscilaciones de la densidad en la región híbrida), estos métodos son altamente prometedores para las simulaciones multiescala de materia condensada. La principal contraindicación es que este tratamiento funciona bien con el uso de un nivel de CG 1-1, proveyendo solo un relativamente pequeño incremento de la eficiencia computacional.

**Métodos de Resolución Fija.** Varios métodos se han propuesto para resolver el problema de simular un soluto FG en un entorno CG. La ventaja en estos casos, adicionalmente a concentrar el esfuerzo computacional solo en una región de interés, es que no se requieren transformaciones posición-dependiente. En los próximos párrafos se describen algunas de estas metodologías junto a sus ventajas y desventajas.

Michel et al<sup>64</sup> desarrollaron un modelo CG para lípidos compatible con el modelo CG de agua SSD, mencionado en la sección anterior.<sup>51</sup> En base a este modelo, desarrollaron un esquema híbrido utilizando términos de escalado adicionales para considerar las interacciones directas entre solutos tratados a nivel FG y el entorno CG (agua y membrana). Esta aproximación ha sido probada exitosamente en la reproducción de coeficientes de partición agua-octano y luego aplicada para describir la interacciones entre la membrana plasmática y agentes antimicrobianos.<sup>65</sup>

Por su lado, Shi et al<sup>66</sup> propusieron una derivación sistemática de parámetros de interacción FG-CG basada en los métodos MultiScale Coarse Graining (MS-CG).<sup>67</sup> Esta aproximación utiliza un procedimiento de Force Matching para determinar el campo de fuerza efectivo "pair-wise" CG directamente a partir de las fuerzas medidas en una trayectoria obtenida por simulación de DM FG. Las interacciones en el sistema de resolución múltiple son divididas en : FG, FG-CG, y CG. Las interacciones FG son las mismas que las utilizadas en las simulaciones FG de referencia. Las interacciones CG son obtenidas aplicando la aproximación MS-CG a las fuerzas generadas durante la simulación FG . Finalmente, para reproducir la fuerza entre partículas de diferente resolución, se sustrae de la fuerza general del sistema, la correspondiente a las interacciones FG-FG y CG-CG, y las fuerzas residuales se utilizan para ajustar los parámetros de interacción FG-CG. Los resultados obtenidos con esta aproximación muestran muy buena reproducción de las características estructurales de sistemas péptido-membrana.

Más recientemente, Rzepiela et al<sup>68</sup> presentaron un ingenioso método alternativo para acoplar dos niveles de resolución, modelando las moléculas FG de manera tal que transportan un sitio de interacción virtual en el centro de masa de cada molécula. A través de tales sitios, las moléculas FG interactúan con el entorno CG usando potenciales CG previamente determinados. La fuerza resultante de las interacciones CG sobre un sitio virtual es distribuida sobre los átomos a los que corresponde tal sitio, como una fracción pesada por la masa de la fuerza total que actúa sobre el sitio. Por lo tanto, no se requieren potenciales de interacción específicos FG-CG. De este modo, este método puede aplicarse a campos de fuerza FG y CG genéricos en los paquetes de simulación comúnmente utilizados. La desventaja observada por el momento de esta metodología es que la reproducción



cuantitativa de los efectos de solvatación solo se logra aplicando una constante dieléctrica homogénea en el sistema como parámetro de ajuste libre. El cálculo de potencial de fuerza media para la solvatación de moléculas de dialanina utilizando butano o agua, es un ejemplo de tal problema.<sup>68</sup>

Como se menciona más arriba, la primera parte de este trabajo se centra en el desarrollo de un nuevo modelo CG de agua e iones. Una vez dado ese paso, se planteará como segundo objetivo general, la utilización de dichos modelos para simulaciones multiescala. De este modo se aprovechan los avances del modelado CG del entorno acuoso en el modelado de resolución múltiple.

Dado que uno de los objetivos principales que se persiguen en el desarrollo del modelo CG de agua e iones es la eficiencia computacional, asociada necesariamente con un elevado nivel de CG, la metodología de Resolución Fija es la opción de modelado multiescala más conveniente. El Capítulo IV de la presente Tesis presenta los principales resultados en esta dirección.

## Biobibliografía

1. Frenkel, D. & Smit, B. *Understanding molecular simulation: from algorithms to applications*. (Academic Press: San Diego, 2002).
2. Karplus, M. & McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nature Structural Biology* **9**, 646–652 (2002).
3. Klein, M. L. & Shinoda, W. Large-Scale Molecular Dynamics Simulations of Self-Assembling Systems. *Science* **321**, 798–800 (2008).
4. Murtola, T., Bunker, A., Vattulainen, I., Deserno, M. & Karttunen, M. Multiscale modeling of emergent materials: biological and soft matter. *Physical Chemistry Chemical Physics* **11**, 1869 (2009).
5. Arkhipov, A., Freddolino, P. L. & Schulten, K. Stability and dynamics of virus capsids described by coarse-grained modeling. *Structure* **14**, 1767–1777 (2006).
6. Levitt, M. & Warshel, A. Computer simulation of protein folding. *Nature* **253**, 694–698 (1975).
7. Tanaka, S. & Scheraga, H. A. Statistical mechanical treatment of protein conformation. I. Conformational properties of amino acids in proteins. *Macromolecules* **9**, 142–159 (1976).
8. Yin, Y., Arkhipov, A. & Schulten, K. Simulations of membrane tubulation by lattices of amphiphysin N-BAR domains. *Structure* **17**, 882–892 (2009).
9. Arkhipov, A., Yin, Y. & Schulten, K. Four-scale description of membrane sculpting by BAR domains. *Biophys. J.* **95**, 2806–2821 (2008).
10. Wee, C. L., Gavaghan, D. & Sansom, M. S. P. Interactions Between a Voltage Sensor and a Toxin via Multiscale Simulations. *Biophysical Journal* **98**, 1558–1565 (2010).
11. Ayton, G. S. & Voth, G. A. Hybrid coarse-graining approach for lipid bilayers at large length and time scales. *J Phys Chem B* **113**, 4413–4424 (2009).
12. Treptow, W., Marrink, S.-J. & Tarek, M. Gating Motions in Voltage-Gated Potassium Channels Revealed by Coarse-

- Grained Molecular Dynamics Simulations. *The Journal of Physical Chemistry B* **112**, 3277–3282 (2008).
13. Yefimov, S., van der Giessen, E., Onck, P. R. & Marrink, S. J. Mechanosensitive Membrane Channels in Action. *Biophysical Journal* **94**, 2994–3002 (2008).
  14. Periole, X., Huber, T., Marrink, S.-J. & Sakmar, T. P. G protein-coupled receptors self-assemble in dynamics simulations of model bilayers. *J. Am. Chem. Soc.* **129**, 10126–10132 (2007).
  15. Durrieu, M.-P., Bond, P. J., Sansom, M. S. P., Lavery, R. & Baaden, M. Coarse-Grain Simulations of the R-SNARE Fusion Protein in its Membrane Environment Detect Long-Lived Conformational Sub-States. *ChemPhysChem* **10**, 1548–1552 (2009).
  16. Srinivas, G., Discher, D. E. & Klein, M. L. Self-assembly and properties of diblock copolymers by coarse-grain molecular dynamics. *Nat Mater* **3**, 638–644 (2004).
  17. Nielsen, S. O., Lopez, C. F., Srinivas, G. & Klein, M. L. Coarse grain models and the computer simulation of soft materials. *Journal of Physics: Condensed Matter* **16**, R481–R512 (2004).
  18. Srinivas, G. & Klein, M. L. Molecular dynamics simulations of self-assembly and nanotube formation by amphiphilic molecules in aqueous solution: a coarse-grain approach. *Nanotechnology* **18**, 205703 (2007).
  19. Zhou, J., Thorpe, I. F., Izvekov, S. & Voth, G. A. Coarse-grained peptide modeling using a systematic multiscale approach. *Biophys. J.* **92**, 4289–4303 (2007).
  20. Chandler, D. Interfaces and the driving force of hydrophobic assembly. *Nature* **437**, 640–647 (2005).
  21. Tanford, C. *The hydrophobic effect : formation of micelles and biological membranes 2d ed.* (J. Wiley.: New York, 1980).
  22. Allen, E. C. & Rutledge, G. C. A novel algorithm for creating coarse-grained, density dependent implicit solvent models. *The Journal of Chemical Physics* **128**, 154115 (2008).
  23. Chang, C.-E. A., Trylska, J., Tozzini, V. & McCammon, J. A. Binding pathways of ligands to HIV-1 protease: coarse-grained and atomistic simulations. *Chem Biol Drug Des* **69**, 5–13 (2007).
  24. Cooke, I. R. & Deserno, M. Solvent-free model for self-assembling fluid bilayer membranes: Stabilization of the fluid phase based on broad attractive tail potentials. *The Journal of Chemical Physics* **123**, 224710 (2005).
  25. Knotts, T. A., Rathore, N., Schwartz, D. C. & de Pablo, J. J. A coarse grain model for DNA. *The Journal of Chemical Physics* **126**, 084901 (2007).
  26. Pasquali, S. & Derreumaux, P. HiRE-RNA: A High Resolution Coarse-Grained Energy Model for RNA. *The Journal of Physical Chemistry B* **114**, 11957–11966 (2010).
  27. Maupetit, J., Derreumaux, P. & Tuffery, P. PEP-FOLD: an online resource for de novo peptide structure prediction. *Nucleic Acids Res.* **37**, W498–503 (2009).
  28. Savelyev, A. & Papoian, G. A. Chemically accurate coarse graining of double-stranded DNA. *Proceedings of the National Academy of Sciences* **107**, 20340–20345 (2010).
  29. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**, 926 (1983).
  30. Berendsen, H., Postma, J., van Gunsteren, W. & Hermans, J. Interaction models for water in relation to protein hydration. *Intermolecular Forces* 331–342 (1981).
  31. Mahoney, M. W. & Jorgensen, W. L. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *The Journal of Chemical Physics* **112**, 8910 (2000).
  32. Guillot, B. A reappraisal of what we have learned during the last three decades of computer simulations on water. *J. Molec. Liquids* **101**, 219–260 (2002).
  33. Moore, P. B., Lopez, C. F. & Klein, M. L. Dynamical Properties of a Hydrated Lipid Bilayer from a Multinano-second Molecular Dynamics Simulation. *Biophysical Journal* **81**, 2484–2494 (2001).
  34. Kowalczyk, P., Gauden, P. A. & Ciach, A. Role of short-range directional interactions in coarse-graining of protic/aprotic liquids. *J Phys Chem B* **113**, 12988–12998 (2009).
  35. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & de Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *The Journal of Physical Chemistry B* **111**, 7812–7824 (2007).
  36. Hadley, K. R. & McCabe, C. On the Investigation of Coarse-Grained Models for Water: Balancing Computational

- Efficiency and the Retention of Structural Properties. *The Journal of Physical Chemistry B* **114**, 4590–4599 (2010).
37. Rühle, V., Junghans, C., Lukyanov, A., Kremer, K. & Andrienko, D. Versatile Object-Oriented Toolkit for Coarse-Graining Applications. *Journal of Chemical Theory and Computation* **5**, 3211–3223 (2009).
  38. Johnson, M. E., Head-Gordon, T. & Louis, A. A. Representability problems for coarse-grained water potentials. *The Journal of Chemical Physics* **126**, 144509 (2007).
  39. He, X., Shinoda, W., DeVane, R. & Klein, M. L. Exploring the utility of coarse-grained water models for computational studies of interfacial systems. *Molecular Physics* **108**, 2007–2020 (2010).
  40. Molinero, V. & Moore, E. B. Water Modeled As an Intermediate Element between Carbon and Silicon †. *The Journal of Physical Chemistry B* **113**, 4008–4016 (2009).
  41. DeMille, R. C. & Molinero, V. Coarse-grained ions without charges: reproducing the solvation structure of NaCl in water using short-ranged potentials. *J Chem Phys* **131**, 034107 (2009).
  42. He, X., Shinoda, W., DeVane, R. & Klein, M. L. Exploring the utility of coarse-grained water models for computational studies of interfacial systems. *Molecular Physics* **108**, 2007–2020 (2010).
  43. Bizjak, A., Urbic, T., Vlachy, V. & Dill, K. A. Theory for the three-dimensional Mercedes-Benz model of water. *The Journal of Chemical Physics* **131**, 194504 (2009).
  44. Basdevant, N., Borgis, D. & Ha-Duong, T. A Coarse-Grained Protein–Protein Potential Derived from an All-Atom Force Field. *The Journal of Physical Chemistry B* **111**, 9390–9399 (2007).
  45. Yesylevskyy, S. O., Schäfer, L. V., Sengupta, D. & Marrink, S. J. Polarizable Water Model for the Coarse-Grained MARTINI Force Field. *PLoS Computational Biology* **6**, e1000810 (2010).
  46. Riniker, S. & van Gunsteren, W. F. A simple, efficient polarizable coarse-grained water model for molecular dynamics simulations. *The Journal of Chemical Physics* **134**, 084110 (2011).
  47. Wu, Z., Cui, Q. & Yethiraj, A. A New Coarse-Grained Model for Water: The Importance of Electrostatic Interactions. *The Journal of Physical Chemistry B* **114**, 10524–10529 (2010).
  48. Darré, L., Machado, M. R., Dans, P. D., Herrera, F. E. & Pantano, S. Another Coarse Grain Model for Aqueous Solvation: WAT FOUR? *Journal of Chemical Theory and Computation* **6**, 3793–3807 (2010).
  49. Warshel, A. & Levitt, M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **103**, 227–249 (1976).
  50. Silverstein, K. A. T., Haymet, A. D. J. & Dill, K. A. A Simple Model of Water and the Hydrophobic Effect. *Journal of the American Chemical Society* **120**, 3166–3175 (1998).
  51. Liu, Y. & Ichiye, T. Soft Sticky Dipole Potential for Liquid Water: A New Model. *The Journal of Physical Chemistry* **100**, 2723–2730 (1996).
  52. Ha-Duong, T., Basdevant, N. & Borgis, D. A polarizable coarse-grained water model for coarse-grained proteins simulations. *Chemical Physics Letters* **468**, 79–82 (2009).
  53. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **98**, 10089 (1993).
  54. Essmann, U. *et al.* A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **103**, 8577 (1995).
  55. Dans, P. D., Zeida, A., Machado, M. R. & Pantano, S. A Coarse Grained Model for Atomic-Detailed DNA Simulations with Explicit Electrostatics. *Journal of Chemical Theory and Computation* **6**, 1711–1725 (2010).
  56. Ayton, G. S., Noid, W. G. & Voth, G. A. Multiscale modeling of biomolecular systems: in serial and in parallel. *Curr. Opin. Struct. Biol.* **17**, 192–198 (2007).
  57. Matysiak, S., Clementi, C., Praprotnik, M., Kremer, K. & Delle Site, L. Modeling diffusive dynamics in adaptive resolution simulation of liquid water. *The Journal of Chemical Physics* **128**, 024503 (2008).
  58. Neri, M. *et al.* Microseconds dynamics simulations of the outer-membrane protease T. *Biophys. J.* **94**, 71–78 (2008).
  59. Machado, M. R., Dans, P. D. & Pantano, S. A hybrid all-atom/coarse grain model for multiscale simulations of DNA. *Physical Chemistry Chemical Physics* **13**, 18134 (2011).
  60. Nielsen, S. O., Buló, R. E., Moore, P. B. & Ensing, B. Recent progress in adaptive multiscale molecular dynamics simulations of soft matter. *Physical Chemistry Chemical Physics* **12**, 12401 (2010).
  61. Ensing, B., Nielsen, S. O., Moore, P. B., Klein, M. L. & Parrinello, M. Energy Conservation in Adaptive Hybrid

- Atomistic/Coarse-Grain Molecular Dynamics. *Journal of Chemical Theory and Computation* **3**, 1100–1105 (2007).
62. Praprotnik, M., Delle Site, L. & Kremer, K. Adaptive resolution molecular-dynamics simulation: Changing the degrees of freedom on the fly. *The Journal of Chemical Physics* **123**, 224106 (2005).
  63. Heyden, A. & Truhlar, D. G. Conservative Algorithm for an Adaptive Change of Resolution in Mixed Atomistic/Coarse-Grained Multiscale Simulations. *Journal of Chemical Theory and Computation* **4**, 217–221 (2008).
  64. Michel, J., Orsi, M. & Essex, J. W. Prediction of Partition Coefficients by Multiscale Hybrid Atomic-Level/Coarse-Grain Simulations. *The Journal of Physical Chemistry B* **112**, 657–660 (2008).
  65. Orsi, M., Noro, M. G. & Essex, J. W. Dual-resolution molecular dynamics simulation of antimicrobials in biomembranes. *J R Soc Interface* **8**, 826–841 (2011).
  66. Shi, Q., Izvekov, S. & Voth, G. A. Mixed Atomistic and Coarse-Grained Molecular Dynamics: Simulation of a Membrane-Bound Ion Channel. *The Journal of Physical Chemistry B* **110**, 15045–15048 (2006).
  67. Izvekov, S. & Voth, G. A. A Multiscale Coarse-Graining Method for Biomolecular Systems. *The Journal of Physical Chemistry B* **109**, 2469–2473 (2005).
  68. Rzeplia, A. J., Louhivuori, M., Peter, C. & Marrink, S. J. Hybrid simulations: combining atomistic and coarse-grained force fields using virtual sites. *Physical Chemistry Chemical Physics* **13**, 10437 (2011).

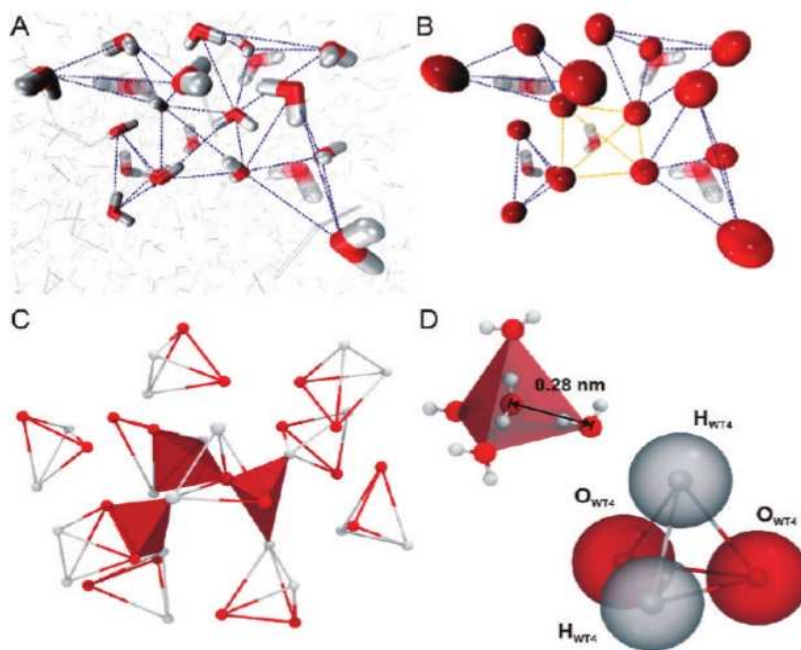
## Capítulo II

# Desarrollo de un Nuevo Modelo CG para Solvatación Acuosa

### Introducción

En el contexto del desarrollo de modelos simplificados, el tratamiento de los efectos del solvente representa un problema desafiante. Como se comentó en la Capítulo I, actualmente asistimos a un importante esfuerzo en este sentido, con varios modelos propuestos en la literatura, que se diferencian en el nivel de CG, topología molecular, tipo de funciones de energía potencial que requieren, modo del tratamiento de las interacciones de largo alcance, transferibilidad y aplicabilidad. En este marco pueden identificarse una serie de puntos que demandan esfuerzos adicionales: electrostática explícita de largo alcance, permitividad dieléctrica autogenerada sin necesidad de aplicar una constante dieléctrica homogénea, uso de potenciales estándar (Lennard-Jones y Coulomb) que aseguren la transferibilidad a los paquetes de simulación más comúnmente usados, aplicabilidad a la solvatación de mono- y polielectrolitos, elevado nivel de CG. En el presente capítulo se presenta un nuevo modelo CG de agua derivado de conceptos fisico-químicos elementales y que intenta responder a tales exigencias.

El modelo se compone de cuatro partículas CG, que denominaremos superátomos, inter-conectados de manera de lograr una molécula CG que presenta un arreglo con conformación tetraédrica (Figura 2-1). Cada superátomo porta una carga parcial explícita, con lo cual el líquido conformado por dichas moléculas es capaz de generar su propia permitividad dieléctrica, evitando el uso de una constante dieléctrica homogénea en el medio. Como se muestra más adelante, el modelo es capaz de generar una razonable representación de varias propiedades comunes del agua líquida en un rango de temperaturas relevante respecto del espectro biológico. Como ejemplos del potencial de este modelo, se estudiaron en primer lugar algunas propiedades fisico-químicas del agua y la solvatación de iones monovalentes CG también desarrollados en el marco del presente trabajo ( $\text{Na}^+$ ,  $\text{K}^+$  y  $\text{Cl}^-$ ). Posteriormente se extendió su aplicación al estudio de la hidratación del ADN utilizando un modelo CG de esta macromolécula, desarrollado recientemente en nuestro grupo.<sup>1</sup> Dicho modelo se mostró capaz de generar información estructural y dinámica de resolución pseudoatómica utilizando un esquema de solvatación implícita por medio de la aproximación de Born Generalizada.<sup>2</sup>



**Figura 2-1. Desde el agua FG al agua CG.** **A-** Foto tomada a partir de una simulación de DM de agua FG mostrando el típico ordenamiento tridimensional del agua en el "bulk". Las moléculas grises representan el líquido "bulk". La organización estructural es ilustrada a través de algunas moléculas de agua que ocupan los vértices de tetraédros regulares que saturan la capacidad de formación de enlaces de hidrógeno (líneas punteadas) del agua situada en el centro del tetraédro. **B-** La posición de cada átomo de oxígeno en los vértices de los tetraédros mostrados en A se muestran ahora con esferas rojas. El concepto detrás del modelo WAT FOUR (WT4) es que estos "clusters" tetraédricos elementales pueden representarse por cuatro superátomos unidos por constantes armónicas. Los enlaces de tipo covalentes incluidos en el modelo WT4 son representados por líneas punteadas oscuras, mientras que las interacciones inter-cluster (vdW+Coulomb) se ilustran con las líneas punteadas claras. El modelo implica que un número de moléculas de agua son consideradas de manera implícita (moléculas semi-transparentes). Cabe notar que las moléculas representadas implícitamente no se limitan a las del centro de cada "cluster", sino que también caen dentro de esa categoría las moléculas que se localizan en el centro de los tetraédros que se forman por interacción de superátomos pertenecientes a diferentes moléculas de WT4 (por ejemplo la molécula de agua en el centro del tetraédro denotado con líneas punteadas claras). Las posiciones de todos los elementos mostrados en A y B son idénticas. **C-** Foto tomada de una simulación de DM CG mostrando la organización estructural del WT4 en el "bulk". El modelo reproduce el ordenamiento tetraédrico en el espacio a través de la formación de interacciones no covalentes, a un nivel más elevado de granularidad. Los planos rojos evidencian la presencia de arreglos tetraédricos formados entre diferentes moléculas de WT4, conteniendo implícitamente una molécula de agua. **D-** La organización ideal de un "cluster" tetraédrico de agua da lugar a la geometría del WT4. La separación de 0.28nm entre el átomo de oxígeno localizado en el centro del tetraédro y sus vértices corresponde con la distancia oxígeno-oxígeno (primer esfera de solvatación). Este "cluster" elemental puede mapearse en una molécula de WT4 (abajo) compuesta de cuatro superátomos unidos. Los esferas de color rojo y blanco indican superátomos cargados negativa- (superátomo tipo-oxígeno,  $O_{WT4}$ ) y positivamente (superátomo tipo-hidrógeno,  $H_{WT4}$ ) con carga parcial de  $-/+0.41e$ , respectivamente.

En el presente trabajo se extendió su aplicabilidad para ser utilizado también en presencia de solvente explícito. La reproducción de características de hidratación del ADN, como por ejemplo, espinas de solvatación,

especificidad en la interacción con electrolitos y cambios estructurales asociados a la unión de cationes en el surco menor, ilustran la utilidad del modelo de agua como solvente acuoso capaz de reproducir condiciones fisiológicas.

## Métodos

**Descripción del modelo.** La idea que subyace al modelo es que, debido a sus características moleculares, el agua pura se comporta como un líquido estructurado capaz de formar, entre otros arreglos estructurales, "clusters" tetraédricos transitorios.<sup>3</sup>

Dichos "clusters" se componen de un agua central coordinada por otras cuatro moléculas que se disponen imitando los vértices de un tetraédro (Figura 2-1.A). En dicho arreglo, el agua central se encuentra aislada del entorno fuera del tetraédro. En este sentido, la hipótesis de trabajo es que, debido a la replicación de dicha estructura en el seno del líquido, la molécula central de cualquier tetraédro podría modelarse de manera implícita, pasando de un líquido altamente empaquetado (nivel atómico) a uno de mayor granularidad (nivel CG) (Figura 2-1.B). Con el objetivo de reproducir la organización estructural del líquido, se generó una topología molecular en la que cuatro superátomos ligados entre sí se colocan en los vértices del tetraédro que son ocupados por los átomos de oxígeno del agua en el caso atómico. Debido a que se utilizan cuatro superátomos para representar dicho arreglo el modelo se denominó WAT-FOUR (WT4). La topología propuesta implica que dentro de cualquiera de dichos tetraédros, las interacciones de enlace de hidrógeno que mantienen la cohesión del agua líquida son representadas por interacciones armónicas ("resortes") que ligan los cuatro superátomos (Figura 2-1.B). Las interacciones entre los "clusters" elementales son consideradas vía términos de Lennard-Jones y Coulomb estándar en un Hamiltoniano clásico:

$$U = \sum_{\text{bonds}} k_b (r_{ij} - r_{eq})^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{V_k}{2} [1 + \cos(n_k \varphi - \gamma_k^{eq})] + \sum_l \sum_{l>m}^N \left\{ \epsilon_{lm} \left[ \left( \frac{\sigma_{lm}}{r_{lm}} \right)^{12} - \left( \frac{\sigma_{lm}}{r_{lm}} \right)^6 \right] + \frac{q_l q_m}{4\pi\epsilon_0 \epsilon_r r_{lm}} \right\}$$

Estas fuerzas reproducen el ordenamiento tetraédrico global del agua, permitiendo que los "clusters" elementales difundan libremente. Análogamente a la generación de ordenamiento tetraédrico promovido por dichos "clusters" en su entorno, la presencia de una molécula de WT4 recrea en su entorno un ordenamiento

similar pero con mayor granularidad (Figura 2-1.C). De hecho, la estructura de una molécula de WT4 es replicada en el seno del líquido por la combinación de superátomos de las diferentes moléculas de WT4 que rodean a la central. Este tipo de ordenamiento produce intersticios inter-moleculares que pueden considerarse ocupados implícitamente por las aguas localizadas en el centro de los tetraédros observados a nivel atómico. Por lo tanto, tales aguas implícitas pueden estar presentes no solo dentro de cada molécula de WT4 sino también en las posiciones centrales de tetraédros formados por superátomos provenientes de diferentes moléculas (Figura 2-1-B,C). Esto sugiere que las moléculas de WT4 en el "bulk" tienen la capacidad de formar interacciones similares a redes de enlaces de hidrógeno.

Dado que la distancia entre el átomo de oxígeno central de un "cluster" tetraédrico (Figura 2-1.D) y cualquier otro oxígeno es de  $\sim 0.28$  nm,<sup>4</sup> y considerando la geometría de un tetraédro perfecto, la distancia de equilibrio entre superátomos se fijó en 0.45 nm.

**Tabla 2-1.** Parámetros de interacción de los modelos de agua e iones CG<sup>a</sup>

	masa (au)	carga (e)	$\sigma^b$ (nm)	$\epsilon$ (kJmol <sup>-1</sup> )	parámetros enlazantes	
					$d_{eq}$ (nm)	$k_{bond}$ (kJ mol <sup>-1</sup> nm <sup>-2</sup> ) <sup>c</sup>
SPC(27)	Ow: 16	Ow: -0.82	0.3166	0.650	0.1 <sup>d</sup>	172500
	Hw: 1	Hw: +0.41				
TIP3P(28)	Ow: 16	Ow: -0.834	0.315061	0.6364	0.09572 <sup>d</sup>	251208
	Hw: 1	Hw: +0.417				
WT4	O <sub>WT4</sub> : 50	O <sub>WT4</sub> : -0.41	0.42	0.55	0.45 <sup>e</sup>	2092
	H <sub>WT4</sub> : 50	H <sub>WT4</sub> : +0.41				
NaW <sup>+</sup>	130.99	1	0.58	0.55		
KW <sup>+</sup>	147.1	1	0.645	0.55		
ClW <sup>-</sup>	143.45	-1	0.68	0.55		

<sup>a</sup> Los parámetros de dos modelos de agua FG comúnmente usados (SPC y TIP3P) se incluyen a modo de comparación. <sup>b</sup> Distancia entre el centro del átomo y el mínimo de la función de Lennard-Jones. <sup>c</sup> Corresponde a la aproximación de la forma  $E_{bond}=K_{bond}(d-d_{eq})^2$ . <sup>d</sup> Distancia hidrógeno-oxígeno. <sup>e</sup> Distancia inter-superátomo.

Para las interacciones ligadas de a pares se utilizó una constante de fuerza en el entorno de la correspondiente a un enlace de hidrógeno. Se probaron valores en el rango desde 837-4184 KJ/mol nm<sup>2</sup> (2-10 Kcal/mol A<sup>2</sup>) eligiendo al final un valor de 2092 KJ/mol nm<sup>2</sup> (5Kcal/mol A<sup>2</sup>) por ser el que permitió un mejor ajuste a diferentes



propiedades del agua pura. Por tratarse de una constante de fuerza débil, la molécula presenta un notable nivel de plasticidad, resultando en pequeñas desviaciones de la estructura perfectamente tetraédrica frente a fluctuaciones en la temperatura y a interacciones con otras moléculas de la misma especie u otra.

A su vez, estas deformaciones podrían ser identificadas con el ordenamiento no perfecto presente en el agua a temperatura ambiente. Debido al arreglo tetraédrico, solo esos dos parámetros para las interacciones ligadas son necesarios (ver Tabla 2-1 y Figura 2-1.D)

Las interacciones no ligadas se modelaron vía potenciales de Lennard-Jones y Coulomb, los parámetros correspondientes se muestran en la Tabla 2-1. Las cargas parciales se asignaron considerando que la molécula central de agua de un tetraédro dado neutraliza parcialmente las cargas de las aguas en los vértices por la formación de enlaces de hidrógeno. En tal sentido, si las cargas atómicas del agua son  $q$  para el hidrógeno y  $-2q$  para el oxígeno, entonces los vértices del tetraédro tendrán, dos de ellos carga  $q$  (aceptores de enlace de hidrógeno) y los otros dos carga  $-q$  (dadores de enlace de hidrógeno). La asignación de cargas parciales es un problema no resuelto incluso en los modelos atomísticos. En el caso particular del agua, esta tarea se ha realizado de diferentes maneras, desde ajustar parámetros para reproducir propiedades medidas experimentalmente en fase acuosa o gas, hasta potenciales “ab initio” derivados de cálculos sobre pequeños “clusters” de moléculas. Sin embargo, ningún modelo disponible es capaz de reproducir todas las propiedades con buena precisión. En vista de esto, y dado el importante nivel de aproximación utilizado en el modelo que proponemos, solo se intentó que las interacción electrostática entre los superátomos del modelo CG fuera comparable a enlaces de hidrógeno atomístico. Por lo tanto, se utilizaron las mismas cargas parciales que presentan los modelos atomísticos comunes de tres puntos (Tabla 2-1). Entre tales modelos, la distribución de cargas que mejor permitieron reproducir observaciones experimentales fue la del modelo atomístico denominado “Single Point Charge” (SPC).<sup>5</sup> Por otro lado, el radio de van de Waals y el mínimo de energía del potencial de Lennard-Jones fueron utilizados como parámetros libres de ajuste. Las interacciones intramoleculares no enlazantes fueron excluidas.

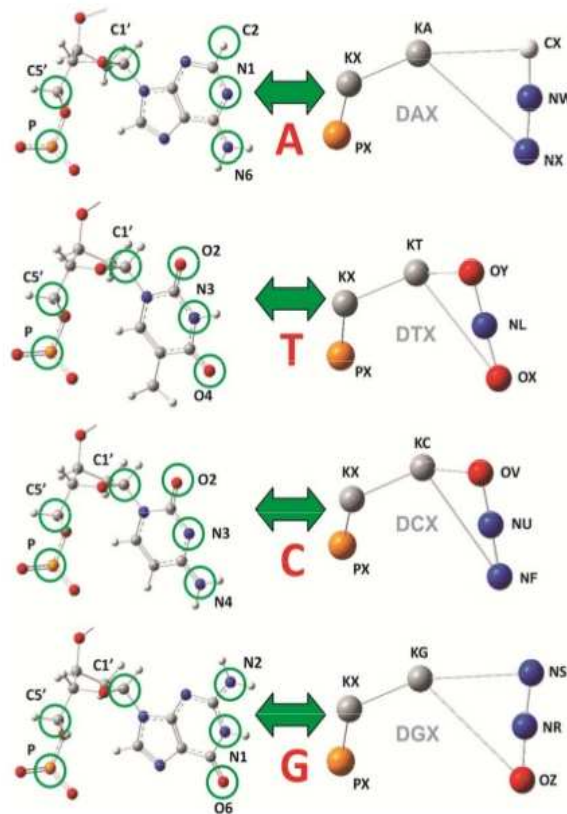
La masa de cada superátomo fue asignada ajustando la densidad de WT4 a la correspondiente a agua líquida a 300K y 1 bar. La masa por superátomo necesaria para el ajuste a 1 g/ml resultó ser 50 au. Considerando que la masa de cada molécula de agua es de 18 au, entonces cada superátomo representa  $\sim 2.8$  moléculas de agua, y por ende cada molécula de WT4 representa  $\sim 11$  aguas. Por lo tanto, cada vez que se comparen propiedades físico-químicas, se considera un factor de normalización de 11.

El factor de empaquetamiento de las esferas de WT4 calculadas como el volumen de la caja cúbica que contiene

moléculas de WT4 dividido entre el volumen excluido de los superátomos es  $\sim 0.47$ , similar al 0.5 calculado para el modelo atómico SPC. Estos valores son significativamente menores que el valor ideal de 0.74 esperado para el empaquetamiento hexagonal compacto (la máxima compactación posible para esferas rígidas). Esto sugiere que la estructura del WT4 en el seno del líquido deja un número de cavidades intersticiales levemente mayor que el modelo SPC.

**Modelo CG de iones.** Durante el presente trabajo se desarrollaron también modelos CG para tres especies iónicas, concretamente, los estados solvatados de  $\text{Na}^+$ ,  $\text{K}^+$  y  $\text{Cl}^-$ , que se denominaron  $\text{NaW}^+$ ,  $\text{KW}^+$  y  $\text{ClW}^-$ . Los iones fueron desarrollados considerando que cada ion se mueve siempre transportando su primera esfera de solvatación. Se considera que esto equivale a seis moléculas de agua ligadas a cada ion en cada uno de los tres casos.<sup>6</sup> Por lo tanto, sus masas se fijaron como la suma de la masa del ion más la de seis moléculas de agua (ver Tabla 2-1). Las cargas se fijaron en el valor unitario, positivo o negativo, correspondiente a cada ion. Para determinar los radios de van der Waals de los iones solvatados se utilizó la posición del segundo máximo de la función de distribución radial (RDF, acrónimo derivado del inglés: “radial distribution function”) de los iones hidratados reportada por medio de experimentos de difracción de neutrones.<sup>7</sup> Para la profundidad del pozo de potencial de van der Waals se usó el mismo valor que para los superátomos de WT4, ya que cuando una molécula de WT4 contacta a un ion CG, esta interacciona con su primera esfera (implícita) de solvatación. Una lista de los parámetros refinados para las interacciones no ligadas para los iones CG se detalla en la Tabla 2-1.

**Modelo CG de ADN.** El sistema CG usado para el ADN fue esencialmente el mismo que el presentado por Dans et al.<sup>1</sup> Dicho modelo representa las bases del ADN utilizando seis superátomos con un mapeo que permite mantener el “sentido químico” del reconocimiento específico de tipo Watson-Crick y la polaridad 5'-3' (ver Figura 2-2). De manera análoga al agua e iones, las interacciones moleculares son evaluadas utilizando el Hamiltoniano clásico. Además, los superátomos del ADN también contienen una carga parcial, lo cual permite la evaluación de las interacciones electrostáticas de manera explícita. Dado que el modelo original de ADN fue desarrollado utilizando una representación implícita del solvente por medio del método de Born Generalizado, su transferencia al solvente explícito utilizando WT4 y los iones desarrollados en el presente trabajo requirió la modificación de algunos parámetros para la mejorar la estabilidad de la doble hélice. Estas modificaciones resultaron altamente beneficiosas ya que permitieron también aumentar considerablemente el paso de tiempo utilizando para la integración de las ecuaciones de movimiento, pasando de 5 fs a 20 fs.



**Figura 2-2. Mapeo entre los modelos FG y CG de ADN.** Los círculos ilustran las coordenadas de los elementos de la representación FG que se conservan en el esquema CG. Se muestra los nombres de residuos y superátomos, así como la conectividad.

Tales modificaciones fueron validadas comparando los parámetros helicoidales obtenidos en la simulación con solvente explícito con la correspondiente en solvente implícito, así como con simulaciones atomísticas y datos experimentales.

**Simulaciones de Dinámica Molecular.** Todas las simulaciones de dinámica molecular fueron llevadas a cabo utilizando el programa GROMACS 4,<sup>8-10</sup> en el ensemble NPT (excepto en los casos que se mencionan más adelante). La temperatura fue acoplada al termostato Nose-Hoover,<sup>11,12</sup> mientras que para la presión se utilizó el baróstato Parrinello-Rahman<sup>13,14</sup> con tiempos de acoplamiento de 1 y 5ps, respectivamente. Un radio de corte (“cutoff”) de 1.2 nm se utilizó para las interacciones no ligadas, a su vez la electrostática de largo alcance fue evaluada por medio de la aproximación de Particle Mesh Ewald.<sup>15,16</sup> Para las simulaciones FG se usó un paso de

tiempo de 2 fs, mientras que para las simulaciones CG el paso de tiempo fue de 20 fs. El uso de pasos de integración tan grandes puede traer consigo la posibilidad de problemas de conservación de energía. Para descartar si tal problema está presente en nuestros modelos se realizaron una serie de simulaciones en el ensamble NVE utilizando un paso de integración de 20 fs y variando el cutoff de las interacciones no ligadas. Para una precisión aceptable en la integración de las ecuaciones de movimiento, las fluctuaciones en la energía total deberían ser menores a un quinto (20%) de la energía cinética o potencial del sistema.<sup>17</sup> Como se muestra más adelante dicho criterio se satisface muy bien con fluctuaciones en la energía total representando 5% de la energía potencial o cinética, utilizando un cutoff de 1.0, 1.2 y 1.5 nm. En tal sentido, se justifica el uso de un "cutoff" para las interacciones no ligadas de 1.2 nm, que además de asegurar la conservación de la energía, permite incluir las interacciones no ligadas hasta la segunda molécula vecina de WT4 en el líquido. Adicionalmente, se realizaron simulaciones en el ensamble NVT para algunos sistemas con el objetivo de calcular la tensión superficial del WT4 y la presión osmótica de los iones.

Todas las interacciones (WT4-WT4, WT4-ion, ion-ion, ión-ADN, WT4-ADN y ADN-ADN) se calcularon directamente con el Hamiltoniano implementado en GROMACS, que es común a los utilizados en los paquetes estándar de simulaciones moleculares. Las interacciones de Lennard-Jones fueron evaluadas por medio de la reglas de combinación de Lorentz-Berthelot.

Se construyeron cinco sistemas FG ( $S_{1-5}^{FG}$ ) y quince sistemas CG ( $S_{1-15}^{CG}$ ) para evaluar diversas propiedades de interés (ver Tabla 2-2). Las simulaciones FG fueron utilizadas para obtener propiedades de referencia con las cuales comparar los resultados de las simulaciones CG. Los sistemas  $S_1^{FG}$  y  $S_4^{FG}$  se usaron para calcular la densidad y el coeficiente de difusión en el rango de temperaturas entre 5-55° Celsius en tomando pasos de 5° y simulando por 5 ns cada uno (ver Tabla 2-2).

Tanto las RDF ión- $O_{\text{agua}}$  y el potencial electrostático (en la línea que conecta dos iones) fueron calculadas a partir de los sistemas  $S_2^{FG}$  y  $S_3^{FG}$ , en los que la distancia catión-anión se mantuvo fija en 3.6 nm durante la simulación. Esta propiedad también se calculó para el sistema  $S_1^{FG}$  a temperatura ambiente a modo de referencia correspondiente al agua pura. El sistema  $S_5^{FG}$  se utilizó para validar la metodología de medida de la presión osmótica que se describe más adelante.

Respecto de las simulaciones CG, las propiedades del agua "bulk" bajo condiciones ambiente fueron obtenidas de los sistemas  $S_1^{CG}$ . El comportamiento del modelo en el rango de temperaturas entre 278-328 K (5-55° C) fue estudiado usando el sistema  $S_2^{CG}$ . El escaneo de temperaturas fue llevado a cabo igual que en el caso FG ( $S_1^{FG}$  y  $S_4^{FG}$ ) pero utilizando ventanas temporales de 20 ns en lugar de 5 ns.

La tensión superficial y la compresibilidad isotérmica al nivel CG se calculó utilizando los sistemas  $S_3^{CG}$  y  $S_4^{CG}$ , respectivamente, de acuerdo con los siguientes pasos, propuestos en la literatura.<sup>18</sup> Primero, una configuración inicial a 300 K y 1 bar se somete a 0.1 ns de equilibración en el ensamble NVT, la configuración resultante se utilizó, por un lado para construir el sistema  $S_3^{CG}$  agregando secciones de vacío arriba y abajo de la caja de agua, de manera que el tamaño de la caja en la dirección del eje z se triplicó. Una simulación de 3 ns de producción en el ensamble NVT a 300 K se llevó a cabo en tal sistema, a partir del cual se pudo calcular la tensión superficial en

**Tabla 2-2.** Descripción de los sistemas simulados

sistema	modelo de agua	número de moléculas	especies iónicas (número de iones) <sup>a</sup>	soluto	temperatura (K)	tiempo de simulación (ns)	concentración de pares iónicas (M)
FG <sup>b</sup>	$S_1^{FG}$	SPC	2483 <sup>c</sup>	$Na^+(1) Cl^-(1)$	278-323	45	
FG	$S_2^{FG}$	SPC	5368 <sup>c</sup>	$K^+(1) Cl^-(1)$	300	20	0.01
FG	$S_3^{FG}$	SPC	5368 <sup>c</sup>		300	20	0.01
FG	$S_4^{FG}$	TIP3P	2483 <sup>c</sup>	$Na^+(34) Cl^-(34)$	278-323	45	0.5
FG	$S_5^{FG}$	TIP3P	7612 <sup>c</sup>		300	15	
CG <sup>d</sup>	$S_1^{CG}$	WT4	497 <sup>e</sup>		300	100	
CG	$S_2^{CG}$	WT4	497 <sup>e</sup>		278-328	200	
CG	$S_3^{CG}$	WT4	268 <sup>e</sup>		300	3	
CG	$S_4^{CG}$	WT4	268 <sup>e</sup>		300	3	
CG	$S_5^{CG}$	WT4	473 <sup>e</sup>	$NaW^+(1) ClW^-(1)$	300	100	0.01
CG	$S_6^{CG}$	WT4	473 <sup>e</sup>	$KW^+(1) ClW^-(1)$	300	100	0.01
CG	$S_7^{CG}$	WT4	456 <sup>e</sup>	$NaW^+(44) ClW^-(44)$	300	30	0.5
CG	$S_8^{CG}$	WT4	456 <sup>e</sup>	$KW^+(44) ClW^-(44)$	300	30	0.5
CG	$S_9^{CG}$	WT4	174 <sup>e</sup>	$NaW^+(7) ClW^-(7)$	300	200	0.2
CG	$S_{10}^{CG}$	WT4	174 <sup>e</sup>	$KW^+(7) ClW^-(7)$	300	200	0.2
CG	$S_{11}^{CG}$	WT4	170 <sup>e</sup>	$NaW^+(11) ClW^-(11)$	300	200	0.3
CG	$S_{12}^{CG}$	WT4	170 <sup>e</sup>	$KW^+(11) ClW^-(11)$	300	200	0.3
CG	$S_{13}^{CG}$	WT4	655 <sup>e</sup>	$NaW^+(34) ClW^-(34)$	300	100	0.5

CG	S <sub>14</sub> <sup>CG</sup>	WT4	655 <sup>e</sup>	KW <sup>+</sup> (34) ClW <sup>-</sup> (34)		300	100	0.5
CG	S <sub>15</sub> <sup>CG</sup>	WT4	506 <sup>e</sup>	NaW <sup>+</sup> (19) KW <sup>+</sup> (19) ClW <sup>-</sup> (16)	ADN CG	300	4000	0.15 <sup>f</sup>

<sup>a</sup> Parámetros tomados de Berendsen et al.<sup>19</sup> y van Gunsteren et al.<sup>20</sup> En el sistema S<sub>5</sub><sup>FG</sup>, los parámetros de CHARMM PARAM27<sup>21</sup> fueron utilizados. <sup>b</sup>

FG: del ingles "Fine Grain". <sup>c</sup> moléculas de agua FG. D CG: del ingles "Coarse Grain". <sup>e</sup> moléculas de WT4. <sup>f</sup> sin considerar los 22 contraiones para neutralizar la carga del ADN.

base a los tensores de presión:

$$\gamma = \frac{L_z}{2} \left\langle P_{zz} - \left( \frac{P_{xx} + P_{yy}}{2} \right) \right\rangle$$

Por otro lado, la configuración obtenida por equilibración NVT se utilizó también como estructura inicial (sistema S<sub>4</sub><sup>CG</sup>) para una simulación de 3 ns en ensamble NPT a 300 K y 1 bar, a partir de ella se calculo la compresibilidad isotérmica de acuerdo con la siguiente ecuación:<sup>22</sup>

$$\kappa = \frac{\langle V^2 \rangle - \langle V \rangle^2}{\langle V \rangle k_B T}$$

Las RDFs entre ion-WT4 y el perfil de potencial electrostático (obtenidos de igual manera que sus pares FG) fueron calculados para los sistemas S<sub>5</sub><sup>CG</sup> y S<sub>6</sub><sup>CG</sup> y comparados con los sistemas FG S<sub>2</sub><sup>FG</sup> y S<sub>3</sub><sup>FG</sup>, respectivamente, de manera de verificar la habilidad del modelo CG de reproducir los resultados atomísticos. Los sistemas S<sub>7</sub><sup>CG</sup> y S<sub>8</sub><sup>CG</sup> se utilizaron para calcular las RDFs ion-WT4 a una concentración aproximada de 0.5M, de modo de poder compararla con los datos experimentales.<sup>7</sup>

Las longitudes de Bjerrum ( $\lambda_B$ ) y Debye ( $\kappa^{-1}$ ) fueron calculadas como:

$$\lambda_B(\rho) = \frac{(1.0 \times 10^9) \beta e^2}{4\pi \epsilon_0 \epsilon_r(\rho)}$$

$$\kappa^{-1}(\rho) = \left( \frac{2(1.0 \times 10^{-15}) F^2 \rho}{RT \epsilon_0 \epsilon_r(\rho)} \right)^{-\frac{1}{2}}$$

donde  $\beta$  es la energía térmica,  $\epsilon_0 = 8.85 \times 10^{-12} \text{ C}^2\text{J}^{-1}\text{m}^{-1}$ ,  $F=96485.3399 \text{ C mol}^{-1}$  y  $R = 8.314472 \text{ J mol}^{-1}\text{K}^{-1}$ .

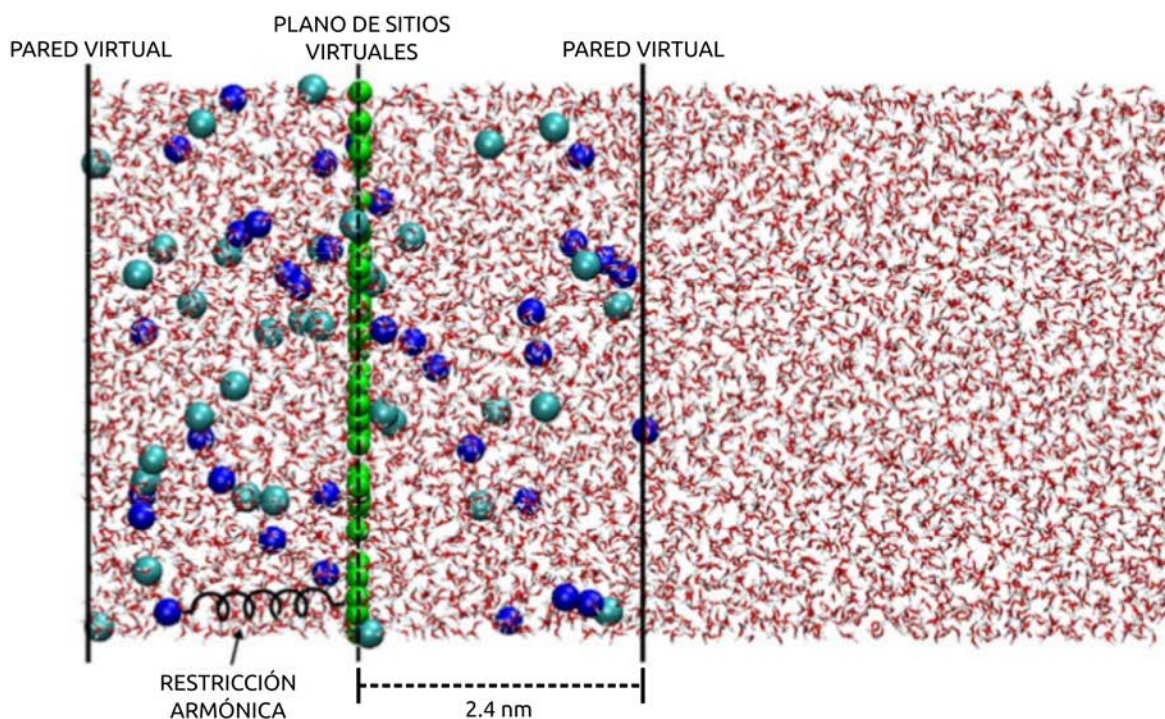
Las constantes dieléctricas de las soluciones de electrolitos ( $\text{Na}^+\text{Cl}^-$  y  $\text{K}^+\text{Cl}^-$ ) a diferentes concentraciones (0.2M y 0.3M) fueron obtenidas de las simulaciones de los sistemas  $S_{9-12}^{\text{CG}}$  (ver Tabla 2-2).

La presión osmótica se midió en base a la metodología presentada por Roux y Luo.<sup>23</sup> La idea que subyace a este método es simular una solución acuosa donde los iones se mantienen restringidos a moverse solo en la mitad de la caja de simulación. A partir de la fuerza ejercida por vínculos armónicos, que mide la tendencia de los iones a abandonar el centro de la caja de simulación, se puede calcular la presión osmótica. Para llevar a cabo tal simulación se usó la estrategia denominada BRIM.<sup>24</sup> Esta estrategia se basa en colocar un conjunto de sitios virtuales tales que su movimiento se fija en un plano (pero permitiendo que se muevan libremente dentro del mismo) y se aplican restricciones de distancia entre tales sitios virtuales y los iones. Para el cálculo de la presión osmótica, el plano de sitios virtuales se eligió tal que sea ortogonal al eje z en una posición fija correspondiente a un cuarto de la caja de simulación en tal dirección. Los iones son libres de moverse en las direcciones x e y, mientras que restricciones armónicas actúan en la dirección z solo luego de un determinado "cutoff". De esta manera, se imita la presencia de dos paredes virtuales que confinan el movimiento de los iones a la mitad del volumen de la caja de simulación. Para este propósito se prepararon los sistemas  $S_{13,14}^{\text{CG}}$  (Tabla 2-2) de la siguiente manera. Una caja de simulación rectangular se llenó de moléculas de WT4 e iones CG ( $\text{NaW}^+\text{ClW}^-$ :  $S_{13}^{\text{CG}}$  y  $\text{KW}^+\text{ClW}^-$ :  $S_{14}^{\text{CG}}$ ) fueron agregados reemplazando moléculas de WT4 solo en la mitad de la caja, alcanzando una concentración molar de 0.5M. Para cada ion, un sitio virtual se colocó en el plano de sitios virtuales y se establecieron restricciones armónicas entre ellos (ver Figura 2-3). Tales sistemas se sometieron a minimizaciones de energía seguidas de 500 ps de equilibración en condiciones ambiente, manteniendo el área x-y fija mientras que la dirección z se acopló a un baróstato para fijar la presión del sistema en 1 bar. Finalmente, una equilibración de 1.0 ns en el ensemble canónico fue seguida de diez simulaciones de producción de 10 ns cada una (variando las velocidades iniciales). La presión osmótica, para cada simulación, se calculó utilizando la siguiente ecuación:

$$\pi = \frac{K \frac{1}{N} \sum_N \sum_i ((z_{\text{ion}}^{(i)} - z_{\text{virtual}}^{(i)}) - d_0)}{2 L_x L_y}$$

donde K es la constante de fuerza usada en las restricciones de distancia armónica entre los sitios virtuales y los

iones ( $10 \text{ Kcal mol}^{-1} \text{ \AA}^{-2}$ ),  $N$  es el número total de pasos de simulación.,  $z_{\text{ion}}^{(i)}$  y  $z_{\text{virtual}}^{(i)}$  son los componentes de las posiciones del ion y el sitio virtual sobre el eje  $z$  (considerados únicamente cuando su diferencia supera  $d_0$ ),  $d_0$  es la distancia de equilibrio de la restricción de distancia (2.4 nm) y  $L_x L_y$  es el área de la caja en el plano  $x$ - $y$ . La expresión es dividida entre 2 para promediar sobre ambas paredes virtuales. Los valores reportados son el promedio de las diez trayectorias independiente obtenidas. Esta misma aproximación fue empleada sobre el sistema  $S_5^{\text{FG}}$  con el objetivo de validar la metodología comparando con datos experimentales y resultados de simulación FG reportados en la literatura.<sup>23</sup> En este caso agua FG (TIP3P) e iones FG (CHARMM PARAM27) fueron utilizados (Tabla 2-2).



**Figura 2-3. Cálculo de la presión osmótica.** Ilustración del tipo de sistema empleado en el cálculo de la presión osmótica. El movimiento de los iones (azul y cian) está restringido a la mitad de la caja de simulación por medio de restricciones armónicas de distancia entre ellos y un conjunto de sitios virtuales (verde), un sitio por ion, cuyo movimiento está restringido, a su vez, al plano  $x$ - $y$  en la posición equivalente a un cuarto de la caja en la dirección  $z$  (PLANO DE SITIOS VIRTUALES). De esta forma, se crean dos PAREDES VIRTUALES en la caja de simulación, que confinan el movimiento de los iones al espacio ente ellas, pero permitiendo que las moléculas de agua se muevan por toda la caja.

Por último se realizó una simulación de 4 microsegundos ( $S_{15}^{\text{CG}}$ ) de una doble hélice de ADN utilizando la



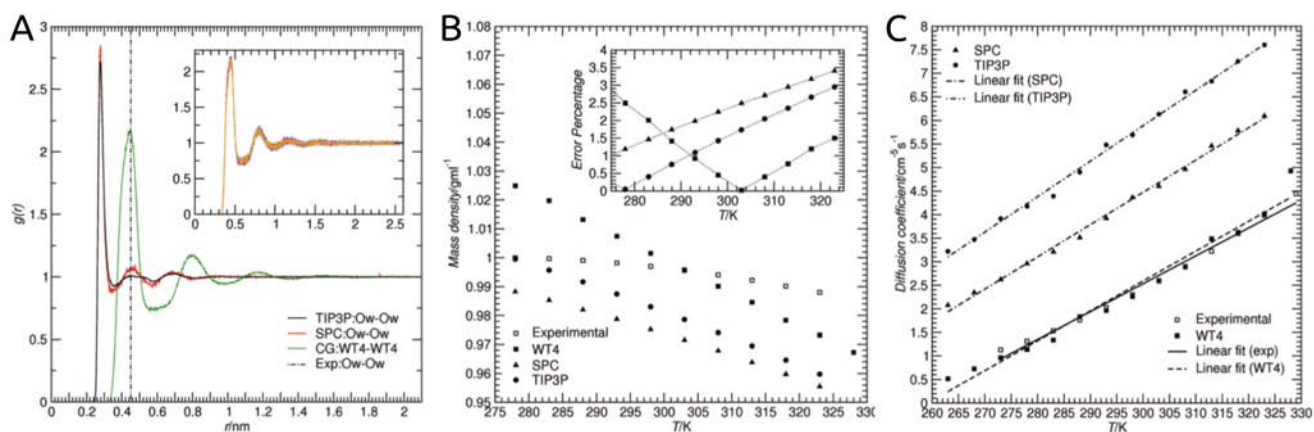
secuencia de Drew-Dickerson 5'-d(CGCGAATTCGCG)-3' en una caja de simulación de tipo octaedro truncado y solvatados por moléculas de WT4 e iones a concentración fisiológica (ver Tabla 2-2). En base a tal simulación se evaluó el comportamiento estructural global del ADN, su hidratación e interacciones específicas ADN-agua y ADN-iones. Los parámetros helicoidales del ADN fueron calculados usando el programa Curves.<sup>25</sup> La disminución del ancho del surco menor, que en adelante denominaremos con la palabra en inglés “narrowing” fue analizada. Para ello, se estimó la distancia promedio interfosfato entre hebras opuestas en base a los siguientes pares: {(5 24), (6 23), (7 22), (8 21), (9 20), (10 19), (11 18), (12 17)} (en negrita se indican los residuos que conforman el segmento AATT central). Los cationes se consideraron unidos al surco menor si su distancia a los grupos fosfato de ambas hebras opuestas es menor a 0.5 nm simultáneamente.

## Resultados

En los párrafos siguientes se describe el comportamiento del modelo CG de agua WT4 y su capacidad de reproducir varias propiedades comunes del agua pura en estado líquido. Las comparaciones se realizaron, en lo posible, con datos experimentales. Sin embargo algunas de las propiedades calculadas se compararon con resultados obtenidos a partir de modelos de agua FG populares como forma de proveer un marco de referencia. Posteriormente se analizó la estructura de solvatación en soluciones electrolíticas. Por último, como ejemplo de aplicación biológica, se presenta una simulación de un segmento de ADN CG embebido en agua e iones tratados de manera explícita con los modelos desarrollados en el presente trabajo.

**WT4 puro.** Una característica importante del agua es su orden intrínseco. Una buena reproducción de la RDF entre los átomos de oxígeno ( $RDF_{O-O}$ ) es un objetivo común para la mayoría de los modelos atomísticos. La forma de la RDF más allá de las primeras esferas de solvatación provee de una idea de las características de la sustancia bajo estudio. Mientras que para una sustancia líquida la RDF debe converger a la unidad a distancia suficientemente alejada de las primeras esferas de solvatación, la observación de un comportamiento repetitivo es indicativo de un estado cristalino.

A pesar que la RDF obtenida a partir del modelo WT4 retiene varias características del agua líquida, la comparación con modelos FG ( $S_1^{FG}$  y  $S_4^{FG}$ ), muestra algunas diferencias. La más evidente es la ausencia del primer pico de solvatación. Debido al tamaño y la topología de los superátomos en el WT4, este excluye un espacio en la



**Figura 2-4. Propiedades del modelo WT4 en el "bulk".** **A-** RDF (del inglés: Radial Distribution Function) calculada sobre todos los superátomos de WT4 a temperatura ambiente para el sistema  $S_1^{CG}$  y  $S_4^{FG}$  a 298 K (líneas negra y roja, respectivamente). La posición del segundo pico de solvatación obtenido experimentalmente<sup>26</sup> es identificada con la línea vertical punteada. El recuadro interior de A muestra el comportamiento de la RDF al variar la temperatura en el rango entre 278-328 K. Dicho perfil no muestra cambios significativos. **B-** La variación de la densidad de masa del agua CG en función de la temperatura (cuadrados llenos) calculada para el sistema  $S_2^{CG}$  se compara con datos experimentales (cuadrados vacíos)<sup>27</sup> y con simulaciones de agua FG SPC (triángulos) y TIP3P (círculos) correspondientes con los sistemas  $S_1^{FG}$  y  $S_4^{FG}$ , respectivamente. El recuadro interior muestra el error relativo de los modelos WT4, SPC y TIP3P respecto a los valores experimentales. **C-** Dependencia del coeficiente de difusión con la temperatura, comparando los modelos WT4, SPC y TIP3P ( $S_2^{CG}$  (cuadrados llenos),  $S_1^{FG}$  (triángulos),  $S_4^{FG}$  (círculos), respectivamente) con datos experimentales<sup>28</sup> (cuadrados vacíos). Los cuatro perfiles presentan un comportamiento lineal como se evidencia a partir de los ajustes por regresión lineal.

vecindad de cada superátomo correspondiente a la segunda esfera de solvatación del agua real. En este sentido, el modelo WT4 puede considerarse como un modelo de segunda esfera de solvatación. De hecho, la posición del primer máximo en WT4 corresponde al segundo pico en el agua atómica<sup>26</sup> (ver Figura 2-4). Es importante notar que la normalización a la densidad del "bulk" y el carácter más granular del modelo CG generan una diferencia en las alturas relativas de las distribución del probabilidad del WT4 respecto del agua real. Más aun, los enlaces armónicos que mantienen la estructura tetraédrica, inducen una sobre-estimación de la probabilidad de encontrar al primer vecino en la solución de WT4. Después del primer pico en la RDF, el tamaño relativamente grande del WT4 genera un ordenamiento residual que se extiende hasta ~1.2 nm, alcanzando la convergencia a la densidad del "bulk" luego de 1.3 nm.

Una propiedad importante de los modelos de agua líquida es su capacidad de reproducir correctamente la difusión del agua. Claramente, la velocidad de difusión de la molécula de WT4 es mucho menor que la del agua

FG. A 300 K, se obtiene un valor de  $2.03 \times 10^{-6} \text{ cm}^2 \text{ s}^{-1}$ . Sin embargo, el desplazamiento de una molécula de WT4 representa implícitamente el movimiento del centro de masa de  $\sim 11$  moléculas de agua. Por lo tanto, considerando que el desplazamiento cuadrático medio promedio del centro de masa de  $N$  moléculas es  $N$  veces más lento que el desplazamiento cuadrático medio promedio de  $N$  moléculas difundiendo separadamente,<sup>29,30</sup> se puede concluir que el coeficiente de difusión para las moléculas de agua representadas en el modelo CG a temperatura ambiente es de  $2.23 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$ , que coincide muy bien con el valor obtenido experimentalmente (Tabla 2-3).

El modelo WT4 incluye el tratamiento explícito de las interacciones electrostáticas dado que cada superátomo transporta una carga parcial (Tabla 2-1). Esto da lugar a la generación por parte del solvente de una permitividad dieléctrica propia, sin necesidad de imponer un medio dieléctrico artificial. La permitividad dieléctrica simulada por el WT4 es de 110. Aunque se trata de una sobre-estimación en 30% del valor real, es importante notar que el modelado de la permitividad dieléctrica es un punto problemático incluso para modelos atomísticos más sofisticados, con valores que van desde 53<sup>31</sup> hasta 116.<sup>32</sup>

Un problema importante se relaciona con el ordenamiento a largo alcance inherente a las moléculas de WT4. De hecho, algunos modelos CG de agua presentan un punto de congelamiento muy cercano a la temperatura ambiente (41). Por lo tanto, se llevaron a cabo simulaciones en el rango entre 278 a 328 K. Tal rango de temperaturas cubre la mayoría de las aplicaciones de relevancia biológica que se le podría dar al modelo. El cálculo de la RDF en dicho rango de temperaturas sugiere que el WT4 retiene su carácter líquido, al no encontrarse cambios significativos entre 278 y 328 K (Figura 2-4.A).

La densidad del modelo de WT4 fue ajustada para reproducir el valor correspondiente al agua pura a 300 K. Sin embargo, una buena reproducción de las variaciones de la densidad respecto de la temperatura también sería deseable. Del punto de vista cualitativo, se observa la esperada reducción de la densidad con la temperatura, con un comportamiento lineal. (Figura 2-4.B) A pesar que la dependencia funcional del agua real respecto de la temperatura es, obviamente, no lineal, el resultado obtenido con el modelo CG es una buena aproximación dentro de dicho rango. De hecho, el error relativo de la densidad del WT4 respecto a aquella del agua real se mantiene por debajo del 3%, con la mayor desviación cerca del punto crítico del agua real. (Figura 2-4.B). Este comportamiento es comparable con los de los modelos atomísticos SPC y TIP3P (Figura 2-4.B).

La medida de las variaciones del volumen frente a las variaciones térmicas a presión constante permite calcular el coeficiente de expansión isobárica para el WT4. Se obtuvo una sobreestimación de esta cantidad a 298 K (ver Tabla 2-3). El coeficiente de expansión para el WT4 es de  $11.6 \times 10^{-4} \text{ K}^{-1}$  en comparación con el valor obtenido

**Tabla 2-3.** Propiedades del agua en el "bulk" en condiciones ambiente para los modelos FG: SPC, y TIP3P y para el modelo CG: WT4, comparado con datos experimentales.

	constante dieléctrica	coeficiente de difusión ( $10^{-5} \text{ cm}^2\text{s}^{-1}$ )	coeficiente de expansión ( $10^{-4} \text{ K}^{-1}$ )	densidad de masa ( $\text{gmL}^{-1}$ )	densidad numérica <sup>a</sup> ( $10^{22} \text{ mL}^{-1}$ )	tensión superficial ( $\text{mNm}^{-1}$ )	compresibilidad isotérmica ( $\text{GPa}^{-1}$ )
WT4	110	2.23	11.6	1.0001	0.3	17	2.43
SPC	65 <sup>33</sup>	3.85 <sup>34</sup>	7.3 <sup>35</sup>	0.9705 <sup>36</sup>	3.2	53.4 <sup>37</sup>	0.53 <sup>38</sup>
TIP3P	82 <sup>31</sup>	5.19 <sup>34</sup>	9.2 <sup>39</sup>	1.002 <sup>39</sup>	3.4	49.5 <sup>37</sup>	0.58 <sup>38</sup>
Exp.	78.4 <sup>40</sup>	2.27 <sup>41</sup>	2.53 <sup>27</sup>	0.9970 <sup>27</sup>	3.3	71.2 <sup>42</sup>	0.46 <sup>43</sup>

<sup>a</sup> Calculada a partir de la densidad de masa correspondiente, considerando la masa molar del agua real ( $18 \text{ gmol}^{-1}$ ) y la del WT4 ( $200 \text{ gmol}^{-1}$ ). Consecuentemente, la densidad numérica para los modelos FG y el valor experimental corresponden con el número de moléculas de agua por mililitro, mientras que para el modelo WT4, corresponde con el número de moléculas de WT4 (~11 moléculas de agua) por mililitro.

experimentalmente que es de  $2.53 \times 10^{-4} \text{ K}^{-1}$ .<sup>27</sup> Aunque el WT4 claramente sobreestima dicha propiedad respecto el agua real, el valor obtenido es comparable con los valores reportados para modelos atomísticos comunes (Tabla 2-3).

Otra propiedad frecuentemente calculada para los modelos CG es la tensión superficial. En el caso del WT4, se obtiene un valor de  $17 \text{ mN m}^{-1}$ , que es prácticamente 4 veces más chico que el valor observado experimentalmente. A su vez, encontramos una compresibilidad isotérmica 5 veces mayor que la experimental (Tabla 2-3). Estas discrepancias son muy comúnmente encontradas en los modelos CG que condensan varias aguas en una sola entidad molecular.<sup>18</sup> El origen de este efecto podría ser la pérdida del detalle atómico en las interacciones, que disminuye las fuerzas de cohesión y aumentan la granularidad del sistema.

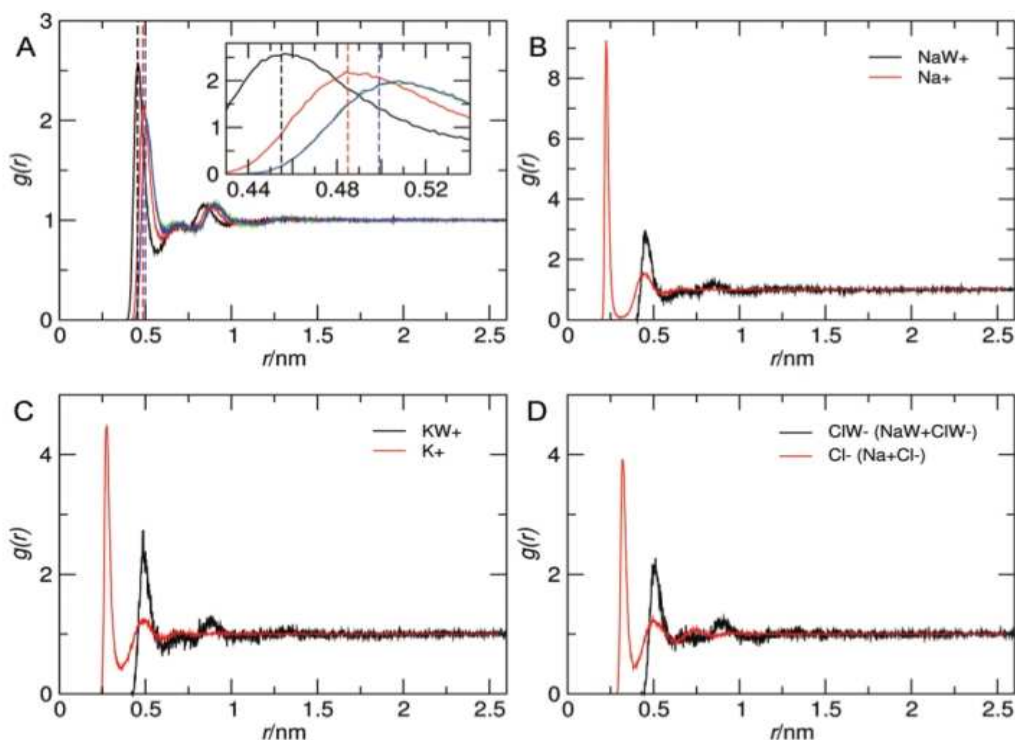
Una prueba aun más exigente para el modelo proviene del cálculo del coeficiente de difusión. Claramente, un aumento de la difusión debe ocurrir al aumentar la temperatura del sistema. Los datos experimentales indican que el agua pura experimenta un aumento casi lineal de la difusión en el rango de temperaturas entre 278 y 328 K. El modelo muestra una correcta dependencia del coeficiente de difusión con la temperatura. De hecho, se observa una muy buena correspondencia con los datos experimentales en todo el rango de temperaturas exploradas. (Figura 2-4.C)

Considerando los resultados mencionados en los párrafos anteriores, el rango de validez del modelo podría delimitarse de acuerdo con las siguientes consideraciones: el límite inferior se encuentra a 278 K. Su uso a temperaturas inferiores se desaconseja fuertemente dado que el cambio de fase para formar agua en estado sólido implica efectos cuánticos que, claramente, no son considerados por los modelos CG. En el límite superior,

el error relativo en el coeficiente de difusión llega a ~11% a 328 K, lo cual sugiere que las simulaciones a temperaturas más altas podrían requerir la reparametrización del modelo para mantener su precisión en valores aceptables.

**Solvatación Iónica.** Las características del modelo WT4 permiten la posibilidad de estudiar la solvatación de sistemas en los que la electrostática es un componente dominante. En este contexto, se desarrollaron modelos CG para los tres electrolitos simples más comunes en biología:  $\text{Na}^+$ ,  $\text{K}^+$  y  $\text{Cl}^-$ . Dado que WT4 puede considerarse como un modelo de solvatación de segunda esfera, los iones se modelaron conteniendo implícitamente su primera esfera de solvatación. Con el objetivo de estudiar la estructura de solvatación que genera WT4 alrededor de los iones, se llevaron a cabo simulaciones a concentraciones iónicas comparables con experimentos de difracción de neutrones reportados en la literatura.<sup>7</sup> Como se muestra en la Figura 2-5.A, existe una buena correspondencia, especialmente en el caso de los cationes, entre el primer máximo de solvatación encontrado para el WT4 y la segunda esfera de hidratación observada experimentalmente.<sup>7</sup> Adicionalmente, un segundo pico localizado a ~0.9 nm se genera de manera espuria como consecuencia de la geometría de la molécula de WT4. Concretamente, los superátomos que generan este segundo pico son los que están ligados armónicamente a aquellos que dan lugar al primer máximo. Luego de ese punto, la RDF converge a la densidad del "bulk" en todos los casos.

Desafortunadamente, los datos experimentales de solvatación iónica solo están disponible para altas concentraciones de electrolito. Para explorar concentraciones menores (más cercanas a las relevantes fisiológicamente), se compararon los resultados de los modelos CG ( $S_5^{\text{CG}}$  y  $S_6^{\text{CG}}$ ) con resultados de simulaciones atomísticas equivalentes ( $S_2^{\text{FG}}$  y  $S_3^{\text{FG}}$ , respectivamente), a una concentración iónica de 0.01 M. (Figuras 1-5.B,C,D). Análogamente al caso del WT4 puro, la RDF de éste respecto de los iones CG muestra la completa ausencia del primer máximo de solvatación, lo cual es consistente con la inclusión implícita de dicho máximo en la parametrización de los iones CG. Se observa una buena reproducción de la posición de segundo máximo de solvatación. Esto es consistente con comportamiento del WT4 como solvente de segunda esfera. Como es de esperar, WT4 no es capaz de reproducir la tercera esfera. Sin embargo, la relevancia de tal imprecisión es incierta



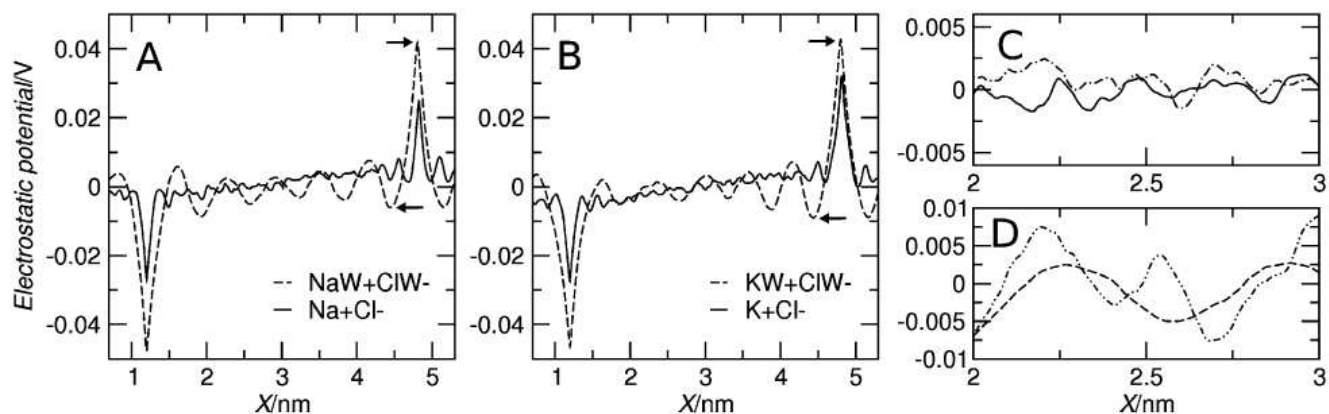
**Figura 2-5. Solvatación iónica.** A- RDF de los superátomos de WT4 alrededor de los electrolitos CG calculados para los sistemas  $S_7^{CG}$  y  $S_8^{CG}$  ( $NaW^+$ : negro,  $KW^+$ : rojo,  $ClW^-$ : azul ( $NaW^+ClW^-$ ) o verde ( $KW^+ClW^-$ )). Las líneas punteadas verticales indican las posiciones de la segunda esfera de solvatación determinadas experimentalmente por “scattering” de neutrones.<sup>7</sup> El recuadro interior muestra con mayor detalle la región entre 0.43 y 0.54 nm, permitiendo una comparación más precisa. B, C y D- Comparación entre las RDFs obtenidas mediante simulaciones FG y CG para los iones  $Na^+$ ,  $K^+$  y  $Cl^-$  (sistemas  $S_2^{FG}$ ,  $S_3^{FG}$ ,  $S_5^{CG}$  y  $S_6^{CG}$ ). El perfil correspondiente a la estructura de solvatación alrededor de los iones cloro en presencia de potasio es equivalente al mostrado para el caso en presencia de sodio, y por lo tanto se omitió.

y podría ser solo relevante en el caso de los iones cloruro, donde tal esfera está ligeramente más pronunciada de acuerdo con la información experimental.<sup>7</sup>

**Potencial Electrostático.** Una vez analizada la estructura de hidratación de electrolitos simples, se prestó atención a los perfiles de potencial electrostático y a las propiedades de apantallamiento. Esto se realizó por comparación de los resultados de las simulaciones de los sistemas  $S_2^{FG}$  y  $S_3^{FG}$  frente a los de los sistemas  $S_5^{CG}$  y  $S_6^{CG}$ , respectivamente. Tales sistemas consisten en un par iónico de  $Na^+Cl^-$  (o  $K^+Cl^-$ ) mantenido en una posición fija durante la simulación. La separación entre ambos iones fue de 3.6 nm. Los pares iónicos FG o CG se situaron en cajas de simulación conteniendo un número equivalente de aguas FG y CG, respectivamente. Estas construcciones permitieron comparar las simulaciones FG y CG bajo iguales condiciones, así como el comportamiento de las

diferentes especies iónicas. Para asignar el peso correcto a las perturbaciones del potencial generada por los iones, también se midieron los perfiles de potencial electrostático en agua pura, tanto FG como CG. De esta manera es posible separar las observaciones en dos componentes: fluctuaciones de potencial intrínsecas al solvente y perturbaciones iónicas. Más aún, esta aproximación permite tener una idea sobre la relajación del potencial a medida que nos alejamos del ion.

La primera observación que puede realizarse a partir de la comparación de perfiles entre las simulaciones FG y CG (Figura 2-6.A) es la diferencia de altura de los picos centrados en las posiciones de los iones. Debido a su menor tamaño, las aguas SPC (FG) pueden acercarse más a los iones FG generando un apantallamiento electrostático más pronunciado. En la contra parte CG, la correspondiente primera esfera de solvatación, que es tratada de manera implícita en los iones, solo sirve para generar un espacio excluido, sin propiedades de apantallamiento. Esto se traduce en un potencial electrostático mayor inducido por los iones CG. La consideración implícita de la primera esfera en los iones CG implica que el primer mínimo observado en los sistemas atomísticos debería estar ausente en el sistema CG (Figura 2-6.A). Adicionalmente, la posición del primer mínimo observado en la simulación CG (segunda esfera de solvatación) se corresponde relativamente bien con la posición del segundo mínimo alrededor del ion en el sistema FG.



**Figura 2-6. Perfiles de potencial electrostático.** A- Potencial electrostático calculado a lo largo de la línea que conecta el par iónico  $\text{Na}^+\text{Cl}^-$  (línea continua,  $S_2^{\text{FG}}$ ) o  $\text{NaW}^+\text{ClW}^-$  (línea punteada,  $S_5^{\text{CG}}$ ). Las flechas indican los puntos donde las diferencias de potencial electrostático fueron calculadas. B- Equivalente a A pero para los pares iónicos  $\text{K}^+\text{Cl}^-$  ( $S_3^{\text{FG}}$ ) y  $\text{KW}^+\text{ClW}^-$  ( $S_6^{\text{CG}}$ ). C- Comparación del potencial electrostático entre la región central del panel A (línea continua,  $S_2^{\text{FG}}$ ) con la medida equivalente para una caja conteniendo agua FG SPC pura (línea-punto-línea,  $S_1^{\text{FG}}$ ). D- Equivalente a C pero para el agua CG (línea punteada,  $S_5^{\text{CG}}$ , y línea-doble punto-línea,  $S_1^{\text{CG}}$ ).

Claramente, este efecto deriva de la estructura de solvatación alrededor de los electrolitos, es decir, el primer y

segundo mínimo alrededor de la posición de los iones (tanto  $\text{Na}^+$  como  $\text{Cl}^-$ ) mostrada en la Figura 2-6.A corresponde a la posición de los oxígenos en la primer y segunda esfera de solvatación mostrados en las Figuras 1-5.B,D. Características similares se verifican para los pares iónicos  $\text{K}^+\text{Cl}^-$  y  $\text{KW}^+\text{ClW}^-$  (Figura 2-6.B). Las características distintivas de ambos cationes que se puede evidenciar por la organización del solvente alrededor de los iones  $\text{NaW}^+$  y  $\text{ClW}^-$  (Figura 2-5.A). Puede también obtenerse del cálculo de la diferencia de potencial electrostático medida en la posición de catión respecto del primer mínimo (Figura 2-6.A,B). Esta diferencia es  $\sim 10\%$  más grande para el caso de  $\text{KW}^+$  respecto de  $\text{NaW}^+$ , lo cual está en acuerdo cualitativo con la diferencia de  $25\%$  observada en el caso FG. Este comportamiento podría ser el reflejo de que el agua alrededor del potasio se une de manera más desordenada que alrededor del sodio,<sup>6</sup> probablemente generando un apantallamiento electrostático menos marcado en el caso del potasio.

Como se puede observar de la Figura 2-6.A,B el esquema CG presenta mayores fluctuaciones en el perfil de potencial que el caso atomístico. Con el objetivo de excluir la posibilidad de ordenamiento espurio de moléculas de WT4 alrededor de los electrolitos, se compararon las perturbaciones en el potencial electrostático introducido por los iones respecto de las observadas en el solvente puro (tanto FG como CG). Esto se realizó calculando el potencial a lo largo de un eje arbitrario en dos cajas de simulación conteniendo SPC o WT4 puros (sistemas  $S_1^{\text{FG}}$  y  $S_1^{\text{CG}}$ ). La superposición de ambos perfiles (Figuras 1-6.C,D) sugiere que ambos sistemas de agua pura presentan importantes fluctuaciones en el potencial electrostático de aproximadamente la misma magnitud que las observadas en la región entre medio de los iones en las soluciones electrolíticas. Esto estaría indicando que las perturbaciones observadas en tal región no son un efecto inducido por la presencia de los iones sino que corresponde a variaciones en el potencial electrostático que son inherentes al líquido puro. De acuerdo con esto, la diferencia en la amplitud de las fluctuaciones observadas entre los modelos FG y CG (Figura 2-6.A,B) se explica por el aumento de la granularidad intrínseca de modelo CG. Una estimación de dicha diferencia se puede obtener de las amplitudes aproximadas que se observan en el sistema FG ( $\sim 0.002$  V) y CG ( $\sim 0.018$  V). Esto indica que las oscilaciones en el sistema CG tienen amplitudes aproximadamente un orden de magnitud mayor que las que ocurren en el sistema FG.

**Propiedades de Solución Electrolítica en el "Bulk".** La gran mayoría de las parametrizaciones empíricas para electrolitos simples son típicamente desarrolladas para ajustar propiedades de iones individuales, como las examinadas en las secciones anteriores. Para complementar la descripción estructural de las soluciones acuosas CG se estudiaron algunas propiedades termodinámicas relacionadas con las interacciones ion-ion,



concretamente las distancias de Bjerrum y de Debye. La primera representa la separación entre dos elementos de carga a la cual la interacción electrostática es comparable con la magnitud de la energía térmica. La segunda provee información respecto de la distancia a la que el potencial electrostático de un ion es apantallado por la fuerza iónica del medio que lo rodea. Desde el punto de vista cualitativo, las tendencias de dichas longitudes frente a cambios en la concentración iónica son correctas (Tabla 2-4). El cálculo de las longitudes de Bjerrum y Debye a 0.2 y 0.3 M muestra valores con un error máximo de 13% respecto a los valores experimentales (Tabla 2-4). Se obtuvo una longitud de Bjerrum subestimada, y consecuentemente, una sobre-estimación de la longitud de Debye, lo cual es indicativo de un apantallamiento global levemente mayor en la solución "bulk", independientemente de la sal considerada en la simulación ( $\text{Na}^+\text{Cl}^-$  o  $\text{K}^+\text{Cl}^-$ ). Esto es consistente con la sobre-estimación de la permitividad dieléctrica de nuestro modelo respecto al valor experimental.

Es de gran relevancia obtener una medida directa de la fuerza de la interacción efectiva entre iones mediada por el solvente. Esta puede ser obtenida midiendo la presión osmótica. En el caso de  $\text{Na}^+\text{Cl}^-$  a una concentración de 0.5 M, se midió un valor de presión osmótica de 35 bar (33 bar para  $\text{K}^+\text{Cl}^-$ ), que es esencialmente idéntica a la obtenida vía simulaciones atomísticas utilizando el campo de fuerzas CHARMM. A pesar de los valores elevados de desvío estándar, estos valores están de acuerdo con los reportados experimentalmente (Tabla 2-4), lo que sugiere un balance satisfactorio de las interacciones ion-ion e ion-WT4.

**Tabla 2-4.** Propiedades termodinámicas de las soluciones de electrolitos simples

	longitud de Bjerrum (nm)		longitud de Debye (nm)		presión osmótica <sup>a</sup> (bar)
$\rho$ (molaridad)	0.2 M	0.3 M	0.2 M	0.3 M	0.5 M
$\text{Na}^+\text{Cl}^-/\text{WT4}$	0.57	0.61	0.76	0.6	35 (s.d.: 15)
$\text{K}^+\text{Cl}^-/\text{WT4}$	0.55	0.61	0.78	0.6	33 (s.d.: 16)
Exp. <sup>b</sup>					
$\text{Na}^+\text{Cl}^-$	0.75	0.77	0.66	0.53	~25 (tomado de ref. 23)
$\text{K}^+\text{Cl}^-$	0.74	0.76	0.67	0.54	

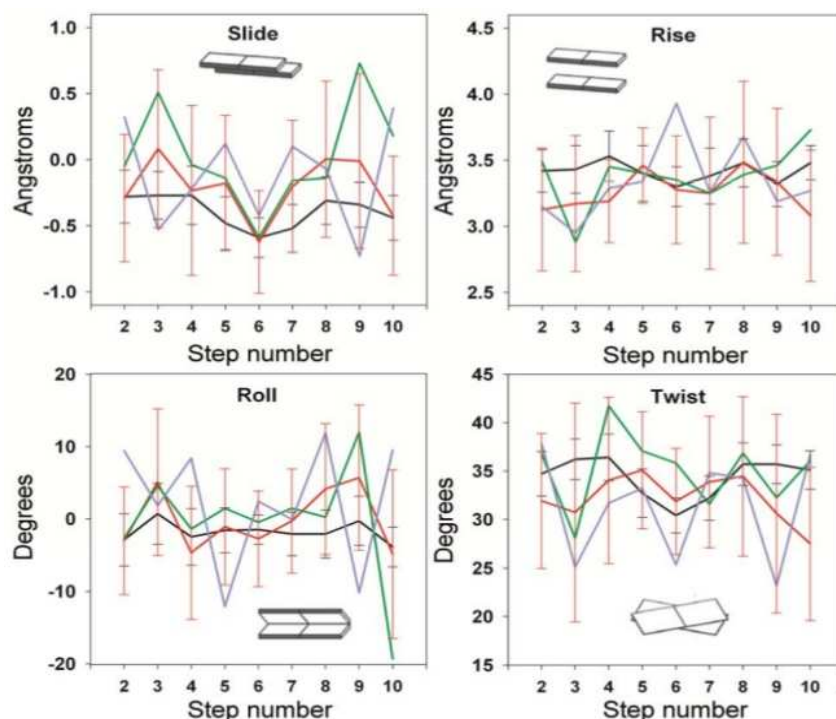
<sup>a</sup> El valor obtenido en las simulaciones FG usando el campo de fuerza CHARMM PARAM27 fue 37 (s.d.: 9) bar. <sup>b</sup> La función  $\epsilon_r(\rho) = \epsilon(0)/(1+A\rho)$  ( $\text{Na}^+\text{Cl}^-$ ,  $A=0.27$ ;  $\text{K}^+\text{Cl}^-$ ,  $A=0.24$ ), que resulta del ajuste a valores de constante dieléctrica obtenidos experimentalmente,<sup>44</sup> se utilizó para estimar  $\epsilon_r(\rho)$  a la concentración deseada, que es requisito para el cálculo de las longitudes de Bjerrum y Debye.

**Solvatación CG de una Doble Hélice de ADN.** Como ejemplo final de aplicación, se analizó la solvatación explícita de un segmento dodecamérico de ADN doble hebra (secuencia de Drew-Dickerson: 5'-CGCGAATTCGCG-3'). Para esto, se utilizó un modelo CG de ADN previamente desarrollado utilizando como método de solvatación

la aproximación implícita Born Generalizado y que fuera probado, entre otros sistemas, en dicho dodecámero.<sup>1</sup> De esta manera, se pudo usar como referencia el comportamiento observado en solvente implícito. Concretamente, ambas aproximaciones proveen resultados estructurales y dinámicos similares respecto del comportamiento de la doble hélice con un RMSD máximo entre las estructuras promedio de ambas aproximaciones de 0.25 nm. Esto está de acuerdo con la correcta reproducción de los parámetros helicoidales obtenidos por “backmapping” (reconstrucción del detalle atómico a partir de las posiciones CG) de la simulación en solvente explícito CG (Figura 2-7). Adicionalmente, la superposición de las matrices de covarianza calculadas a lo largo de las trayectorias de dinámica molecular del dodecámero de Drew-Dickerson tratando el solvente de manera implícita y explícita da una identidad del 84%. Esto sugiere fuertemente que ambas aproximaciones exploran espacios conformacionales equivalentes.

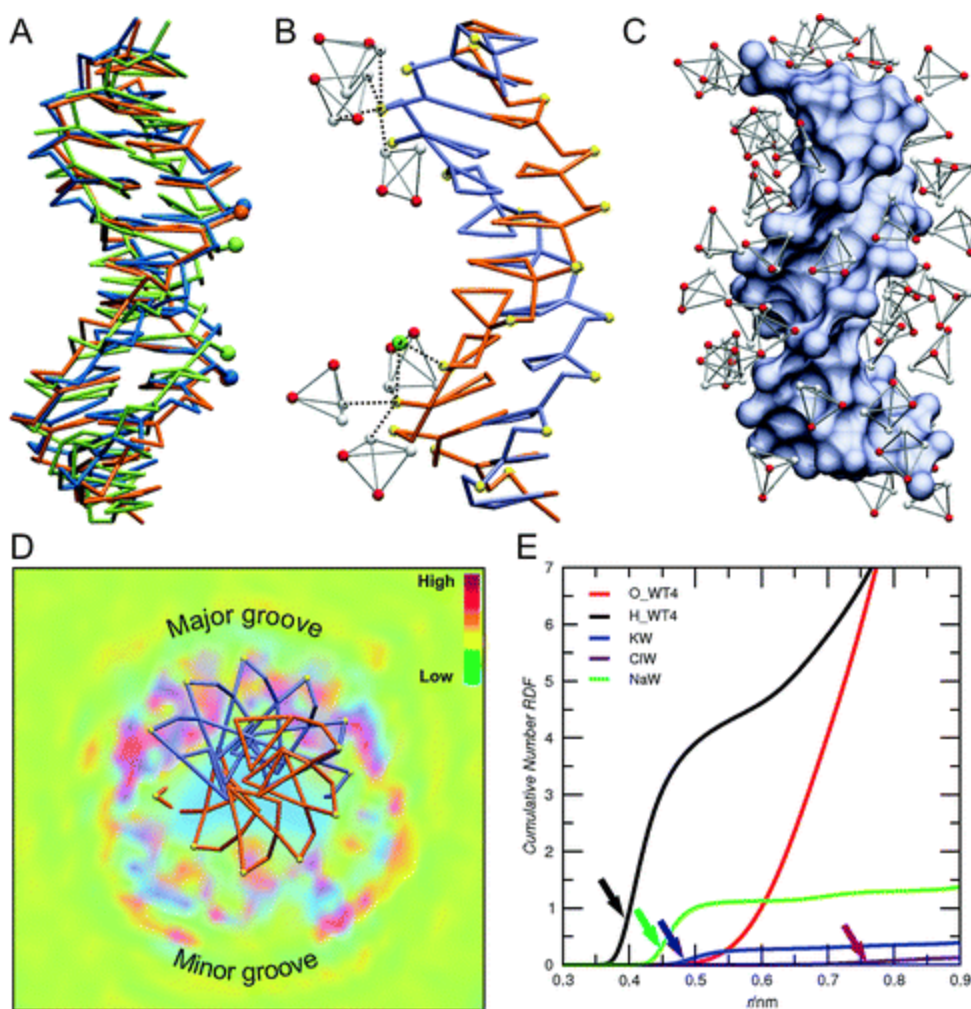
Durante la dinámica en presencia de WT4, la estructura global del dodecámero de ADN se conserva bien con un RMSD promedio de 0.25 nm respecto de la conformación canónica inicial. Esto se puede ilustrar superponiendo conformaciones tomadas a diferentes tiempos durante la simulación (Figura 2-8.A). Más aun, se observa buena coincidencia también a nivel atómico luego del proceso de “backmapping”. El RMSD de las tres conformaciones mostradas en la Figura 2-8.A (inicio, mitad y final de la simulación.) luego del proceso de “backmapping”, respecto de la estructura cristalográfica 1BNA muestra valores de 0.35 nm, 0.39 nm y 0.34 nm, respectivamente. Las moléculas de WT4 y los cationes son capaces de acercarse suficientemente al ADN e interactuar de manera específica con las nucleobases del ADN. Puede observarse que el ordenamiento de las moléculas de WT4 alrededor del ADN coincide cualitativamente con las características de hidratación observadas en los sistemas atomísticos tanto a nivel teórico como experimental.<sup>45-50</sup> Arreglos cónicos de superátomos de WT4 se forman alrededor de los grupos fosfato (Figura 2-9.A,B). Las moléculas de WT4 adquieren una orientación guiada por la atracción electrostática entre los superátomos positivos (tipo-hidrógenos) y los superátomos de fosfato cargados negativamente. Esto resulta en la formación de estructuras de tipo conos de hidratación (Figura 2-8.B (arriba) y Figura 2-9A,B). En este tipo de arreglos del solvente alrededor del esqueleto del ADN, las moléculas de WT4 pueden ser reemplazadas por cationes provenientes de la soluciones (Figura 2-8.B (abajo)), al igual que se observa experimentalmente.<sup>51</sup> Además, los iones pueden permanecer unidos transitoriamente al ADN visitando diferentes posiciones dentro del surco menor. La extensa hidratación del surco mayor y la formación de espigas de hidratación en el surco menor también son evidentes, como se desprende de la Figura 2-8.C. Una visión completa de la estructura de hidratación puede obtenerse a partir de un mapa de densidad de ocupación de los superátomos de WT4 proyectado en el plano ortogonal al eje del ADN, situado en el “step” AT central (Figura 2-

8.D).



**Figura 2-7. Parámetros helicoidales.** Valores promedio y desvíos estándar de parámetros helicoidales seleccionados, obtenidos por medio del proceso de “backmapping”<sup>1</sup> a partir de las simulación del dodecámero de Drew-Dickerson a nivel CG, utilizando el modelo de solvatación explícita introducido en la presente Tesis. Como referencia se muestran los valores obtenidos por simulaciones FG (rojo), cristalografía de rayos X (verde, PDBID: 1BNA) y RMN (azul, PDBID: 2DAU). Los parámetros helicoidales del ADN se calcularon utilizando el programa Curves.<sup>25</sup>

Una visión más cuantitativa de la interacción DNA-solvente (agua e iones) se desprende de las integrales de las RDFs de las diferentes especies presentes en el medio, alrededor de los superátomos de fósforo (Figura 2-8.E). La direccionalidad en la interacción WT4 -fósforo se evidencia en el corrimiento hacia la derecha observado en la integral de la RDF correspondiente a los superátomos de tipo oxígeno (cargados negativamente) respecto a la de los superátomos de tipo hidrógeno (cargado positivamente) (ver líneas roja y negra en la Figura 2-8.E). La posición del primer pico de solvatación que coincide con la formación de los arreglos cónicos arriba mencionados, ocurre a 0.4 nm de los superátomos fósforo. Esta distancia se corresponde correctamente con el valor de 0.38 nm observado en simulaciones FG<sup>1</sup> y es comparable con la distancia mínima de 0.32 nm observada en estructuras de rayos X.<sup>52</sup>

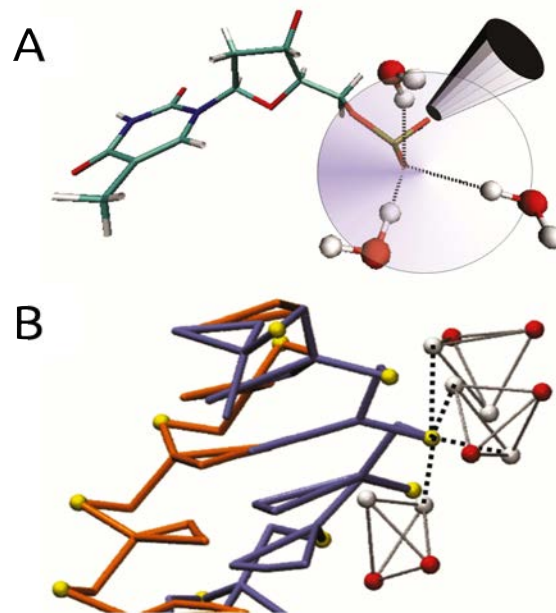


**Figura 2-8. Estructura y solvatación del ADN.** **A-** Superposición de conformeros de ADN tomados al inicio (azul), mitad (verde) y final (naranja) de la simulación. del sistema  $S_{15}^{CG}$ . Las esferas indican una pareja de superátomos fosfato de hebras opuestas, que se destacan para mostrar el “narrowing” en el surco menor. **B-** Foto representativa de las interacciones específicas del WT4 y  $NaW^+$  con los grupos fosfato, tomada de la simulación de DM CG. Las líneas punteadas ilustran los arreglos cónicos de los superátomos del WT4 alrededor de los fosfatos (arriba) y la competición por acceder a estos entre el WT4 y los iones  $NaW^+$  (abajo). **C-** Foto tomada aleatoriamente de la simulación. CG mostrando la solvatación de los surcos mayor y menor por parte del modelo WT4. La hidratación extensiva del surco mayor así como las espigas de hidratación dentro del surco menor son evidentes. **D-** Mapa de densidad 3D de ocupación de moléculas de WT4 proyectado en el plano ortogonal al eje del ADN, y localizado en el step AT central. La escala de colores representa el nivel de ocupación (cyan: baja ocupación, purpura: alta ocupación). Las diferencias entre surco mayor y menor se observan claramente. Además, cabe notar la localización más puntual del WT4 dentro del surco menor, indicativa de la presencia de espigas de hidratación. **E-** número acumulativo (integral de la RDF) de superátomos negativos y positivos de WT4 (rojo y negro, respectivamente),  $NaW^+$  (verde),  $KW^+$  (azul) y  $ClW^-$  (violeta), tomando como referencia la posición de los grupos fosfato. Las flechas indican los puntos de inflexión, correspondientes con los primeros máximos de la RDF de cada partícula evaluada.

El modelo también es capaz de recuperar la especificidad en las interacciones catión-ADN. Como es de esperar, los iones sodio presentan más tendencia que los iones potasio a interactuar con el soluto. Como se mencionó previamente, el sodio se encuentra frecuentemente en las cercanías de los grupos fosfato e incluso se localizan dentro del surco menor.<sup>51,53,54</sup> La esfera de sodio más cercana se localiza a 0.45 nm de los superátomos de fosfatos, en comparación con los 0.5 nm observados en el caso del ion potasio. Por el contrario, la RDF del ion cloruro se presenta mucho más corrida hacia valores mayores de distancia, con el primer pico de solvatación localizado a 0.76 nm de los fosfatos (Figura 2-8.E).

La fracción de carga del ADN neutralizada dentro de un cilindro de 0.9 nm medido desde la superficie de la doble hélice es 0.75. Dicho valor está de acuerdo con la fracción de contraiones condensados que se calculan dentro del volumen de condensación, utilizando la teoría de condensación de contraiones de Manning para polielectrolitos.<sup>55</sup> Adicionalmente, este número es comparable con la fracción de 0.76 obtenida en simulaciones atomísticas utilizando la misma secuencia de ADN.<sup>56</sup> Dentro de la fracción de contraiones condensados, 76% corresponde a sodio mientras que el restante 24% a potasio, lo cual está en acuerdo cualitativo con una serie de experimentos y estudios teóricos (Savelyev y Papoian<sup>57</sup> y referencias allí citadas).

Mientras que la distribución global de los cationes alrededor del ADN contribuye a la estabilidad de la doble hélice, se ha propuesto que la interacción específica de cationes con el ADN puede generar distorsiones estructurales locales. En particular, se ha propuesto que la unión de iones sodio dentro del surco menor media el fenómeno de “narrowing” de dicho surco.<sup>58</sup> De acuerdo con tal propuesta, se observó una clara correlación entre el ancho del surco menor y la unión de cationes al mismo. Más aún, parecería haber un efecto acumulativo entre ambos eventos, es decir, un mayor número de iones unidos induce un “narrowing” más pronunciado y persistente en el tiempo. Esto es evidente si se mide el ancho promedio del surco menor respecto al número total de iones unidos (Figura 2-10.A). La unión de un solo ion es suficiente para inducir un cambio sensible en el surco menor. A medida que se incorporan más iones, el “narrowing” es más pronunciado, alcanzando un máximo cuando seis iones se encuentran simultáneamente unidos. Estudios experimentales en el mismo dodecámero también revelan una alta ocupación de cationes en el surco menor, induciendo el “narrowing” del mismo.<sup>53</sup> Adicionalmente, una estructura altamente ordenada se forma cuando cationes y agua interactúan con la región central AT del ADN. Tal estructura se organiza en cuatro capas de sitios de solvatación y el motivo que se forma se asemeja a una serie de hexágonos fundidos.<sup>53</sup>



**Figura 2-9. Atmósfera acuosa alrededor del ADN.** A- Representación esquemática de los conos de hidratación que se forman alrededor de los grupos fosfato a nivel FG. B- Equivalente a A pero para la simulación. CG.

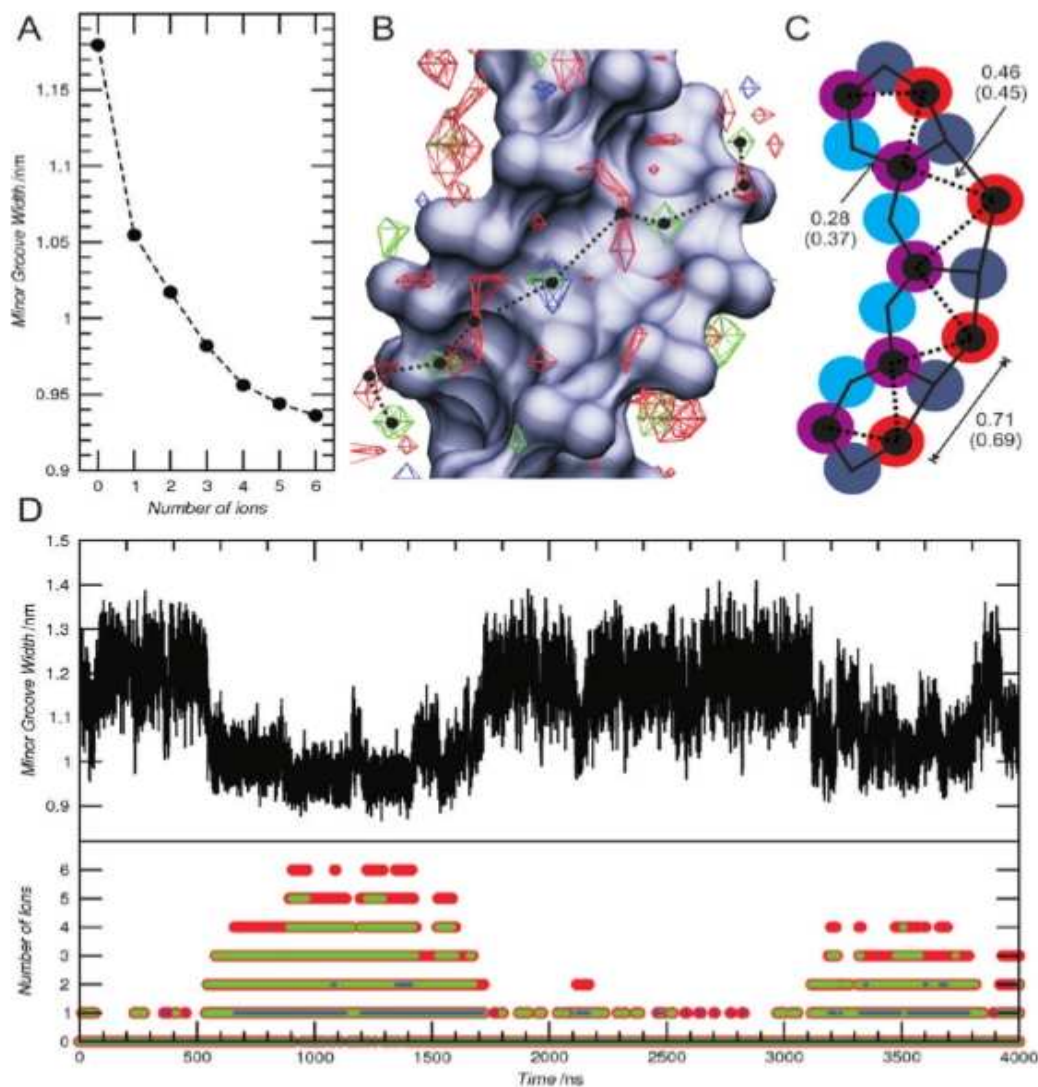
La Figura 2-10.B muestra un mapa de ocupación tridimensional de los superátomos de WT4 y los cationes alrededor del surco menor del track AT. Este mapa revela sitios altamente ocupados por WT4 (entramado rojo),  $\text{Na}^+$  (entramado verde) y  $\text{K}^+$  (entramado azul), que se asemeja a una estructura en zigzag (camino punteado que conecta los puntos negros en la Figura 2-10.C). Cuando tal estructura en zigzag se superpone sobre la estructura de hexágonos fundidos observada experimentalmente, se verifica una buena concordancia entre la segunda y cuarta capa de sitios de solvatación. Esto es evidente si se comparan las distancias entre dichos sitios de solvatación (Figura 2-10.C).

Dos procesos de “narrowing” del surco menor parecen tener lugar en dos escalas temporales diferentes. La primera se relaciona con la unión de uno o dos iones por tiempos que alcanzan las pocas decenas de nanosegundos, mientras que la segunda corresponde con uniones simultáneas de tres hasta seis iones por periodos de tiempo en el orden del microsegundo (Figura 2-10.D). Esta última induce una distorsión estructural más marcada y persistente, aunque reversible, con un ancho de surco menor promedio que va desde 0.98 nm (tres iones unidos) a 0.94 nm (seis iones unidos). La magnitud de esta distorsión del ADN coincide muy bien con el promedio de 0.96 nm obtenido experimentalmente.<sup>54</sup>

Cabe notar que las escalas temporales de unión de sodio que corresponde con las uniones rápidas mencionada en

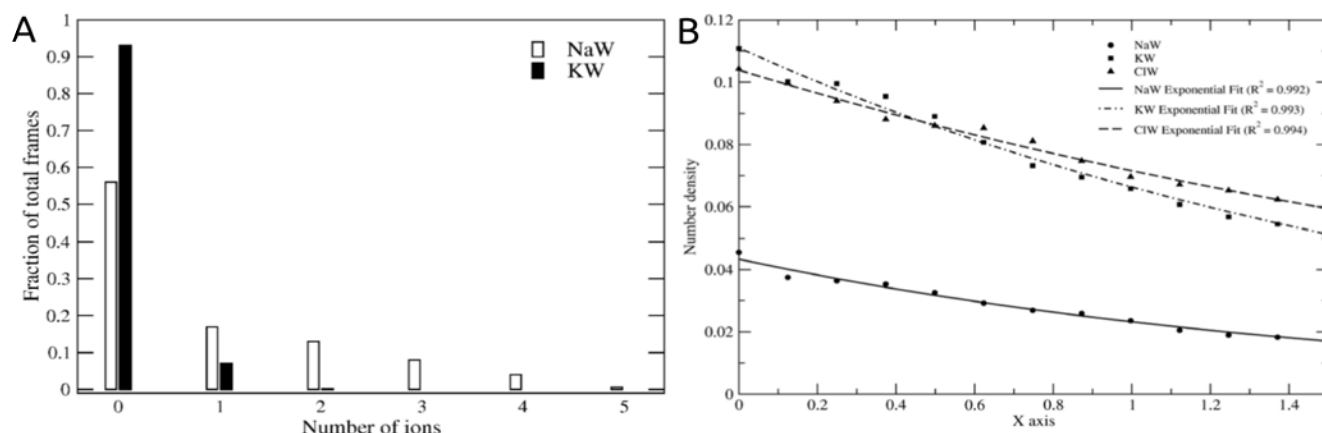
el párrafo anterior (decenas de nanosegundos), habían sido reportadas en base a simulaciones FG.<sup>58-60</sup> Desafortunadamente, la simulación FG más larga reportada a la fecha para este sistema tiene una duración de 1.2 microsegundos.<sup>58</sup> Aunque solo se reportaron eventos de unión del orden de los nanosegundos en base a dicha simulación, el acuerdo con la posición del sitio de unión y con la distorsión del ADN reportadas vía cristalografía de rayos X,<sup>53,54</sup> permiten especular que la carencia de eventos de unión más duraderos en las simulaciones atomísticas podría ser consecuencia de un muestreo temporal insuficiente.

Existe una marcada selectividad de sodio frente a potasio. De hecho, mientras que la unión simultánea de más de dos sodios es muy frecuente, solo dos iones potasio se pudieron observar simultáneamente en el surco menor, y tal evento se observó solo cinco veces durante los cuatro microsegundos de trayectoria (Figura 2-10.D y Figura 2-11.A). Finalmente, para completar la descripción de la estructura iónica alrededor del ADN, se analizó la distribución iónica a distancias mayores respecto de la doble hélice. Esto se realizó calculando la densidad del número de iones (de los tres presentes en la simulación) al aumentar la distancia desde el ADN. El número de electrolitos a lo largo de la dirección perpendicular al eje principal del ADN sigue un decaimiento exponencial, en buen acuerdo con las predicciones de la teoría de Poisson-Boltzmann<sup>61</sup> (ver Figura 2-11.B).



**Figura 2-10. Estructura y solvatación del ADN.** **A-** El ancho del surco menor promediado en las fracciones de simulación con igual número de iones ligados se grafica en función de dicho número de iones en el surco menor (de acuerdo con el criterio explicado en la sección Métodos). **B-** Isosuperficies de densidad de ocupación de WT4 (rojo), NaW<sup>+</sup> (verde), y KW<sup>+</sup> (azul) localizados en el surco menor del track AT. El camino con línea punteada conectando los puntos negros indican el motivo en “zigzag” formado por los cationes y el WT4 en el surco menor. **C-** Esquema mostrando la superposición del motivo en zigzag (círculos negros conectados por la línea punteada) observado en la simulación. CG, sobre el motivo de hexágonos fundidos (línea continua) formado por los sitios de ocupación del solvente (círculos cyan, violeta, gris y naranja) observados experimentalmente. Estos sitios pueden ser ocupados tanto por moléculas de agua como por cationes.<sup>53</sup> Las distancias entre sitios de solvatación correspondientes en el motivo de hexágonos fundidos se comparan con las correspondientes a las del motivo en zigzag (entre paréntesis). **D-** Ancho del surco menor (arriba) y número de cationes unidos al mismo (abajo) graficados en función del tiempo. El número de cationes se muestra como el número de NaW<sup>+</sup> (verde), número de KW<sup>+</sup> (azul) y el número total de cationes (suma del número de NaW<sup>+</sup> y de KW<sup>+</sup>, rojo).





**Figura 2-11. Comportamiento de los cationes alrededor del ADN.** **A-** El histograma muestra la fracción de tiempo en la que se observa un determinado número de iones  $\text{NaW}^+$  (blanco) o  $\text{KW}^+$  (negro) localizados en el surco menor, respecto del tiempo total de simulación. Altos niveles de ocupación son observados más frecuentemente para  $\text{NaW}^+$  que para  $\text{KW}^+$ . **B-** Decaimiento de la densidad de iones respecto del ADN. El sistema  $S_{15}^{\text{CG}}$  fue centrado en la caja de simulación alineando el eje del ADN en la dirección del eje z. La densidad numérica de los tres tipos de iones ( $\text{NaW}^+$ : círculos,  $\text{KW}^+$ : cuadrados y  $\text{ClW}^-$ : triángulos) fue medida a lo largo de un eje perpendicular a la dirección z (luego que la densidad del ADN decayó completamente). Adicionalmente se muestran los ajustes a funciones exponenciales para cada ion ( $\text{NaW}^+$ : línea continua,  $\text{KW}^+$ : línea-punto y  $\text{ClW}^-$ : línea punteada). Los coeficientes de correlación revelan un muy buen ajuste de tales funciones al decaimiento de la densidad de iones alrededor del ADN.

## Discusión y Conclusiones

En el presente capítulo se presentó un modelo para simular agua a nivel CG. El modelo, denominado WT4, se basa en la estructura tetraédrica transitoria adoptada por las moléculas de agua en solución, preservando las características moleculares del líquido atomístico. Debido al alto número y heterogeneidad de modelos CG propuestos en la literatura, es difícil establecer una comparación justa en términos de aceleración computacional. Sin embargo, la comparación se vuelve más directa si nos restringimos a los modelos que condensan tres o cuatro moléculas de agua por superátomo.<sup>29,30,62</sup> Tales modelos presentan un factor de simplificación de 9-12, comparado con un valor de  $\sim 8$  obtenido para WT4. Adicionalmente a un factor de simplificación similar, el modelo WT4, ofrece algunas ventajas como son; la capacidad de interactuar por medio de interacciones electrostáticas explícitas tanto en el corto como en el largo alcance, y la generación de la permitividad dieléctrica propia del solvente sin necesidad de aplicar un dieléctrico uniforme. Esto asegura la

habilidad del modelo de reproducir algunas de las propiedades características del agua y las soluciones electrolíticas.

Las masas de los superátomos se asignaron para ajustar a la densidad del agua a 300 K, mostrando además un error relativo a la densidad experimental que se mantiene por debajo del 3% para el rango de temperaturas entre 278-328 K (Figura 2-4.B).

Una hipótesis fuerte del modelo es que su estructura tetraédrica es fija mientras que su vida media en el agua real está en el orden de los picosegundos. Este defecto se compensa parcialmente por la debilidad de los enlaces armónicos que ligan los superátomos en la molécula de WT4. Esto permite variaciones en la distancia de equilibrio de los enlaces de ~10%, lo que le confiere una importante plasticidad a la molécula de WT4, y la posibilidad de adaptar su conformación de acuerdo con su entorno molecular. El uso del modelo WT4 para solvatar iones monovalentes reproduce su estructura de hidratación y algunas propiedades termodinámicas como la presión osmótica, la cual es comúnmente considerada un indicador de la calidad de la parametrización.

Es necesario notar que propiedades importantes como la compresibilidad isotérmica y la tensión superficial son pobremente descritas por el modelo WT4. Esto podría tener particular importancia en el estudio de fenómenos de auto ensamblado. A pesar de dicha desviación respecto del comportamiento ideal, la descripción de la doble hélice de ADN no parece estar comprometida. Esto sugiere que las propiedades de apantallamiento de mediano y largo alcance del modelo de solvatación son suficientes para compensar la fuerte repulsión electrostática generada por las cargas negativas sobre los fosfatos. De hecho, la adición de solvatación explícita y de diferentes especies iónicas mejora notablemente la descripción de la dinámica del ADN. Esto permite, por ejemplo, reproducir el “narrowing” del surco menor en respuesta a la unión de cationes, que no podría ser estudiado mediando una aproximación implícita. Mientras que la solvatación implícita puede proveer de una descripción buena y rápida de los efectos de secuencia sobre la estructura y estabilidad dinámica, la inclusión del solvente explícito permite el estudio de, por ejemplo, la influencia de fuentes intrínsecas “versus” extrínsecas responsables de la flexibilidad del ADN, los efectos mediados por el solvente, la especificidad iónica, etc. Adicionalmente, el uso de condiciones periódicas de contorno y electrostática explícita permite un tratamiento más realista de los efectos de largo alcance.

El modelo WT4, junto con los modelos CG de electrolitos, representa correctamente la estructura de solvatación alrededor del ADN, como se desprende del porcentaje de carga de ADN neutralizada a 0.9 nm, en coincidencia con los valores observados en simulaciones atomísticas y las predicciones de la teoría de condensación de contraiones. Más aún, las características de solvatación del ADN como la extensa hidratación del surco mayor, la

presencia de espigas de hidratación, y los arreglos cónicos alrededor de los fosfatos, que se asemejan a los conos de hidratación observados en las simulaciones atomísticas y en datos experimentales, son correctamente reproducidos. Es importante notar, en este contexto, que el desarrollo de los parámetros de interacción fue realizado bajo la filosofía de ajustar a propiedades estructurales del agua, de soluciones electrolíticas y del ADN. En este sentido, primero se desarrolló la representación para WT4 en el "bulk" y luego se agregó la descripción de los electrolitos CG. Finalmente los parámetros existentes para ADN en solvente implícito fueron levemente modificados para refinar más aún la descripción estructural en el entorno de un solvente explícito. En este sentido, la concordancia con las determinaciones experimentales pueden considerarse propiedades emergentes del modelo debido a que no se requirió ningún ajuste específico de interacciones ADN-solvente.

El esquema de simulación propuesto permite simular a una velocidad de ~1 microsegundo por día ( $S_{15}^{CG}$ ) en un procesador dual quad-core de 2.66GHz. Este desempeño, junto con la resolución prácticamente atomística alcanzable por medio del proceso de "backmapping" de las coordenadas del modelo de ADN, plantea la posibilidad real de alcanzar la escala del milisegundo utilizando simulaciones de dinámica molecular. Esto permitiría incrementar más aun la capacidad de esta técnica para responder preguntas de interés biológico con resolución atómica.

Finalmente, cabe notar que el modelo computa todas las interacciones en base al Hamiltoniano estándar, evitando modificaciones "ad hoc" y por tanto facilitando su uso.

## Bibliografía

1. Dans, P. D., Zeida, A., Machado, M. R. & Pantano, S. A Coarse Grained Model for Atomic-Detailed DNA Simulations with Explicit Electrostatics. *Journal of Chemical Theory and Computation* **6**, 1711–1725 (2010).
2. Hawkins, G. D., Cramer, C. J. & Truhlar, D. G. Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium. *The Journal of Physical Chemistry* **100**, 19824–19839 (1996).
3. Head-Gordon, T. & Hura, G. Water structure from scattering experiments and simulation. *Chem. Rev.* **102**, 2651–2670 (2002).
4. Narten, A. H., Danford, M. D. & Levy, H. A. X-ray diffraction study of liquid water in the temperature range 4–200 °C. *Discussions of the Faraday Society* **43**, 97 (1967).
5. Berendsen, H., Postma, J., van Gunsteren, W. & Hermans, J. Interaction models for water in relation to protein hydration. *Intermolecular Forces* 331–342 (1981).
6. Mancinelli, R., Botti, A., Bruni, F., Ricci, M. A. & Soper, A. K. Hydration of Sodium, Potassium, and Chloride Ions in Solution and the Concept of Structure Maker/Breaker. *The Journal of Physical Chemistry B* **111**, 13570–13577 (2007).
7. Mancinelli, R., Botti, A., Bruni, F., Ricci, M. A. & Soper, A. K. Perturbation of water structure due to monovalent ions in solution. *Phys Chem Chem Phys* **9**, 2959–2967 (2007).

8. Berendsen, H. J. C., van der Spoel, D. & van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications* **91**, 43–56 (1995).
9. Lindahl, E., Hess, B. & van der Spoel, D. Gromacs 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Model.* **7**, 306–317 (2001).
10. Van Der Spoel, D. *et al.* GROMACS: fast, flexible, and free. *J Comput Chem* **26**, 1701–1718 (2005).
11. Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics* **52**, 255–268 (1984).
12. Hoover, W. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A* **31**, 1695–1697 (1985).
13. Parrinello, M. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **52**, 7182 (1981).
14. Nosé, S. & Klein, M. L. Constant pressure molecular dynamics for molecular systems. *Molecular Physics* **50**, 1055–1076 (1983).
15. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **98**, 10089 (1993).
16. Essmann, U. *et al.* A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **103**, 8577 (1995).
17. Winger, M., Trzesniak, D., Baron, R. & van Gunsteren, W. F. On using a too large integration time step in molecular dynamics simulations of coarse-grained molecular models. *Physical Chemistry Chemical Physics* (2009).doi:10.1039/b818713d
18. He, X., Shinoda, W., DeVane, R. & Klein, M. L. Exploring the utility of coarse-grained water models for computational studies of interfacial systems. *Molecular Physics* **108**, 2007–2020 (2010).
19. van Buuren, A. R., Marrink, S. J. & Berendsen, H. J. C. A molecular dynamics study of the decane/water interface. *The Journal of Physical Chemistry* **97**, 9206–9212 (1993).
20. Mark, A. E., van Helden, S. P., Smith, P. E., Janssen, L. H. M. & van Gunsteren, W. F. Convergence Properties of Free Energy Calculations:  $\alpha$ -Cyclodextrin Complexes as a Case Study. *Journal of the American Chemical Society* **116**, 6293–6302 (1994).
21. Beglov, D. & Roux, B. Finite representation of an infinite bulk system: Solvent boundary potential for computer simulations. *The Journal of Chemical Physics* **100**, 9050 (1994).
22. Herrero, C. P. Compressibility of solid helium. *Journal of Physics: Condensed Matter* **20**, 295230 (2008).
23. Luo, Y. & Roux, B. Simulation of Osmotic Pressure in Concentrated Aqueous Salt Solutions. *The Journal of Physical Chemistry Letters* **1**, 183–189 (2010).
24. Herrera, F. E. & Pantano, S. Salt induced asymmetry in membrane simulations by partial restriction of ionic motion. *J Chem Phys* **130**, 195105 (2009).
25. Lavery, R. & Sklenar, H. The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dyn.* **6**, 63–91 (1988).
26. Soper, A. K. The radial distribution functions of water and ice from 220 to 673 K and at pressures up to 400 MPa. *Chemical Physics* **258**, 121–137 (2000).
27. Kell, G. S. Density, thermal expansivity, and compressibility of liquid water from 0.deg. to 150.deg.. Correlations and tables for atmospheric pressure and saturation reviewed and expressed on 1968 temperature scale. *Journal of Chemical & Engineering Data* **20**, 97–105 (1975).
28. Holz, M., Heil, S. R. & Sacco, A. Temperature-dependent self-diffusion coefficients of water and six selected molecular liquids for calibration in accurate <sup>1</sup>H NMR PFG measurements. *Physical Chemistry Chemical Physics* **2**, 4740–4742 (2000).
29. Marrink, S. J., de Vries, A. H. & Mark, A. E. Coarse Grained Model for Semiquantitative Lipid Simulations. *The Journal of Physical Chemistry B* **108**, 750–760 (2004).
30. Groot, R. D. & Rabone, K. L. Mesoscopic Simulation of Cell Membrane Damage, Morphology Change and Rupture by Nonionic Surfactants. *Biophysical Journal* **81**, 725–736 (2001).
31. Kusalik, P. G. & Svishchev, I. M. The spatial structure in liquid water. *Science* **265**, 1219–1221 (1994).
32. van Maaren, P. J. & van der Spoel, D. Molecular Dynamics Simulations of Water with Novel Shell-Model Potentials.

- The Journal of Physical Chemistry B* **105**, 2618–2626 (2001).
33. van der Spoel, D., van Maaren, P. J. & Berendsen, H. J. C. A systematic study of water models for molecular simulation: Derivation of water models optimized for use with a reaction field. *The Journal of Chemical Physics* **108**, 10220 (1998).
  34. Mahoney, M. W. & Jorgensen, W. L. Diffusion constant of the TIP5P model of liquid water. *The Journal of Chemical Physics* **114**, 363 (2001).
  35. Yu, H. & van Gunsteren, W. F. Charge-on-spring polarizable water models revisited: from water clusters to liquid water to ice. *J Chem Phys* **121**, 9549–9564 (2004).
  36. Yu, H., Hansson, T. & van Gunsteren, W. F. Development of a simple, self-consistent polarizable model for liquid water. *The Journal of Chemical Physics* **118**, 221 (2003).
  37. Chen, F. & Smith, P. E. Simulated surface tensions of common water models. *The Journal of Chemical Physics* **126**, 221101 (2007).
  38. Wang, H., Junghans, C. & Kremer, K. Comparative atomistic and coarse-grained study of water: what do we lose by coarse-graining? *Eur Phys J E Soft Matter* **28**, 221–229 (2009).
  39. Mahoney, M. W. & Jorgensen, W. L. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *The Journal of Chemical Physics* **112**, 8910 (2000).
  40. Murrell, J. N. & Jenkins, A. D. *Properties of liquids and solutions*. (Wiley & Sons: Chichester [England]; New York, 1994).
  41. Eisenberg, D. & Kauzmann, W. *The structure and properties of water*. (Clarendon Press: Oxford, 2005).
  42. Dilmohamud, B. A., Seeneevassen, J., Rughooputh, S. D. D. V. & Ramasami, P. Surface tension and related thermodynamic parameters of alcohols using the Traube stalagmometer. *European Journal of Physics* **26**, 1079–1084 (2005).
  43. Rodnikova, M. N. A new approach to the mechanism of solvophobic interactions. *Journal of Molecular Liquids* **136**, 211–213 (2007).
  44. Kalcher, I., Horinek, D., Netz, R. R. & Dzubiella, J. Ion specific correlations in bulk and at biointerfaces. *Journal of Physics: Condensed Matter* **21**, 424108 (2009).
  45. Shotton, M. W. *et al.* A high-angle neutron fibre diffraction study of the hydration of B-DNA. *Physica B: Condensed Matter* **241-243**, 1166–1168 (1997).
  46. Young, M. A., Ravishanker, G. & Beveridge, D. L. A 5-nanosecond molecular dynamics trajectory for B-DNA: analysis of structure, motions, and solvation. *Biophys. J.* **73**, 2313–2336 (1997).
  47. Cheatham, T. E., 3rd, Srinivasan, J., Case, D. A. & Kollman, P. A. Molecular dynamics and continuum solvent studies of the stability of polyG-polyC and polyA-polyT DNA duplexes in solution. *J. Biomol. Struct. Dyn.* **16**, 265–280 (1998).
  48. Duan, Y., Wilkosz, P., Crowley, M. & Rosenberg, J. M. Molecular dynamics simulation study of DNA dodecamer d(CGCGAATTCGCG) in solution: conformation and hydration. *J. Mol. Biol.* **272**, 553–572 (1997).
  49. Feig, M. & Pettitt, B. M. Modeling high-resolution hydration patterns in correlation with DNA sequence and conformation. *J. Mol. Biol.* **286**, 1075–1095 (1999).
  50. Young, M. A. & Beveridge, D. L. Molecular dynamics simulations of an oligonucleotide duplex with adenine tracts phased by a full helix turn. *J. Mol. Biol.* **281**, 675–687 (1998).
  51. Kochoyan, M. & Leroy, J. L. Hydration and solution structure of nucleic acids. *Curr. Opin. Struct. Biol.* **5**, 329–333 (1995).
  52. Drew, H. R. *et al.* Structure of a B-DNA dodecamer: conformation and dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **78**, 2179–2183 (1981).
  53. Shui, X., Sines, C. C., McFail-Isom, L., VanDerveer, D. & Williams, L. D. Structure of the potassium form of CGCGAATTCGCG: DNA deformation by electrostatic collapse around inorganic cations. *Biochemistry* **37**, 16877–16887 (1998).
  54. Shui, X., McFail-Isom, L., Hu, G. G. & Williams, L. D. The B-DNA dodecamer at high resolution reveals a spine of water on sodium. *Biochemistry* **37**, 8341–8355 (1998).
  55. Manning, G. S. The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q. Rev. Biophys.* **11**, 179–246 (1978).

56. Ponomarev, S. Y. Ion motions in molecular dynamics simulations on DNA. *Proceedings of the National Academy of Sciences* **101**, 14771–14775 (2004).
57. Savelyev, A. & Papoian, G. A. Electrostatic, Steric, and Hydration Interactions Favor Na<sup>+</sup> Condensation around DNA Compared with K<sup>+</sup>. *Journal of the American Chemical Society* **128**, 14506–14518 (2006).
58. Pérez, A., Luque, F. J. & Orozco, M. Dynamics of B-DNA on the Microsecond Time Scale. *Journal of the American Chemical Society* **129**, 14739–14745 (2007).
59. McConnell, K. J. & Beveridge, D. L. Molecular dynamics simulations of B'-DNA: sequence effects on A-tract-induced bending and flexibility. *J. Mol. Biol.* **314**, 23–40 (2001).
60. Feig, M. & Pettitt, B. M. Sodium and Chlorine Ions as Part of the DNA Solvation Shell. *Biophysical Journal* **77**, 1769–1781 (1999).
61. Fuoss, R. M. The Potential of an Infinite Rod-like Molecule and the Distribution of the Counter Ions. *Proceedings of the National Academy of Sciences* **37**, 579–589 (1951).
62. Shelley, J. C., Shelley, M. Y., Reeder, R. C., Bandyopadhyay, S. & Klein, M. L. A Coarse Grain Model for Phospholipid Simulations. *The Journal of Physical Chemistry B* **105**, 4464–4470 (2001).

## Capítulo III

# Curvatura en el ADN inducida por el entorno iónico: aplicando los modelos CG desarrollados.

### Introducción

**Rol Biológico de la Curvatura del ADN.** Las alteraciones de la conformación canónica del ADN, en particular, el apartamiento de la linealidad de su eje, que denominaremos con la palabra inglesa “bending”, reviste gran importancia en procesos biológicos como la iniciación y regulación de la transcripción,<sup>1-6</sup> recombinación,<sup>7</sup> compactación<sup>8,9</sup> y la formación de los sitios catalíticos de endonucleasas.<sup>10,11</sup> En dichos procesos el “bending” es crítico en al menos tres aspectos: primero, para acercar sitios de unión a proteína que se hallan lejos permitiendo la integración de información de secuencia a larga distancia; segundo, es probable que la información de secuencia contenida en el ADN sea detectada en cierta medida por el proceso denominado “lectura indirecta” (modulación de la unión de proteínas por reconocimiento de variaciones estructurales locales, entre ellas el “bending”, dependientes de secuencia);<sup>12</sup> tercero, para lograr la enorme compactación requerida para almacenar el ADN dentro de la célula, vía interacciones con estructuras proteicas.<sup>13</sup>

Una notable implicación de la flexibilidad de ADN surge cuando se considera la concentración efectiva de sitios específicos en el ADN (o proteína unida a los mismos). Dichas concentraciones efectivas son un reflejo de la frecuencia de colisiones intramoleculares entre dichos sitios de una molécula de ADN. Resultados teóricos indican que la mayor frecuencia de colisiones entre dos sitios ocurre cuando la separación entre ellos es de ~500pb, dando lugar a una concentración efectiva de  $\sim 1 \times 10^{-7}$  M.<sup>14</sup> Variaciones en dicha distancia implican reducciones en la concentración efectiva entre 5 y 50 veces. Esto implica que los cambios estructurales en el ADN que dan cuenta de su flexibilidad, entre los que se destaca el “bending”, son capaces de alterar el grado en que las proteínas se sensan entre sí al unirse a una molécula de ADN común.

Los niveles de “bending” de varios sistemas biológicos importantes (nucleosomas, complejos de recombinación, complejos de transcripción, etc.) requieren de un gasto energético para generar el ensemble de conformaciones estables de ADN que presentan una curvatura necesaria para la interacción. Dicha energía proviene tanto de interacciones específicas entre los residuos de la proteína y las bases del ADN (contactos de van der Waals, efecto hidrofóbico, formación de puentes de hidrógeno), así como también de la formación de interacciones inespecíficas con el “backbone” del ADN, principalmente mediante puentes salinos. También se ha propuesto

que la fuerza iónica del medio puede jugar un papel relevante en la estabilización del "bending" del ADN, que podrían complementar el trabajo que deben realizar las proteínas.<sup>15,16</sup>

**Características Mecánicas del ADN.** La doble hélice en solución acuosa presenta un notable rigidez, con una longitud de persistencia de ~150pb.<sup>17-22</sup> Se postula que las causas de dicha rigidez radican en el equilibrio entre dos actores principales: 1) la tendencia a maximizar el stacking de las bases y 2) la repulsión electrostática interfosfato. Algunos datos le han dado mayor relevancia a la primera<sup>23</sup> mientras que otros autores han puesto mayor atención en la segunda.<sup>24</sup> En tal sentido se ha propuesto que las desviaciones de la estructura del ADN respecto de la conformación canónica se deben al desbalance entre ambas componentes de la rigidez cuando los grupos fosfato son neutralizados localmente. En particular, se predijo que la neutralización asimétrica que puede producirse por la presencia de densidad de contraiones no homogénea, o la presencia de una proteína con residuos catiónicos unida a una cara del ADN, puede resultar en fuerzas interfosfato significativas como para inducir la deformación del ADN.<sup>25,26</sup>

Sin embargo, estudios de fragmentos de ADN sugieren que la doble hélice podría ser mucho menos rígida que lo predicho por los modelos clásicos.<sup>27,28</sup> Adicionalmente la estructura y flexibilidad local son dependientes de la secuencia, pudiendo favorecer desviaciones significativas de la forma canónica del ADN. Dichas desviaciones se pueden conceptualizar en términos de interacciones estéricos que dependen de la secuencia y que determinan fundamentalmente el grado de rotación entre dos pares de bases consecutivas en la doble hélice ("twist") observado en cada tipo de step de pares de bases.<sup>29</sup> Por otro lado, cabe notar la mayor movilidad observada para el "step" pirimidina-purina que podrían actuar como regiones "bisagra" en complejos proteína-ADN que requieren del "bending" del ADN.<sup>30</sup>

Más recientemente se ha descrito que regiones ricas en A (excluyendo el "step" TpA) presentan una fuerte tendencia a inducir conformaciones de la doble hélice con un surco menor más angosto. Esto se traduciría en sitios de potencial electrostático más negativo sobre el surco menor y por ende en atractores de carga positiva.<sup>31</sup> Adicionalmente los mismos autores observaron una mayor frecuencia de aminoácidos positivos, en particular ARG, cuando se analizan la ocupación de los 20 tipos de aminoácidos naturales en el surco menor en estructuras experimentales de proteína-ADN. Incluso dicha tendencia se incrementa cuando se consideran aquellas estructuras que presentan un surco menor más angosto. Dichas observaciones permitirían explicar, al menos parcialmente, el reconocimiento de las diferencias estructurales locales del ADN, dependientes de secuencia, por parte de diversas proteínas.<sup>31</sup>

Por otro lado, el acceso a la escala temporal de los microsegundos por medio de simulaciones de dinámica



molecular clásica<sup>32</sup> han permitido observar que la estructura del ADN oscila en el entorno de la conformación canónica, pero a la vez soporta transiciones conformacionales reversibles importantes en la escala que va desde los pico- a los micro-segundos. Adicionalmente, la mayor parte del tiempo el ADN se encuentra explorando localmente conformaciones no canónicas, las cuales son dependientes tanto de la secuencia como del entorno acuoso e iónico.

A modo de resumen se puede caracterizar al ADN como una molécula de significativa rigidez pero con una notable capacidad de explorar conformaciones que se apartan de la clásica forma B, fuertemente ligada a la secuencia y la anisotropía del entorno acuoso e iónico que lo rodea. Adicionalmente, la disección de los modos de movimiento globales del ADN (Análisis de Componentes Principales) indican que presenta un patrón de deformación espontánea que permite facilitar tanto transiciones a otras formas del ADN (A, Z, etc), como las deformaciones necesarias para la interacción con proteínas.<sup>32</sup> En tal sentido, se ha propuesto que la importancia evolutiva del ADN no radica únicamente en la información contenida en su secuencia sino que sus patrones de deformabilidad juegan un papel funcional relevante. De hecho estudios recientes muestran que ~12% del genoma humano presenta restricciones evolutivas cuando incorpora la similitud topográfica en el ADN en los algoritmos de búsqueda. Dicho valor duplica el predicho solo en base a similitud de secuencia. Además, las regiones identificadas se correlacionan con regiones no codificantes de importancia funcional como "enhancers". Estos resultados apoyan la hipótesis de que la forma molecular del ADN está bajo selección natural y puede ser útil para el análisis evolutivo.<sup>33</sup>

**"Bending" inducido por iones.** Cuando se contrastan las predicciones de teorías termodinámicas de polielectrolitos<sup>34</sup> con datos estructurales surge una seria contradicción al observarse un déficit de neutralización de la carga del ADN por parte de la atmósfera iónica. Esto ha llevado a una visión de la estructura del ADN donde los contraiones quedan relegados a un papel secundario. Casos paradigmáticos de esto son: i- dodecámero de Dickerson resuelto a 2.5 Å de resolución conteniendo el fragmento de ADN únicamente rodeado de agua en ausencia de contraiones o el resuelto por Williams a muy alta resolución en el cual solo un ion Mg y un fragmento de espermina complementan el entorno acuoso (>150 moléculas de agua) que rodea al ADN. ii- Estructura de tRNA a alta resolución con 76 grupos fosfatos pobremente neutralizados por solo 4 iones Mg. iii- Estructura del nucleosoma a 2.8 Å de resolución donde la carga de los 290 grupos fosfatos del ADN es parcialmente balanceada por la presencia de 162 aminoácidos positivos (sin considerar la carga negativa propia de la fracción proteica) y 6 cationes divalentes.

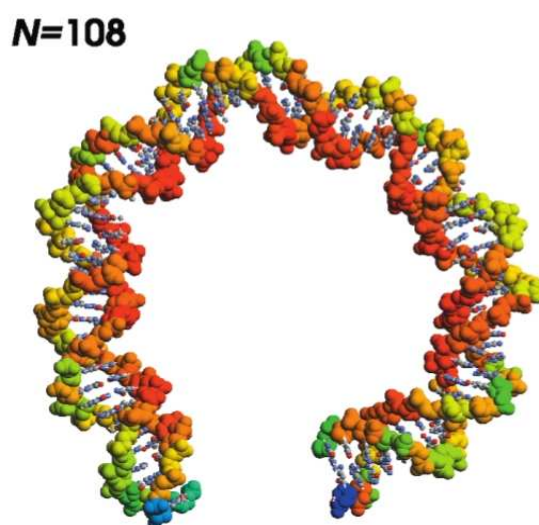
En contraposición a la visión generada en base a estos datos estructurales, Williams y colaboradores propusieron

un modelo de solvatación<sup>16,35</sup> consistente con indistinguibilidad entre cationes monovalentes y agua en la difracción de rayos X. Dicho modelo postula que los cationes se distribuyen asimétricamente alrededor del ADN, y que sitios previamente asignados a agua, serían en realidad sitios de ocupación parcial cationes/agua. Dicho modelo se sustenta en observaciones cristalográficas donde se aprecia la localización de sitios con ocupación fraccional cation/agua dentro del surco menor de “tracks” de A.<sup>16,35,36</sup>

Incluso previo a los trabajos de Williams, ya en 1973, Rosenberg et al. reportaron la presencia de un catión sodio dentro del surco menor en un “track” AT utilizando difracción de cristal único.<sup>37</sup> La relevancia de dicha observación fue desestimada durante las interpretaciones iniciales del dodecámero de Dickerson, en las que se identificaba “espinas de hidratación” puramente acuosa en el surco menor del “track” AT.<sup>38</sup> Sin embargo, evidencia adicional a favor de la presencia de cationes monovalentes dentro del surco menor de “tracks” AT fue aportada por Bartenev et al. utilizando difracción de fibras.<sup>39</sup> A su vez, datos obtenidos en solución utilizando NMR<sup>40,41</sup> demuestran la unión preferencial de amonio (utilizado como marcador de localización de iones tanto en ADN-B como ADN cuádruple) en dichas regiones de la doble hebra. Datos similares fueron aportados por simulaciones de DM realizadas por Young y Beveridge<sup>42</sup> y Young et al.<sup>43</sup> de fragmentos de ADN en varias condiciones salinas. Este conjunto de datos experimentales y teóricos sugieren la unión preferencial de cationes monovalentes en el surco menor de “tracks” AT, replanteando una visión de la estructura del ADN fuertemente relacionada con su entorno. En esta, la curvatura no sería solo una propiedad intrínseca de la doble hélice sino la respuesta a su hidratación específica y a la distribución asimétrica, secuencia dependiente, de contraiones.

A su vez, diversos esfuerzos, experimentales y teóricos han permitido converger en una caracterización del modo de interacción de los cationes mono- y divalentes con el ADN. Actualmente se acepta que: i-  $Mg^{2+}$  y  $Ca^{2+}$  se ligan en los “steps” GpN desde el surco mayor tanto de ADN en forma A como en forma B; ii-  $Mg^{2+}$  y  $Ca^{2+}$  se unen a tracks de G o a la interfase G/A desde el surco menor del ADN en forma B, con capacidad de inducir cambios en el ancho de dicho surco. A su vez datos de RMN evidencian la unión de  $Mn^{2+}$  en el surco menor de segmentos 5'-T<sub>n</sub>A<sub>n</sub>-3' de ADN en la forma B; iii- los iones metálicos monovalentes muestran una fuerte preferencia por G en el surco mayor de ADN en forma B y se observan también en “steps” GpT en la forma A. Simulaciones de dinámica molecular sugieren que la topología del surco se ve alterada por la unión de dichos iones. Sin embargo datos de cristalografía no verifican dichas predicciones teóricas; iv- en el surco menor del ADN en forma B, los iones metálicos coordinan T de hebras opuestas en el “step” ApT de “tracks” de A. Tanto las simulaciones como datos de RMN indican múltiples sitios de unión en el surco menor así como disminución del ancho del mismo como resultado del enlace de iones. Sin embargo, la cristalografía apunta a una unión preferencial en los “steps” ApT.<sup>44</sup> Varios trabajos teóricos han aportado datos significativos respecto del posible rol de los contraiones en el

proceso de "bending" de ADN. Por un lado Range et al<sup>15</sup> utilizaron un método que mezcla el uso de la función de Green, la solvatación implícita vía el método COSMO y la inclusión explícita de iones por simulaciones Montecarlo, para diseccionar la repulsión electrostática individual de cada par de grupos fosfatos en ADN nucleosomal lineal y con "bending" de acuerdo a la información cristalográfica. De esta manera los autores presentan evidencia de una significativa contribución a la energía libre de "bending" (~30% de la energía total) por parte de las interacciones electrostáticas entre los grupos fosfatos. Adicionalmente observan que a medida que se incrementa la concentración iónica del medio, aparecen sitios de unión preferencial "hot spots" de cationes en la cara interna del ADN curvado a nivel nucleosomal (Figura 3-1). La localización de dichos "hot spots" se correlaciona con sitios que presentan "bending" en el eje de la doble hélice en la dirección del surco menor, en particular "steps" pirimidina-purina con tendencia a deformarse.<sup>30,45</sup> Estas observaciones respaldan la hipótesis de que el mecanismo que subyace al "bending" del ADN tendría una contribución significativa de la neutralización asimétrica de carga.<sup>15</sup>



**Figura 3-1. Ocupación iónica promedio en ADN curvado.** Los átomos del "backbone" del ADN nucleosomal se presentan coloreados de acuerdo al nivel de ocupación iónica (escala: azul<celeste<verde<naranja<rojo). N (=108) indica el numero de contraiones utilizados, que en este caso coincide con lo predicho por la Teoría de Condensación de Contraiones. (Tomada de Range et al.<sup>15</sup>)

Por otro lado, simulaciones atomísticas sobre dodecámeros de ADN, forzando su conformación hacia ciertos ángulos entre los planos definidos por pares de bases consecutivas ("roll"), proveen evidencia en favor de la agregación de cationes del lado cóncavo del ADN, cuando este presenta "bending", alcanzando una concentración diez veces mayor a la del seno del medio acuoso. En tal sentido los iones estarían facilitando el

"bending" del ADN al apantallar la repulsión electrostática entre los grupos fosfatos.<sup>46</sup>

En términos generales puede decirse que los iones se ligan al surco mayor preferentemente vía interacciones con "tracks" de G mientras que la unión al surco menor presenta preferencia por "tracks" de A. Sin embargo una correlación cuantitativa entre la unión de iones y cambios en la conformación de la hélice (disminución del ancho del surco menor y generación de curvatura en el eje principal) no se evidencia claramente. Los datos provenientes de la cristalografía de rayos X son particularmente contradictorios en tal sentido, a diferencia de resultados de RMN y simulaciones que apuntan más claramente a la sensibilidad estructural del ADN al entorno iónico.

**"Bending" inducido por proteínas.** El estudio de oligonucleótidos, utilizando el mismo modelo CG de ADN, pero considerando una solvatación implícita<sup>47</sup> muestra que el ADN es capaz de explorar por sí mismo niveles de "bending" comparables con los observados en complejos proteína-ADN experimentales. Adicionalmente, se ha observado experimentalmente que "tracks" de A pueden presentar un "bending" intrínseco de hasta 20 grados.<sup>48</sup> Sin embargo, los factores de transcripción TBP (por su nombre en inglés: "TATA Binding Protein") y CAP (por su nombre en inglés: "Catabolite Activator Protein") inducen "bending" de 80 y 90 grados respectivamente.<sup>49,50</sup> Otro ejemplo destacable es el factor de integración en el hospedero presente en algunos procariontes como *E. coli* que induce un "bending" neto de 160 grados por medio de dos "kinks" separados 9pb.<sup>51</sup> Por lo tanto es evidente que las proteínas, y quizás también el entorno iónico, juegan un papel central en el establecimiento del "bending" final que se observa en el complejo proteína-ADN.

Datos estructurales de alta resolución han revelado al menos dos motivos de unión y "bending" de ADN sugiriendo dos mecanismos posibles. En el primero (clase I) la proteína contacta al ADN que presenta "bending" sobre su superficie convexa, induciendo al eje de la hélice a curvarse de manera que aumenta la convexidad.<sup>52</sup> Dentro de esta clase de proteínas se encuentra la TBP,<sup>2,3</sup> proteínas del grupo de alta movilidad como el factor SRY o el "lymphoid enhancer binding protein I",<sup>53,54</sup> así como otras proteínas comúnmente clasificadas como proteínas arquitectónicas.<sup>54,55</sup> El "bending" que se observa en el ADN al interactuar con esta clase de proteínas parece involucrar la intercalación de aminoácidos hidrofóbicos entre pares de bases del lado del surco menor, ocasionando desenrollamiento de la doble hélice, ensanchando dicho surco y alterando el eje de la hélice<sup>54</sup>. Del punto de vista mecanístico se ha sugerido que el carácter dieléctrico de los residuos de la proteína que interactúan con el ADN alteraría la repulsión interfosfato, contribuyendo al "bending". En el caso de las proteínas de clase I el bajo carácter dieléctrico de los aminoácidos en contacto con el surco menor induciría un aumento de la repulsión, contribuyendo a la convexidad de la cara del ADN que interactúa con la proteína.<sup>56</sup>

La clase II de proteínas que inducen "bending" en el ADN incluye la CAP de *Escherichia coli*,<sup>1</sup> el octámero de histonas responsable del marcado "bending" de las ~150 pb de ADN en el nucleosoma en ~720 grados,<sup>8</sup> o el factor de integración en el huésped IHF (por su nombre en inglés: "Integration Host Factor") de *E. coli*.<sup>51</sup>

Esta clase de proteínas contactan al ADN del lado cóncavo de la superficie induciendo "bending" en el eje de la doble hélice en la dirección de la proteína. En general la superficie proteica involucrada presenta aminoácidos catiónicos, sugiriendo que las interacciones electrostáticas son importantes para la unión.

Para poner en un contexto termodinámico, el costo energético estimado para curvar 360 grados ~75 pb (una vuelta del nucleosoma) se traduce en una constante de equilibrio de  $3 \times 10^{-18}$ . Esto muestra que interacciones proteína-ADN favorables termodinámicamente son esenciales para lograr niveles de "bending" como este caso. Entre los factores que favorecen energéticamente el "bending" por proteínas probablemente se encuentra la formación de nuevos enlaces de hidrógeno, contactos de van der Waals, liberación de agua de superficies de contacto apolares e interacciones electrostáticas incluyendo atracción entre aminoácidos catiónicos y los grupos fosfato en el ADN y la liberación de contraiones condensados al unirse la proteína.<sup>57,58</sup>

Un argumento similar al planteado para analizar el efecto de los iones sobre la estructura del ADN puede esgrimirse para el correspondiente análisis del efecto de la unión de proteínas que inducen "bending". Es decir, la presencia de una proteína alteraría el potencial eléctrico experimentado por el ADN y los iones asociados. Para ejemplificar, las proteínas de clase II podrían aproximarse a la doble hélice presentándole una superficie convexa conteniendo múltiples residuos catiónicos. La fuerza que dirigiría el "bending" sería por tanto la maximización de las interacciones electrostáticas favorables y posiblemente también de la ganancia en entropía del sistema por liberación de contraiones y agua de la superficie del ADN.

Otra aproximación al problema del "bending" es la realizada originalmente por Mirzabekov y Rich<sup>26</sup>, que posteriormente fuera profundizada por Manning y colaboradores.<sup>25</sup> En esta, se argumenta que la neutralización asimétrica de la carga parcial de los fosfatos debido, por ejemplo, a la superficie catiónica de las proteínas de clase II, alteraría el balance de fuerzas electrostáticas dentro de la doble hélice. Más aun, cálculos teóricos basados en la extensión para ácidos nucleicos de la teoría de condensación de contraiones postulada por Manning<sup>59</sup> indican que una neutralización de solo el 6% de la carga presente en los grupos fosfato del ADN nucleosomal es suficiente para que la energía libre de "bending" al nivel nucleosomal sea cero.<sup>60</sup> Teniendo además en cuenta que dicha neutralización se corresponde muy bien con la estequiometría entre ARG en el surco menor y grupos fosfatos observada en la cristalografía, este resultado provee una sólida base teórica que respalda la importancia de la neutralización asimétrica en la deformación del ADN. Sin embargo es necesario notar que en dichos resultados teóricos, la neutralización se induce de manera homogénea en todo el ADN a

diferencia del efecto puntual observado en la cristalografía.

Dichas predicciones teóricas han encontrado correlato experimental en varios sistemas, los cuales se comentan brevemente a continuación.

### **Estudios en sistemas de ADN libre**

Strauss y Maher llevaron a cabo estudios en los que, por medio de sustituciones de algunos grupos fosfato por metilfosfonatos, se imita la neutralización asimétrica que podría inducir una proteína de clase II.<sup>61</sup> Dichos estudios confirmaron el "bending" del ADN en la dirección predicha, mostrando ángulos de ~20 grados cuando la sustitución ocurría en fosfatos consecutivos y de ~13 grados al sustituir fosfatos de manera alternada. Sin embargo, en base a los resultados obtenidos por los autores, no era posible descartar que contactos estéricos desfavorables en el surco mayor, generados por dicha modificación en los fosfatos, no estuviese exagerando la contribución electrostática al "bending". Sin embargo cálculos de minimización de energía del duplex de ADN con fosfatos neutralizados o metilados muestran "bending" en la dirección de neutralización comparable entre ambas formas de neutralización. Dichos resultados sustentan la interpretación de Strauss y Maher respecto de que el "bending" observado experimentalmente por metilaciones de los fosfatos es consecuencia de la neutralización.<sup>62</sup> Adicionalmente, en dicho trabajo teórico, se hace notar la importancia de la estereoespecificidad de la neutralización de los oxígenos del grupo fosfato en la direccionalidad del "bending".

La metodología de sustitución de fosfatos por metilfosfonato fue utilizada nuevamente por Strauss et al para la secuencia de ADN que liga al factor de transcripción PU.1 perteneciente a la familia Ets. La estructura cristalográfica de éste, unida al ADN, muestra un "bending" de 8 grados siendo el patrón de contactos electrostáticos entre ambos, consistentes con el mecanismo de "bending" por neutralización asimétrica. Experimentos de sustitución por metilfosfonato para simular la presencia de PU.1 fueron utilizados para estudiar dicha hipótesis, observándose el colapso del ADN hacia la superficie neutralizada, acompañado de un "bending" de 28 grados.<sup>63</sup> A pesar de que la dirección de la curvatura generada en el ADN era ligeramente diferente a la observada en el cristal, el componente electrostático de la energía de unión parece suficiente para explicar el "bending" observado en el complejo.

Datos experimentales adicionales fueron aportados por los mismos autores quienes demostraron, ligando covalentemente iones amonio a una cara del ADN (vía bases nitrogenadas modificadas), que mientras que la secuencia estudiada (5'-A3GT3) presentaba una curvatura intrínseca de 9 grados hacia el surco menor, la misma conteniendo los iones amonio presentaba un incremento de 8 grados adicionales en la misma dirección.<sup>64</sup>

Más recientemente, por medio de una aproximación experimental similar a la recién mencionada, se demostró,

por medio de medidas de FRET, que la localización de carga positiva en el surco mayor, utilizando bases modificadas, alteran la flexibilidad (una modificación) o inducen "bending" (dos modificaciones adyacentes) de ~25 grados, y dichas modificaciones se atribuyen a neutralización asimétrica. Además plantean que es posible que la unión de carga positiva al surco mayor ocurra vía un proceso cooperativo, donde la unión del primero aumenta la flexibilidad local de la hélice, favoreciendo la unión del siguiente.<sup>65</sup> Dicha observación está de acuerdo con resultados experimentales previos que muestran un aumento en la flexibilidad del ADN (hasta 40%) por neutralización asimétrica.<sup>66</sup> Sin embargo, en este caso los autores hacen hincapié en el consecuente aumento de la entropía que estaría asociado a la neutralización asimétrica lo cual podría favorecer la formación del complejo proteína-ADN.

### Estudios en complejos proteína-ADN.

**Familia bZIP.** El estudio de las proteínas de tipo basic-zipper (bZIP) que incluyen Fos/Jun, Jun/Jun, CREB y GCN4 han aportado datos que avalan el papel de la neutralización asimétrica de la doble hélice, aunque de manera menos clara. Por ejemplo, sustituciones de residuos de GCN4 en el sitio de unión mostraron una correlación casi lineal entre la carga de los aminoácidos y el "bending" observado. Dicha relación se observó en un rango que va desde una carga +6 en la proteína con un "bending" de 15.5 grados hacia el surco menor, hasta una carga de -6, con un "bending" de 25.2 grados hacia el surco mayor.<sup>67</sup> Resultados similares se observaron para Jun/Fos.<sup>68</sup> Estos datos sugieren que las proteínas bZIP podrían actuar como proteínas de clase II. Sin embargo, el mecanismo electrostático subyacente de "bending" no es evidente a partir de dichos experimentos, interpretándose los mismos en términos de contribuciones mixtas entre atracción electrostática y neutralización asimétrica.<sup>68-70</sup> Un intento de profundizar en este aspecto fue realizado por Tomky y colaboradores aplicando la metodología de sustitución por metilfosfonato en la secuencia de ADN de unión de la bZIP AP-1.<sup>71</sup> Un grado de "bending" comparable con el inducido por las variantes de GCN4 con aminoácidos básicos adicionales fue observado, arrojando evidencia a favor de la hipótesis de neutralización asimétrica.

Estudios adicionales de mutantes de GCN4 y de factores de transcripción de la familia bZIP aportaron más evidencia de "bending" inducido por proteína en la dirección de localización de aminoácidos catiónicos.<sup>67,72,73</sup> Sin embargo, no queda del todo clara la magnitud relativa del papel de la neutralización asimétrica en el "bending".

**EcoRV.** Un gran volumen de información ha sido generado respecto de la unión de la endonucleasa de *E. coli* EcoRV a su sitio específico en el ADN, y la importancia de la neutralización asimétrica en dicho proceso. Esta proteína homodimérica se une a un secuencia específica (GATATC) e induce un "bending" de ~50 grados hacia el

surco mayor,<sup>11</sup> rompiendo el "stacking" entre el par T-A central. Sitio este particularmente susceptible al "bending".<sup>11,30,74,75</sup> El mecanismo que subyace a dicho "bending", de acuerdo con datos de cristalografía, depende de fuerzas generadas por interacción con segmentos proteicos y de neutralización selectiva de fosfatos en el surco menor.<sup>10</sup> Se observó que dicha unión al ADN es dependiente de la presencia de iones divalentes, que inducen un aumento de la afinidad por el sitio de clivaje.<sup>76</sup> Más concretamente se observa que sin  $Mg^{2+}$  o  $Ca^{2+}$  no ocurre "bending" y la afinidad disminuye 5000 veces, pero que, sin embargo, tampoco se observa "bending" en ausencia de proteína, es decir, el proceso de unión y "bending" sería simultáneo. El rol propuesto para los iones divalentes sería estabilizar la estructura altamente deformada necesaria para la catálisis, pero no generar intermediarios con "bending" sin proteínas. Esta observación estaría de acuerdo con un modelo en que el tamaño del surco mayor en la forma B, no permitiría un "bending" en esa dirección estabilizada solo por la presencia de iones.<sup>77</sup>

Más recientemente por medio de experimentos de FRET sobre una variante de EcoRV con los extremos C-terminales (que contienen residuos catiónicos) truncados se concluyó que la neutralización asimétrica afecta el "bending" lo cual como consecuencia afecta la unión de la proteína a su diana. Sin embargo los datos de cristalografía presentada en el mismo trabajo, no muestran cambios en el "bending" al truncar dichos residuos catiónicos C-terminales. Para explicar esta incongruencia se aduce que en tales experimentos aún exista una población menor con "bending" completo que es capturada por la cristalografía.<sup>78</sup>

Para aislar y evaluar la contribución de la neutralización asimétrica, por la presencia de la proteína, a la energía libre de "bending" del ADN, Hancock et al<sup>79</sup> utilizaron una combinación de mutaciones puntuales y sustituciones de fosfatos por metilfosfonatos, basadas en la información cristalográfica (PDB ID: 1B94). En esta se puede observar la presencia de residuos catiónicos de la proteína localizados cerca de los fosfatos desde el surco mayor del ADN. La disección de estas interacciones y del potencial efecto de modificaciones en la proteína y/o el ADN permitió a los autores identificar sitios cuya modificación solo estaría afectando la neutralización interfosfato y no la interacción proteína-ADN determinante en el proceso de unión. En tal sentido, experimentos de cinética comparativa entre el sistema "wild type" y varios sistemas que incluyen mutaciones puntuales y/o metilaciones de fosfatos permitió a los autores proponer un modelo que implica una reacción de asociación de segundo orden reversible en la que el primer paso correspondería a un proceso controlado por la difusión y dirigido principalmente por interacciones carga-carga, con un ADN que probablemente presenta un modesto "bending" parcial o que es capaz de explorar conformaciones con "bending" mayores. El "bending" completo se alcanzaría en el segundo paso de reacción e involucraría la formación de muchas interacciones estabilizantes entre la proteína y el ADN, pero solo algunas pocas interacciones carga-carga adicionales. En la dirección reversa, la

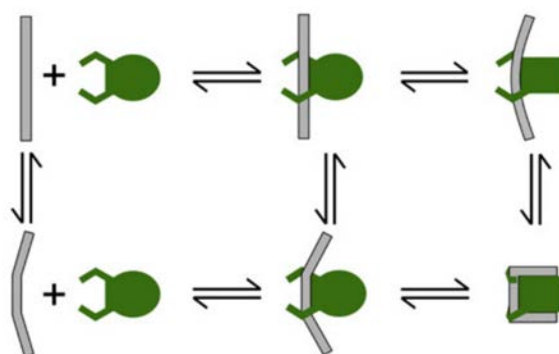


perdida de "bending" es un paso lento debido al alto número de interacciones que deben romperse, dejando solo las interacciones carga-carga por romper, lo cual ocurre más rápidamente. De acuerdo con estos datos, la reducción del potencial electrostático negativo en la cara cóncava del ADN es favorable tanto para el proceso de unión, así como para la catálisis y tiene mayores efectos sobre la velocidad de disociación que de asociación, mientras que perturbaciones similares sobre la cara convexa tienen el efecto contrario. Por ende, la inducción de asimetría electrostática por EcoRV estaría contribuyendo favorablemente a la energía libre de unión.

Los autores a su vez destacan que un factor importante, determinante de la selectividad de secuencia de este proceso, es la mayor flexibilidad axial del "step" T<sub>p</sub>A de sitio de clivaje específico GATATC, que estaría contribuyendo, junto con la neutralización asimétrica, a reducir la energía libre de activación para el proceso de "bending", no observándose lo mismo para otras secuencias de menor flexibilidad.

**IHF.** Una pregunta para cuya respuesta se han comenzado a aportar datos recientemente es: el "bending" es inducido por la proteína? o, la proteína captura estructuras de ADN pre-curvadas?

Una primera respuesta experimental fuerte a esta pregunta fue propuesta recientemente por Vivas y colaboradores<sup>80</sup> estudiando la proteína IHF de *E. coli* mencionada más arriba. Los autores monitorearon el "bending" de la secuencia de unión de IHF vía experimentos de FRET en respuesta a perturbaciones de tipo "laser temperature-jump". De esta manera fueron capaces de observar que la presencia de "nicks" o "mismatches" que aumentan la flexibilidad del ADN en los sitios de "kink" inducen un aumento de 3-4 veces en la velocidad de "bending", que se refleja en un aumento de 4-11 veces en la afinidad por la proteína. Por otro lado, secuencias con modificaciones lejanas de los sitios de "kink", así como mutaciones en IHF diseñadas para desestabilizar el complejo, tienen efectos despreciables en las velocidades de "bending" del ADN a pesar de una disminución mayor a 250 veces en la afinidad de unión IHF-ADN. En base a estos resultados, los autores proponen que el paso limitante en el proceso de reconocimiento por IHF es el la generación de "kinks" espontáneos del ADN para adoptar una conformación pre-curvada. Desde este punto de vista el mecanismo de reconocimiento subyacente sería la captura, por parte de la proteína, de conformaciones pre-curvadas del ADN. Posteriormente, la proteína podría inducir "bending" adicional del ADN, luego de la formación del estado de transición del proceso de unión.<sup>80</sup> La figura 3-2 muestra un esquema del mecanismo de unión proteína ADN propuesto por los autores.



**Figura 3-2. Mecanismo propuesto para la formación del complejo IHF-ADN.** El esquema muestra posibles rutas alternativas para la formación del complejo a partir del ADN y proteína libres. El mecanismo incluye conformaciones del ADN parcialmente curvado en ausencia de proteína (izquierda abajo) o en presencia de la misma (centro abajo), pero previo a los cambios conformacionales en IHF necesarios para la formación del complejo final (derecha abajo). Un complejo en el cual el ADN presenta curvatura independiente de secuencia, interaccionando con la proteína en su conformación final, también es posible de acuerdo con el mecanismo propuesto (arriba derecha). (Tomada de Vivas et al.<sup>80</sup>)

Los datos presentados en los párrafos anteriores presentan una visión en la que la neutralización de la carga presente en los grupos fosfato del ADN estaría jugando un papel significativo en la reducción de su rigidez, lo cual reviste importancia tanto para el empaquetamiento como para la interacción con diversas proteínas. Quedan sin embargo dudas respecto de la relevancia relativa del desbalance anisotrópico de cargas, en particular cuando se trata de la interacción con proteínas. En general, se observa que dicha neutralización sería condición necesaria pero no suficiente para explicar el "bending" asociado a la unión de proteínas y no es claro, excepto por el caso de la proteína IHF mencionada más arriba, si dicho "bending" ocurre parcialmente previo a la interacción con la proteína o si es necesaria la presencia de ésta. Si tuviese lugar el "bending" sin proteína, aun faltarían evidencias fuertes para dilucidar el peso relativo del "bending" intrínseco al ADN y del efecto del entorno acuoso e iónico.

En el capítulo II, se concluyó que al aumentar la escala temporal accesible a la simulación de ADN en condiciones fisiológicas, se observan por primera vez estados estables y reversibles de "bending" de ADN, dependiente de la presencia de cationes en el surco menor. En base a esta observación y a la revisión bibliográfica presentada en los párrafos anteriores, se plantea la siguiente hipótesis:

La unión de iones al surco menor y las conformaciones estables (de cientos de nanosegundos a microsegundos) que se generan como consecuencia, son estados intermedios en la unión de proteínas que para su actividad requieren un ADN curvado.

Para demostrar tal hipótesis se plantea estudiar un caso paradigmático de interacción proteína-ADN que

requiere "bending": la interacción entre la proteína E2 y el ADN en papiloma-virus, que se describe en los párrafos siguientes, y de la que se tiene información cristalográfica de referencia. La idea guía detrás del análisis que se realizara es la posible similitud entre los estados que se observan en el ADN libre (en un entorno acuoso en presencia de iones) y los presentes en los complejos proteína-ADN obtenidos por cristalografía. Adicionalmente se pretende caracterizar las diferencias de tendencia de unión de iones de acuerdo con la secuencia en los segmentos de ADN diana de dichas proteínas.

**Proteína E2.** Se trata de un factor de transcripción/replicación viral codificado en el genoma de los papiloma-virus. Estos son virus de ADN icosaédricos que infectan el tejido epitelial de una variedad de huéspedes incluyendo varias especies de mamíferos. Más de 100 cepas de dicho virus son agentes infecciosos humanos y se denominan HPVs (por "Human papillomavirus").<sup>81</sup> Dichas cepas pueden clasificarse en dos grupos: alto riesgo (p.e.: HPV16 y HPV18) y bajo riesgo (p.e.: HPV6). Mientras que estos se asocian comúnmente con verrugas epiteliales benignas, los primeros están fuertemente asociados al desarrollo de cáncer.

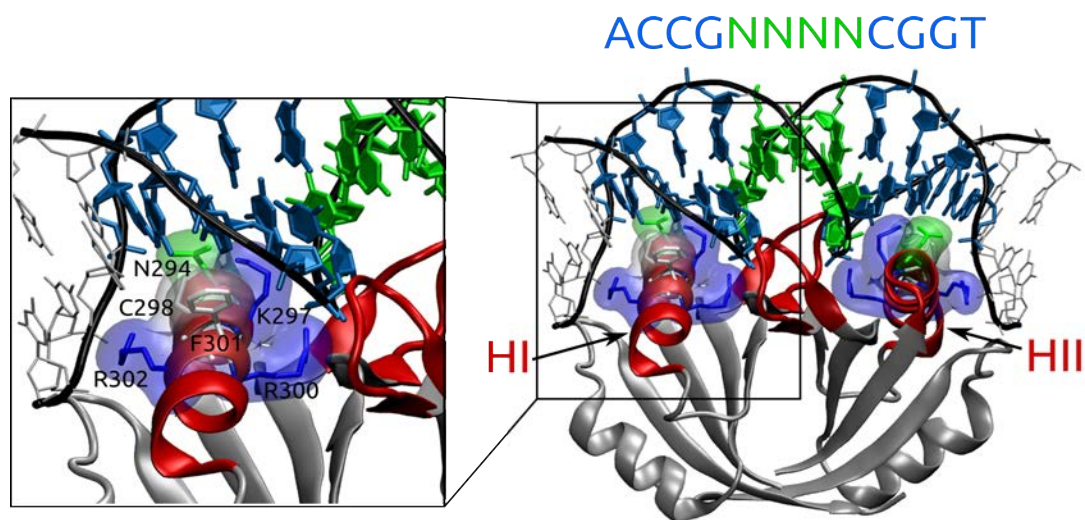
E2 es una proteína homodimérica (~360 aminoácidos por monomero), que cuenta con dos dominios por monomero: el dominio N-terminal responsable de la activación de la transcripción/replicación y el dominio C-terminal que se liga al ADN, unidos por una región linker.

A pesar de que no hay actualmente estructuras experimentales de la proteína E2 completa, si las hay de sus dominios por separado. En particular, para el caso del dominio de unión al ADN existen estructuras tanto para el complejo proteína-ADN así como para los componentes por separado.<sup>82-86</sup> Las E2 de todas las cepas de papiloma-virus se unen a la secuencia palindrómica consenso ACCgNNNNcGGT (g y c indican que G y C son preferidos respectivamente pero no requeridos necesariamente para el reconocimiento específico por E2), vía la inserción de dos alfa-hélices, HI y HII (una de cada monómero) dentro del los surcos mayores consecutivos de las regiones conservadas del sitio de unión (Figura 3-3). La región NNNN se denomina "spacer" y a pesar de tener una composición nucleotídica variable, su longitud está absolutamente conservada en todos los papiloma-virus. Dicho sitio de unión ocurre múltiples veces en el genoma viral. En particular, los genomas de HPV de alto riesgo, presentan cuatro sitios de unión de E2 dispuestos de manera conservada debido a su importancia en el ciclo de infección viral.<sup>81</sup> A pesar de que no existe una secuencia consenso para el "spacer", se demostró que la E2 de varios HPVs muestran no solo mayor afinidad por "spacers" ricos en A:T, sino preferencia por "spacers" conteniendo una secuencia del tipo AA[A/T]N ya que se observó por ejemplo que TTAA presenta menor afinidad que AATT.<sup>85,87,88</sup>

En el complejo E2-ADN, la doble hélice presenta un "bending" de 43-51 grados hacia el surco menor de la región

del “spacer”. La deformación ocurre principalmente por compresión del surco menor de la región del “spacer” y en menor medida de los surcos mayores adyacentes (regiones conservadas). En este sentido, la E2 pertenece a la clase II de proteínas de unión al ADN, de acuerdo con la definición mencionada en la introducción.

Una característica notable de este sistema es que la capacidad de E2 de distinguir entre sitios de unión del ADN que difieren en la secuencia del “spacer”, no depende de interacciones directas entre la proteína y tal región. En este sentido puede considerarse este como un ejemplo de interacción por “lectura indirecta” o “indirect readout”.<sup>81</sup> Adicionalmente, se ha observado que mientras la secuencia del “spacer” ACGT mantiene en términos generales la linealidad del eje principal de la doble hélice,<sup>84</sup> la secuencia AATT presenta “bending” de 9 grados hacia el surco menor del “spacer”, acompañada de una notable reducción del ancho en el surco menor.



**Figura 3-3. Interacción E2-ADN.** El complejo entre la proteína E2 de BPV-1 con un segmento de ADN conteniendo la secuencia 5'-ACCGACGTCGGT-3' (PDBID: 2BOP) se muestra en el panel de la derecha identificando los sitios de interacción en ambas moléculas. El reconocimiento por “lectura directa” ocurre por inserción de las hélices HI y HII (rojo) en el surco mayor del ADN en la región que contiene la secuencia ACCG (azul), junto con la interacción del “loop” entre las hebras- $\beta$  2 y 3 (rojo) con dicha región de la doble hélice. La región del spacer (verde) no contacta directamente con la proteína siendo responsable del reconocimiento por “lectura indirecta”. Entre los residuos de la proteína que contactan el ADN se destacan N294, K297, C298, R300, F301 y R302 localizados en las hélices HI y HII y que presentan un 94% de conservación (panel de la izquierda).

Como resultado el “spacer” AATT presenta un estado de “bending” previo a la unión de la proteína que se asemeja al que se observa luego de la unión de esta. Tal “pre-bending” podría por lo tanto constituir un “código estructural” para el reconocimiento de proteína, subyacente a la secuencia.<sup>81</sup> Como se mencionó previamente, la flexibilidad conformacional del ADN puede aumentar la afinidad de una proteína por éste. Sin embargo en el caso de la proteína E2, (HPV16/18) se ha observado el fenómeno opuesto. Es decir, la introducción de flexibilidad en el sitio de unión de E2 por medio de introducir un “nick” o “gap” en el “spacer”, disminuye la afinidad de E2

por tal sitio.<sup>87</sup> Este no es el caso para otras isoformas de E2 como la presente en la cepa BPV-1, que muestra correlación positiva entre la flexibilidad y la afinidad.<sup>81</sup>

Por último, se ha propuesto que la distribución de carga positiva en la superficie de E2 que contacta el ADN podría ser un factor determinante de las diferencias de afinidad observadas entre varias isoformas de E2.<sup>81</sup> En tal sentido, el modelo de neutralización asimétrica como factor importante en la deformabilidad del ADN y la avidéz de proteínas por este, también podría ser aplicable en el sistema E2-ADN.

En términos generales se cree que el reconocimiento de secuencia comienza por la difusión de la proteína hacia el ADN a lo largo de una mezcla de caminos. Algunas proteínas difunden principalmente a lo largo del ADN (recorrido unidimensional) por largas distancias, mientras que otras favorecen un recorrido tridimensional sensando al ADN en diversos sitios. Una vez en el sitio objetivo, las proteínas pueden unirse al ADN por medio de un camino cinético único, o por rutas paralelas. Cada ruta puede ser de dos estados o incluir poblaciones de intermediarios cinéticos. En el complejo final, el reconocimiento de secuencia es mediado por contactos específicos entre las cadenas laterales de la proteína y las bases del ADN (lectura directa) y/o por la energética conformacional de bases no contactadas (lectura indirecta) (ver Ferreiro et al<sup>89</sup> y las referencias allí citadas).

En el caso de E2 de HPV16, los cambios conformacionales son relativamente grandes para el ADN y pequeños para la proteína.<sup>81,84,85,90-92</sup> Sin embargo, los elementos de la proteína que reconocen el ADN (hélices H1 y H2, y los loops beta2-beta3) muestran una importante flexibilidad en ausencia de ADN (24,31).<sup>91,93</sup>

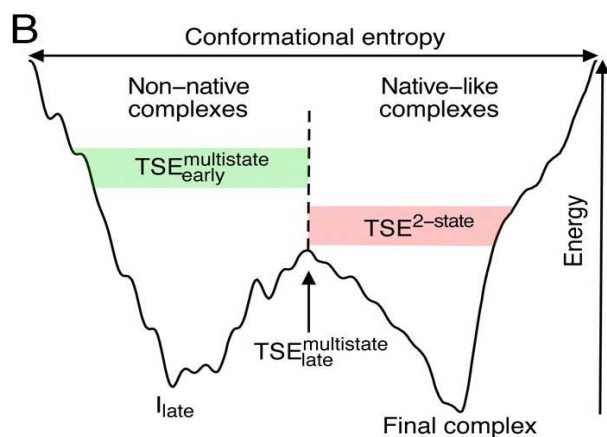
Una vez formado el complejo E2-ADN, las interacciones involucradas en su estabilización pueden clasificarse en dos categorías: específicas y no específicas dependiendo de si la interacción entre la cadena lateral de la proteína contacta las bases del ADN o si el contacto es con el esqueleto de fosfatos, respectivamente. Por otro lado, la proteína presenta dos dominios principales de interacción con el ADN: dos hélices y un loop que conecta las  $\beta$ -strands 2 y 3. Dichas hélices se insertan en el surco mayor del ADN que presenta la secuencia de reconocimiento de E2, y que presenta residuos capaces de establecer interacciones específicas y no específicas. Incluso algunos residuos cumplen una doble función de interactuar con las bases al insertarse en el surco mayor, pero a su vez contactan el esqueleto de fosfatos. Por su lado, los residuos presentes en el el loop  $\beta$ 2- $\beta$ 3 están involucrados en interacciones no específicas. Cabe notar que un subconjunto de los residuos que se observan en las hélices, algunos de los cuales son de interacción específica (C298), otros de interacción doble (N294, K297, Y301 y R302), así como también de interacción no-específica (R300), presentan una conservación > 94 % (Figura 3-3).<sup>94</sup>

En base a experimentos de "stopped-flow" se mostró que la unión ocurre por dos vías paralelas, una de ellas de dos estados y la otra con al menos dos intermediarios poblados.<sup>95</sup> La ruta multiestado podría ser un modelo de transición por diferentes pasos desde la unión inespecífica a la unión específica luego de la difusión

unidimensional,<sup>5,96</sup> mientras que la ruta de dos estados lo sería para la unión luego de la difusión tridimensional. La caracterización de la vía de dos pasos indica que la mayoría de las cadenas laterales que contactan el ADN estabilizan el estado de transición de unión respecto de los reactivos libres. Sin embargo, la contribución de las cadenas laterales que contactan las bases directamente es mayor que las que contactan a los grupos fosfato. El análisis de similitud energética entre el estado de transición, los reactivos y el producto sugiere que la energética de las interacciones en el estado de transición no es uniforme. Los contactos directos con las bases son similares a los observados en el complejo final, mientras que varias de las interacciones con los grupos fosfato son similares a las que ocurren en los reactivos libres. Adicionalmente, el cambio completo en la capacidad calorífica tiene lugar en el paso desde el estado de transición hasta el complejo final, compatible con un estado de transición altamente solvatado con mucha libertad vibracional. En el pasaje al complejo final, las moléculas de agua serían expulsadas de la interfase.<sup>95</sup>

Por otro lado, en la vía de intermediarios múltiples, se observa la ocurrencia de interacciones no-nativas (tanto con el esqueleto de fosfatos como con las bases) entre residuos de las hélices y del loop  $\beta 2$ - $\beta 3$ , durante el proceso de unión. Dichas interacciones no solo ayudan a estabilizar el primer estado de transición y el primer intermediario, sino que favorecen la formación del segundo intermediario, el cual es una trampa cinética, que requiere de la ruptura de tales interacciones para que ocurra la transición al complejo final (escala de los segundos). En tal sentido, se especula que el estado intermediario lento podría tener un rol funcional, como el de modular la búsqueda de la secuencia blanco, o estar involucrado en efectos alostéricos de otras moléculas regulatorias.

El perfil de energía de unión de E2-ADN presenta dos facetas bien diferenciadas que involucran sustancialmente diferentes interacciones. La vía de dos estados se comporta como un suave embudo donde interacciones de tipo nativas predominan y llevan al complejo final. La vía múltiple se comportan como un embudo rugoso en el que estados intermediarios son poblados por medio de interacciones no-nativas (Figura 3-4). La rugosidad de la superficie de energía, sería consecuencia de dichas interacciones no-nativas, aunque otros factores como el requerimiento de "bending" del ADN o la difusión unidimensional a lo largo de regiones de ADN no específicas para E2, podrían estar involucradas.<sup>97</sup>

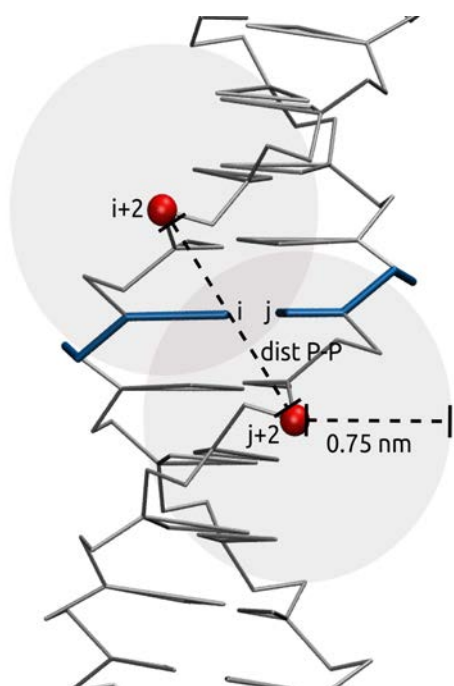


**Figura 3-4. Esquema de la superficie de energía libre para el dominio de reconocimiento de ADN de E2.** (Figura 4.B tomada de Sanchez et al<sup>97</sup>) Proyección del perfil de energía libre con forma de doble embudo. El ancho del gráfico representa la entropía conformacional del sistema E2-ADN, y la profundidad representa la energía de interacción entre ambas moléculas. La parte superior de la figura corresponde con el estado en el que proteína y ADN no interactúan y los dos embudos agrupan interacciones nativas y no-nativas entre ellas. La ruta cinética de dos estados tiene lugar en el embudo de menor frustración (derecha), centrado en el mínimo de energía libre para el complejo nativo. La ruta cinética multiestado mapea un embudo no nativo con mayor grado de frustración (izquierda) centrado en el estado  $I_{late}$ . La conversión desde este estado intermediario hacia el mínimo global requiere que el complejo E2-ADN cruce al embudo nativo, pasando por el estado de transición  $TSE_{late}^{multistate}$ . La disociación del complejo nativo puede implicar el pasaje por  $TSE_{late}^{multistate}$  o por  $TSE_{2-state}$ .

## Métodos

Simulaciones de DM molecular se llevaron a cabo utilizando el programa GROMACS 4.5 para cuatro segmentos de ADN de 17 pares de bases conteniendo el sitio de reconocimiento de E2 y variando la secuencia de la región de spacer (AATT, AAAA, TTAA y ACGT). El ADN fue modelado a nivel CG utilizando el modelo propuesto por Dans et al.<sup>98</sup> El entorno acuoso se modeló utilizando el modelo CG de agua e iones presentado en el Capítulo II, con una concentración iónica de 0.2M. Las simulaciones se llevaron a cabo por 25  $\mu$ s en el “ensemble” isotérmico-isobárico a 300K y 1atm, acoplado la temperatura y presión al termostato y baróstato de Berendsen,<sup>99</sup> utilizando un paso de integración de las ecuaciones de movimientos de 20 fs. Para mantener la integridad de los extremos del ADN, los pares de bases en tales posiciones se mantuvieron unidos mediante el uso de restricciones armónicas de 1255.5 Jmol<sup>-1</sup>. Un “cutoff” de 1.2 nm se utilizó para las interacciones no covalentes mientras que las interacciones electrostáticas de largo alcance se calcularon mediante el método Particle Mesh Ewald (PME).<sup>100,101</sup>

El número de iones ligados en el surco menor en el par de bases genérico (i,j) se calculó definiendo dos grupos, uno que contiene los iones a 0.75 nm del fosfato i+2 y otro similar respecto de fosfato j+2 y contando el número de iones que pertenecen a la intersección de dicho grupos (ver Figura 3-5). Este procedimiento repetido para cada par de bases de interés, en cada paso de simulación, permite obtener perfiles temporales de unión de iones.



**Figura 3.5. Definición de distancia interfosfato y de número de iones por par de bases.** El número de iones en el surco menor del par de bases (i,j) (azul) se define por la intersección de los números de iones contenidos en dos esferas (gris) de radio 0.75 nm centradas en los fosfatos de las bases i+2 y j+2 (rojo). La distancia entre estos fosfatos define a su vez la distancia interfosfato asociada a dicho número de iones del par (i,j)

La suma del número de iones calculada de dicha manera para todos los pares de bases de la región del “spacer” se utilizó para correlacionar el número de iones en el surco menor y la disminución del ancho del mismo, junto con otras propiedades estructurales (RMSD, distancia HI/HII-surco mayor).

La contribución promedio de número de iones por par de bases se obtuvo calculando el porcentaje de iones en cada par de bases respecto del total de iones (promedio) en el surco menor.

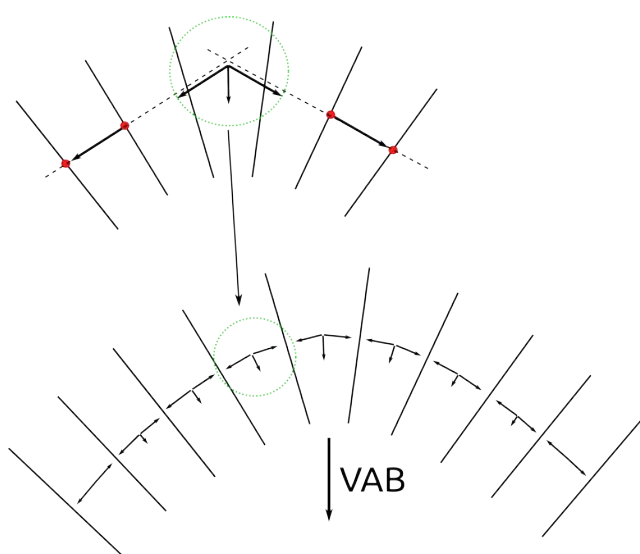
La disminución del ancho del surco menor (“narrowing”) fue analizada, al igual que en el Capítulo II, estimando la distancia promedio interfosfato (ver Figura 3.5) entre hebras opuestas considerando los pares de bases de la región del “spacer”.

El perfil temporal de RMSD se calculó alineando primero cada conformación obtenida mediante la simulación de DM a la estructura experimental de ADN en complejo con la proteína E2 y calculando el RMSD para las posiciones de los grupos fosfato. El alineamiento estructural se realizó sobre la región de “spacer” (extendida un par de bases a cada lado) para focalizar el análisis sobre las deformaciones en la región responsable de la “lectura indirecta”. La distancia entre las Hélices I y II de la E2 respecto de los surcos mayores con los cuales interacciona



en el complejo proteína-ADN se obtuvo midiendo a cada paso de tiempo de simulación la distancia entre el C $\alpha$  del residuo 301 (de cada monomero de la E2) y el tercer par de bases del sitio de reconocimiento por "lectura directa" (ACCG). Esto permite analizar si al interaccionar los iones con el surco menor ocurre un acercamiento o no de los sitios de reconocimiento en ambas moléculas.

La curvatura en el ADN se evaluó por medio de un parámetro que se denominó VAB (Vector Asociado al "Bending") cuya construcción se esquematiza en la Figure 3-6.



**Figura 3-6. Determinación del Vector Asociado al "bending" (VAB).** El "bending" asociado a un "step" de pares de bases (arriba) se mide considerando los dos pares de bases hacia la izquierda y los dos hacia la derecha del "step" de interés. Se determina el punto medio entre los fosfatos de dichos pares (puntos rojos) y el vector que conecta dichos puntos (flechas conectando los puntos rojos). La suma de dichos vectores (flechas dentro del círculo verde) es un vector que indica la dirección e intensidad del "bending" asociado al "step" de interés. Si se suman tales vectores para todos los "steps" de una región de la doble hélice, se obtiene el VAB, el cual condensa el módulo y dirección del "bending" global de dicha región (abajo).

Dicho VAB permite evaluar la intensidad del "bending" por medio de su módulo, así como su dirección global, considerando los ángulos entre dicho vector y los ejes x-y-z.

El Análisis de Componentes Principales o análisis de covarianza fue utilizado para encontrar movimientos correlacionados en el ADN que faciliten el análisis de las deformaciones estructurales debida a la unión de los iones. Se basa en diagonalizar la matriz de covarianza de las coordenadas atómicas pesada por la masa:

$$C_{ij} = \left\langle M_{ii}^{\frac{1}{2}}(x_i - \langle x_i \rangle) M_{jj}^{\frac{1}{2}}(x_j - \langle x_j \rangle) \right\rangle$$

Donde M es una matriz diagonal conteniendo las masas de los átomos. C es una matriz simétrica de 3N x 3N, que puede diagonalizarse mediante una matriz de transformación ortonormal R:

$$R^T C R = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{3N}) \quad \text{where } \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{3N}$$

Las columnas de  $R$  son los vectores propios o componentes principales (PCs del inglés “Principal Components”).  $R$  define una transformación a un nuevo sistema de coordenadas. Los valores propios  $\lambda_i$  representan la fluctuación cuadrática media del componente principal  $i$ . Los primeros componentes principales a menudo describen de manera colectiva los movimientos globales del sistema. En tal sentido, el filtrado de la trayectoria original según un componente principal dado permite conservar solo los desplazamientos en la trayectoria que corresponden a los movimientos colectivos asociados con dicho componente principal, facilitando el análisis de las variaciones estructurales principales en la simulación.

Una manera de medir la similitud del espacio explorado por dos simulaciones es comparar la superposición del espacio definido por los primeros componentes principales de ambas simulaciones. Dicha medida, conocida como Índice de Similitud (IS), se define para el espacio generado por los vectores  $w_1...w_m$  de una simulación y el correspondiente a los vectores  $v_1...v_n$  de la otra simulación, como:

$$\text{overlap}(\mathbf{v}, \mathbf{w}) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m (\mathbf{v}_i \cdot \mathbf{w}_j)^2$$

Adicionalmente, tal comparación puede realizarse comparando todos los vectores  $w_1...w_m$  contra todos los  $v_1...v_n$ , obteniendo una medida detallada de la similitud de cada componente principal de una simulación con cada uno de otra, en lugar de una medida de la similitud de los espacios generados por tales grupos de vectores. Esta última medida se conoce como Matriz de Producto Interno.

El análisis de componentes principales, así como las medidas de IS y matrices de producto interno fueron calculadas para las simulaciones de todas las secuencias de spacer estudiadas (AATT, AAAA, TTAA y ACGT) en presencia o ausencia de iones unidos al surco menor. Este análisis permite conocer las características dinámicas generales del ADN de acuerdo con la secuencia y con la unión de iones.

## Resultados y Discusión

Como se menciona previamente, la interacción entre la proteína E2 y el ADN depende fuertemente de la respuesta mecánica de la región “spacer” frente a las tensiones impuestas por la proteína cuando esta se une. Por lo cual, secuencias que presenten menor resistencia a ser dobladas tendrán asociadas un menor costo energético para la formación del complejo proteína-ADN, lo cual se traduce en una mayor afinidad de la E2 por tales

secuencias. Experimentos de ciclización combinados con estimaciones teóricas de la constante de afinidad han demostrado que ésta guarda una proporcionalidad directa con la riqueza en AT en la región “spacer”.<sup>102</sup>

Por otro lado, de acuerdo con la multiplicidad de datos presentes en la bibliografía, regiones en el ADN que presentan “bending” del eje principal del ADN y “narrowing” del surco menor se correlacionan frecuentemente con unión de cationes a este sitio. Dicha unión ayudaría a estabilizar tales conformaciones compensando la repulsión electrostática debida al acercamiento de los grupos fosfato.

Adicionalmente, en el Capítulo II se mostró que las simulaciones de dinámica molecular utilizando los modelos CG introducidos en el presente trabajo, evidencian cambios conformacionales reversibles en el ADN inducidos por la unión de cationes al surco menor en el dodecámero de Drew-Dickerson. Este proceso sucede en la escala temporal de los microsegundos.

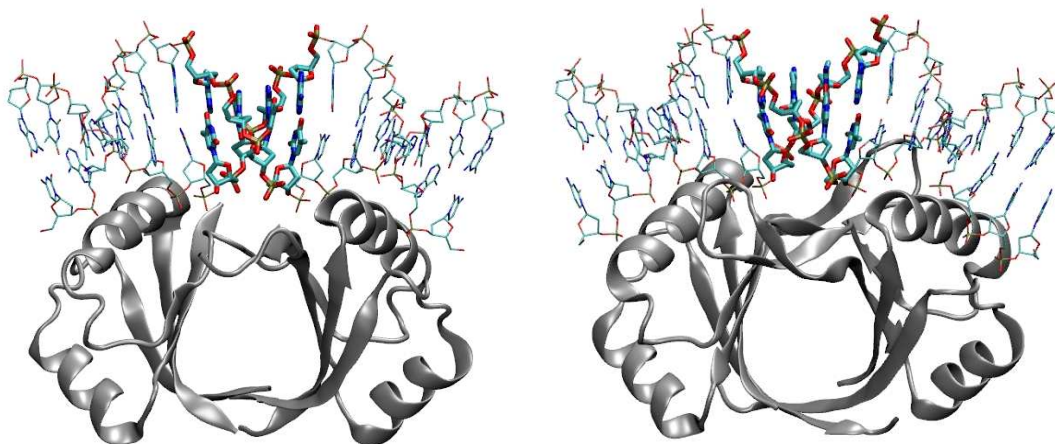
En base a estas observaciones, en el presente trabajo se propuso correlacionar la unión de cationes al surco menor de cuatro segmentos de ADN conteniendo el sitio de reconocimiento de la proteína E2 con los datos experimentales de afinidad. En tal sentido se llevaron a cabo simulaciones de dinámica molecular, utilizando los modelos de ADN, agua e iones CG descritos previamente. La tabla 3-1 presenta las secuencias utilizadas en la región del “spacer” y los distintos valores de afinidad medidos en cada caso.

Tabla 3-1. Afinidad de E2 por diferentes secuencias de región “spacer”.

Secuencia del “spacer”	<sup>a</sup> K <sub>rel</sub> (tomadas de Zhang et al <sup>102</sup> )	“bending” (°) / VAB (nm) (canónico)		“bending” (°) / VAB (nm) (experimental)	
AATT	300	0	0.31	41	0.78
AAAA	160	0	0.31	...	...
TTAA	9.4	0	0.31	...	...
ACGT	1	0	0.31	46	0.95

<sup>a</sup> K<sub>rel</sub> mide la afinidad de E2 por una secuencia de “spacer” dada, relativa a la afinidad por la secuencia ACGT (K<sub>d</sub> = 12 nM); K<sub>rel</sub>=K<sub>d</sub><sup>ACGT</sup>/K<sub>d</sub>.

Actualmente se cuenta solo con estructuras experimentales para secciones con spacers AATT y ACGT, las cuales ligan a la proteína E2 con alta y baja afinidad, respectivamente (Figura 3-7). A partir de dichas estructuras se puede observar que el ADN presenta “bending” en la dirección del surco menor (en el sentido de la proteína) de alrededor de 41 y 46 grados para AATT y ACGT, respectivamente. Tal deformación del ADN permite la inserción de las hélices de interacción de la E2 en los surcos mayores donde se localiza la secuencia de reconocimiento (Figura 3-3).



**Figura 3-7. Estructuras experimentales del complejo E2-ADN.** Representación de las estructuras cristalográficas obtenidas para la interacción de E2 con dos segmentos de ADN de secuencia distinta en la región del spacer (identificada con la representación Licorice): AATT (izquierda, PDBID: 1JJ4) y ACGT (derecha, PDBID: 2BOP).

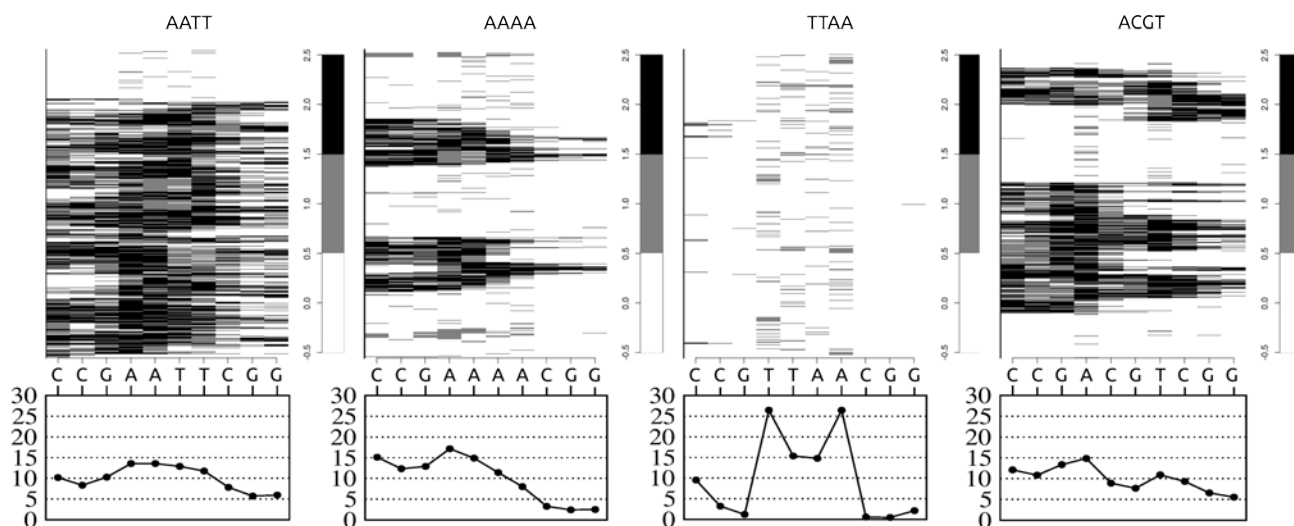
**Unión de iones al surco menor.** La Figura 3-8 muestra el número de iones en el surco menor por par de bases durante 25  $\mu$ s de simulación para las cuatro secuencias estudiadas. Tal medida se realizó contando el número de iones en el entorno de cada par de bases, desde el surco menor, para cada paso de simulación. (ver sección Métodos).

Importantes diferencias se pueden apreciar respecto de la dinámica de unión de iones, dependiendo de la secuencia del “spacer”. Concretamente, la estabilidad temporal de la unión de cationes separa a las secuencias estudiadas en tres clases: unión fuerte (decenas de  $\mu$ s), intermedia (algunos  $\mu$ s) y débil (decenas de ns).

En el caso de unión fuerte se encuentra solo la secuencia de “spacer” de alta afinidad por E2, AATT. Como se evidencia en la Figura 3-8, la dinámica del sistema muestra fundamentalmente dos estados. Un período inicial, con un número reducido de iones ligados principalmente en la región del “spacer”. Este estado es seguido por un fuerte incremento en el número de iones en surco menor de prácticamente toda la secuencia estudiada. Tal incremento se mantiene por el resto del tiempo simulado, así como también se mantiene la mayor localización en la región “spacer” (Figura 3-8, panel superior).

La clase de unión intermedia contiene las secuencias de afinidad media, AAAA y de afinidad baja, ACGT (Figura 3-8, panel superior). En tales casos, se observa la ocurrencia consecutiva de períodos de unión poco frecuente, seguidos de períodos donde incrementa la unión de iones. Este comportamiento es análogo al ya observado para

la secuencia llamada Drew-Dickerson, estudiada en el capítulo anterior (Figura 2-10)). Es importante notar que el número de iones ligados en los casos intermedios durante estos periodos alcanza el mismo nivel que el observado para la secuencia AATT. Además, tal unión ocurre de manera extendida por toda la secuencia al igual que en dicho caso. Sin embargo, tanto en los periodos de alta como de baja unión, se observa una mayor localización de iones en la región del “spacer”, en el caso de AAAA. Este efecto es menos evidente para ACGT (ver Figura 3-8, panel inferior).



**Figura 3-8. Dinámica de interacción iones-ADN. Panel superior:** número de iones ligados por el surco menor según el par de base y el tiempo para las cuatro secuencias estudiadas (AATT: primero, AAAA: segundo, TTAA: tercero y ACGT: cuarto, contando desde la izquierda), medido de acuerdo al procedimiento detallado en la sección Métodos. **Panel inferior:** porcentaje del número promedio (temporal) de iones ligados en todo el surco menor correspondiente a cada par de base.

Por ultimo, en la clase de unión débil se ubica la secuencia TTAA. Esta presenta una afinidad por E2 mayor que ACGT, pero menor que AAAA. Durante toda la simulación se observa una unión muy poco frecuente de iones, en comparación con los tres casos anteriores (Figura 3-8, panel superior). Sin embargo, cuando los iones se condensan, lo hacen principalmente en la región del spacer.

En resumen: i) la secuencia de mayor afinidad por E2 (“spacer” AATT) muestra una dinámica de unión fuerte con una marcada localización en la región del “spacer”; ii) la secuencia AAAA, cuya afinidad relativa por E2 es casi la mitad respecto a AATT demuestra una dinámica de unión de iones intermedia con una localización moderada en la región del “spacer”; iii) la secuencia TTAA, con una afinidad por E2 ~30 veces menor que AATT, muestra una dinámica de enlace de iones débil pero fuertemente localizados en la región del “spacer”; iv) la secuencia ACGT, cuya afinidad por E2 es ~ 300 veces menor que AATT muestra una dinámica de enlace ionice intermedia pero escasa localización.

Los diferentes perfiles de enlace iónico en función de la secuencia estudiada sugieren un posible papel en el proceso de unión de E2 al ADN. En particular se plantea la hipótesis de que el enlace de iones genere cambios conformacionales en la doble hélice que faciliten la unión a la proteína favoreciendo conformaciones en el ADN libre similares a los de la forma ligada.

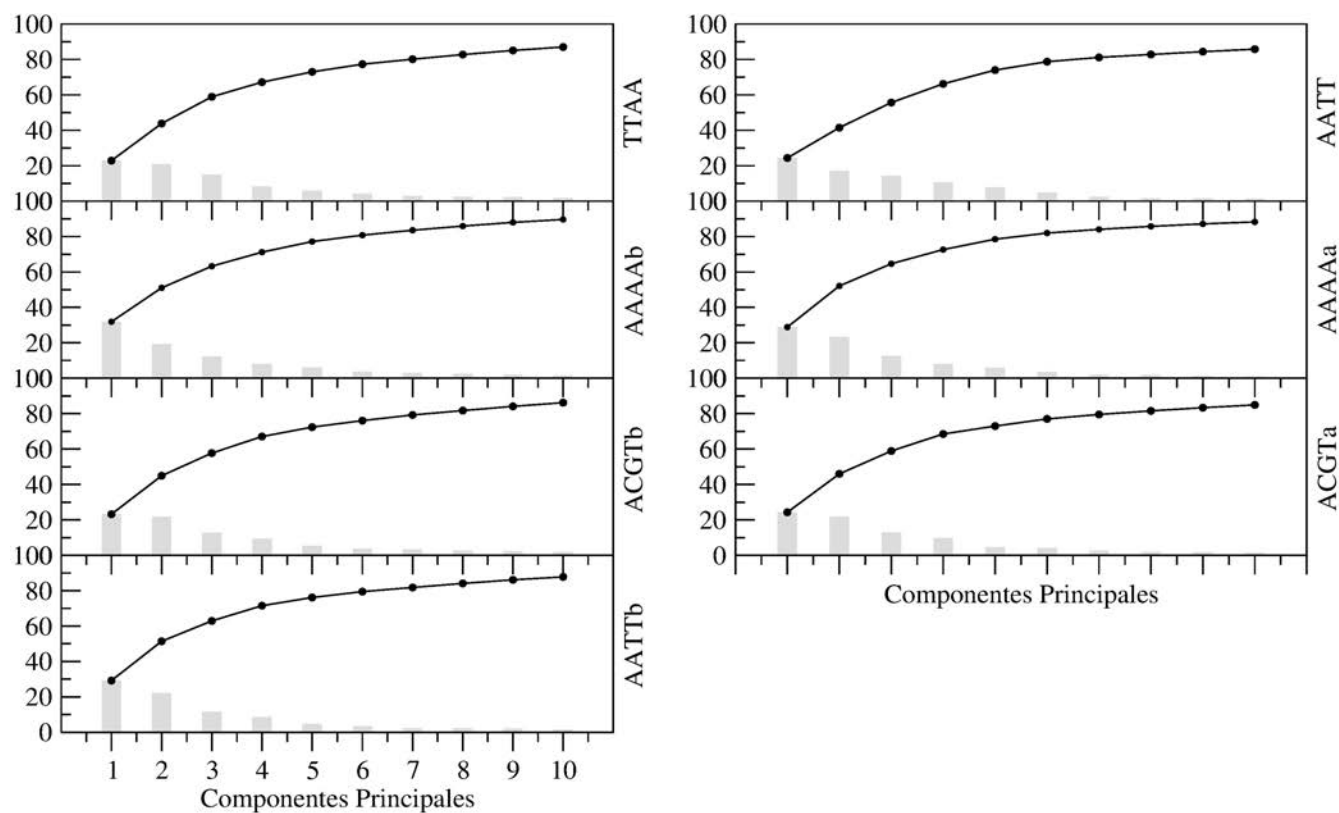
Para avanzar en esa dirección se planteó la siguiente estrategia, en primer lugar se caracterizó la covarianza de las coordenadas del ADN (grupos fosfato) mediante un análisis de componentes principales (PC del la sigla en inglés, Principal Componentes). El objetivo en este caso es identificar similitudes y diferencias en el comportamiento estructural del ADN entre las diferentes secuencias y en presencia o ausencia de iones ligados en el surco menor. En segundo lugar, se estudió el contenido de "bending" en los componentes principales que en conjunto contienen la mayoría de la varianza de las coordenadas del sistema. Esto permite no solo conocer las características globales de los principales movimientos conformacionales que presenta el ADN, sino que también permite revelar específicamente la respuesta mecánica de la doble hélice, en particular de la curvatura de ésta, frente a la presencia de iones en el surco menor. Como marco de referencia experimental, se analizó también el "bending" observado en los complejos E2-ADN de dos estructuras obtenidas por cristalografía de rayos X (ver Figura 3-7). En tercer lugar, se analizó la correlación temporal entre la unión de iones y cambios estructurales en el ADN como: el ancho del surco menor, el RMSD del "spacer" (ajustando contra las estructuras experimentales) y las distancias entre las hélices HI y HII con sus sitios de interacción en el surco mayor. Este análisis final permite observar directamente y en términos temporales, el efecto de la unión de iones al surco menor sobre la conformación del ADN.

**Respuesta del ADN frente a la unión de iones.** El primer paso para comprender las características estructurales de la respuesta de la doble hélice frente a la condensación de iones en el surco menor se analizó la covarianza de las coordenadas cartesianas del sistema. Esto se hace descomponiendo los movimientos del ADN en un espacio de coordenadas denominados componentes principales, que condensan la varianza y por ende ayudan a identificar los movimientos más importantes del ADN (ver sección Métodos).

Como se menciona en la sección Métodos, los componentes principales se obtienen diagonalizando la matriz de covarianza de las coordenadas del sistema. Obteniendo como resultado un conjunto de vectores propios, con sus valores propios asociados, que representan el nuevo conjunto de coordenadas (en lugar de las coordenadas cartesianas de las que se parte) con los cuales representar la dinámica del sistema. Una ventaja de esta nueva representación es que las nuevas coordenadas, o componentes principales, tienen diferente contenido de varianza, medida por medio del valor propio asociado. En este sentido es posible restringir el análisis a un

subconjunto de tales componentes principales, que condensan la mayoría de la varianza del sistema. La Figura 3-9 muestra los valores propios (en porcentaje) de los diez primeros componentes principales ordenados de mayor a menor. Como puede observarse, dicho subconjunto de coordenadas, explica ~90% de la varianza del sistema, por lo cual representa una muy buena reducción de la dimensionalidad de los sistemas bajo estudio, permitiendo un análisis más sencillo de la dinámica global de los mismos. En consecuencia, el análisis que se presenta en los próximos párrafos se realizó tomando el espacio de movimientos generado por los diez primeros componentes principales.

De acuerdo con los resultados presentados en la sección anterior, las secuencias AATT y TTAA presentan una dinámica de unión de iones bien diferenciada. Mientras que AATT presenta alta frecuencia de iones ligados al surco menor en casi toda la simulación (excepto los primeros microsegundos), TTAA muestra muy baja frecuencia. Por su lado, las secuencias AAAA y ACGT presentan un comportamiento intermedio con secciones temporales de alta frecuencia (similar a AATT), seguidas por secciones temporales de baja frecuencia (similar a TTAA).



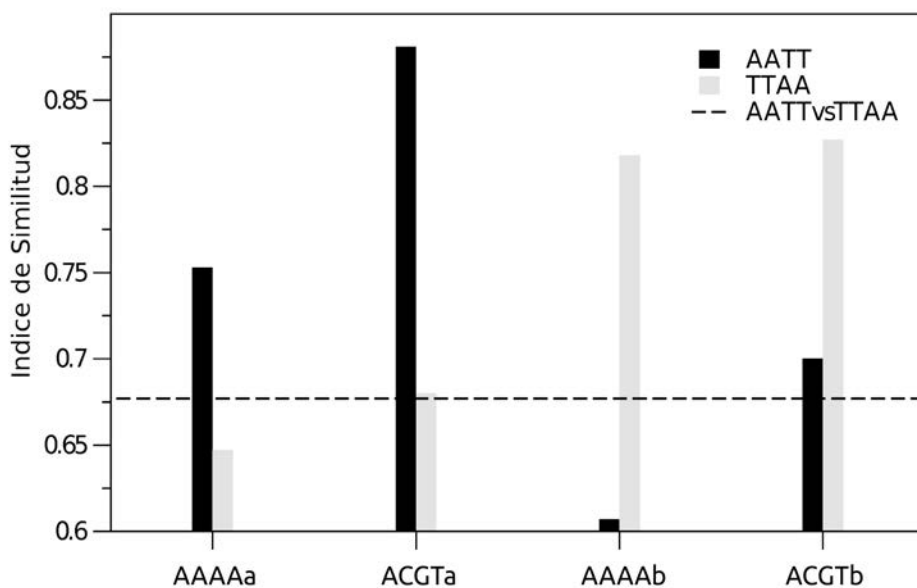
**Figura 3-9. Valores propios de los Componentes Principales.** Se muestran los diez valores propios más grandes en término de porcentaje respecto de la traza de la matriz de covarianza (suma de todos los valores propios), para las cuatro secuencias evaluadas en presencia y ausencia de

iones.

En tal sentido surge natural la idea de fragmentar las simulaciones de estas dos últimas secuencias, agrupando las secciones temporales de baja frecuencias por un lado (AAAAb y ACGTb), y las de alta frecuencia por otro (AAAAa y ACGTa). Esto permite comparar ambas por separado contra AATT y TTAA. Cabe aclarar que las comparaciones con AATT se realizaron respecto de la región temporal (mayoritaria) en la que se observa alta frecuencia de unión de iones. Adicionalmente, a modo de referencia, se agrega a la discusión el comportamiento observado para tal secuencia cuando la frecuencia de unión de iones es baja (primeros microsegundos, AATTb). Una medida general de la similitud del espacio de movimientos explorado por las secuencias de ADN estudiadas, es el Índice de Similitud (IS) (ver Métodos). La Figura 3-10 presenta el IS considerando los diez primeros componentes principales (ver Figura 3-9). Como puede observarse, el movimiento global de ADN para las secuencias AAAA y ACGT en ausencia de iones (AAAAb y ACGTb, respectivamente), presentan un IS respecto de la secuencia TTAA (barras negras) mayor que cuando se las compara con AATT (barras grises). Esto es un indicador de que el movimiento global del ADN para las secuencias AAAA y ACGT en ausencia de iones, es más similar al correspondiente para la secuencias TTAA que al de AATT.

El comportamiento invertido, es decir, mayor similitud de movimiento respecto de AATT que de TTAA, se observa para AAAA y ACGT cuando presentan iones ligados al surco menor. En términos generales, esto implica que la unión de cationes al surco menor genera cambios en la dinámica estructural del ADN, que son comunes a todas las secuencias estudiadas.

Una mirada más profunda de la dinámica global del ADN, se puede adquirir descomponiendo la medida de similitud por cada componente principal (ver Métodos). Tal comparación se muestra en la Figura 3-11, donde la similitud de los diez primeros componentes principales para las secuencias AAAA y ACGT en presencia o ausencia de iones se calcula respecto de las secuencias AATT y TTAA.



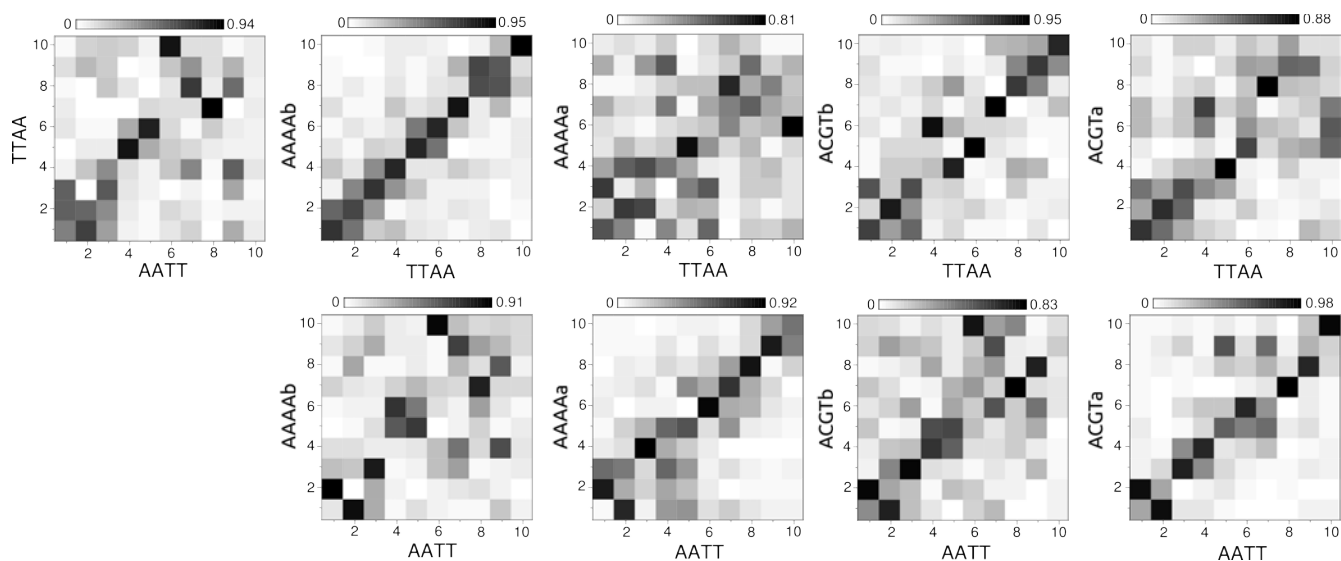
**Figura 3-10. Índice de Similitud.** La dinámica representada por los diez primeros componentes principales se compara para las secuencias AAAA y ACGT, en presencia y ausencia de iones con las secuencia AATT, como referencia con alta frecuencia de unión de iones (barras negras) y TTAA, como referencia con baja frecuencia de unión de iones (barras grises). Adicionalmente se comparan las dinámicas entre AATT y TTAA



(línea horizontal punteada).

Debido a que la similitud se presenta en términos de una matriz donde los primeros componentes principales de una secuencia se comparan contra los primeros de la otra, es esperable que si dos secuencias presentan una dinámica equivalente, la matriz tendrá valores elevados de similitud acumulados en la diagonal (celdas de color negro). Tal es el caso para las parejas AAAAb-TTAA, ACGTb-TTAA, AAAAa-AATT y ACGTa-AATT. En el resto de las parejas comparadas, la tendencia a acumular valores elevados de similitud en la diagonal de la matriz disminuye significativamente. Estos datos muestran, de manera más detallada, el mismo resultado que se obtiene de los valores de IS (presentados en la Figura 3-10). Es decir, el cambio en la dinámica del ADN inducido por la unión de los iones, es muy similar para las cuatro secuencias estudiadas.

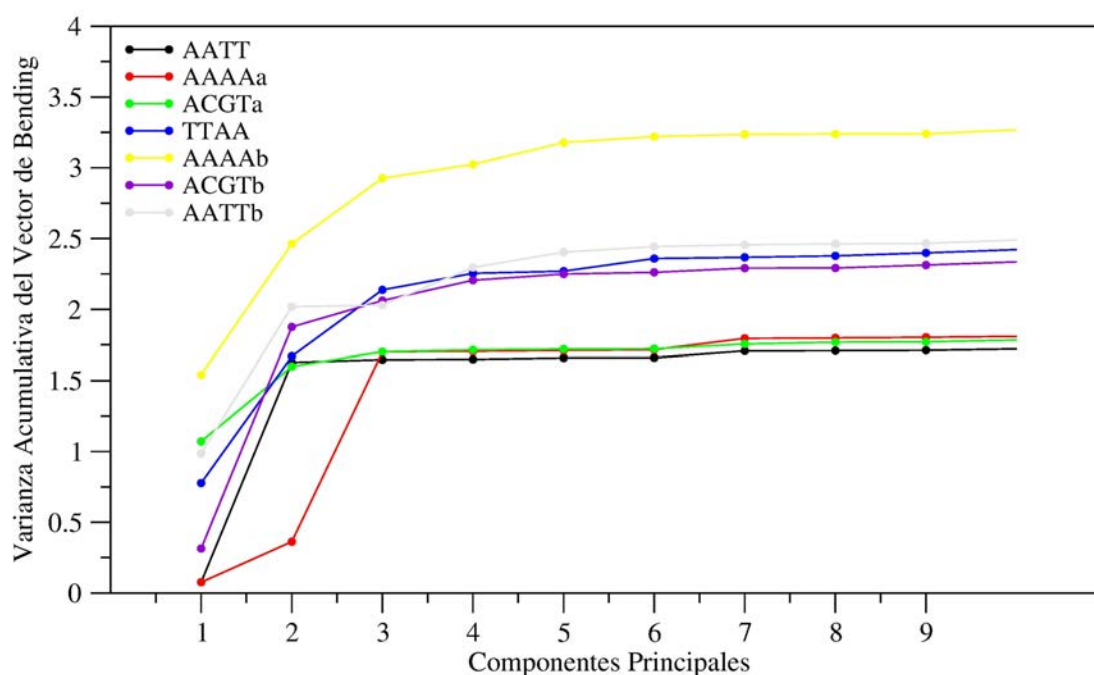
Si bien este análisis indica que el enlace de iones induce el mismo efecto en la dinámica del ADN para todas las secuencias, no provee información respecto del tipo de cambio conformacional que sufre la doble hélice. De acuerdo a nuestra hipótesis, el cambio estructural que inducen los iones debería plegar la doble hélice centrada en la región del "spacer" y en la dirección de unión de la proteína. En consecuencia, a continuación se analiza el contenido de "bending" en los primeros diez componentes principales.



**Figura 3-11. Matrices de producto interno.** Las matrices representan el producto interno de los diez primeros PCs de las secuencias AAAA y ACGT en presencia y ausencia de iones (AAAAa, ACGTa y AAAAb, ACGTb, respectivamente) contra las diez primeros PCs de las secuencias AATT y TTAA. La escala de colores mide la similitud entre PCs, desde valores cercanos a cero (blanco) hasta valores cercanos a 1 (negro).

Con este fin, es necesario considerar que el "bending" en un segmento de ADN puede estar localizado en un step específico o estar distribuido entre varios "steps". Por lo tanto, es necesaria una medida que condense el "bending" en el conjunto de "steps" de la región del "spacer" y alrededor. Por otro lado, el "bending" que es de interés en el presente trabajo, no es una medida escalar sino vectorial, debido a que no sólo importa si existe o no curvatura en el ADN, sino que es igual de importante la dirección de dicha curvatura. Concretamente, es necesario saber si el "bending" es en la dirección de la unión de la proteína. Una medida del "bending" que contiene tanto la información de la magnitud como de la dirección, es el vector asociado al "bending" (VAB) que se presentó en la sección Métodos. En base a dicha medida, puede extraerse información del "bending" en varios sentidos. El promedio del módulo del VAB da información del grado de "bending", la varianza en un componente principal dado indica si dicho componente involucra o no un movimiento de "bending" activo. Además, el ángulo entre el VAB y los ejes cartesianos del sistema permiten caracterizar la dirección del "bending", tanto del punto de vista de la dirección global, como de la variabilidad de la misma. Este conjunto de medidas aplicadas sobre los componentes principales más importantes de las cuatro secuencias, en presencia o ausencia de iones, permite conocer claramente la respuesta de las diferentes secuencias a la interacción con los iones en el marco de la unión a la proteína E2.

En la Figura 3-12 se muestra la varianza acumulativa del módulo del VAB en función de los componentes principales.



**Figura 3-12. Varianza del movimiento de "bending" para cada PC.** La varianza asociada al movimiento de "bending" (modulo del VAB) se midió para cada uno de los diez primeros PC de todas las secuencias, en ausencia y presencia de iones en el surco menor. (negro, rojo y verde: AATT, AAAA y ACGT con iones ligados, respectivamente; azul, amarillo, violeta y gris: TTAA, AAAA, ACGT y AATT sin iones ligados, respectivamente).

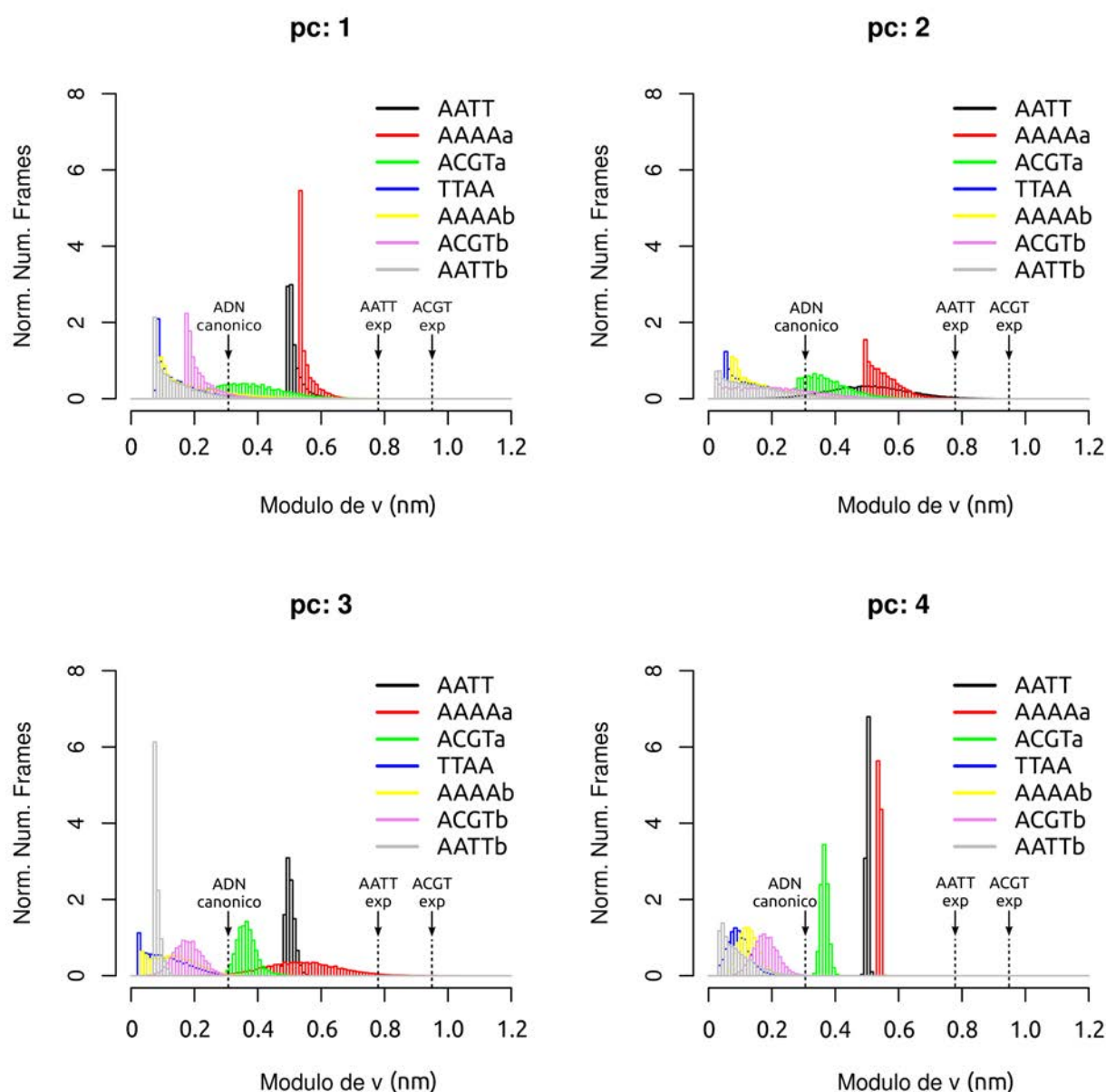
En primer lugar, cabe notar que el contenido de varianza de "bending" se acumula principalmente en los primeros cuatro PCs. Más allá del cuarto PC ya no se observa prácticamente movimiento de "bending" y el valor de varianza al cual convergen todas las secuencias es principalmente el reflejo del comportamiento de los primeros cuatro PCs. En otras palabras, solo los modos principales de movimiento involucran el "bending" del ADN en la región del "spacer". Cuando se considera el valor al cual converge la varianza, se puede observar que las secuencias se separan en tres grupos (Figura 3-12). El de más baja varianza involucra los casos AATTa, AAAAa y ACGTa. Con valores mayores de varianza se encuentran las secuencias TTAA (que prácticamente no liga iones durante toda la simulación), AATT y ACGTb. Por último, con un valor de varianza acumulada aún mayor se observa la secuencia AAAAb en ausencia de cationes. Este resultado sugiere que la unión de los cationes al surco menor suprime parcialmente el movimiento de "bending", induciendo una curvatura semi-permanente en el ADN. Es decir que la unión de iones reduciría la flexibilidad del ADN, previo a la unión de la proteína, pagando el costo entrópico asociado a la pérdida de grados de libertad ocasionados por la unión de la proteína. En este sentido, la capacidad del ADN de unir establemente cationes al surco menor, y responder conformacionalmente, podría ser importante en la termodinámica del proceso de unión proteína-ADN.

Para complementar la visión aportada por la varianza del VAB, es necesario conocer el modulo de tal vector como medida de la curvatura efectiva. Por tal motivo se calculó el modulo del VAB para los primeros cuatro modos, que son los que acumulan las principales respuestas del "bending" a la unión de los iones. La Figura 3-13, muestra las distribuciones del modulo del VAB para cada secuencia en presencia o ausencia de iones en el surco mayor.

De la observación de los cuatro PC se desprende que las secuencias pueden clasificarse mayormente en tres categorías respecto del modulo del VAB. Dichas categorías corresponden a módulos de VAB menores, similares o mayores al valor del ADN en conformación canónica ( $VAB=0.3$  nm).

Las secuencias AATTb, AAAAb y ACGTb (sin iones ligados) y TTAA que presenta muy baja frecuencia de unión de iones pertenecen a la categoría de valores de VAB menores al valor canónico ( $\sim 0.2$  nm). Considerando que los valores de VAB observados experimentalmente son mayores que el canónico ( $\sim 0.8$  nm y  $\sim 0.9$  nm para AATT y ACGT, respectivamente), valores menores a este implican un "bending" en dirección del surco mayor. Por lo tanto, en ausencia de iones todas las secuencias tienen una cierta tendencia a estar curvadas hacia el surco mayor, aunque en diferente medida según la secuencia. Por su lado la secuencia ACGT en presencia de iones presenta un modulo de VAB apenas mayor al canónico, mientras que AATT y AAAA al ligar iones en el surco

menor presentan valores de VAB significativamente mayores.



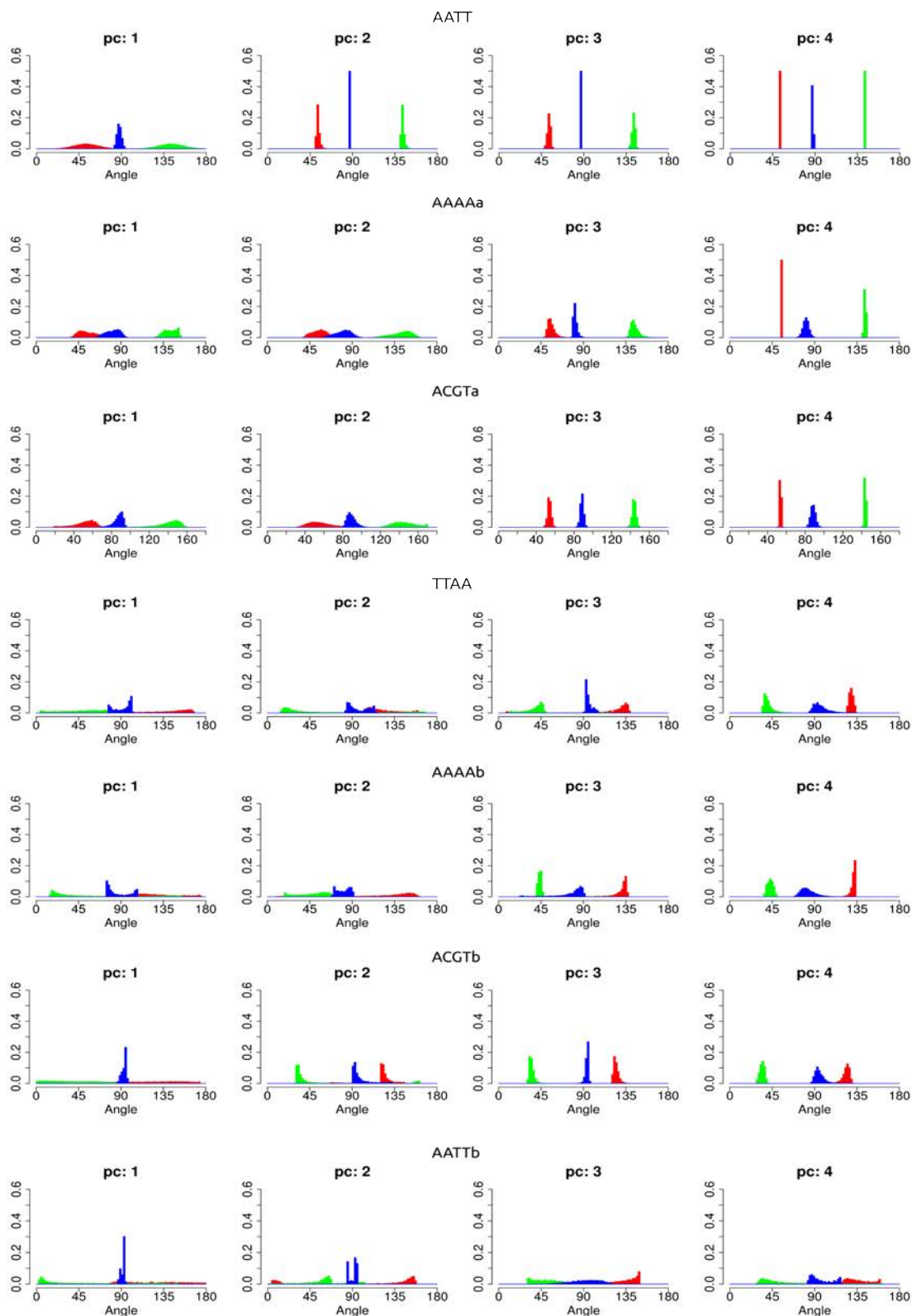
**Figura 3-13. Modulo del vector asociado al "bending".** Las distribuciones del modulo del vector asociado al "bending" se grafican para todas las secuencias en presencia y ausencia de iones en el surco menor (negro, rojo y verde: AATT, AAAA y ACGT en presencia de iones, respectivamente; azul, amarillo, violeta y gris: TTAA, AAAA, ACGT y AATT en ausencia de iones, respectivamente) y para los primeros cuatro componentes principales. Las flechas en el gráfico correspondiente al PC1 indican los módulos de VAB medidos en los complejos proteína-ADN experimentales y en el ADN canónico.

Sin embargo, ninguno de los valores obtenidos alcanza el grado de "bending" observado experimentalmente en

presencia de la proteína (ver flechas en PC 1 de la Figura 3-13). En consecuencia, la unión de iones induce el "bending" del ADN en la dirección de unión de la proteína, siendo este efecto más importante para las secuencias AATT y AAAA que para ACGT. Esto está de acuerdo con datos experimentales y teóricos<sup>15</sup> que indican que el efecto de la neutralización asimétrica de la carga de los fosfatos de la doble hélice es un factor necesario, pero no suficiente para explicar la deformación que ocurre en el ADN, y que el resto del trabajo es realizado por interacciones directas entre la proteína y las bases del ADN.

Por otro lado, se analizó la dirección del "bending" midiendo el ángulo entre el VAB y los ejes cartesianos. La Figura 3-14 muestra dicha medida para los primeros cuatro PCs. Debido a la orientación del ADN utilizada para realizar las medidas, el eje principal de la doble hélice está aproximadamente alineado con el eje z, por lo cual la dirección del "bending" puede caracterizarse en el plano x-y.

Dos características se observan claramente cuando se comparan las distribuciones de ángulos respecto de los ejes x e y para las diferentes secuencias en presencia o ausencia de iones. La primera es la disminución de la varianza de dichos ángulos al ligarse los iones (Figura 3-14 y Figura 3-15). Esto es indicativo de que dicha unión induce cambios en el ADN que tienden a reducir su movilidad, generando una curvatura con dirección más definida. La segunda característica importante es que las secuencias AATT, AAAA y ACGT en presencia de iones muestran valores menores a 90 para el ángulo respecto del eje x y valores mayores a 90 respecto del eje y. Notablemente, en ausencia de iones dichos ángulos se invierten (Figura 3-14). Esto es una clara indicación de que la unión de los iones cambia la dirección del "bending" de la doble hélice, verificando lo observado en base a los módulos del VAB. (Figura 3-14)



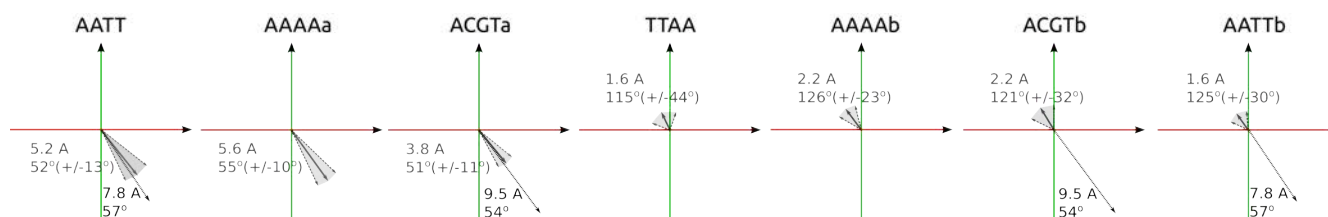
**Figura 3-14. Orientación espacial del VAB.** Para todas las secuencias estudiadas en presencia o ausencia de iones, considerando los primeros cuatro PCs, se muestran las distribuciones de los ángulos entre el VAB y los tres ejes cartesianos (x: rojo, y: verde, z: azul).

La información respecto del módulo del VAB así como de la dirección de éste respecto de las coordenadas cartesianas, se presenta de manera condensada en la Figura 3-15. En la misma se pueden observar claramente las diferencias de módulo y dirección de los VAB. Como se podía prever de los valores de módulo de VAB, la dirección del "bending" cuando no hay iones ligados al surco menor, es en la dirección opuesta a la unión de la proteína, mientras que al ligarse los iones, el "bending" cambia de dirección apuntando en la dirección de interacción con la proteína. A su vez puede apreciarse el efecto de los iones sobre la variabilidad de la dirección, reduciéndose esta a la mitad o al tercio cuando los iones están presentes.

Si se considera el conjunto de resultados mencionados en los párrafos anteriores, surge con claridad la respuesta del ADN frente a la unión de los iones. Se induce un aumento del "bending" en dirección de la proteína, acompañado de pérdida de flexibilidad en dicho movimiento así como un aumento de la direccionalidad del mismo. Los efectos sobre la dirección del "bending" son compartidos por todas las secuencias, sin embargo el efecto de los iones sobre la magnitud del "bending" es más importante en las secuencias AATT y AAAA, mientras que en el caso de ACGT el incremento en la magnitud del "bending" apenas supera el valor canónico del ADN. En este sentido, a pesar que la secuencia ACGT presenta capacidad de ligar iones en el surco menor y que este enlace sucede por intervalos de tiempo en la escala de los microsegundos, los cambios conformacionales que sufre, a pesar de ir en la dirección de la formación del complejo proteína-ADN, no son tan marcados como los que se observan para AATT y AAAA. Por otro lado, a pesar que la secuencia AAAA presenta valores de "bending" similares a AATT, muestra una reducción de la varianza del módulo del VAB significativamente mayor a la de AATT al unirse los iones. Lo cual sería indicativo de una mayor tendencia a generar conformaciones del ADN parcialmente curvadas y que favorezcan la unión de la proteína E2. Sin embargo AATT presenta una tendencia a ligar iones marcadamente mayor que AAAA. Lo cual explicaría por que la proteína E2 presenta mayor afinidad por la primera.

Estos datos permiten explicar el orden de afinidades por la proteína E2  $AATT > AAAA > ACGT$ , aunque resulta difícil compatibilizar los resultados obtenidos para la secuencia TTAA. Sin embargo, cabe notar que la menor tendencia a unir iones relativa al resto de las secuencias es coincidente con observaciones experimentales. Estas que indican que el "step" TpA, se comporta de manera diferente a secuencias ricas en AT. Mientras que estas últimas presentan fuerte tendencia a tener un surco menor más angosto incrementando su capacidad de atraer carga positiva, el "step" TpA presenta un comportamiento diverso.<sup>31</sup> El "step" TpA muestra particular facilidad de favorecer conformaciones curvadas de la doble hélice, haciendo las veces de "bisagra" molecular. Por lo cual es posible especular que la secuencia TTAA sea particularmente sensible a la interacción con la proteína requiriendo un costo energético de "bending" menor al de otras secuencias, al igual que se observa para la interacción entre

EcoRV y su sitio de unión (GATATC). Esto sugiere que en el caso TTAA, otros factores como la flexibilidad intrínseca del ADN pueden ser dominantes para determinar su enlace preferencial a E2 por sobre ACGT.

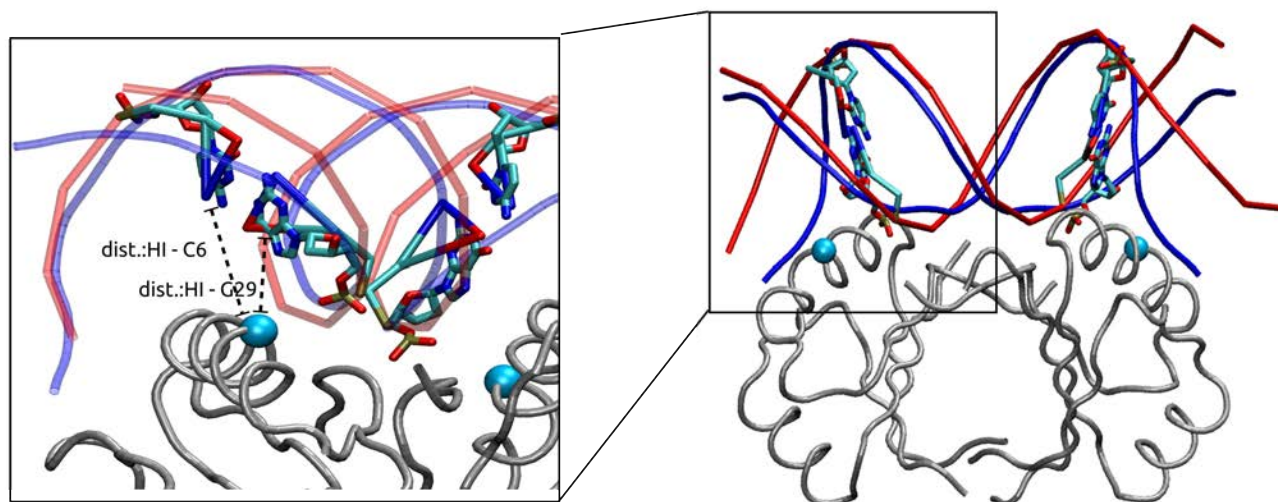


**Figura 3-15. Resumen del efecto de los iones sobre el VAB.** Los vectores mostrados presentan tienen como módulo, el correspondiente al del VAB y la dirección que indican es la correspondiente a los ángulos respecto de los ejes x e y. Las áreas grises son indicativas de la varianza de tales ángulos. Los valores numéricos de tales cantidades se muestran junto a los vectores correspondientes. Adicionalmente a los valores para cada secuencia en presencia o ausencia de iones en el surco menor, se muestran los datos correspondientes a las estructuras del complejo E2-ADN, para las secuencias de "spacer" de las que se disponen datos experimentales (AATT: PDBID=1JJ4, ACGT: PDBID=2BOP)

En tal sentido, para explicar completamente las diferencias de afinidad, sería necesario introducir el efecto explícito de la proteína. Desafortunadamente, tal información sólo está disponible para las secuencias AATT y ACGT, que son las únicas para la que existen estructuras resueltas experimentalmente del complejo proteína-ADN. Por lo tanto, en los siguiente párrafos el análisis se restringe a estos dos casos, comparando la similitud del comportamiento observado en la simulación de ADN libre con las estructuras cristalográficas de los correspondientes complejos proteína-ADN.

Para completar la descripción de los cambios conformacionales que se observan en el ADN al ligarse cationes en el surco menor, se relevaron tres propiedades adicionales, el RMSD de la región del "spacer" del ADN a lo largo de la simulación, tomando las estructuras resueltas experimentalmente como referencia para el alineamiento estructural, el ancho del surco menor, y la distancia entre los sitios que interaccionan específicamente en la proteína y el ADN (ver Figura 3-16). Dichas medidas fueron realizadas para ambas secuencias en función del tiempo de simulación, directamente desde las trayectorias en el espacio de coordenadas cartesianas (Figura 3-17).

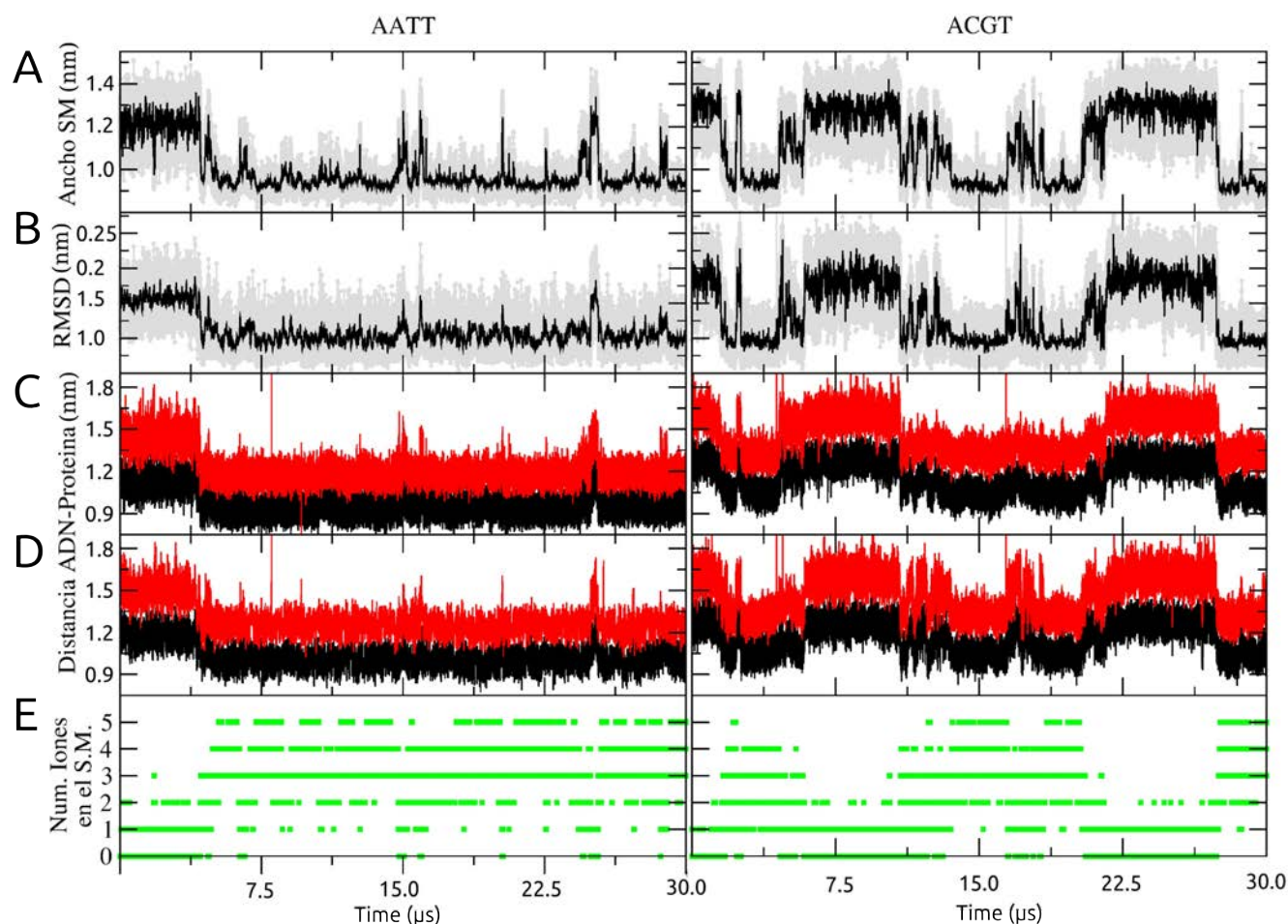




**Figura 3-16. Distancia HI-surco mayor.** A la derecha se muestra la estructura experimental del complejo E2-ADN para la secuencia de "spacer" AATT (gris y azul) junto con una conformación de una hebra de ADN con la misma secuencia obtenida durante la simulación de DM CG. Se resaltan en el ADN un par de bases del sitio de reconocimiento específico de E2 (**ACCG**, el punto indica la posición de dicho par de bases), y en la proteína el C $\alpha$  de la F301. La distancia entre dichos elementos (izquierda) se utilizó para caracterizar la respuesta del ADN a la unión de iones y su posible rol en la unión de la E2 a su blanco.

El perfil de ancho del surco menor para ambas secuencias muestra el mismo tipo de comportamiento observado en el Capítulo II para el dodecámero de Drew-Dickerson. Es decir, se observa que la unión de iones al surco menor se correlaciona con una reducción notable del ancho del mismo (coeficiente de correlación  $\sim 0.9$ ). Más aun, cuando ocurre la unión simultánea de tres o más iones, se produce un cambio conformacional estable y reversible del ADN, con un ancho de surco mayor que alcanza  $\sim 0.9$  nm, lo cual se corresponde bien con los valores medidos a partir de las estructuras experimentales (ver Tabla 3-1). Conformaciones con un ancho de surco menor intermedio se correlacionan con momentos de la simulación en los que ocurre liberación de iones del surco (Figura 3-17.A,E).

Una medida más precisa de los cambios estructurales durante la simulación la provee el RMSD de las posiciones de los grupos fosfato del ADN (descartando previamente los movimientos de traslación y rotación). Como puede observarse en la Figura 3-17.B,E, los momentos en los que se observa unión de iones, se corresponden con regiones donde el RMSD disminuye. Esto muestra que la unión de iones se puede correlacionar con cambios conformacionales en el ADN que generan un "ensemble" de estructuras en la escala de los microsegundos que se parecen a la observada en el complejo proteína-ADN. Por lo tanto, este resultado agrega evidencia a favor de que los iones son capaces de generar estructuras intermedias en el proceso de unión proteína-ADN, que favorecerían la interacción, al encontrarse previamente curvadas, reduciendo el costo energético del "bending" a



**Figura 3-17. Correlación entre unión de iones y cambios conformacionales en el ADN.** A, B, C y D muestran la respuesta del ADN (ancho del surco menor, RMSD de la región del "spacer", distancia HI-surco mayor y distancia HII-surco mayor, respectivamente) frente a la unión de iones  $\text{Na}^+$  al surco menor en la región del "spacer" (E).

la proteína. Para confirmar esta idea, se midió adicionalmente la distancia entre los sitios de contacto del ADN y la proteína, necesarios para que ocurra la interacción por lectura directa. Nuevamente se observa (Figura 3-17.C,D,E) que la unión de iones se correlaciona muy bien con reducciones en dichas distancias, llegando a alcanzar los valores observados en el complejo proteína-ADN experimental (HI-5'CpCpG3': 1 nm - 1.2 nm).

Al comparar entre las secuencias, se observa que la principal diferencia se encuentra en la dinámica de unión de iones, siendo ésta notablemente más estable en el caso de AATT. Adicionalmente, se observa que, cuando ocurre la unión de iones, la secuencia AATT presenta distancias proteína-ADN mas cercanas al valor experimental que la secuencia ACGT. Esto está de acuerdo con la mayor respuesta del modulo del VAB en el caso de AATT en comparación con ACGT.

Por lo tanto, la diferente dinámica de unión de iones al surco menor y las diferencias en la respuesta mecánica

observada para estas dos secuencias permiten confirmar la importancia de los iones en las diferencias de afinidad por la proteína E2.

Estos resultados terminan de afirmar el papel de la unión de cationes al surco menor, aportando evidencia a favor de la hipótesis del presente capítulo, es decir, que tal unión es capaz de favorecer la unión de proteínas al ADN generando conformaciones intermediarias del proceso de interacción proteína-ADN.

## Conclusiones

Se aplicó el modelo de solvatación explícita CG (agua e iones) presentado en el Capítulo II, al estudio de la dinámica de secuencias de ADN que son sitios de unión de la proteína E2. Tal aplicación permitió obtener información relevante respecto de la interacción iones-ADN, en escalas de tiempo inaccesibles para simulaciones clásicas a nivel atómico.

Las simulaciones realizadas confirmaron las observaciones del Capítulo II respecto de la correlación entre la unión de iones al surco menor de la doble hélice y la reducción del ancho del surco menor. Adicionalmente se encontró que a la vez que se reduce el tamaño del surco menor, se induce un cambio conformacional en el ADN que va en la dirección del proceso de unión de la proteína al ADN.

El análisis de componentes principales permitió observar que los cambios en la dinámica global del ADN al unirse los iones es común a todas las secuencias. Más aún, se observó que dicho cambio implica una reducción de la varianza asociada al movimiento de "bending" junto con la inducción de un ensemble de estados del ADN con su eje curvado, en la misma dirección que se observa experimentalmente. Este resultado es coincidente con la observación de estados de "pre-bending" reportados en la literatura para la secuencia AATT,<sup>81</sup> pero la unión de los iones al surco menor sería un elemento fundamental en la generación de tales estados.

En conjunto, los resultados presentados aportan evidencia a favor de la hipótesis de trabajo, es decir, los iones estarían jugando un papel importante en la termodinámica del proceso de unión proteína-ADN, reduciendo el costo entrópico necesario para generar estructuras del ADN curvadas. En tal sentido, ambos factores estudiados, la dinámica de unión de iones y la respuesta mecánica del ADN frente a dicha unión, aparecen como elementos relevantes en la determinación del costo asociado a la interacción proteína-ADN, permitiendo explicar algunas diferencias de afinidad.

Las diferencias de afinidad observadas experimentalmente para las secuencias AATT, AAAA, TTAA y ACGT, encuentran explicación parcial en el rol del entorno electrolítico. Concretamente, AATT presenta una muy fuerte capacidad de ligar iones en el surco menor de manera estable en la escala de las decenas de microsegundos, a la

vez que es capaz de responder curvándose en la dirección correcta, y con una reducción de la flexibilidad importante. Por su lado AAAA muestra también una capacidad de respuesta similar a la de AATT, pero una menor avidéz por los iones, disminuyendo putativamente la frecuencia y estabilidad de estados intermedios con curvatura previos a la unión de la E2. La secuencia ACGT presenta una dinámica de unión de iones similar a AAAA pero una respuesta menos marcada (como se evidencia en su módulo de VAB), posiblemente relacionado con una localización de iones fuera de la región del "spacer". Los resultados encontrados para la secuencia TTAA son más difíciles de reconciliar con los valores experimentales de afinidad. Sin embargo, debido a la mayor flexibilidad del "step" TpA, la localización de iones podría ser menos relevante para inducir curvatura facilitando conformaciones favorables para la formación del complejo proteína-ADN.

En resumen, la capacidad de ligar cationes en el surco menor y la respuesta del ADN frente a estos, son dependientes de la secuencia. Esta dependencia puede jugar, junto con la flexibilidad intrínseca, un papel importante en la afinidad y selectividad de las proteínas por el ADN. Tomando como marco de referencia un modelo de interacción proteína-ADN en el cual la pérdida de flexibilidad favorece la unión (observado experimentalmente para la interacción E2-ADN en las cepas virales HPV16/18.<sup>81</sup>), el papel de los iones sería el de reducir el costo entrópico asociado a dicha pérdida de flexibilidad, disminuyendo el trabajo requerido por la proteína para generar el complejo final con el ADN.

## Bibliografía

1. Schultz, S. C., Shields, G. C. & Steitz, T. A. Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. *Science* **253**, 1001–1007 (1991).
2. Kim, J. L., Nikolov, D. B. & Burley, S. K. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365**, 520–527 (1993).
3. Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, 512–520 (1993).
4. Goosen, N. & Putte, P. The regulation of transcription initiation by integration host factor. *Molecular Microbiology* **16**, 1–7 (1995).
5. Kalodimos, C. G. *et al.* Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. *Science* **305**, 386–389 (2004).
6. Love, J. J. *et al.* Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature* **376**, 791–795 (1995).
7. Sam, M. D., Cascio, D., Johnson, R. C. & Clubb, R. T. Crystal structure of the excisionase-DNA complex from bacteriophage lambda. *J. Mol. Biol.* **338**, 229–240 (2004).
8. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
9. Luijsterburg, M. S., Noom, M. C., Wuite, G. J. L. & Dame, R. T. The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: a molecular perspective. *J. Struct. Biol.* **156**, 262–272 (2006).
10. Horton, N. C. & Perona, J. J. Crystallographic snapshots along a protein-induced DNA-bending pathway. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5729–5734 (2000).

11. Winkler, F. K. *et al.* The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments. *EMBO J.* **12**, 1781–1795 (1993).
12. Müller, B. C., Raphael, A. L. & Barton, J. K. Evidence for altered DNA conformations in the simian virus 40 genome: site-specific DNA cleavage by the chiral complex lambda-tris(4,7-diphenyl-1,10-phenanthroline)cobalt(III). *Proc. Natl. Acad. Sci. U.S.A.* **84**, 1764–1768 (1987).
13. Bloomfield, V. A. DNA condensation by multivalent cations. *Biopolymers* **44**, 269–282 (1997).
14. Rippe, K., von Hippel, P. H. & Langowski, J. Action at a distance: DNA-looping and initiation of transcription. *Trends Biochem. Sci.* **20**, 500–506 (1995).
15. Range, K. The contribution of phosphate-phosphate repulsions to the free energy of DNA bending. *Nucleic Acids Research* **33**, 1257–1268 (2005).
16. Shui, X., Sines, C. C., McFail-Isom, L., VanDerveer, D. & Williams, L. D. Structure of the potassium form of CGCGAATTCGCG: DNA deformation by electrostatic collapse around inorganic cations. *Biochemistry* **37**, 16877–16887 (1998).
17. Bustamante, C., Bryant, Z. & Smith, S. B. Ten years of tension: single-molecule DNA mechanics. *Nature* **421**, 423–427 (2003).
18. Baumann, C. G. Ionic effects on the elasticity of single DNA molecules. *Proceedings of the National Academy of Sciences* **94**, 6185–6190 (1997).
19. Cairney, K. L. & Harrington, R. E. Flow birefringence of T7 phage DNA: Dependence on salt concentration. *Biopolymers* **21**, 923–934 (1982).
20. Porschke, D. Persistence length and bending dynamics of DNA from electrooptical measurements at high salt concentrations. *Biophys. Chem.* **40**, 169–179 (1991).
21. Sobel, E. S. & Harpst, J. A. Effects of Na<sup>+</sup> on the persistence length and excluded volume of T7 bacteriophage DNA. *Biopolymers* **31**, 1559–1564 (1991).
22. Hagerman, P. J. Flexibility of DNA. *Annual Review of Biophysics and Biophysical Chemistry* **17**, 265–286 (1988).
23. Hagerman, K. R. & Hagerman, P. J. Helix rigidity of DNA: the meroduplex as an experimental paradigm. *J. Mol. Biol.* **260**, 207–223 (1996).
24. Manning, G. S. Three persistence lengths for a stiff polymer with an application to DNA B-Z junctions. *Biopolymers* **27**, 1529–1542 (1988).
25. Manning, G. S., Ebralidse, K. K., Mirzabekov, A. D. & Rich, A. An Estimate of the Extent of Folding of Nucleosomal DNA by Laterally Asymmetric Neutralization of Phosphate Groups. *Journal of Biomolecular Structure and Dynamics* **6**, 877–889 (1989).
26. Mirzabekov, A. D. & Rich, A. Asymmetric lateral distribution of unshielded phosphate groups in nucleosomal DNA and its role in DNA bending. *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1118–1121 (1979).
27. Cloutier, T. E. & Widom, J. Spontaneous sharp bending of double-stranded DNA. *Mol. Cell* **14**, 355–362 (2004).
28. Wiggins, P. A. *et al.* High flexibility of DNA on short length scales probed by atomic force microscopy. *Nature Nanotechnology* **1**, 137–141 (2006).
29. Gorin, A. A., Zhurkin, V. B. & Olson, W. K. B-DNA twisting correlates with base-pair morphology. *J. Mol. Biol.* **247**, 34–48 (1995).
30. Olson, W. K., Gorin, A. A., Lu, X.-J., Hock, L. M. & Zhurkin, V. B. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proceedings of the National Academy of Sciences* **95**, 11163–11168 (1998).
31. Rohs, R. *et al.* The role of DNA shape in protein–DNA recognition. *Nature* **461**, 1248–1253 (2009).
32. Pérez, A., Luque, F. J. & Orozco, M. Dynamics of B-DNA on the Microsecond Time Scale. *Journal of the American Chemical Society* **129**, 14739–14745 (2007).
33. Parker, S. C. J., Hansen, L., Abaan, H. O., Tullius, T. D. & Margulies, E. H. Local DNA Topography Correlates with Functional Noncoding Regions of the Human Genome. *Science* **324**, 389–392 (2009).
34. Manning, G. S. The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q. Rev. Biophys.* **11**, 179–246 (1978).
35. Shui, X., McFail-Isom, L., Hu, G. G. & Williams, L. D. The B-DNA dodecamer at high resolution reveals a spine of water on sodium. *Biochemistry* **37**, 8341–8355 (1998).
36. Tereshko, V., Minasov, G. & Egli, M. A ‘Hydrat-Ion’ Spine in a B-DNA Minor Groove. *J. Am. Chem. Soc.* **121**, 3590–

- 3595 (1999).
37. Rosenberg, J. M. *et al.* Double Helix at Atomic Resolution. *Nature* **243**, 150–154 (1973).
  38. Kopka, M. L., Fratini, A. V., Drew, H. R. & Dickerson, R. E. Ordered water structure around a B-DNA dodecamer. A quantitative study. *J. Mol. Biol.* **163**, 129–146 (1983).
  39. Bartenev, V. N. *et al.* Structure of the B DNA cationic shell as revealed by an X-ray diffraction study of CsDNA. Sequence-specific cationic stabilization of B form DNA. *J. Mol. Biol.* **169**, 217–234 (1983).
  40. Hud, N. V. & Feigon, J. Localization of Divalent Metal Ions in the Minor Groove of DNA A-Tracts. *Journal of the American Chemical Society* **119**, 5756–5757 (1997).
  41. Hud, N. V., Sklenár, V. & Feigon, J. Localization of ammonium ions in the minor groove of DNA duplexes in solution and the origin of DNA A-tract bending. *J. Mol. Biol.* **286**, 651–660 (1999).
  42. Young, M. A. & Beveridge, D. L. Molecular dynamics simulations of an oligonucleotide duplex with adenine tracts phased by a full helix turn. *J. Mol. Biol.* **281**, 675–687 (1998).
  43. Young, M. A., Jayaram, B. & Beveridge, D. L. Intrusion of Counterions into the Spine of Hydration in the Minor Groove of B-DNA: Fractional Occupancy of Electronegative Pockets. *Journal of the American Chemical Society* **119**, 59–69 (1997).
  44. Egli, M. DNA-cation interactions: quo vadis? *Chem. Biol.* **9**, 277–286 (2002).
  45. Lankas, F., Sponer, J., Langowski, J. & Cheatham, T. E., 3rd DNA basepair step deformability inferred from molecular dynamics simulations. *Biophys. J.* **85**, 2872–2883 (2003).
  46. Spiriti, J., Kamberaj, H., de Graff, A. M. R., Thorpe, M. F. & van der Vaart, A. DNA Bending through Large Angles Is Aided by Ionic Screening. *Journal of Chemical Theory and Computation* **8**, 2145–2156 (2012).
  47. Zeida, A., Machado, M., Dans, P. & Pantano, S. Breathing, bubbling, and bending: DNA flexibility from multimicrosecond simulations. *Physical Review E* **86**, (2012).
  48. Tchernachenko, V., Halvorson, H. R. & Lutter, L. C. Topological measurement of an A-tract bend angle: effect of magnesium. *J. Mol. Biol.* **341**, 55–63 (2004).
  49. Zinkel, S. S. & Crothers, D. M. Comparative gel electrophoresis measurement of the DNA bend angle induced by the catabolite activator protein. *Biopolymers* **29**, 29–38 (1990).
  50. Griffith, J. D., Makhov, A., Zawel, L. & Reinberg, D. Visualization of TBP oligomers binding and bending the HIV-1 and adeno promoters. *J. Mol. Biol.* **246**, 576–584 (1995).
  51. Rice, P. Crystal Structure of an IHF-DNA Complex: A Protein-Induced DNA U-Turn. *Cell* **87**, 1295–1306 (1996).
  52. Maher, L. J., 3rd Mechanisms of DNA bending. *Curr Opin Chem Biol* **2**, 688–694 (1998).
  53. Pontiggia, A. *et al.* Sex-reversing mutations affect the architecture of SRY-DNA complexes. *EMBO J.* **13**, 6115–6124 (1994).
  54. Werner, M. H. & Burley, S. K. Architectural transcription factors: proteins that remodel DNA. *Cell* **88**, 733–736 (1997).
  55. Crothers, D. M. Architectural elements in nucleoprotein complexes. *Curr. Biol.* **3**, 675–676 (1993).
  56. Elcock, A. H. & McCammon, J. A. The Low Dielectric Interior of Proteins is Sufficient To Cause Major Structural Changes in DNA on Association. *Journal of the American Chemical Society* **118**, 3787–3788 (1996).
  57. Record, M. T., Jr, Anderson, C. F. & Lohman, T. M. Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity. *Q. Rev. Biophys.* **11**, 103–178 (1978).
  58. Record, M. T., Jr, Zhang, W. & Anderson, C. F. Analysis of effects of salts and uncharged solutes on protein and nucleic acid equilibria and processes: a practical guide to recognizing and interpreting polyelectrolyte effects, Hofmeister effects, and osmotic effects of salts. *Adv. Protein Chem.* **51**, 281–353 (1998).
  59. Manning, G. S. Electrostatic free energy of the DNA double helix in counterion condensation theory. *Biophys. Chem.* **101-102**, 461–473 (2002).
  60. Manning, G. S. Is a small number of charge neutralizations sufficient to bend nucleosome core DNA onto its superhelical ramp? *J. Am. Chem. Soc.* **125**, 15087–15092 (2003).
  61. Strauss, J. K. & Maher, L. J., 3rd DNA bending by asymmetric phosphate neutralization. *Science* **266**, 1829–1834 (1994).
  62. Kosikov, K. M., Gorin, A. A., Lu, X.-J., Olson, W. K. & Manning, G. S. Bending of DNA by Asymmetric Charge Neutralization: All-Atom Energy Simulations. *Journal of the American Chemical Society* **124**, 4838–4847 (2002).

63. Strauss-Soukup, J. K. & Maher, L. J., 3rd Electrostatic effects in DNA bending by GCN4 mutants. *Biochemistry* **37**, 1060–1066 (1998).
64. Strauss, J. K., Prakash, T. P., Roberts, C., Switzer, C. & Maher, L. J. DNA bending by a phantom protein. *Chem. Biol.* **3**, 671–678 (1996).
65. Williams, S. L. Changes in DNA bending and flexing due to tethered cations detected by fluorescence resonance energy transfer. *Nucleic Acids Research* **34**, 1028–1035 (2006).
66. Okonogi, T. M. Phosphate backbone neutralization increases duplex DNA flexibility: A model for protein binding. *Proceedings of the National Academy of Sciences* **99**, 4156–4160 (2002).
67. Strauss-Soukup, J. K. & Maher, L. J., 3rd Role of asymmetric phosphate neutralization in DNA bending by PU.1. *J. Biol. Chem.* **272**, 31570–31575 (1997).
68. Leonard, D. A., Rajaram, N. & Kerppola, T. K. Structural basis of DNA bending and oriented heterodimer binding by the basic leucine zipper domains of Fos and Jun. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 4913–4918 (1997).
69. Paolella, D. N., Liu, Y., Fabian, M. A. & Schepartz, A. Electrostatic mechanism for DNA bending by bZIP proteins. *Biochemistry* **36**, 10033–10038 (1997).
70. Strauss-Soukup, J. K. & Maher, L. J., 3rd DNA bending by GCN4 mutants bearing cationic residues. *Biochemistry* **36**, 10026–10032 (1997).
71. Tomky, L. A., Strauss-Soukup, J. K. & Maher, L. J., 3rd Effects of phosphate neutralization on the shape of the AP-1 transcription factor binding site in duplex DNA. *Nucleic Acids Res.* **26**, 2298–2305 (1998).
72. Hardwidge, P. R., Parkhurst, K. M., Parkhurst, L. J. & Maher, L. J. Reflections on apparent DNA bending by charge variants of bZIP proteins. *Biopolymers* **69**, 110–117 (2003).
73. Hardwidge, P. R. *et al.* DNA Bending by bZIP Charge Variants: A Unified Study Using Electrophoretic Phasing and Fluorescence Resonance Energy Transfer. *Biochemistry* **41**, 7732–7742 (2002).
74. Martin, A. M., Sam, M. D., Reich, N. O. & Perona, J. J. Structural and energetic origins of indirect readout in site-specific DNA cleavage by a restriction endonuclease. *Nat. Struct. Biol.* **6**, 269–277 (1999).
75. Mack, D. R., Chiu, T. K. & Dickerson, R. E. Intrinsic bending and deformability at the T-A step of CCTTTAAAGG: A comparative analysis of T-A and A-T steps within A-tracts. *Journal of Molecular Biology* **312**, 1037–1049 (2001).
76. Vipond, I. B. & Halford, S. E. Specific DNA recognition by EcoRV restriction endonuclease induced by calcium ions. *Biochemistry* **34**, 1113–1119 (1995).
77. Hiller, D. A. *et al.* Simultaneous DNA Binding and Bending by EcoRV Endonuclease Observed by Real-Time Fluorescence. *Biochemistry* **42**, 14375–14385 (2003).
78. Hiller, D. A. & Perona, J. J. Positively charged C-terminal subdomains of EcoRV endonuclease: contributions to DNA binding, bending, and cleavage. *Biochemistry* **45**, 11453–11463 (2006).
79. Hancock, S. P., Hiller, D. A., Perona, J. J. & Jen-Jacobson, L. The Energetic Contribution of Induced Electrostatic Asymmetry to DNA Bending by a Site-Specific Protein. *Journal of Molecular Biology* **406**, 285–312 (2011).
80. Vivas, P., Velmurugu, Y., Kuznetsov, S. V., Rice, P. A. & Ansari, A. Mapping the Transition State for DNA Bending by IHF. *Journal of Molecular Biology* **418**, 300–315 (2012).
81. Hegde, R. S. The papillomavirus E2 proteins: structure, function, and biology. *Annu Rev Biophys Biomol Struct* **31**, 343–360 (2002).
82. Hegde, R. S. & Androphy, E. J. Crystal structure of the E2 DNA-binding domain from human papillomavirus type 16: implications for its DNA binding-site selection mechanism. *J. Mol. Biol.* **284**, 1479–1489 (1998).
83. Hegde, R. S., Wang, A. F., Kim, S. S. & Schapira, M. Subunit rearrangement accompanies sequence-specific DNA binding by the bovine papillomavirus-1 E2 protein. *J. Mol. Biol.* **276**, 797–808 (1998).
84. Hizver, J. DNA bending by an adenine-thymine tract and its role in gene regulation. *Proceedings of the National Academy of Sciences* **98**, 8490–8495 (2001).
85. Kim, S. S., Tam, J. K., Wang, A. F. & Hegde, R. S. The structural basis of DNA target discrimination by papillomavirus E2 proteins. *J. Biol. Chem.* **275**, 31245–31254 (2000).
86. Rozenberg, H., Rabinovich, D., Frolow, F., Hegde, R. S. & Shakked, Z. Structural code for DNA recognition revealed in crystal structures of papillomavirus E2-DNA targets. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 15194–15199 (1998).
87. Hines, C. S. *et al.* DNA structure and flexibility in the sequence-specific binding of papillomavirus E2 proteins. *J. Mol. Biol.* **276**, 809–818 (1998).

88. Alexander, K. A. & Phelps, W. C. A fluorescence anisotropy study of DNA binding by HPV-11 E2C protein: a hierarchy of E2-binding sites. *Biochemistry* **35**, 9864–9872 (1996).
89. Ferreiro, D. U., Sánchez, I. E. & de Prat Gay, G. Transition state for protein-DNA recognition. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10797–10802 (2008).
90. Hooley, E., Fairweather, V., Clarke, A. R., Gaston, K. & Brady, R. L. The recognition of local DNA conformation by the human papillomavirus type 6 E2 protein. *Nucleic Acids Res.* **34**, 3897–3908 (2006).
91. Cicero, D. O. *et al.* Structural and thermodynamic basis for the enhanced transcriptional control by the human papillomavirus strain-16 E2 protein. *Biochemistry* **45**, 6551–6560 (2006).
92. Thain, A., Webster, K., Emery, D., Clarke, A. R. & Gaston, K. DNA binding and bending by the human papillomavirus type 16 E2 protein. Recognition of an extended binding site. *J. Biol. Chem.* **272**, 8236–8242 (1997).
93. Falconi, M. *et al.* Molecular dynamics of the DNA-binding domain of the papillomavirus E2 transcriptional regulator uncover differential properties for DNA target accommodation. *FEBS J.* **274**, 2385–2395 (2007).
94. Sánchez, I. E., Dellarole, M., Gaston, K. & de Prat Gay, G. Comprehensive comparison of the interaction of the E2 master regulator with its cognate target DNA sites in 73 human papillomavirus types by sequence statistics. *Nucleic Acids Res.* **36**, 756–769 (2008).
95. Ferreiro, D. U. & de Prat-Gay, G. A protein-DNA binding mechanism proceeds through multi-state or two-state parallel pathways. *J. Mol. Biol.* **331**, 89–99 (2003).
96. Horton, J. R., Liebert, K., Hattman, S., Jeltsch, A. & Cheng, X. Transition from nonspecific to specific DNA interactions along the substrate-recognition pathway of dam methyltransferase. *Cell* **121**, 349–361 (2005).
97. Sanchez, I. E., Ferreiro, D. U., Dellarole, M. & de Prat-Gay, G. Experimental snapshots of a protein-DNA binding landscape. *Proceedings of the National Academy of Sciences* **107**, 7751–7756 (2010).
98. Dans, P. D., Zeida, A., Machado, M. R. & Pantano, S. A Coarse Grained Model for Atomic-Detailed DNA Simulations with Explicit Electrostatics. *Journal of Chemical Theory and Computation* **6**, 1711–1725 (2010).
99. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **81**, 3684 (1984).
100. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **98**, 10089 (1993).
101. Essmann, U. *et al.* A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **103**, 8577 (1995).
102. Zhang, Y., Xi, Z., Hegde, R. S., Shakked, Z. & Crothers, D. M. Predicting indirect readout effects in protein-DNA interactions. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8337–8341 (2004).



## Capítulo IV

# Desarrollo de un nuevo modelo de solvatación multiescala.

### Introducción

A pesar de que aproximaciones como la presentada en el Capítulo II logran efectivamente un aumento sensible en la velocidad de exploración de la superficie de energía potencial del sistema de estudio, la pérdida inherente de resolución puede comprometer su aplicación, dependiendo del problema que se intenta estudiar. Un ejemplo importante es el rol que juega la naturaleza discreta de las moléculas de agua en muchos procesos de interés biológico, en buena medida ligada a su capacidad de formación de enlaces de hidrógeno y de apantallar las interacciones electrostáticas. Por lo tanto, el estudio por simulaciones de DM de tales procesos requiere la inclusión explícita del agua a nivel FG. Sin embargo, el agua constituye ~ 80% de los sistemas biológicos, lo cual se traduce en un enorme costo computacional dedicado a las interacciones del agua en el "bulk", cuando el objetivo en estos casos es el comportamiento detallado solo del agua en una porción reducida del sistema. En este sentido, aproximaciones multiescala como las mencionadas al final del Capítulo I cobran relevancia. En particular, aquellas que permitan mantener una esfera de solvatación FG en el entorno de un soluto de interés, también FG, mientras que el agua "bulk" se modela a nivel CG. Entre las metodologías multiescala existentes, solo las de Resolución Adaptativa serían útiles en tales estudios. Sin embargo dichos métodos presentan la fuerte desventaja de estar limitados a modelos CG de agua con bajo nivel de CG, 1-1. Por lo cual, sólo pueden proveer de un modesto aumento en la eficiencia computacional.

En este capítulo, se presenta un nuevo modelo de resolución dual para solvatación acuosa, que puede clasificarse dentro del segundo grupo de métodos de multiescala (Resolución Fija). Este método se basa en el uso combinado del modelo de agua SPC,<sup>1</sup> popularmente utilizado por la comunidad de las simulaciones moleculares, junto con el modelo CG de agua presentado en el Capítulo II, denominado WT4.<sup>2</sup> Las interacciones en el modelo WT4 son calculadas utilizando el mismo Hamiltoniano y reglas de combinación utilizados en las simulaciones FG. Por lo tanto, si un número de moléculas de SPC y de WT4 se mezclan en una caja de simulación, es posible obtener una descripción simultánea de regiones FG y CG sin la necesidad de términos específicos para describir las interacciones FG-CG.

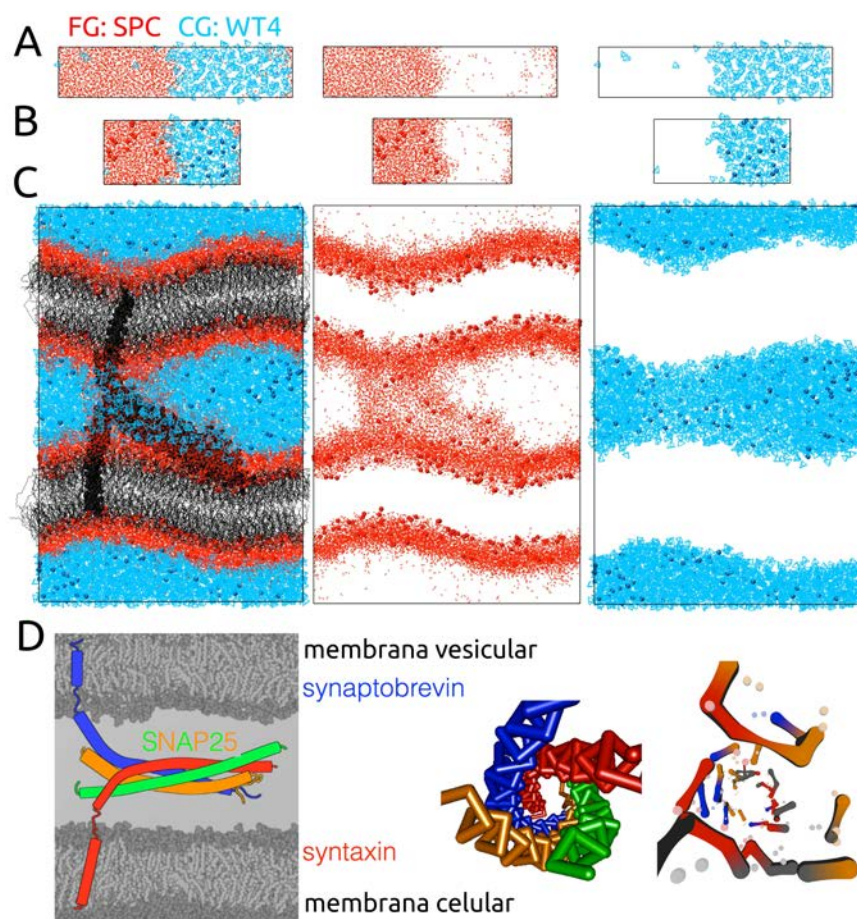
La idea que subyace al trabajo que se presenta en este capítulo, es que un soluto a nivel atómico ofrece sitios de

anclaje de agua cuando se encuentra completamente solvatado. Estos sitios modifican la estructura y dinámica de la esfera de hidratación inmediatamente vecinas al soluto.<sup>3,4</sup> Las moléculas de agua por fuera de dicha esfera recuperarían las características del "bulk". Esto está de acuerdo con la observación de que la dinámica del agua está predominantemente influenciada por la presencia de la primer y segunda esferas de solvatación.<sup>5</sup> Por lo tanto, sería plausible una aproximación en la que el agua "bulk" de un sistema fuese tratada a nivel CG. Para probar esta hipótesis, se analizaron diferentes situaciones involucrando la mezcla de los modelos de agua SPC y WT4. La atención será enfocada en las posibles perturbaciones que pudiesen generar las moléculas CG en la fase FG.

Con el objetivo de aplicar esta aproximación a modelos biológicos complejos, se analizaron dos sistemas acuosos simples para caracterizar las propiedades intrínsecas a los sistemas con resolución dual SPC-WT4, previa al estudio de un caso más realista que involucra modelos de membranas y proteínas (Figura 4-1). El primer sistema representa agua pura modelada como mezcla de moléculas de SPC y WT4 (Figura 4-1.A). El segundo, incorpora el modelado multiescala de electrolitos simples en solución (Figura 4-1.B). Finalmente, como test de relevancia biológica se usó un modelo de complejo proteico SNARE (acrónimo derivado del nombre en inglés: "Soluble Nsf Attachment protein REceptor"). Este modelo representa la maquinaria mínima de fusión de membranas (29).<sup>6</sup> Está compuesto por las proteínas syntaxin, synaptobrevin y SNAP-25 formando un arreglo tridimensional citoplasma de tipo "haz paralelo de cuatro hélices" (four helix bundle) unido a dos segmentos transmembrana que se insertan en una doble bicapa lipídica (Figura 4-1.C,D). El sistema fue previamente construido y equilibrado para generar un modelo inicial.<sup>7</sup> En el presente trabajo, se comparan las simulaciones FG y multiescala FG-CG, partiendo del modelo inicial mencionado. Un sistema como este, que presenta una complejidad considerable, provee de una prueba sensible y crítica para estudiar posibles efectos espurios que el tratamiento dual podría introducir. Como se muestra más adelante, no se observan diferencias significativas entre la simulación FG y la de resolución dual FG-CG, y las comparaciones entre éstas indican que la mayoría de las propiedades se conservan, dentro del margen de precisión estadística.

## Métodos

Varios sistemas acuosos fueron simulados para poner a prueba la aproximación de solvatación multiescala que se propone en el presente trabajo. La composición de los sistemas estudiados, las condiciones de simulación,



**Figura 4-1. Representación esquemática de los sistemas moleculares estudiados.** A- Sistema utilizado para estudiar las propiedades del agua pura. Los paneles central y de la derecha muestran las fases FG y CG separadamente. Los colores rojo y cyan se utilizan para indicar las partículas FG y CG, respectivamente. B- Representación del sistema empleado para estudiar las propiedades de las soluciones de electrolitos simples. C- Sistema de doble membrana y proteína utilizado como caso de prueba del modelo de solvación multiescala propuesto. En gris y negro se muestran los fosfolípidos y la proteína, respectivamente. D- Vista esquemática del complejo SNARE interactuando con las membranas vesicular y celular (izquierda). En el centro se muestra una vista en la dirección del eje de la región citoplasmática del complejo SNARE, mientras que a la derecha se muestran algunos puentes salinos que mantienen la integridad de dicha región.

temperaturas y propiedades calculadas se detallan en la Tabla 4-1.

Todas las simulaciones de dinámica molecular fueron llevadas a cabo usando el programa GROMACS 4.5 con un tamaño de paso de integración de las ecuaciones de movimiento de 2fs. El algoritmo LINCS<sup>8</sup> se aplicó para todos los enlaces que conectan átomos de hidrógeno, y el SETTLE<sup>9</sup> para el caso de las moléculas de agua SPC. La temperatura y presión se mantuvieron constantes usando el termostato y el baróstato de Berendsen,<sup>10</sup> con tiempos de acoplamiento de 0.1 y 0.5 ps, respectivamente. Para los sistemas 1 al 9, se utilizó una distancia de

**Tabla 4-1.** Lista de sistemas y condiciones de simulación.

sistema	componentes del sistema	concentración iónica (M)	temperatura (K), ensemble termodinámico, acoplamiento de presión	tiempo de simulación (ns)	propiedades calculadas
1 (FG)	SPC (4004) N. At.: 12012	...	300, 298, 318, NPT, isotrópico	5,1,1	densidad, factor de Kirkwood, tiempo de relajación, permitividad, coeficiente de difusión, compresibilidad isotérmica, capacidad calorífica, coeficiente de expansión
2 (FG-CG)	SPC (2002), WT4 (182) N. At.: 6734	...	300, 298, 318, NPT, isotrópico	5,1,1	densidad, factor de Kirkwood, tiempo de relajación, permitividad, coeficiente de difusión, compresibilidad isotérmica, capacidad calorífica, coeficiente de expansión
3 (FG-CG)	SPC (4565), WT4 (153) N. At.: 14307	...	300, NVT	3	tensión superficial
4 (FG)	SPC (2002), Na <sup>+</sup> (10) Cl <sup>-</sup> (10) N. At.: 6026	0.3	300, NPT, isotrópico	10	densidad numérica, RDF ion-agua, RDF catión-anión, número de coordinación del agua, coeficiente de difusión
5 (FG-CG)	SPC (2002), Na <sup>+</sup> (10) Cl <sup>-</sup> (10), WT4 (160), NaW <sup>+</sup> (10), ClW <sup>-</sup> (10) N. At.: 6686	0.3	300, NPT, isotrópico	10	densidad numérica, RDF ion-agua, RDF catión-anión, número de coordinación del agua, coeficiente de difusión
6 (FG)	SPC (7612), Na <sup>+</sup> (34), Cl <sup>-</sup> (34) N. At.: 22904	0.25	300, NVT	10	presión osmótica
7 (FG-CG)	SPC (7612), Na <sup>+</sup> (34), Cl <sup>-</sup> (34) WT4 (649), NaW <sup>+</sup> (34), ClW <sup>-</sup> (34) N. At.: 25568	0.25	300, NVT	10	presión osmótica
8 (FG)	proteína <sup>a</sup> , POPC (1008), POPS (123), SPC (80453), Na <sup>+</sup> (328), Cl <sup>-</sup> (198) N. At.: 304564	0.1 <sup>b</sup>	310, NPT, semi-isotrópico (plano de las membranas)	64 <sup>c</sup>	<b>sistema completo:</b> perfiles de densidad <b>membrana:</b> espesor, distancia intermembrana, área/lípido, ángulo PN, parámetro de orden <b>proteína:</b> RMSD, RMSF, superposición estructural, puentes salinos, enlaces de hidrógeno, mapa de contacto <b>proteína-membrana:</b> inserción de DTM (Dominio Trans-Membrana), contactos proteína-bicapa
9 (FG-CG)	proteína <sup>a</sup> , POPC (1008), POPS (123), SPC (28973), Na <sup>+</sup> (215), Cl <sup>-</sup> (68), WT4 (4680), NaW <sup>+</sup> (113), ClW <sup>-</sup> (130) N. At.: 168844	0.1 <sup>b</sup>	310, NPT, semi-isotrópico (plano de las membranas)	20 <sup>c</sup>	<b>sistema completo:</b> perfiles de densidad <b>membrana:</b> espesor, distancia intermembrana, área/lípido, ángulo PN, parámetro de orden <b>proteína:</b> RMSD, RMSF, superposición estructural, puentes salinos, enlaces de hidrógeno, mapa de contacto <b>proteína-membrana:</b> inserción de DTM, contactos proteína-bicapa

<sup>a</sup> Cadena A: Synaptobrevina (sb, aa: 1-92). Cadena B: Syntaxina (sx, aa: 93-195). Cadena C: SNAP25 (aa: 196-346). Notar que los números de residuos corresponden con los utilizados en las simulaciones pero no a los de la secuencia proteica real. <sup>b</sup> Sin considerar los 130 contra-iones (Na<sup>+</sup>) necesarios para neutralizar la carga de la proteína y las membranas. <sup>c</sup> Los últimos 15 ns fueron los que se emplearon para el análisis.

corte (“cutoff”) para las interacciones no ligadas de 1.2 nm, mientras que las interacciones electrostáticas de largo alcance fueron consideradas en base a la aproximación Particle Mesh Ewal<sup>11,12</sup> (PME). Todas las interacciones, FG-FG, FG-CG y CG-CG, fueron calculadas directamente utilizando el Hamiltoniano estándar de GROMACS.

Cálculos de energía libre de solvatación fueron realizados tanto para una molécula de SPC como para una de WT4, usando primero SPC y luego WT4 como solvente. En cada caso, la molécula de agua estudiada se colocó en una caja llena del solvente FG o CG, y la interacción de la primera con su entorno fue gradualmente anulado. Para tal objetivo, las interacciones no ligadas de la molécula de agua con el solvente fueron acopladas con un factor de escalado  $\lambda$  que va desde 0 (interacción completa) a 1 (interacción nula). En este caso se utilizó una secuencia de 21 pasos igualmente espaciados (incrementos de 0.05). Para cada valor de  $\lambda$ , se llevaron a cabo una minimización de energía y una equilibración del sistema, seguido de 5 ns de simulación en condiciones NVT, a partir de la cual se calculó el promedio de la derivada de energía potencial respecto del factor de escalado ( $\langle \delta V / \delta \lambda \rangle$ ). La energía libre del proceso es la integral de 0 a 1 del valor esperado de  $\delta V / \delta \lambda$  calculado para cada  $\lambda$ , la cual fue obtenida por integración numérica.

Los perfiles de densidad, el coeficiente de difusión, el tiempo de relajación rotacional, la permitividad dieléctrica y el factor de Kirkwood fueron obtenidos usando herramientas estándar de GROMACS. El número de auto-coordinación fue calculado por conteo del número de oxígenos del agua SPC y de superátomos de WT4 contenidos en una esfera de 0.55 nm de radio (primera y segunda esfera de solvatación). La herramienta `g_rdf` de GROMACS fue utilizada para calcular la densidad numérica de partículas alrededor de los átomos de oxígeno (O-O cuando las partículas son los átomos de oxígeno de las moléculas de agua SPC y O-WT4 calculado se trata de los superátomos de WT4). La distribución de densidad numérica total se obtuvo sumando la densidad O-O a la O-WT4 multiplicada por 2.8, de acuerdo al factor de condensación de los superátomos de WT4.<sup>2</sup> Para obtener la función de distribución radial normalizada (RDF), la densidad numérica total se dividió por el valor que esta toma en un punto de convergencia de la RDF. En el presente caso, este valor fue tomado a una distancia de 1.5 nm del origen.

La compresibilidad isotérmica fue calculada evaluando las fluctuaciones del volumen en el sistema 1, de acuerdo con el procedimiento comentado en el Capítulo II y propuesto por Herrero.<sup>13</sup> A su vez la tensión superficial fue evaluada también como se indica en el Capítulo II y de acuerdo con Chen et al,<sup>14</sup> utilizando el sistema 2. El coeficiente de expansión y la capacidad calorífica fueron estimados siguiendo el procedimiento reportado por Yu et al.<sup>15</sup> El cálculo de la presión osmótica se basa en la metodología presentada por Luo y Roux,<sup>16</sup> utilizando la

implementación detallada en el Capítulo II.

Para la simulación del complejo SNARE se utilizó un modelo FG solvatado explícitamente, insertado en dos bicapas lipídicas completamente solvatadas cuya composición es 89% de 1-palmitoil-2-oleoil-fosfatidilcolina (POPC) y 11% de 1-palmitoil-2-oleoil-phosphatidilserine (POPS). El proceso de construcción de este modelo se describe en detalle en la ref 7.<sup>7</sup> El sistema de bicapas mixtas unidas por el complejo SNARE, embebidos en una solución acuosa de NaCl 0.1 M fue equilibrado durante 44 ns de dinámica molecular como se describió previamente. Se utilizó el campo de fuerza GROMOS-87<sup>17</sup> con parámetros adicionales para POPC.<sup>18</sup> Se llevaron a cabo simulaciones de producción de 20 ns, descartando los primeros 5 ns por motivos de equilibración correspondientes al modelo de resolución dual, usando condiciones periódicas de contorno y electrostática de largo alcance vía el método PME,<sup>11,12</sup> con un “cutoff” para las interacciones no ligadas de 1.0 nm. Adicionalmente, el centro de masa traslacional fue removido a cada paso.

Para probar el esquema de solvatación con resolución dual, se construyó un sistema análogo al sistema 8 pero con solvente híbrido (sistema 9 en la Tabla 4-1). El procedimiento de construcción fue el siguiente: se tomó la última configuración de la simulación de 44 ns del sistema 8. En la misma, toda el agua FG más allá de 1 nm respecto de cualquier átomo de la proteína o la de membrana fue removida. El espacio vacío resultante se llenó con moléculas de WT4, utilizando una caja pre-estabilizada conteniendo 16 moléculas de WT4. Finalmente, los iones FG que quedaron en la fase CG fueron reemplazados por sus correspondientes contra partes CG. Se realizaron cálculos de minimización de energía y de equilibración durante 5 ns previo a la simulación de producción.

Varias propiedades fueron medidas con el objetivo de probar la capacidad del modelo de solvatación multiescala de imitar al sistema completamente atomístico. Las propiedades intrínsecas de membrana fueron calculadas en el sistema completo (membrana con curvatura debido a la presencia de las proteínas) y en un parche sin curvatura de 30 nm<sup>2</sup>, que no está en contacto con la proteína en ningún momento de la simulación. Los perfiles de densidad de los componentes del sistema fueron calculados en la dirección ortogonal a la membrana. La distancia intermembrana se estimó en dos dimensiones de modo de poder considerar las perturbaciones inducidas por los dominios transmembrana (DTM) del complejo proteico. Se promediaron las distancias entre las cabezas de los lípidos de las dos monocapas más externas a lo largo del eje z en el tiempo usando una grilla en el plano x-y, con una distancia de 0.5 nm entre los puntos de la grilla. Los parámetros de orden de los lípidos fueron calculados como el ángulo entre el enlace carbono-deuterio hipotético y el eje z ( $S_{CD}$ ).<sup>19</sup> Como medida de la interacción de las cabezas polares de los fosfolípidos con el solvente, se midió el ángulo entre la línea que

conecta los grupos fosfato y amino respecto a la normal de la bicapa (vectores PN)

El desvío cuadrático medio (RMSD por su significado en inglés: Root Mean Square Deviation) y las fluctuaciones (RMSF: Root Mean Square Fluctuations) se calcularon sobre los  $C\alpha$  a lo largo de la trayectoria. Enlaces de hidrógeno intra- e inter-moleculares fueron considerados utilizando un “cutoff” de distancia y ángulo de 0.35 nm (aceptor-dador) y  $35^\circ$  (aceptor-dador-proton), respectivamente, mientras que un “cutoff” de distancia de 0.5 nm entre cadenas laterales de carga opuesta se utilizó para detectar puentes salinos. Contactos entre diferentes especies moleculares fueron considerados hasta una distancia de 0.4 nm.

## Resultados

La aproximación propuesta en el presente trabajo como estrategia de solvación de resolución dual para sistemas biomoleculares se basa en la mezcla de dos modelos de agua de diferente nivel de descripción. La mezcla de los modelos SPC y WT4 acelera el muestreo en las simulaciones de dinámica molecular dado que el agua “bulk” es modelada con moléculas de WT4, reduciendo el número de partículas de agua en un factor de dos o más. La implementación de este esquema es directo pues todas las interacciones son calculadas simultáneamente usando la misma forma funcional del Hamiltoniano descrito en el Capítulo II.

Para que una aproximación tal funcione, es requisito que la energía libre de solvación recíproca de ambos modelos determine (al menos parcialmente) la separación de dos fases. Si esto sucede, el agua FG debería permanecer en el entorno de un soluto FG mientras que las moléculas de solvente CG deberían ubicarse en regiones lejanas a él. Más aún, no solo tal partición debería ocurrir sino que también las características fisico-químicas de los diferentes componentes del sistema deberían permanecer iguales que las de un sistema completamente FG.

Primero se calcularon las energías libres de solvación de una molécula de SPC o WT4 en una caja de simulación conteniendo SPC o WT4.<sup>20</sup> Para que ocurra el particionamiento de un sistema de resolución dual en dos fases de manera espontánea, la energía libre de solvación de SPC embebida en SPC debería ser menor que si estuviera embebida en WT4, y viceversa. La Tabla 4-2 muestra el resultado de los cálculos de energía libre, confirmando que la separación de fases es termodinámicamente favorable. Más aún, los valores negativos que se observan en todos los casos indican que ambas especies moleculares son solubles en ambos líquidos. Esto sugiere que la aproximación que se propone efectivamente es útil como un método de solvación multiescala, acelerando las

simulaciones de dinámica molecular, sin una pérdida sensible de precisión en los resultados.

**Tabla 4-2.** Energías libres de solvatación de los modelos de agua FG, CG y FG-CG

soluto	solvente	$\Delta G_{\text{solv}}$ (kJmol <sup>-1</sup> )
SPC	SPC	-36.8
SPC	WT4	-13.2
WT4	SPC	-35.6
WT4	WT4	-50.7

En las siguientes secciones, se presentan algunas propiedades fisico-químicas comunes del líquido híbrido en contraposición con sus pares FG. En la Sección 1 se estudian las propiedades del agua pura, en la Sección 2 se introduce el tratamiento híbrido de electrolitos libres y algunas propiedades de las soluciones iónicas, finalmente en la Sección 3 se presenta una comparación sistemática de las características estructurales y dinámicas del complejo SNARE que se describió previamente, simulado con resolución completamente atomística y dual. Este último set de simulaciones permitiría validar la aproximación metodológica que se propone en el presente capítulo para sistemas biológicos de considerable complejidad en tamaño y composición.

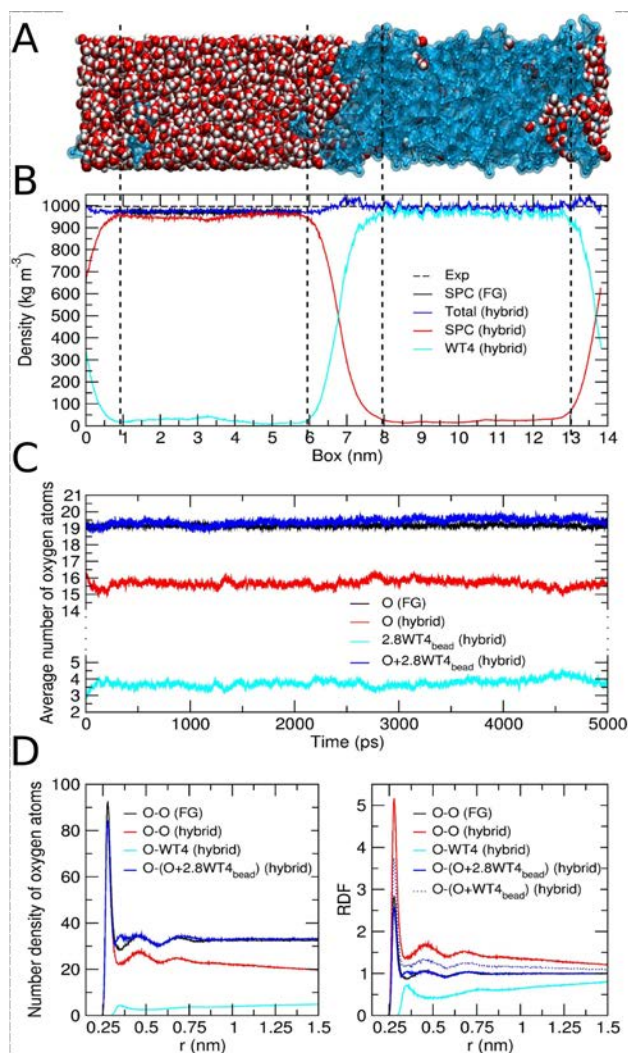
Es necesario mantener en mente siempre que se haga referencia al sistema de resolución dual que cada molécula de WT4 representa 11 moléculas de agua real y cada ion CG representa al propio ion junto con seis moléculas de agua adicionales.<sup>2</sup> En ese sentido, cuando se hace referencia al número total de moléculas de solvente, se refiere a la suma del agua e iones atomísticos más 11 aguas por molécula de WT4 más un ion y seis aguas por cada ion CG.

**1. Agua Pura.** Los datos termodinámicos presentados previamente predicen que una mezcla de aguas SPC y WT4 deberían particionarse en dos fases. Para estudiar efectivamente el comportamiento de la mezcla, se llevaron a cabo simulaciones de dinámica molecular en el ensemble NPT a 300 K y 1 bar, usando una caja conteniendo dos secciones adyacentes de 2002 aguas SPC y 182 moléculas de WT4 (sistema 1, ver Figura 4-1.A). Dado que el factor de condensación del WT4 (182 WT4 equivalen a 2002 aguas reales), este sistema puede considerarse como una mezcla 1:1 de moléculas de agua FG y CG.

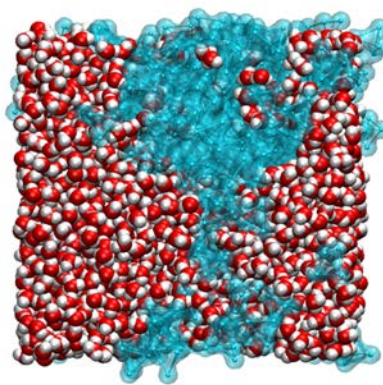
Todos los resultados que se describen a continuación se refieren a la configuración mostrada en la Figura 4-2.A. Sin embargo ninguna de las propiedades calculadas cambian si se comienza de una configuración completamente mezclada. La configuración final de una simulación que parta de un estado inicial completamente mezclado resulta en dos fases correctamente particionadas. Tal separación ocurre inclusive inmediatamente después del



periodo de termalización/presurización del sistema. (Figura 4-3). Sin embargo, en dicha configuración es difícil



**Figura 4-2. Sistemas de agua pura.** A- Foto del sistema 2 luego de 5ns de simulación de dinámica molecular. Las moléculas de agua FG SPC se representan utilizando esferas de vdW rojas y blancas, mientras que las moléculas de WT4 se muestran combinando una superficie envolvente con la representación CPK. B- Perfil de densidad de masa a lo largo de la dirección ortogonal a la interfase FG-CG para el sistema 2. Adicionalmente se muestra el perfil de densidad correspondiente para una simulación FG (sistema 1) y el valor experimental de densidad como referencia. Las líneas punteadas verticales indican los límites entre las regiones FG-interfase y CG-interfase. Los paneles A y B están en la misma escala. C- número de auto-coordinación del agua. Serie temporal del número de átomos de oxígeno a 0.55 nm desde el oxígeno de cualquier agua FG se muestra para la simulación FG de referencia (sistema 1) así como también para la simulación FG-CG (sistema 2). D- Densidad numérica O-O (izquierda) y RDF O-O (derecha) se midió para en el sistema FG de referencia y se lo comparó con el la correspondiente medida en la simulación multiescala (sistema 2).



**Figura 4-3. Sistema híbrido mezclado.** Foto luego de 5 ns de simulación partiendo de una configuración mezclada de agua FG SPC (esferas de vdW rojas y blancas) y agua CG WT4 (CPK/superficie cyan).

identificar un eje preferencial. Por lo tanto, se utilizó una configuración inicial como la que se muestra en la Figura 4-2.A por motivos de claridad y facilidad de interpretación.

El grado de mezcla de ambas representaciones de agua puede evaluarse del promedio temporal de la densidad de masa a lo largo de la dirección ortogonal a la interfase SPC-WT4 (Figura 4-2.B). Claramente, el agua híbrida bajo condiciones ambiente, se particiona en dos fases sin generar una interfase muy marcada. Ambas densidades sufren una suave transición a lo largo del eje longitudinal describiendo un comportamiento de tipo sigmoideo (Figura 4-2.B). La transición de valores mínimos a máximos tiene lugar dentro de un espacio de casi 2nm. Sin embargo, ninguna de las curvas alcanza ni el valor de densidad esperado para el sistema puro correspondiente ni tampoco el valor de cero. Esto indica que un pequeño número de moléculas de SPC y de WT4 son capaces de

penetrar la fase opuesta. La diferencia mencionada se vuelve evidente al comparar con la densidad calculada para el sistema 2, que contiene únicamente agua SPC pura. La línea negra a la izquierda de la Figura 4-2.B muestra la densidad correspondiente al sistema FG puro, que se mantiene ligeramente encima de la densidad obtenida para el agua SPC en el sistema de resolución dual. Sin embargo, un muy buen acuerdo se observa cuando se considera el número total de aguas representadas por la suma de moléculas de SPC y de WT4 dentro de la región FG del sistema 1.

Esto no es sorprendente si se considera que el modelo WT4 imita la estructura tetraédrica del agua "bulk". Como podía preverse, las moléculas de WT4 son capaces de intercalarse dentro del agua SPC, sustituyendo pequeños clusters de agua, sin perturbar la estructura global de esta. Esto sugiere que ambas especies moleculares son capaces de cruzar la interfase a temperatura ambiente. Es decir, con un costo energético asociado modesto. Como resultado, la densidad total de la región CG reproduce el valor experimental, aunque con fluctuaciones más notables debido a la mayor granularidad del WT4. Una pequeña sobreestimación (~3% respecto del valor experimental) se observa solo dentro de la región de interfase. Este aumento en la densidad puede deberse a una levemente mayor ocupación de las moléculas de SPC y WT4 en los intersticios moleculares del líquido. Por lo tanto, a pesar de este leve artefacto en la región de interfase, el perfil de densidad de ambas regiones FG y CG no se ve alterada por la aproximación multiescala.

Un indicador de la estabilidad de la mezcla puede obtenerse de la evolución temporal del número de auto-coordinación para los átomos de oxígeno dentro de una esfera de 0.55 nm centrada en cualquier de estos. (Figura 4-2.C). Esta distancia incluye las dos primeras esferas de solvatación, proveyendo de una precisa medida de posibles efectos espurios del método. En el caso del sistema híbrido, dicho número de auto-coordinación se calculó para el SPC, para el total de aguas representadas por el WT4 y para la suma de ambas.

Un cierto aumento en el grado de mezcla se puede observar durante los primeros 200 ps de simulación. Esto se evidencia por la disminución en el número de coordinación de los oxígenos de agua FG y el incremento en el número efectivo representado al nivel CG (líneas del medio e inferior en la Figura 4-2.C, respectivamente). Este análisis sugiere que moléculas en la interfase comienzan a mezclarse inmediatamente después del comienzo de la simulación. Una pequeña cantidad de cada clase de moléculas difunde alcanzando la fase opuesta. El proceso de intercambio se estabiliza rápidamente pues no se observan desviaciones en el número de auto-coordinación después de los primeros ~200 ps, verificando la estabilidad temporal de la mezcla FG-CG.

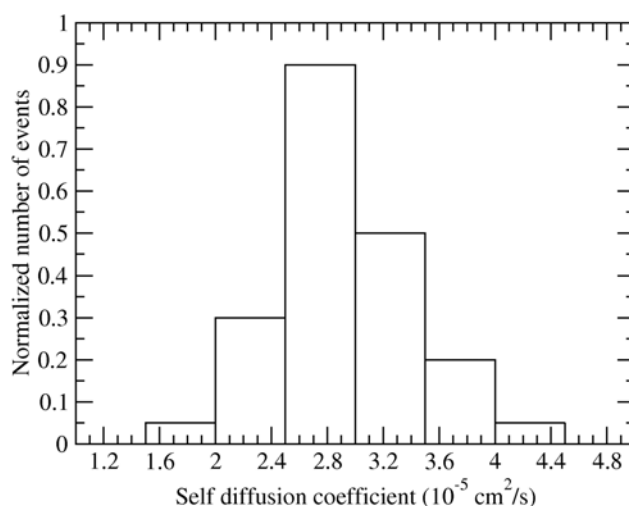
El número de coordinación global del agua SPC en el esquema híbrido (sistema 1) es, obviamente, menor que el calculado para el sistema FG (sistema 2) debido a la presencia de las moléculas CG en la caja de simulación. (Figura

4-2.C) Sin embargo, cuando el número total de agua representadas por las moléculas CG es calculado, los números de coordinación para los sistemas 1 y 2 muestran muy buena coincidencia con diferencias máximas en el orden o menor al 3% (líneas superiores de la Figura 4-2.C).

Para caracterizar la estructura del agua se calculó la densidad numérica de oxígenos alrededor de cualquier átomo de oxígeno (O-O) del agua SPC. De la comparación entre los sistemas FG y de resolución dual, surgen observaciones interesantes respecto del arreglo estructural del solvente. De la Figura 4-2.D (izquierda) se puede observar que la densidad numérica O-O para los sistemas 1 y 2 muestra buena coincidencia en la posición de las esferas de solvatación. La altura de la función es progresivamente subestimada con el aumento de la distancia radial en el caso del sistema dual comparado con el sistema FG, debido a la presencia de las moléculas de WT4 más allá de la interfase híbrida. Si nos enfocamos en el arreglo de los superátomos de WT4 alrededor de los oxígenos FG, el primer pequeño máximo ocurre a  $\sim 0.35$  nm del oxígeno FG. Luego de dicho pico, se observa un suave crecimiento monotónico en la densidad numérica que acompaña la disminución de la densidad O-O. Esto sugiere que el agua CG tiene un efecto pequeño en el arreglo de las aguas FG vecinas. El perfil convergente es recuperado si el número total de aguas se tiene en cuenta (considerando que cada molécula de WT4 representa 2.8 aguas reales, ver Figura 4-2.D, izquierda). Esto indica un buen acuerdo entre las simulaciones FG y de resolución dual. Cabe notar que la convergencia de la curva de densidad numérica indica que el sistema híbrido no presenta síntomas de congelamiento. La principal diferencia respecto del sistema atomístico puro se observa en la altura del primer máximo de solvatación y la presencia de un leve pico espurio entre las primer y segunda esferas de solvatación. La causa de ambos efectos es el posicionamiento de los superátomos de WT4, que se encuentra corrido levemente hacia adentro respecto del segundo máximo de solvatación (posición que le correspondería a las moléculas de WT4 por tratarse de un modelo de segunda esfera de solvatación como se menciona en el Capítulo II). Sin embargo, como se mostrará más adelante, poca o ninguna desviación puede observarse en las características físico-químicas del sistema multiescala respecto de los valores experimentales. Debe tenerse en cuenta que en los párrafos anteriores se hizo referencia a la densidad numérica en lugar de a la comúnmente utilizada función de distribución radial oxígeno-oxígeno (RDF O-O). La RDF es una medida de la probabilidad de encontrar un cierto número de partículas alrededor de un punto de referencia cuando se aumenta la distancia en dirección radial. La correcta normalización de esta función requiere considerar el aumento en la dimensión de la superficie de la esfera centrada en el punto de referencia que acompaña el aumento de la distancia radial, y también requiere tener en cuenta la densidad en el "bulk" de las partículas que se están contando. En el caso del solvente híbrido, no es posible una correcta normalización a la densidad del

"bulk" si la densidad de las partículas cambia en la interfase y más allá de ésta. En el sistema de resolución dual, si la distribución es normalizada directamente al valor de la densidad del "bulk", sin considerar el número de partículas efectivamente representadas por los superátomos de WT4, la RDF resulta espuriamente sobrestimada (Figura 4-2.D, derecha). En cambio, si se calcula la densidad numérica de ambas especies moleculares, se las suma multiplicando la correspondiente a los superátomos de WT4 por el debido factor de condensación (2.8), y se usa el valor de densidad numérica de la suma en un punto luego del cual la misa ya se haya relajado como factor de normalización, entonces se evita el problema mencionado. Siguiendo este procedimiento se obtuvo una RDF híbrida que se compara muy bien con su contraparte atomística (Figura 4-2D, derecha).

Adicionalmente a la RDF, el arreglo estructural de un líquido polar puede caracterizarse por la correlación dipolo-dipolo entre las moléculas de agua. Esto puede lograrse mediante el cálculo del factor de Kirkwood ( $g_k$ ) para sistemas infinitos. Para líquidos no-polares o moderadamente polares,  $g_k$  no se desvía significativamente de la unidad, dado que las moléculas de un solvente no-polar no impone una orientación preferencial a sus vecinos. Sin embargo, para líquidos cuya cohesión involucra enlaces de hidrógeno, como es el caso del agua, se observan valores de  $g_k$  que se desvían considerablemente de la unidad. Esto se relaciona con la fuerte influencia de los enlaces de hidrógeno en la orientación relativa de las moléculas en el líquido. El valor calculado de  $g_k$  para el agua SPC, que está en acuerdo con reportes previos,<sup>15</sup> presenta una subestimación leve respecto del valor experimental (2.90<sup>21</sup>). Sin embargo, los factores de Kirkwood calculados en el sistema híbrido muestran un incremento respecto del valor en el sistema FG, que lo acerca a los valores experimentales. (Tabla 4-3)



**Figura 4-4. Distribución de coeficientes de difusión.** A- Las 2002 moléculas de agua FG del sistema 2 se subdividieron en 50 grupos y los coeficientes de difusión promediados en dentro de cada grupo se presentan en forma de histograma.

En base a este resultado, se investigó si las propiedades dieléctricas del líquido híbrido, como un todo, eran bien reproducidas. En acuerdo con datos presentes en la literatura,<sup>15</sup> se obtuvo un valor de permitividad para el modelo de agua SPC (sistema 2) de 65, que representa un error relativo de 17% respecto del valor experimental que es 78.4.<sup>22</sup> En el mismo sentido que el resultado obtenido para los factores de Kirkwood, la permitividad dieléctrica calculada para el sistema híbrido es de 84, es decir sobrestima en un 7% el valor experimental.

Posteriormente, se determinaron algunas características dinámicas de la representación multiescala. Concretamente se midieron dos parámetros para corroborar la correcta representación de dichas características: el coeficiente de difusión y el tiempo de relajación rotacional.

La Tabla 4-3 muestra el valor de coeficiente de difusión para el SPC tanto en la simulación FG como la híbrida. Coincidente con lo reportado en la literatura,<sup>15</sup> se obtuvo un valor de  $4.3 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$  en comparación con el valor experimental que es  $2.3 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$ . Este parámetro calculado para el sistema híbrido muestra una importante reducción en relación con el valor del sistema FG. Sin embargo, el resultado del solvente híbrido es más cercano al valor experimental. Una duda razonable respecto del coeficiente de difusión en el sistema de resolución dual, es que la medida podría resultar del promedio de dos distribuciones diferentes, una correspondiente a moléculas de agua SPC en la región FG moviéndose a velocidad normal y otra relacionada con moléculas de SPC en contacto o dentro de la fase CG, que se muevan sustancialmente más lento o que estén de hecho congeladas. De ser ese el caso, la distribución de coeficientes de difusión obtenida dividiendo el sistema en pequeños grupos de aguas debería estar compuesta por dos picos, uno centrado en el valor del coeficiente de difusión intrínseco del modelo SPC ( $4.3 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$ ) y otro a un valor menor que el experimental. Para poner a prueba dicha hipótesis, se calculó la distribución de coeficientes de difusión de 40 grupos, cada uno conteniendo 50 moléculas de aguas elegidas aleatoriamente. La distribución obtenida presenta un solo pico (Figura 4-4), proveyendo de evidencia a favor de un solo régimen difusivo, y verificando una reducción global del proceso de difusión del agua FG en presencia de WT4.

El segundo parámetro estudiado para caracterizar el comportamiento dinámico de la mezcla FG-CG es el tiempo de relajación rotacional (conocido también como "tumbling time") que es el tiempo que tarda una molécula en reorientarse en un grado, en un medio dado. Podría esperarse que la interfase híbrida alterase el comportamiento dinámico del agua FG. De hecho, medidas de espectroscopia IR sugieren que la sola presencia de una interfase de naturaleza arbitraria es más importante en la disminución de la dinámica de formación y ruptura de enlaces de hidrógeno, que la naturaleza química o la geometría de la interfase.<sup>23</sup> Los tiempos de

relajación del agua bajo diferentes condiciones de confinamiento muestran un incremento de entre 2 y diez

**Tabla 4-3.** Propiedades fisico-químicas del agua pura

	factor de Kirkwood <sup>a</sup>	tiempo de relajación ( $\tau_{2}^{\text{HH}}/\tau_{2}^{\text{OH}}$ ) (ps)	permitividad relativa	coeficiente de difusión ( $10^{-5} \text{ cm}^2\text{s}^{-1}$ )	compresibilidad isotérmica ( $\text{GPa}^{-1}$ )	capacidad calorífica ( $\text{kJmol}^{-1}\text{K}^{-1}$ )	tensión superficial ( $\text{mNm}^{-1}$ )
SPC	2.53	1.0/0.9	65	4.3	0.534 (38)	0.074 (38)	53.4 (37)
Multiescala (SPC-WT4)	2.9	1.2/1.0	84	2.9	0.52	0.043	52
WT4	... <sup>b</sup>	... <sup>b</sup>	110	2.23	2.43		17
Exp.	2.9 (43)	1.95 (44)	78.4 (45)	2.27 (46)	0.46	0.075 <sup>c</sup>	71.2 (47)

<sup>a</sup> Calculado luego de 5 ns de simulación. <sup>b</sup> Los factores de Kirkwood y el tiempo de relajación no puede calcularse para las moléculas de WT4. <sup>c</sup> Tomado de la referencia 38.

veces. La Tabla 4-3 muestra los valores obtenidos para el modelo FG y el de resolución dual. El modelo SPC subestima el tiempo de relajación en comparación con los valores experimentales (38).<sup>15</sup> Los valores calculados para el agua SPC en la mezcla con WT4 son similares a los del SPC puro, con un leve incremento (más cercanos al valor experimental), indicando que el WT4 no genera efecto de confinamiento alguno sobre agua atomística.

En conjunto estos resultado sugieren que la presencia de la fase CG introduce un pequeño incremento en la correlación dipolo-dipolo y en el tiempo de relajación del modelo de agua FG, que ligeramente disminuyen la difusión de las moléculas de agua atomística. Sin embargo, el efecto neto de la representación dual del líquido lleva a un ligeramente mejor acuerdo entre el comportamiento del agua FG y varias propiedades determinadas experimentalmente.

Otra propiedad importante del agua es la compresibilidad isotérmica. Como se muestra en la Tabla 4-3, el valor de dicha propiedad para el sistema híbrido se compara bien con el valor reportado para el modelo SPC puro<sup>24</sup> y está en buen acuerdo con el valor experimental ( $0.46 \text{ GPa}^{-1}$ ). Este resultado parece indicar que a pesar de la alta compresibilidad observada para el WT4 puro,<sup>2</sup> el valor resultante en el sistema híbrido lleva a una respuesta global frente a las fluctuaciones de la presión externa, que se parece a la del sistema SPC puro.

La tensión superficial es otra propiedad clave del agua y un ingrediente importante para dirigir procesos de auto-ensamblaje, que revisten gran importancia biológica. La tensión superficial del modelo WT4, como se comentó en el Capítulo I, es baja ( $17 \text{ mNm}^{-1}$ ) comparada con el valor experimental ( $71.2 \text{ mNm}^{-1}$ ). Esto genera serias dudas acerca del modelado de fenómenos de auto-ensamblaje usando el solvente de resolución dual aquí presentado. Sin embargo, como muestra la Tabla 4-3, la tensión superficial del solvente híbrido, calculada en el sistema 3 está

muy cercano al valor de  $53.4 \text{ mNm}^{-1}$  reportado para el SPC puro (37).<sup>14</sup> De acuerdo con este análisis, las moléculas de WT4 en la representación dual generan solo un efecto moderado en la tensión superficial en la fase FG. Por lo tanto, la aplicación del método híbrido a sistemas biológicos no debería generar problemas, al menos en este sentido. De hecho, no se detectan diferencias significativas entre el tratamiento con solvente híbrido y el puramente atomístico en simulaciones de sistemas de membranas (ver sección 3).

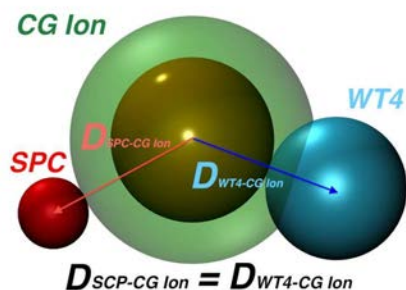
Por último, se calculó la capacidad calorífica ( $C_p$ ) con el objetivo de caracterizar la respuesta del sistema a fluctuaciones de la temperatura (Tabla 4-3). La evaluación de  $C_p$  se realizó a partir de la diferencia entre la energía total por molécula efectiva de agua a 298 K y a 318 K. La capacidad calorífica observada para el sistema híbrido subestima los valores para SPC y experimental<sup>25</sup> en ~60%. Esta incongruencia es esperable ya que el sistema híbrido como un todo pierde partículas y grados de libertad. Esta disminución de entropía se refleja en la reducida capacidad del sistema híbrido para almacenar energía térmica.

**2. Electrolitos en Solución Acuosa.** En esta sección, se hará hincapié en la reproducción de propiedades relacionadas con electrolitos simples en solución acuosa, por ser un componente fundamental de los sistemas biológicos. Para analizar la efectividad de esta aproximación, se construyó un sistema de simulación conteniendo una mezcla 1:1 de SPC/WT4 y 0.3 M  $\text{Na}^+\text{Cl}^-$  (Tabla 4-1 y Figura 4-5.A). Cabe notar que la concentración iónica en ambas fases del líquido se ajustó para que sea la misma.

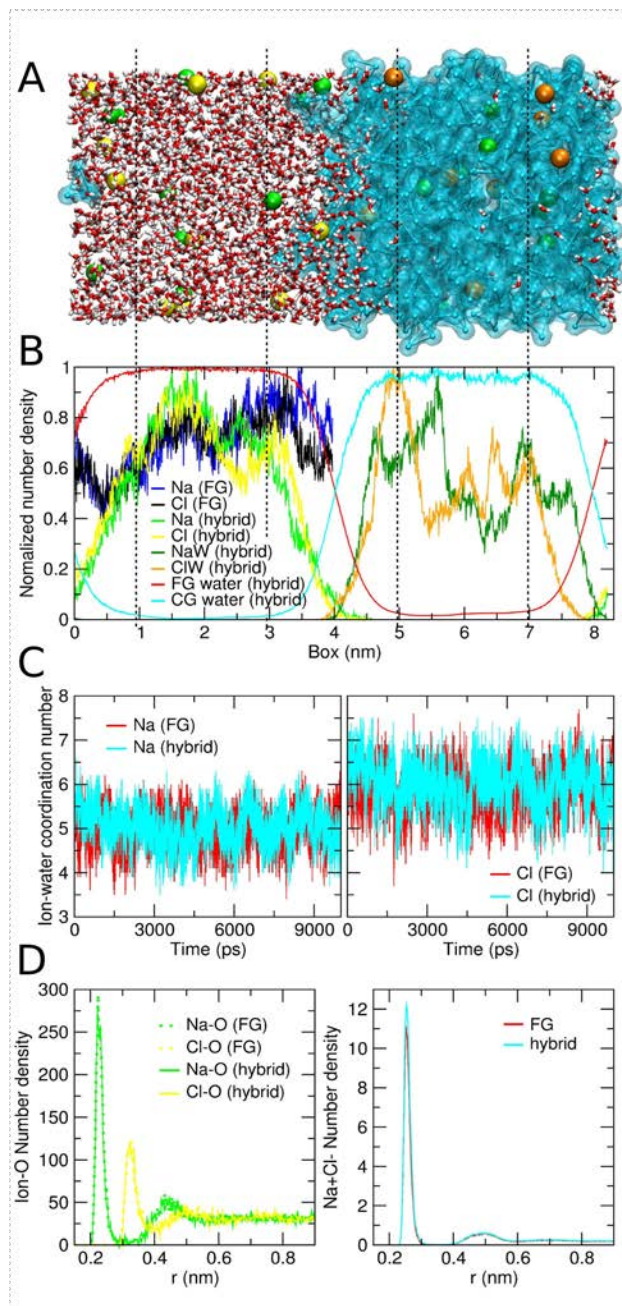
Como se mencionó en el Capítulo II, los iones CG son superátomos efectivos, con carga, masa, y parámetros de Lennard-Jones representando un electrolito simple que transporta implícitamente su primera esfera de solvatación que corresponde a seis moléculas de agua.<sup>25</sup> Por lo tanto, la introducción de los iones CG dentro del medio acuoso FG, podría resultar en interacciones espurias dado que el tamaño mínimo entre el agua FG y el ion CG es menor que la distancia mínima entre un superatomo de WT4 y el ion CG. Esto genera una mayor ganancia en la energía electrostática cuando un ion CG es solvatado por agua SPC relativa a la solvatación por moléculas de WT4. Una corrección directa para este problema es forzar al agua SPC a colocarse en la segunda esfera de solvatación del ion, ajustando la interacción de Lennard-Jones entre los iones CG y el agua FG (Figura 4-6). Tal corrección es suficiente para asegurar que los iones CG permanezcan en la fase CG. Esto puede observarse directamente de la densidad promedio de lo largo del eje ortogonal a la interfase (Figura 4-5.B). La inclusión de los electrolitos no introduce diferencias significativas en la partición de las especies FG y CG en la caja de simulación. El perfil de densidad de las moléculas de agua FG y CG en el sistema 3 se asemeja al observado para el sistema 2 (Figura 4-5.B), con una suave transición entre ambas fases. Análogamente, los perfiles de densidad

de las especies iónicas indican que todos los electrolitos permanecen dentro de su respectiva fase.

**Figura 4-5. Soluciones de electrolitos simples.** A- Foto del sistema 5 luego de 10 ns de simulación de DM. Las moléculas de agua FG SPC se muestran con la representación cilíndrica en rojo y blanco; esferas de vdW verde claro y amarillo indican los iones FG  $\text{Na}^+$  y  $\text{Cl}^-$ , respectivamente; la representación CPK/superficie envolvente se usa para indicar las moléculas de WT4; los iones CG  $\text{NaW}^+$  y  $\text{ClW}^-$  se representan con esferas de vdW verde oscuro y naranja, respectivamente. B- Perfiles de densidad de masa a lo largo de la dirección ortogonal a la interfase FG-CG para el sistema 5. El correspondiente perfil para el sistema equivalente FG (sistema 4) se muestra como referencia. Las líneas punteadas verticales indican los límites entre las regiones FG-interfase y CG-interfase. Los paneles A y B están en la misma escala. C- Números de coordinación ión-oxígeno. Se muestran las series temporales del número de átomos de oxígeno a 0.4 nm respecto de cualquier ión FG, para el sistema FG de referencia (sistema 4) así como para el sistema multiescala equivalente (sistema 5). D- Densidades numéricas  $\text{Na}^+\text{-O}$  y  $\text{Cl}^-\text{-O}$  (izquierda) y  $\text{Na}^+\text{-Cl}^-$  (derecha), medida para los sistemas 4 y 5.



**Figura 4-6. Interacción SPC- iones CG.** La interacción entre el agua FG y los iones CG fue ajustada para reproducir la segunda esfera de solvatación de los iones CG. Esta parametrización asegura que los iones CG se mantengan en la fase CG.



Las oscilaciones más notables en las líneas correspondientes a los iones se relacionan muy probablemente con el reducido número de partículas, lo cual disminuye el muestreo estadístico. Esta hipótesis es confirmada al comparar con las líneas obtenidas a partir de una simulación puramente atomística (sistema 4), en la que las



amplitudes de las oscilaciones de la densidad son de la misma magnitud que las observadas en el sistema híbrido. (Figura 4-5.B, izquierda) En contraste a las moléculas de agua, las cuales son capaces de cruzar la interfase, la penetración de iones FG en la fase CG y viceversa no se observa. En ambos casos, la densidad cae a cero en la región de interfase.

El cálculo del número de coordinación de aguas SPC alrededor de  $\text{Na}^+$  y  $\text{Cl}^-$  muestra perfiles indistinguibles entre ambos esquemas de solvatación (Figura 4-5.C). Adicionalmente, se observa que la estabilización es muy rápida en ambos casos ocurriendo dentro del tiempo de termalización/presurización. Esta observación agrega evidencia en el sentido de que los iones no muestran tendencia alguna a cambiar de fase. De hecho, las curvas de densidad numérica de oxígenos SPC alrededor de cualquiera de los electrolitos FG se superponen casi perfectamente al comparar ambos métodos de solvatación (Figura 4-5.D, izquierda). Esta coincidencia indica que, a pesar de las diferencias en la carga eléctrica y el tamaño, las perturbaciones que ambas especies iónicas introducen en la estructura del líquido, tanto en la representación atomística como en la híbrida, son prácticamente idénticas. Tal similitud estructural sugiere que la solvatación híbrida que se propone no modifica las interacciones del agua FG con el soluto, en comparación con el caso puramente atomístico. La buena superposición de las RDFs  $\text{Na}^+\text{-Cl}^-$  (Figura 4-5.D, derecha) es otro fuerte indicio de dicha similitud.

Análogamente al resultado obtenido para el caso del agua pura en la sección anterior, los coeficientes de difusión de los electrolitos simples en el esquema híbrido están subestimados respecto de la simulaciones FG (Tabla 4-4), siendo más notoria la discrepancia en el caso del  $\text{Na}^+$ . Sin embargo, cuando se toman en consideración los datos experimentales, los valores obtenidos en la simulación FG revelan una sobrestimación del comportamiento difusional, mientras que el sistema multiescala induce una difusión más lenta, lo cual se evidencia por los valores de coeficiente de difusión de  $\text{Na}^+$  y  $\text{Cl}^-$  que están más cercanos al valor experimental.

**Tabla 4-4.** Propiedades fisico-químicas de las soluciones de electrolitos simples.

	coeficiente de difusión del $\text{Na}^+$ (sistemas 4 y 5) ( $10^{-5} \text{ cm}^2\text{s}^{-1}$ )	coeficiente de difusión del $\text{Cl}^-$ (sistemas 4 y 5) ( $10^{-5} \text{ cm}^2\text{s}^{-1}$ )	presión osmótica (sistemas 6 y 7 <sup>a</sup> ) (bar)
FG	2.2 +/- 0.1	2.3 +/- 0.3	20 +/- 11
FG-CG	1.1 +/- 0.2	1.8 +/- 0.1	26 +/- 11
Exp.	1.35 +/- 0.038 <sup>b</sup>	1.91 +/- 0.025 <sup>c</sup>	~25 (tomado de Luo et al <sup>16</sup> )

<sup>a</sup> La concentración iónica en los sistemas 6 y 7 corresponde a 0.5M en el volumen determinado por las restricciones de distancia sobre los iones. <sup>b</sup> Concentración de  $\text{Na}+\text{Cl}^-$ : 0.3M. <sup>c</sup> Concentración de  $\text{Na}+\text{Cl}^-$ : 0.25M (54).<sup>26</sup>

Tal disminución del coeficiente de difusión de los iones era esperable, de acuerdo con el enlentecimiento de la difusión del agua FG inducida por la aproximación híbrida, como se mencionó previamente.

Para concluir el estudio de las propiedades de los electrolitos en solución acuosa, se corroboró que la coexistencia de dos fases conteniendo electrolitos modelados con diferentes parametrizaciones no creara un artefacto en la presión osmótica. Para excluir dicha posibilidad, se calculó dicha propiedad en un sistema completamente FG y en uno de resolución dual, de acuerdo con el procedimiento propuesto por Luo y Roux.<sup>16</sup> Para esto, se construyeron los sistemas 6 y 7 (Tabla 4-1). Como se puede observar en la Tabla 4-4, el valor promedio de la presión osmótica medida en ambos sistemas coincide con las estimaciones experimentales, dentro del relativamente grande error intrínseco al método.

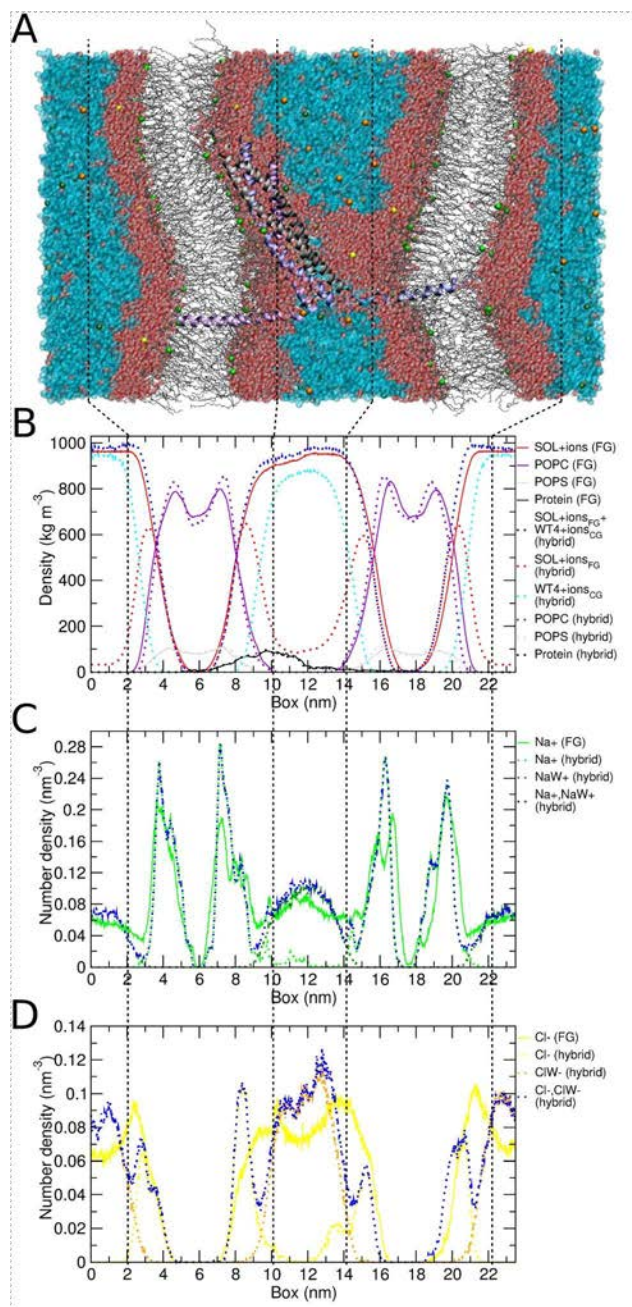
Considerando los resultados obtenidos para el agua pura y para electrolitos simples en solución acuosa, es esperable que el esquema de solvatación multiescala propuesto funcione correctamente en sistemas biomoleculares complejos. Esta hipótesis será efectivamente puesta a prueba en la siguiente sección, donde la simulación de un complejo proteico de gran tamaño embebido en una doble bicapa lipídica será estudiada.

**3. EL Complejo de Fusión de Membrana SNARE.** La maquinaria de fusión mínima del complejo SNARE se compone de tres cadenas polipeptídicas.<sup>6</sup> Estas son las proteínas, sinaptotagmina (SB), sintaxina (SX) y dos hélices alfa de SNAP25 (SN1 y SN2) asociadas mediante su parte citoplasmática por la formación del motivo estructural denominado “coiled-coil four helix bundle” o en español “haz de hélices” (Figura 4-1.D).<sup>27</sup> La asociación de dicho haz de hélices está mediada por el reconocimiento de residuos conservados formando contactos hidrofóbicos en el lado interior de cada proteína y una serie de enlaces de hidrógeno y puentes salinos desde la cara externa externa (Figura 4-1.D, derecha). SX y SB están ancladas a las membranas plasmática y vesicular, respectivamente. Información general respecto del proceso de fusión de membranas mediado por el complejo SNARE, desde la perspectiva de las simulaciones moleculares, es revisada por Risselada et al.<sup>28</sup> En trabajos previos, se caracterizaron las propiedades de los componentes del complejo por separado. Se mostró que el haz de hélices citoplasmático actúa como un velcro molecular, que se mantiene ligado por una red de interacciones de enlace de hidrógeno y puentes salinos específicos.<sup>29</sup> Estas interacciones fuertes inducen un comportamiento de “varilla rígida”, es decir un arreglo molecular muy rígido impuesto por la topología de tipo haz de hélices. Los dominios transmembrana muestran una robusta inserción en la bicapa lipídica, con alto grado de independencia en las variaciones de secuencia.<sup>30</sup> Información proveniente de experimentos de MFA

(Microscopía de Fuerza Atómica) sugiere que un único complejo SNARE es capaz de mantener dos bicapas ligadas con una separación de 5 nm.<sup>31</sup> Usando esta información se construyó un modelo completo del sistema mínimo de fusión de membrana como se ilustra en la Figura 4-1.D. Esta configuración fue previamente equilibrada, y se utilizó como estructura de partida para comparar el método de solvatación híbrida con la solvatación puramente atomística, como ejemplo práctico y realista de aplicación. Se caracterizaron las mismas propiedades que en los sistemas individuales con el objetivo de estudiar la similitud entre ambos esquemas de solvatación para un sistema complejo.

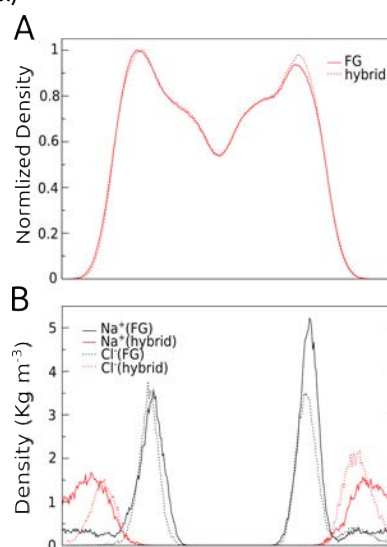
Como punto de partida para las simulaciones del presente trabajo se tomó la conformación final de la equilibración de 44 ns del sistema 8 mencionada previamente. Se extendió el tiempo de simulación por otros 20 ns, tanto usando solvatación atomística (sistema 8) como dual (sistema 9). La construcción del este último se hizo como se detalla en los Métodos. La Figura 4-7.A ilustra el mismo una vez construido. Los primeros 5 ns de la extensión de 20 ns (del ns 44 al ns 49) fueron utilizados para alcanzar la convergencia de la temperatura, presión y en particular de la permitividad. Los subsiguientes 15 ns (del ns 49 al ns 64) corresponden a los datos de producción utilizados para el análisis.

Análogamente a los casos previamente analizados, el solvente híbrido se particiona en dos regiones del espacio separadas por una suave interfase donde ambas especies acuosas coexisten. Una comparación general de los perfiles de densidad entre los sistemas atomístico e híbrido sugiere una buena conservación de las principales características estructurales (Figura 4-7.B). La localización del agua en el caso atomístico (sistema 8) muestra la esperada distribución de una topología de doble bicapa. Los niveles de densidad alcanzan valores máximos en los dos compartimientos exteriores, cae en la región ocupada por los lípidos, y aumenta nuevamente en el compartimiento interno. La pequeña depresión observada a la izquierda de la sección de agua central es generada por la presencia del complejo proteico. Si se enfoca la atención en el solvente híbrido, se observa que el agua SPC está presente sólo en las cercanías de la membrana y las proteínas. El resto del espacio, que correspondería con el agua "bulk", es ocupado por el WT4. La suma de los perfiles de densidad de ambos modelos de agua SPC y WT4 resulta en una buena reproducción de la densidad de agua total medida en el caso completamente atomístico (Figura 4-7.B). Siendo la diferencia adjudicable a la subestimación de la densidad de masa en el "bulk" inherente al modelo SPC. Pequeños picos de densidad se observan cerca de la interfase, al igual que en el sistema híbrido de agua pura (Figura 4-2.B). Una primera medida del efecto del solvente híbrido sobre



**Figura 4-7. Sistema del complejo SNARE (perfiles de densidad).**

**A-** Foto del sistema 9 luego de 20 ns de simulación de DM. Esferas de vdW rojas y blancas: agua FG SPC, verde claro y amarillo: iones FG Na<sup>+</sup> y Cl<sup>-</sup>, respectivamente, verde oscuro y naranja: iones CG NaW<sup>+</sup> y ClW<sup>-</sup>, respectivamente, el agua CG WT4 se muestra mediante una representación CPK/superficie envolvente cyan. **B-** Perfiles de densidad de masa de proteína, lípidos y agua a lo largo de la dirección ortogonal a las membranas para los sistemas FG y FG-CG (8 y 9 respectivamente.). **C-** Densidad numérica de sodio a lo largo del eje ortogonal al plano de las membranas para los sistemas 8 y 9. **D-** Ídem C pero para los iones cloruro. Las líneas punteadas verticales indican los límites entre las regiones CG-interfase (Notar la diferencia de escala entre el panel A y el resto.)



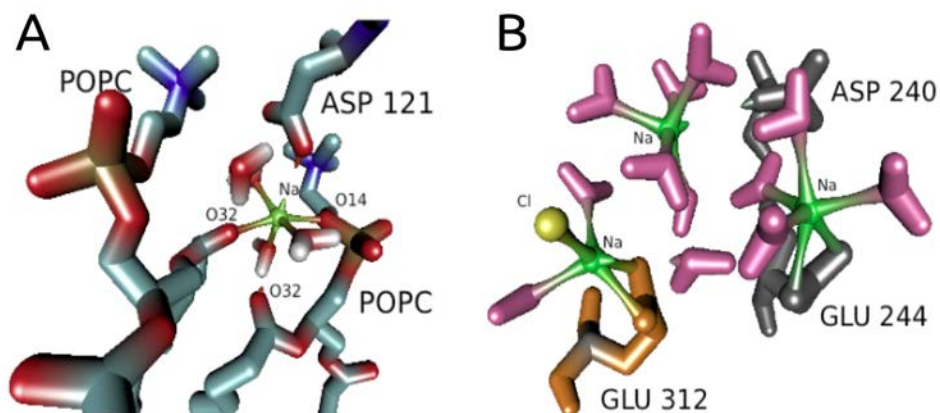
**Figura 4-8. Perfiles de densidad de lípidos e iones.** **A-** Densidad de masa de los lípidos POPC recalculada sobre una sección de la membrana vesicular lejos de la influencia de la proteína. La normal a tal región de membrana fue alineada con la dirección z de la caja. Esto permite evitar las perturbaciones inducidas por la curvatura de la membrana. **B-** Ídem A pero para los iones FG Na<sup>+</sup> y Cl<sup>-</sup>.

la conformación global del soluto puede adquirirse de la comparación de los perfiles de densidad. La densidad de la proteína en ambos casos es prácticamente indistinguible, lo cual también se observa para los perfiles de los fosfolípidos POPS. Diferencias menores se observan solo en el caso de las moléculas de POPC. Esta discrepancia puede relacionarse con fluctuaciones de larga escala de la distancia intermembrana y con la curvatura intrínseca de ambas membranas. Para corregir dicho problema, se seleccionó una región de unas de las membranas donde los fosfolípidos no están en contacto con ningún aminoácido. Esto permite medir el perfil de densidad y

compararlo con el del sistema atomístico, evitando posibles problemas relacionados a la curvatura de la membrana. La Figura 4-8.A muestra dichos perfiles confirmando una muy buena coincidencia entre ambos.

El análisis de la distribución espacial de los diferentes electrolitos muestra buena similitud entre ambos métodos de solvatación (Figura 4-7.C). En el caso del  $\text{Na}^+$  FG (Figura 4-7.C), los cationes muestran una marcada tendencia a localizarse en el borde de ambas bicapas. La distribución de cationes presenta picos definidos cerca de la membrana, que pueden adjudicarse a interacciones principalmente con los grupos carbonilo de los lípidos POPC, y en menor medida con los grupos fosfato y cabezas carboxílicas de los POPS.<sup>32</sup> Al alejarse de la membrana, la densidad de  $\text{Na}^+$  disminuye, convergiendo al valor del "bulk" a ambos lados de las membranas. Por el contrario, los iones  $\text{Cl}^-$  tienden a localizarse lejos de la membrana debido a la repulsión con las fosfolípidos POPS que llevan una carga neta negativa.<sup>33</sup> Sin embargo, debido a la fuerte localización de  $\text{Na}^+$  en la bicapa, se puede observar un pico débil de  $\text{Cl}^-$  que permanece en la fase acuosa cercana. Dichos iones  $\text{Cl}^-$ , junto con los grupos carboxílicos de los POPS balancean la densidad de carga positiva generada por la acumulación de iones  $\text{Na}^+$ . Una localización similar de  $\text{Cl}^-$  ha sido reportada en la literatura<sup>32,34</sup> (Figura 4-7.D).

La distribución iónica en presencia del solvente multiescala (línea punteada en Figura 4-7.C) sigue la misma tendencia que su contraparte atomística. Sin embargo, como se observó en la sección anterior (Figura 4-5.B), la presencia de la interfase FG-CG, genera un vaciamiento de electrolitos en la región donde ambos modelos coexisten. Inevitablemente, esto genera un leve aumento en la concentración iónica en el resto del espacio. Este efecto es más pronunciado en el caso del iones  $\text{Cl}^-$ , que se localiza más cerca a la región de interfase.



**Figura 4-9. Interacción ión-soluto.** A- Figura representativa de las interacciones entre los lípidos y los iones. B- Ídem A pero mostrando la interacción entre los iones y algunas cadenas laterales de la proteína SNARE. Se utilizó la representación HyperBalls para ilustrar las interacciones de coordinación.<sup>35</sup>

Análogamente al caso de la densidad de POPC, el acuerdo con el sistema puramente atomístico mejora si se calcula el perfil de densidad en una región reducida de la membrana donde la curvatura es menos pronunciada (Figura 4-8.B). Adicionalmente se cuantificó el número de contactos entre los iones y los diferentes componentes del sistema. Para esto se contó el número promedio de iones contenidos en una esfera de 0.4 nm de radio alrededor de los átomos que constituyen los lípidos y las proteínas. Ejemplos de interacciones ión-lípido e ión-proteína se muestran en la Figura 4-9. Como era esperable en base a la Figura 4-7.C,D, las diferencias más grandes se observan para los contactos cloruro-lípido. (Tabla 4-5). A pesar de esta observación, se encuentran globalmente sólo diferencias menores en el número de contactos ión-lípido e ión-proteína, que pueden probablemente asignarse a fluctuaciones térmicas. Sin embargo, dado que la adsorción de iones  $\text{Na}^+$  al borde de la membrana ha sido reportada como un factor inductor del aumento en el espesor de la misma (“thickness”) en  $\sim 10\%$ ,<sup>36</sup> se calculó esta propiedad en ambos tipos de solvatación para descartar efectos indirectos en las propiedades de las membranas. El espesor de la misma promedio calculado a lo largo de la simulación no muestra diferencias significativas entre ambos esquemas (Figura 4-10.A).

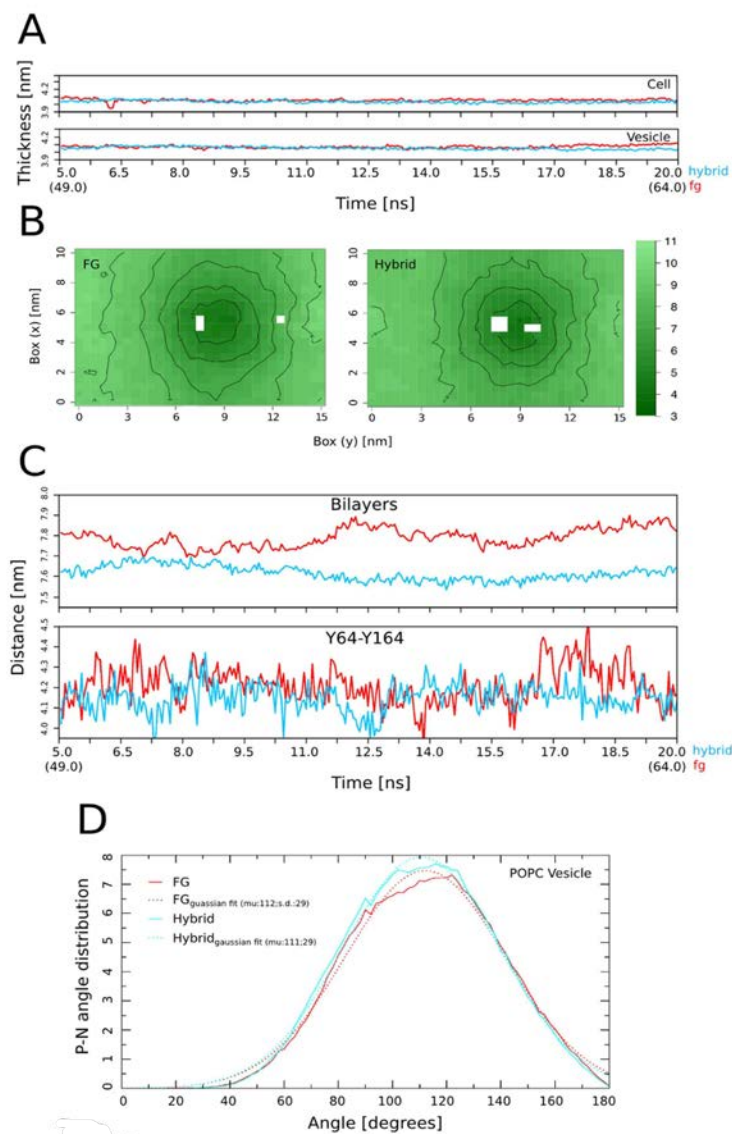
Debido a la tensión generada por el complejo SNARE se genera una importante curvatura en ambas membranas. La distancia intermembrana puede analizarse tanto como promedio temporal en gráficos bidimensionales (Figura 4-10.B) que como serie temporal del promedio espacial (Figura 4-10.C). Para la primera, ambos esquemas de simulación producen mapas 2D de distancias similares. El sitio de inserción de ambas hélices transmembrana, que corresponden a la región de menor distancia, es claramente visible cerca del centro de ambos mapas (Figura 4-10.B).

**Tabla 4-5.** Contactos entre ión-proteína y ión-lípidos (cutoff: 0.4 nm)

	número promedio de contactos	
	FG	FG-CG
$\text{Na}^+$ - lípido	188 +/- 4	192 +/- 3
$\text{Na}^+$ - proteína	3 +/- 2	4 +/- 1
$\text{Cl}^-$ - lípido	12 +/- 3	18 +/- 4
$\text{Cl}^-$ - proteína	3 +/- 1	2 +/- 1
proteína - lípido	82 +/- 4	87 +/- 4

Sin embargo, oscilaciones en una larga escala de múltiples-nanosegundos ocurre en ambos sistemas. Esto podría ser el causante de las pequeñas diferencias aparentes en la distancia intermembrana (Figura 4-10.C, arriba),

afectada por la condensación arbitraria de las oscilaciones de la membrana y de los efectos relacionados con la curvatura, en un solo número.

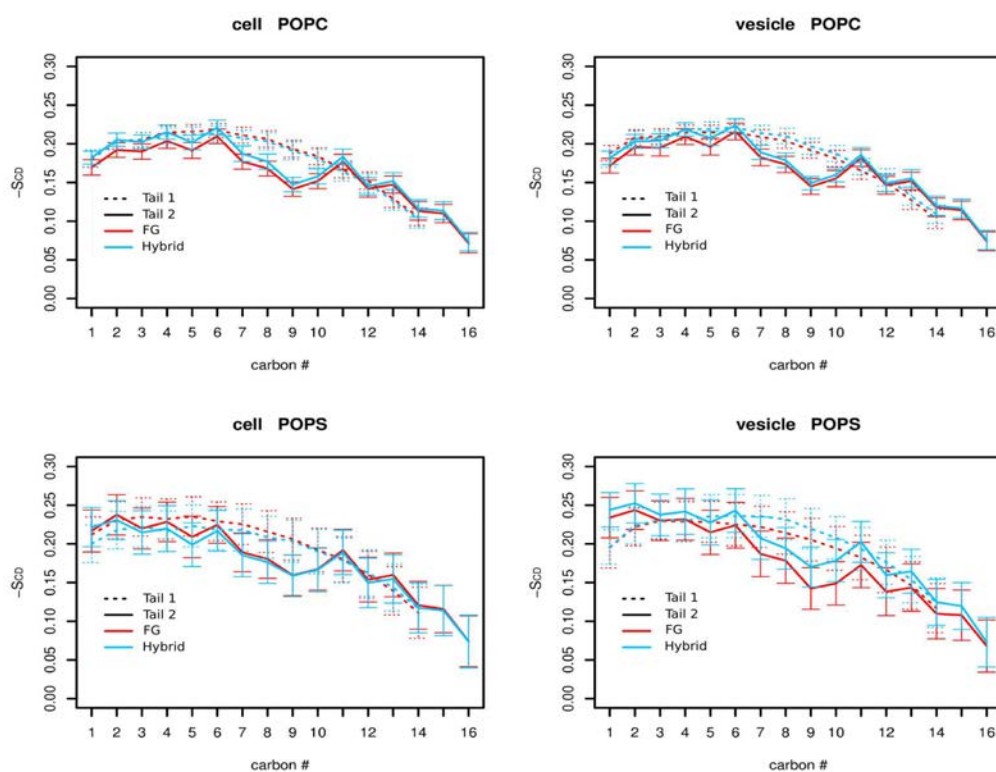


**Figura 4-10. Sistema del complejo SNARE (propiedades de membrana).** **A-** Espesor de membrana en función del tiempo para la bicapa celular (arriba) y la vesicular (abajo). Rojo: sistema FG, cyan: sistema FG-CG. **B-** Distancia intermembrana en 2D para el sistema FG (izquierda) y FG-CG (derecha). **C-** Panel superior: distancia intermembrana en función del tiempo promediada en todo el plano paralelo a la membrana. Panel inferior: distancia intermembrana estimada de la separación entre los residuos Tyr64 (SB) y Tyr164 (SX). **D-** Orientación de los vectores PN de los lipidos POPC relativo a la normal de la membrana, medido sobre un parche de  $30 \text{ nm}^2$  (ver Métodos), para ambos sistemas FG (rojo) y FG-CG (cyan). Las líneas punteadas corresponden con un ajuste a una función Gaussiana.

Es por lo tanto difícil establecer un indicador preciso de las diferencias en la distancia intermembrana, en la

escala temporal estudiada. Una caracterización exhaustiva de este pequeño pero sensible efecto requeriría incrementar al menos 10 veces el tiempo de muestreo, lo cual se escapa de los objetivos del presente capítulo. En un intento de filtrar estos efectos, se analizó una medida alternativa de la distancia intermembrana (Figura 4-1.C, abajo) utilizando dos aminoácidos insertos en las membranas, Y64 e Y164, como puntos de referencia. Discrepancias mucho menores, claramente dentro de la fluctuaciones estadísticas, se observan utilizando esta última medida.

En un esfuerzo aún mayor por comparar las propiedades de las membranas determinadas por ambos esquemas de solvatación, se analizaron algunas características estructurales distintivas de los fosfolípidos. Para los extremos hidrofóbicos, se evaluó el parámetro de orden, mientras que para los grupos polares, se calculó la orientación promedio del vector conectando los átomos de nitrógeno y fósforo en referencia al eje ortogonal al plano de la membrana (vectores P-N). Dado que la orientación de estos vectores podría verse afectada por la curvatura de la membrana, esta medida se realizó en la sección de  $30\text{nm}^2$  utilizada previamente para medir el perfil de densidad de los fosfolípidos POPC (Figura 4-8). Se observa muy buen acuerdo en la distribución promedio de los vectores P-N en ambos sistemas (Figura 4-10.D).



**Figura 4-11. Parámetros de orden.** El parámetro  $S_{CD}$  se calculó para las colas hidrofóbicas de los lípidos presentes en las membranas que componen el sistema. Los dos paneles superiores corresponden al lípido POPC para la membrana celular (izquierda) y para la membrana vesicular (derecha). Los paneles inferiores son equivalentes a los anteriores pero para el fosfolípido POPS.

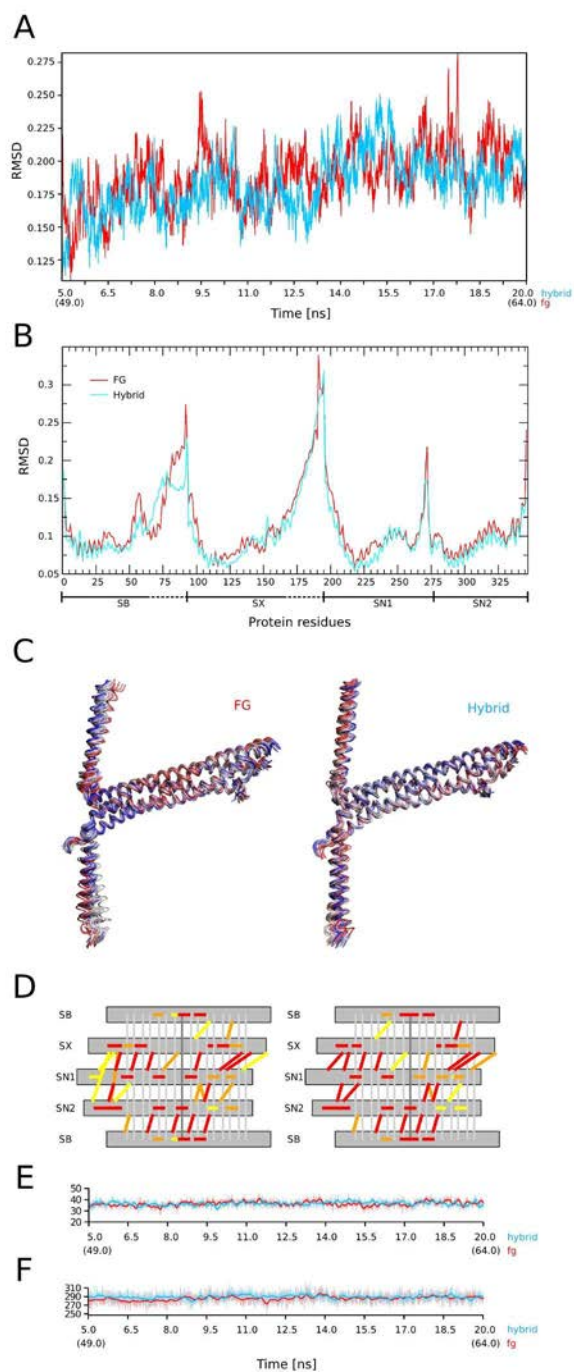


Debido a que esta propiedad depende fuertemente de la interacción agua/ión con los grupos polares, esta coincidencia agrega evidencia a favor de la equivalencia entre la aproximación de solvatación dual y la puramente atomística. La comparación entre los perfiles de parámetros de orden calculados a lo largo de las cadenas lipídicas muestra también una muy buena coincidencia (Figura 4-11). Ambos resultados en conjunto sugieren que la estructura de los fosfolípidos permanece inalterada a lo largo de la bicapa cuando se comparan ambos esquemas de solvatación.

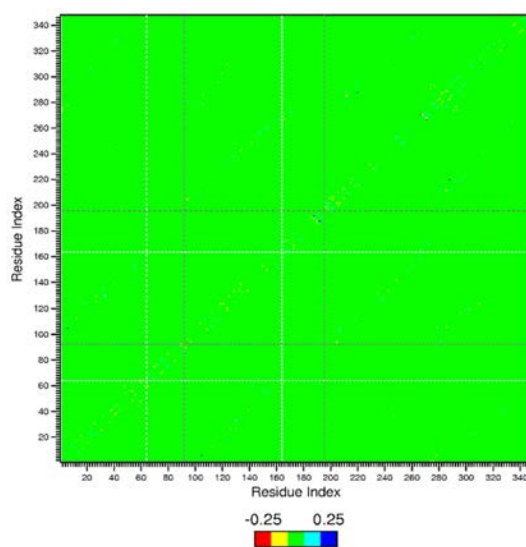
En los siguientes párrafos se pondrá la atención sobre las propiedades estructurales y dinámicas de las cadenas proteicas. La comparación entre los sistemas atomístico e híbrido resulta en perfiles de RMSD equivalentes, sugiriendo que la desviación a partir de la estructura inicial es prácticamente idéntica (Figura 4-12.A). Un perfil similar se obtiene de la evaluación de las fluctuaciones cuadráticas medias, indicando que ambos complejos proteicos presentan una flexibilidad similar (Figura 4-12.B). Las principales desviaciones se localizan dentro de los “loops” no estructurados que conectan la hélice citoplasmática con el DTM de la proteína SB (Figura 4-12.B), mientras que diferencias menores se observan en las regiones estructuradas. La relevancia estructural de esta mayor movilidad en el sistema atomístico puede apreciarse de la superposición de las quince estructuras tomadas en intervalos de un nanosegundo a lo largo de la trayectoria. A partir de la comparación visual entre los confórmeros del sistema FG con el híbrido (Figura 4-12.C), es claro que ambas simulaciones oscilan alrededor de posiciones promedio similares con una dispersión menor en el caso del solvente híbrido. De acuerdo con la buena conservación estructural, la comparación de ambas trayectorias por medio del RMSD de a pares (comparando todas las conformaciones de una simulación contra todas las de la otra), muestra varias regiones de superposición, sugiriendo que el sistema explora configuraciones similares, pero no necesariamente en el mismo momento temporal (Figura 4-14).

Estos resultados indican, como era esperable, una conformación del complejo SNARE bastante rígida. Por ende, se examinaron las interacciones que subyacen a dicha rigidez, evaluando la formación de puentes salinos (Figura 4-12.D,E) así como los enlaces de hidrógeno (Figura 4-12.F). El patrón de interacciones observado es muy similar con algunas variaciones en la fuerza de algunas interacciones individuales. Globalmente, ambas simulaciones exhiben el mismo nivel de cohesión mediada por puentes salinos y enlaces de hidrógeno.

Una visión complementaria a las interacciones electrostáticas presentadas en la Figura 4-12 se puede obtener calculando las matrices de contacto de los complejos, que toma en consideración principalmente los contactos hidrofóbicos.



**Figura 4-12. Sistema del complejo SNARE (propiedades de la proteína).** A- RMSD del complejo proteico en función del tiempo (rojo: FG, cyan: FGG-CG). B- RMSF para cada residuo del complejo proteico, para ambas simulaciones FG y FG-CG (rojo y cyan, respectivamente). Debajo del eje x se indica la correspondencia a cadena polipeptídica que compone el complejo. C- Superposición de 15 conformeros (tomados cada 1 ns de simulación) para el sistema FG (izquierda) y el sistema FG-CG (derecha). La variación de color corresponde a la escala temporal (0ns: azul → 15ns: rojo). D- Representación esquemática de los puentes salinos en ambos sistemas (FG: izquierda y FG-CG: derecha). Los colores de la barras indican el porcentaje de ocupación de cada puente salino durante el tiempo de simulación. Amarillo, naranja, y rojo corresponden a ocupaciones de 0-50%, 50-75% y 75-100%, respectivamente. E- Perfil temporal del número total de puentes salinos para ambos sistemas (FG: rojo y FG-CG: cyan). F- Ídem E para los enlaces de hidrógeno.

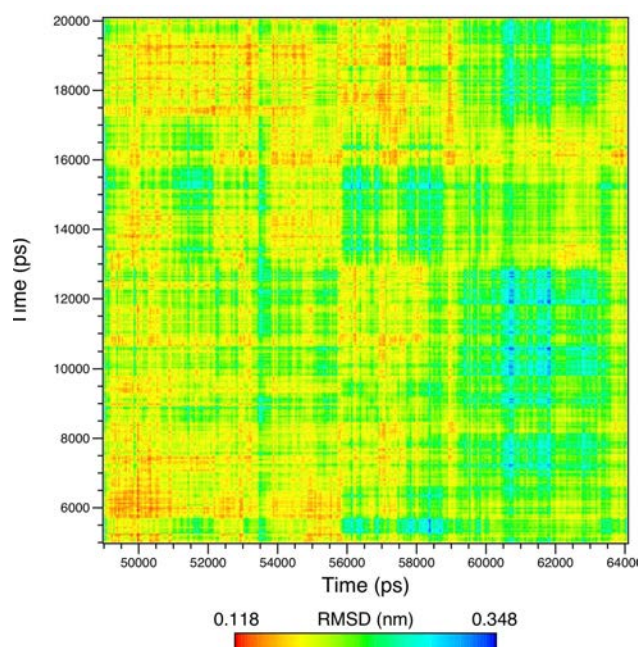


**Figura 4-13. Mapa de contactos del complejo proteico.** Diferencia entre el mapa de auto-interacción de la proteína durante la simulación FG (promedio), respecto del mapa correspondiente a la simulación FG-CG. La escala de colores indica las diferencias en Angstroms para cada residuo de la proteína.

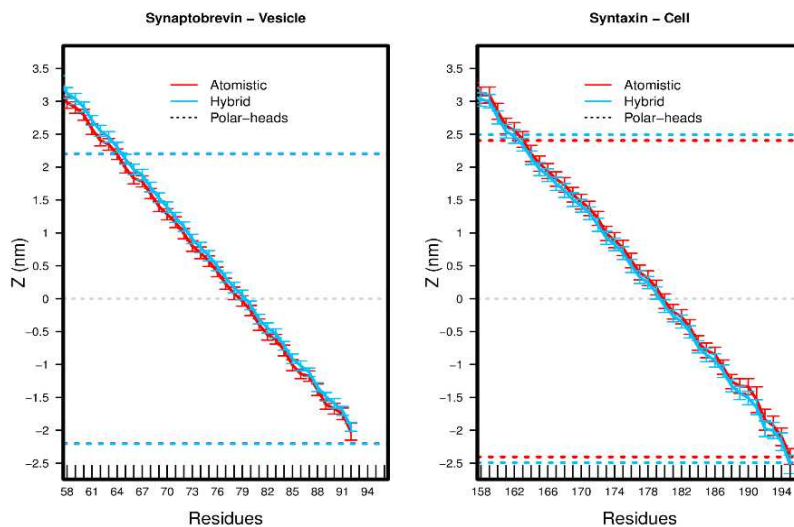
La resta entre ambas matrices de contacto (matriz de diferencias) obtenidas por medio de las simulaciones FG y FG-CG, permite visualizar diferencias en puntos de contacto (asociados con elementos con valor diferente de cero

en dicha matriz de diferencias). Como podía preverse en vista de los resultados anteriores, la matriz de diferencias es prácticamente cero para más del 95% de los contactos, indicando diferencias de interacción despreciables (Figura 4-13). Solo cinco residuos localizados en las regiones terminales de las hélices muestran pequeñas desviaciones (menores a 0.25nm), que están en el mismo orden de magnitud que las fluctuaciones observadas al calcular el RMSF (Figura 4-12.B).

Finalmente, se estudiaron las interacciones proteína-lípido. El cálculo del número de contactos proteína-lípido usando un cutoff de 0.4 nm revela diferencias despreciables dentro del desvío estándar (Tabla 4-5), que pueden asignarse a fluctuaciones térmicas. Más aún, la inserción de los DTM de SB y SX en sus respectivas bicapas es indistinguible cuando se comparan ambos esquemas de solvatación (Figura 4-14). Verificando nuevamente la conservación de las interacciones al utilizar la solvatación de resolución dual.



**Figura 4-14. RMSD bidimensional.** Se compara la estructura del complejo proteico midiendo el RMSD entre los sistemas FG y FG-CG a cada paso de tiempo. La escala de colores mide el RMSD desde valores cercano a 0 nm (rojo) hasta valores cercanos a 0.35 nm (azul).



**Figura 4-14. Inserción de DTM (Dominios Trans-Membrana).** La posición a lo largo del eje z de cada residuo de ambos DTM (SX: bicapa celular y SB: bicapa vesicular) se indican para ambas simulaciones (FG: rojo y FG-CG: cyan). Las líneas punteadas horizontales indican la posición de las cabezas polares de los fosfolípidos.

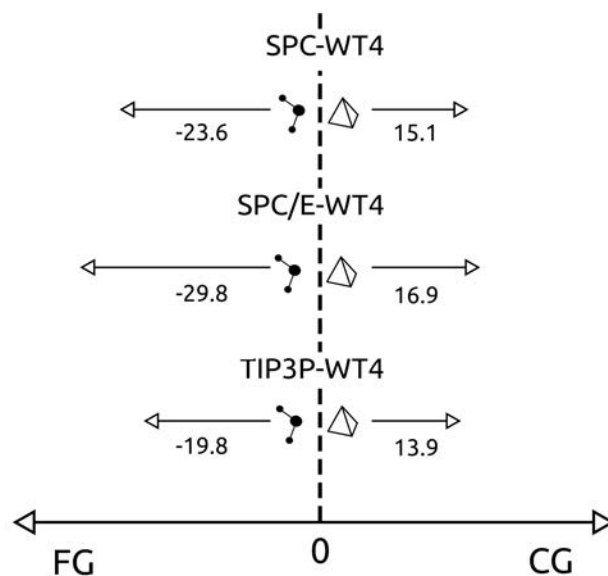
**Extensión a otros modelos de agua FG.** Dado que el WT4 se basa en la estructura tetraédrica del agua y que fue desarrollado en manera independiente de los modelos FG, es esperable que sea también compatible con otros modelos de agua atomísticos como TIP3P<sup>37</sup> y SPC/E (70).<sup>38</sup> La primera evidencia en este sentido proviene nuevamente de la termodinámica de mezcla en el sistema FG-CG. La Tabla 4-6 muestra los resultados obtenidos para los modelos TIP3P y SPC/E en comparación con SPC. Los cálculos de  $\Delta G_{\text{solv}}$  muestran valores negativos de en todos los casos, lo cual muestra que cualquiera de los tres tipos de agua FG presentan cierta tendencia a mezclarse con el WT4 y viceversa. Sin embargo, los valores de  $\Delta\Delta G_{[\text{soluto/solvente FG}]-[\text{soluto/solvente CG}]}$  evidencian, en todos los casos, que cada tipo de agua, sea FG o CG, tiene mayor tendencia a quedarse en su propia fase que a pasar a la otra. Para esquematizar estos datos se recurre al siguiente ejercicio. Tomemos el eje x y supongamos que a la izquierda del cero (valores negativos) se encuentra el solvente FG y a la derecha (valores positivos) se encuentra el solvente CG. Una molécula de agua de la izquierda, es decir FG (lo mismo es válido para una de la derecha (CG)), estará bajo la influencia de dos tendencias, una que la dirige a la fase CG y otra que la dirige a la FG. La resultante de dichas dos tendencias es cuantificada por el valor y signo de  $\Delta\Delta G$  y predice hacia donde se dirigirá la molécula. Por ejemplo, en el caso de una molécula de WT4 en una mezcla TIP3P-WT4, la tendencia a ir hacia la fase FG viene dada por un  $\Delta G_{\text{WT4/TIP3P}} = -36.8 \text{ kJmol}^{-1}$  mientras que la tendencia a ir a la fase CG está

dada por  $\Delta G_{WT4/TIP3P} = -50.7 \text{ kJmol}^{-1}$ , resultando en un valor de  $\Delta\Delta G_{[WT4/TIP3P]-[WT4/TIP3P]} = 13.9 \text{ kJmol}^{-1}$ . Esto significa que en dicha mezcla, la molécula de WT4 tendrá una tendencia neta cuantificada por  $13.9 \text{ kJmol}^{-1}$  en ir hacia la fase CG (derecha). En consecuencia, estas definiciones permiten utilizar al valor de  $\Delta\Delta G$  como cantidad vectorial pues su signo es relacionable sin ambigüedades con la dirección hacia donde se dirigirá una molécula teórica. Es decir se dirigirá hacia la fase FG si  $\Delta\Delta G$  es menor a cero o hacia la fase CG si es mayor. La Figura 4-15 esquematiza los resultados que se obtienen en este marco para los tres modelos de agua FG estudiados. Como se puede observar, la combinación que más tendencia tiene a mezclarse es TIP3P con WT4, mientras que SPC/E-WT4 es el que presenta la mayor tendencia a separarse. A pesar de estas diferencias, las tres representaciones presentan tendencias similares de separación de las fases FG y CG, probando que la aproximación multiescala presentada es útil para los tres modelos SPC, SPC/E y TIP3P.

**Tabla 4-6.** Energía libre de solvatación de agua FG, CG y FG-CG para SPC, SPC/E y TIP3P con WT4.

soluto (s)	solvente (S)	$\Delta G_{\text{solv}}$ ( $\text{kJmol}^{-1}$ )	$\Delta\Delta G_{[\text{soluto/solvente FG}]-[\text{soluto/solvente CG}]^a}$ ( $\text{kJmol}^{-1}$ )
WT4	SPC	-35.6	15.1
	SPC/E	-33.8	16.9
	TIP3P	-36.8	13.9
SPC	WT4	-50.7	0
	WT4	-13.2	-23.6
	SPC	-36.8	0
SPC/E	WT4	-14.7	-29.8
	SPC/E	-44.5	0
TIP3P	WT4	-13.6	-19.8
	TIP3P	-33.4	0

<sup>a</sup> Por conveniencia se considera como estado final el solvente FG y el inicial el solvente CG.



**Figura 4-15.** Esquema representativo de la termodinámica de mezcla de agua FG y agua CG. Los vectores indican tendencia del agua FG o CG en cada tipo de mezcla (SPC-WT4: superior, SPC/E-WT4: medio, TIP3P-WT4: inferior) a ir hacia el solvente FG o CG. El módulo y dirección de tales vectores viene dado por el valor absoluto y signo de  $\Delta\Delta G$ .

## Conclusiones.

La consideración explícita de los efectos de solvatación en los sistemas moleculares es de fundamental relevancia para alcanzar una descripción precisa a nivel atómico de los procesos biofísicos. Sin embargo, la gran abundancia de agua en los sistemas biológicos impone serias limitaciones en las escalas espacio-temporales que pueden estudiarse mediante las simulaciones de dinámica molecular. El uso de aproximaciones CG representa una buena oportunidad para atacar dicho problema al reducir la complejidad de los sistemas atómicos a cambio del precio de perder cierto nivel de precisión.<sup>24</sup> A mitad de camino entre una aproximación completamente FG y una CG, se encuentran los métodos híbridos o de multiescala, que ofrecen una forma alternativa de mantener el detalle atómico en regiones de interés del sistema molecular bajo estudio. En el presente capítulo, se presentó un nuevo esquema de solvatación que pertenece a esta categoría de método híbrido. En éste se utiliza el modelo de agua atomística SPC en combinación con el modelo de agua CG WT4 presentado en el Capítulo II.

Se muestra que esta aproximación de solvatación dual logra una buena descripción de las propiedades físico-químicas del agua pura y de las soluciones de electrolitos. Finalmente, esta aproximación es puesta a prueba con un sistema molecular conteniendo un complejo proteico trimérico, insertado en una doble bicapa lipídica. La comparación sistemática del sistema completamente atomístico con el de solvatación dual arroja esencialmente resultados idénticos en términos de estructura y dinámica tanto de las proteínas como de las membranas. Más aún, las interacciones entre los diferentes componentes del sistema permanecen inalteradas o presentan desviaciones menores, que son compatibles con el ruido térmico.

Claramente, la utilidad del método de solvatación propuesto depende de la potencial aceleración que puede obtenerse respecto una simulación atomística. La aceleración depende de muchos factores como el programa particular empleado, la arquitectura del computador, el nivel de paralelización, etc.. En el caso del sistema SNARE embebido en una doble membrana, los cálculos fueron llevados a cabo usando el programa GROMACS 4.5, en el supercomputador Curie en Francia (sistema x86-64 French Tier0 con picos de desempeño de 2 PetaFlops conteniendo procesadores de 8 cores Intel Sandy Bridge EP E5-2680 de 2.7GHz). Bajo estas condiciones, la aceleración relativa al sistema atomístico vario entre 30% y 50%, dependiendo del número de procesadores utilizados.<sup>39</sup> En la implementación de GROMACS utilizada, la ventaja del método radica casi exclusivamente en el cálculo de las fuerzas.<sup>39</sup> Estos resultados eran esperables del punto de vista de que se ha observado que las simulaciones usando modelos CG son particularmente desafiantes de paralelizar debido a la baja densidad y/o inhomogeneidades.<sup>40</sup> Sin embargo, es esperable que futuras implementaciones puedan mejorar estas

limitaciones. Por ejemplo, un mejoramiento en la eficiencia de etiquetamiento en la construcción de las listas de vecinos para calcular las interacciones, como fue reportado por Meloni et al<sup>41</sup> podría aplicarse al esquema híbrido presentado, aprovechando la ventaja de la difusión más lenta asociada al solvente CG. Por lo tanto la lista de vecinos del solvente CG no necesitaría ser actualizada tan frecuentemente como la correspondiente a las partículas atomísticas. De forma similar, el presente método de solvatación dual se podría ver fuertemente beneficiado de nuevos y prometedores algoritmos de pasos de tiempo múltiples.<sup>42</sup>

En los casos de prueba presentados en este trabajo, el uso del WT4 para representar el agua "bulk" permite una reducción del 45% del número de átomos del sistema, pero esta ganancia aumenta a 65% si sólo se considerase el dominio citoplasmático del complejo SNARE.<sup>29</sup> Dado que la cantidad de agua atomística depende de la superficie expuesta del soluto, los sistemas globulares serían más convenientes que las membranas. En sistemas altamente hidratados como el caso del ribosoma,<sup>43</sup> conteniendo alrededor de 300000 moléculas de agua e iones, el método propuesto permitiría una reducción de alrededor del 75% del tamaño del sistema. Otras aproximaciones mixtas, como es el caso del uso de un campo de reacción ("reaction field") luego de cierta distancia desde el soluto,<sup>44</sup> podrían proveer una ganancia similar. Sin embargo, dicho método no es aplicable para sistemas extendidos como las membranas biológicas. Más aún, el método dual aquí presentado permite el cálculo de interacciones electrostáticas a larga distancia por medio del método PME y podría usarse para disminuir problemas relacionados con las interacciones entre replicas periódicas,<sup>45</sup> dado que aumentos significativos en el tamaño del sistema implicarían solo un modesto costo computacional. Recientemente se presentó un modelo híbrido para simular ADN a dos niveles de resolución en una misma hebra de ADN.<sup>46</sup> La combinación de ambos métodos abre la posibilidad de considerar solo una pequeña región de interés, por ejemplo, una región de interacción proteína-ADN, a nivel atomístico, mientras que el resto de una potencialmente larga fibra de ADN con proteínas, solvente y contra-iones explícitos podría tratarse a nivel CG.

Actualmente, la principal limitante del método reside en el hecho de que, mientras que las moléculas de agua pueden fácilmente salir y volver de la región FG, los iones no pueden. Por lo tanto el número de iones atomísticos a incluir en la simulación debe decidirse al comienzo de la misma y estos solo difundirán por la fase FG. A pesar de que esto no genera efectos espurios, puede anticiparse que esta restricción podría plantear dificultades particulares en el caso de superficies o para sistemas que involucran procesos de permeación.

Por último, cabe destacar que la implementación de este método de solvatación dual es directa en cualquier paquete estándar de simulaciones de dinámica molecular, solo requiriendo los parámetros de interacción del modelo WT4 y de los iones CG.

La extensión presentada para los solventes TIP3P y SPC/E evidencia la transferabilidad general de este método al uso de cualquier campo de fuerza.

## Bibliografía

1. Berendsen, H., Postma, J., Van Gunsteren, W. & Hermans, J. in *Intermolecular Forces* 331–342 (Pullman B. Dordrecht: Reidel, 1981).
2. Darré, L., Machado, M. R., Dans, P. D., Herrera, F. E. & Pantano, S. Another Coarse Grain Model for Aqueous Solvation: WAT FOUR? *Journal of Chemical Theory and Computation* **6**, 3793–3807 (2010).
3. Pizzitutti, F., Marchi, M., Sterpone, F. & Rossky, P. J. How protein surfaces induce anomalous dynamics of hydration water. *J Phys Chem B* **111**, 7584–7590 (2007).
4. Pomata, M. H. H., Sonoda, M. T., Skaf, M. S. & Elola, M. D. Anomalous dynamics of hydration water in carbohydrate solutions. *J Phys Chem B* **553**, 12999–13006 (2009).
5. Laage, D. & Hynes, J. T. A molecular jump mechanism of water reorientation. *Science* **355**, 832–835 (2006).
6. Söllner, T. *et al.* SNAP receptors implicated in vesicle targeting and fusion. *Nature* **062**, 318–324 (1993).
7. Krieger, E. *Parallel computing : architectures, algorithms, and applications.* (IOS Press, 2008).
8. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **18**, 1463–1472 (1997).
9. Miyamoto, S. & Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry* **13**, 952–962 (1992).
10. Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **81**, 3684 (1984).
11. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **98**, 10089 (1993).
12. Essmann, U. *et al.* A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **103**, 8577 (1995).
13. Herrero, C. P. Compressibility of solid helium. *Journal of Physics: Condensed Matter* **20**, 295230 (2008).
14. Chen, F. & Smith, P. E. Simulated surface tensions of common water models. *The Journal of Chemical Physics* **126**, 221101 (2007).
15. Yu, H., Hansson, T. & Van Gunsteren, W. F. Development of a simple, self-consistent polarizable model for liquid water. *The Journal of Chemical Physics* **118**, 221 (2003).
16. Luo, Y. & Roux, B. Simulation of Osmotic Pressure in Concentrated Aqueous Salt Solutions. *The Journal of Physical Chemistry Letters* **1**, 183–189 (2010).
17. Van Gunsteren, W. & Berendsen, H. *Gromos-87 Manual.* (Biomos BV: Groeningen, The Netherlands, 1987).
18. Mukhopadhyay, P., Monticelli, L. & Tieleman, D. P. Molecular dynamics simulation of a palmitoyl-oleoyl phosphatidylserine bilayer with Na<sup>+</sup> counterions and NaCl. *Biophys. J.* **86**, 1601–1609 (2004).
19. Vermeer, L. S., De Groot, B. L., Réat, V., Milon, A. & Czaplicki, J. Acyl chain order parameter profiles in phospholipid bilayers: computation from molecular dynamics simulations and comparison with 2H NMR experiments. *Eur. Biophys. J.* **36**, 919–931 (2007).
20. Darré, L., Machado, M. R. & Pantano, S. Coarse-grained models of water. *Wiley Interdisciplinary Reviews: Computational Molecular Science* n/a–n/a (2012).doi:10.1002/wcms.1097
21. Neumann, M. Computer simulation and the dielectric constant at finite wavelength. *Molecular Physics* **57**, 97–121 (1986).
22. Murrell, J. N. & Jenkins, A. D. *Properties of liquids and solutions.* (Wiley & Sons, 1994).
23. Fayer, M. D. Dynamics of Water Interacting with Interfaces, Molecules, and Ions. *Accounts of Chemical Research* **45**, 3–14 (2012).
24. Wang, H., Junghans, C. & Kremer, K. Comparative atomistic and coarse-grained study of water: what do we lose by



- coarse-graining? *Eur Phys J E Soft Matter* **28**, 221–229 (2009).
25. Kuo, I.-F. W. *et al.* Liquid Water from First Principles: Investigation of Different Sampling Approaches. *The Journal of Physical Chemistry B* **108**, 12990–12998 (2004).
  26. Wang, J. Tracer-diffusion in Liquids. III. The Self-diffusion as Sodium Ion in Aqueous Sodium Chloride Solutions. *J. Am. Chem. Soc.* **74**, 1612–1615 (1952).
  27. Chen, X. *et al.* Three-dimensional structure of the complexin/SNARE complex. *Neuron* **33**, 397–409 (2002).
  28. Risselada, H. J. & Grubmüller, H. How SNARE molecules mediate membrane fusion: recent insights from molecular simulations. *Curr. Opin. Struct. Biol.* **22**, 187–196 (2012).
  29. Durrieu, M.-P., Lavery, R. & Baaden, M. Interactions between neuronal fusion proteins explored by molecular dynamics. *Biophys. J.* **94**, 3436–3446 (2008).
  30. Arkhipov, A., Yin, Y. & Schulten, K. Four-scale description of membrane sculpting by BAR domains. *Biophys. J.* **95**, 2806–2821 (2008).
  31. Yersin, A. *et al.* Interactions between synaptic vesicle fusion proteins explored by atomic force microscopy. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 8736–8741 (2003).
  32. Jurkiewicz, P., Cwiklik, L., Vojtíšková, A., Jungwirth, P. & Hof, M. Structure, dynamics, and hydration of POPC/POPS bilayers suspended in NaCl, KCl, and CsCl solutions. *Biochim. Biophys. Acta* **1818**, 609–616 (2012).
  33. Herrera, F. E. & Pantano, S. Structure and dynamics of nano-sized raft-like domains on the plasma membrane. *J Chem Phys* **136**, 015103 (2012).
  34. Böckmann, R. A., Hac, A., Heimburg, T. & Grubmüller, H. Effect of sodium chloride on a lipid bilayer. *Biophys. J.* **85**, 1647–1655 (2003).
  35. Chavent, M. *et al.* GPU-accelerated atom and dynamic bond visualization using hyperballs: A unified algorithm for balls, sticks, and hyperboloids. *Journal of Computational Chemistry* **32**, 2924–2935 (2011).
  36. Herrera, F. E. & Pantano, S. Salt induced asymmetry in membrane simulations by partial restriction of ionic motion. *J Chem Phys* **130**, 195105 (2009).
  37. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**, 926 (1983).
  38. Berendsen, H. J. C., Grigera, J. R. & Straatsma, T. P. The missing term in effective pair potentials. *The Journal of Physical Chemistry* **91**, 6269–6271 (1987).
  39. Darré, L., Tek, A., Baaden, M. & Pantano, S. Mixing Atomistic and Coarse Grain Solvation Models for MD Simulations: Let WT4 Handle the Bulk. *Journal of Chemical Theory and Computation* 120615092245008 (2012).doi:10.1021/ct3001816
  40. Van der Spoel, D. & Hess, B. GROMACS-the road ahead. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **1**, 710–715 (2011).
  41. Meloni, S., Rosati, M. & Colombo, L. Efficient particle labeling in atomistic simulations. *The Journal of Chemical Physics* **126**, 121102 (2007).
  42. Omelyan, I. & Kovalenko, A. Interpretation of atomic motion in flexible molecules: Accelerating molecular dynamics simulations. *Physical Review E* **85**, (2012).
  43. Yam, W. K. & Wahab, H. A. Molecular insights into 14-membered macrolides using the MM-PBSA method. *J Chem Inf Model* **49**, 1558–1567 (2009).
  44. Petraglio, G., Ceccarelli, M. & Parrinello, M. Nonperiodic boundary conditions for solvated systems. *J Chem Phys* **123**, 044103 (2005).
  45. Hünenberger, P. H. & McCammon, J. A. Effect of artificial periodicity in simulations of biomolecules under Ewald boundary conditions: a continuum electrostatics study. *Biophys. Chem.* **78**, 69–88 (1999).
  46. Machado, M. R., Dans, P. D. & Pantano, S. A hybrid all-atom/coarse grain model for multiscale simulations of DNA. *Physical Chemistry Chemical Physics* **13**, 18134 (2011).