

DEPARTAMENTO DE BIOQUÍMICA Y BIOLOGÍA MOLECULAR

**Dinámica de sistemas de interés biológico.
Estudios de flexibilidad y estabilidad en sistemas
de puente de hidrógeno.**

JOSÉ RAMÓN BLAS PASTOR

2006

Programa de doctorado en Biotecnología
Departamento de Bioquímica i Biología Molecular
Universidad de Barcelona
Bienio 2000-2001

**Dinámica de sistemas de interés biológico. Estudios de
flexibilidad y estabilidad en sistemas de puente de hidrógeno.**

Esta tesis ha sido realizada por el licenciado José Ramón Blas Pastor bajo la dirección del Dr. Modesto Orozco López, catedrático del Departamento de Bioquímica y Biología Molecular de la Universidad de Barcelona, y del Dr. Francisco Javier Luque Garriga, catedrático del Departamento de Físico-Química de la Universidad de Barcelona, para optar al título de doctor por la Universidad de Barcelona en el Programa de Doctorado en Biotecnología.

Dr. Modesto Orozco López

Dr. F. Javier Luque Garriga

A mis padres y a Inma

Índice

| | | |
|-------------------|---|-------------------|
| <u>I</u> | <u>Introducción.....</u> | <u>1</u> |
| <u>II</u> | <u>Métodos: Modelado molecular de sistemas de interés biológico...5</u> | |
| 2.1 | Métodos basados en la teoría cuántica..... | 6 |
| 2.2 | Descripción de sistemas químicos basada en la mecánica clásica..... | 24 |
| 2.3 | Simulación y análisis del movimiento de los sistemas químicos..... | 31 |
| 2.4. | Cálculo de la energía libre..... | 39 |
| 2.5. | Análisis de los sistemas simulados..... | 42 |
| 2.6. | Referencias..... | 49 |
| <u>III</u> | <u>Sistemas de puente de hidrógeno en biología.....</u> | <u>57</u> |
| 3.1 | El puente de hidrógeno..... | 58 |
| 3.2. | Uniones dirigidas por puente de hidrógeno en sistemas modelo: interacción anión \cdots calix-[4]-pirrol..... | 65 |
| 3.3. | Comportamiento de sistemas biológicos basados en puente de hidrógeno: estructura y flexibilidad de los ácidos nucleicos | 73 |
| 3.4 | Referencias | 105 |
| <u>IV</u> | <u>Resultados.....</u> | <u>117</u> |
| 4.1 | Estudio de la interacción de puente de hidrógeno en sistemas modelo: unión de aniones a calix-[4]-pirrol..... | 118 |
| 4.2. | Dinámica esencial del ADN en disolución..... | 123 |
| 4.3. | Tautomería de bases no canónicas del ADN..... | 126 |
| <u>V</u> | <u>Discusión de resultados.....</u> | <u>131</u> |
| 5.1 | Estudio de la unión de aniones a calix-[4]-pirrol..... | 132 |
| 5.2 | Estudio de la flexibilidad y las propiedades tautoméricas del ADN..... | 139 |
| 5.3 | Referencias bibliográficas..... | 153 |
| <u>VI</u> | <u>Conclusiones.....</u> | <u>157</u> |
| <u>VII</u> | <u>Otras publicaciones.....</u> | <u>161</u> |
| | Agradecimientos..... | 165 |

1. Introducción

Las interacciones por puente de hidrógeno resultan cruciales en la definición de numerosos procesos de relevancia biológica. Su reconocida importancia como elementos moduladores de la estructura de macromoléculas, su papel en el reconocimiento molecular de especies químicas (iones, fármacos,...), o su actuación en los mecanismos catalíticos de enzimas, resultan ejemplos evidentes de ello.

Numerosas metodologías experimentales aportan descripciones muy valiosas del comportamiento de los sistemas modulados por puente de hidrógeno. Junto a ellas, los métodos teóricos constituyen un complemento necesario de cara a obtener una descripción detallada de los aspectos estructurales, energéticos y dinámicos de estos sistemas, así como para simular condiciones de entorno y situaciones químicas concretas que no resultan de fácil acceso a la metodología experimental.

La presente tesis se ha centrado en la descripción desde una perspectiva teórica de una serie de sistemas de interés biológico cuyas propiedades vienen determinadas en alto grado por la presencia de puentes de hidrógeno. Los estudios realizados pueden estructurarse en tres bloques temáticos.

En un primer bloque, se analiza el papel que desempeña la interacción por puente de hidrógeno como elemento clave en la modulación del reconocimiento selectivo de especies aniónicas por receptores orgánicos. En este punto, nuestro interés está orientado esencialmente a determinar por un lado la estructura del complejo de interacción entre el receptor y el anión, así como la estabilidad relativa de dicha interacción frente a diversos aniones. Por otra parte, un objetivo complementario consiste en determinar la flexibilidad intrínseca del receptor orgánico, con el fin de determinar el posible coste conformacional asociado a la reorganización de su estructura como paso previo al reconocimiento del anión.

En un segundo bloque, se estudia el comportamiento dinámico de los ácidos nucleicos, cuya estructura, estabilidad y flexibilidad viene determinado en parte por la presencia de interacciones de puente de hidrógeno. A su vez, dichas interacciones son claves en el mantenimiento del código genético, así como en la posibilidad de poder reconocer la secuencia de bases a lo largo de los surcos del DNA, lo cual subyace en el reconocimiento selectivo frente a pequeños ligandos o proteínas. En particular, nuestro

interés se orienta al estudio de la flexibilidad conformacional del dúplex de DNA mediante una serie de herramientas teóricas que permitan extraer información directa acerca de su naturaleza dinámica.

Finalmente, en un tercer bloque se examina el efecto que tiene el mantenimiento de las pautas de puente de hidrógeno en la conservación del código genético y en el funcionamiento de aplicaciones biotecnológicas basadas en ácidos nucleicos sintéticos. Así, nuestro objetivo se orienta en ese caso hacia la comprensión de las propiedades mutagénicas de la isoguanina, que está estrechamente relacionadas con su enorme versatilidad tautomérica, que le permite definir distintos patrones de interacción basada en puentes de hidrógeno. Asimismo, también ha sido nuestro objetivo examinar las características químicas de nucleobases expandidas, basadas en la modificación de las bases naturales mediante inserción o adición de un anillo de benzeno, como paso previo al estudio de su posible utilización como elementos estructurales en el diseño de dúplexes modificados de potencial valor biotecnológico.

2. MÉTODOS: modelado molecular de sistemas de interés biológico

2.1. Métodos basados en la teoría cuántica

2.1.1. Introducción y métodos *ab initio*

Seguidamente se describirá brevemente una serie de métodos basados en la mecánica cuántica que han sido utilizados en la realización de esta tesis. Una presentación detallada y didáctica de dichos métodos puede encontrarse en otras fuentes (Szabo and Ostlund 1982; Leach 1996; Cramer 2002).

La mecánica cuántica se basa en una serie de postulados desarrollados a principios del siglo XX. Un buen punto de partida para su explicación es el concepto de función espectral, que presenta la siguiente forma,

$$\nu\psi = e\psi \quad (2.1)$$

y establece que para todo sistema existe una función de onda o *función propia* (Ψ) y que determinados operadores (ν), al actuar sobre ella, devuelven en forma de *valor propio* (e) las propiedades observables del sistema. El operador cuyo valor propio es la energía del sistema se denomina Hamiltoniano y la correspondiente ecuación se bautizó como *ecuación de Schrödinger*.

En ausencia de factores adicionales (presencia de campos eléctricos o magnéticos externos, efectos de tipo relativista,...) este operador se expresa como

$$\hat{H} = -\sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{A=1}^M \frac{1}{2m_A} \nabla_A^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>1}^N \frac{1}{r_{ij}} + \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{R_{AB}} \quad (2.2)$$

donde ∇^2 es el operador laplaciano, que contiene la derivada respecto a las coordenadas de cada electrón (i) y de cada núcleo (A), m_A es la relación de la masa del núcleo A con la masa del electrón y Z_A es el número atómico del núcleo A . Los dos primeros términos en la ecuación 2.2 expresan la energía cinética de electrones y núcleos, respectivamente. El tercer término contiene la energía de interacción entre núcleos y electrones, mientras

que los dos últimos albergan las contribuciones correspondientes a la interacción entre electrones y entre núcleos, respectivamente.

2.1.1.1. Aproximación de Born-Oppenheimer

La ecuación 2.2 evalúa conjuntamente el movimiento de núcleos y electrones. Sin embargo, dado que los electrones presentan enormes diferencias en masa y velocidad con respecto a los núcleos, para el estudio de un gran número de propiedades de interés químico resulta lícito desacoplar los movimientos nuclear y electrónico. Esta simplificación, conocida como aproximación de Born-Oppenheimer (Born and Oppenheimer 1927) conduce al Hamiltoniano electrónico (ec. 2.3).

$$\hat{H}_{el} = -\sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>1}^N \frac{1}{r_{ij}} \quad (2.3)$$

Empleando este operador, podemos expresar la ecuación de Schrödinger electrónica como

$$(H_{el} + V_N)\Psi_{el}(r_i; R_A) = E_{el}\Psi_{el}(r_i; R_A) \quad (2.4)$$

donde el término V_N se refiere a la energía de repulsión entre núcleos, que viene dada por

$$V_N = \sum_a \sum_{b>a} \frac{Z_a Z_b}{R_{ab}} \quad (2.5)$$

Para una configuración nuclear dada, el valor de V_N es constante, y la ecuación 2.4 se resuelve considerando las coordenadas electrónicas (r_i) como variables independientes y las nucleares como parámetros fijos, prescindiendo del término V_N , que se añade al final al valor propio obtenido para la energía.

El error introducido por la aproximación de Born-Oppenheimer en el cálculo de la energía de los estados fundamentales de moléculas poliatómicas es relativamente

pequeño, siendo algo mayor, aunque aceptable, en el estudio de estados electrónicos excitados. Por ello, su uso está muy generalizado.

2.1.1.2. El tratamiento del problema electrónico: método Hartree-Fock

Al no existir solución analítica exacta de la ecuación de Schrödinger para un sistema polielectrónico, su resolución se basa en la construcción de funciones de onda aproximadas. El procedimiento más empleado, que establece la base para estrategias de cálculo de más alto nivel, es el método de campo autoconsistente de Hartree-Fock.

2.1.1.2.1. Creación de una función de prueba: el determinante de Slater

La función de onda para un sistema polielectrónico se construye como una combinación de funciones monoeléctricas, denominadas orbitales moleculares, escogidas de forma que sean ortonormales entre sí. Cada orbital molecular (χ_i) es el producto de una función espacial (φ) y una de espín (α o β) y constituye un *spinorbital*.

$$\chi_i(q_1) = \varphi_i(r_1)\alpha(\omega_1) \quad \text{o} \quad \chi_i(q_1) = \varphi_i(r_1)\beta(\omega_1) \quad (2.6)$$

Considerando un sistema de N electrones, su función de onda se aproxima como una combinación de N spinorbitales monoeléctricas sujetos al principio de antisimetría, constituyendo el determinante de Slater, que en su notación compacta se expresa como

$$\Psi(q_1, q_2, \dots, q_N) = |\chi_i(q_1), \chi_j(q_2), \dots, \chi_k(q_N)| \quad (2.7)$$

Este determinante ubica N electrones en N spinorbitales, pero no especifica cuál de las $N!$ posibles combinaciones de electrones describe el sistema polielectrónico, sino que cada una de ellas contribuye a la función de onda. El determinante se anula si existe más de un electrón en un *spinorbital*, justificando el principio de exclusión de Pauli. Finalmente, también recoge la “correlación de intercambio electrónico”, que correlaciona el movimiento de pares de electrones con spin idéntico.

2.1.1.2.2. Aplicación del principio variacional

Existen infinitas funciones propias (Ψ_i), con su valor propio asociado (E_i), que resuelven la ecuación de Schrödinger para un operador Hamiltoniano. Dado que éste es hermítico (sus valores propios son reales y las funciones propias ortonormales) y que sus funciones propias constituyen un conjunto completo, cualquier función Φ que cumpla las condiciones de contorno se puede escribir como una combinación lineal de las funciones propias Ψ_i :

$$\Phi = \sum_i c_i \Psi_i \quad (2.8)$$

Según el principio variacional, la energía asociada a cualquier función propia Φ será siempre superior a la energía del estado fundamental del sistema (E_0),

$$E = \frac{\int \Phi H \Phi d\tau}{\int \Phi^2 d\tau} \geq E_0 \quad (2.9)$$

Ello, pues, permite definir un criterio para evaluar la calidad de la función de onda de prueba en base a la minimización de la energía del estado fundamental.

2.1.1.2.3. El método de campo autoconsistente

La resolución de la ecuación de Schrödinger mediante el principio variacional hace uso del método de campo autoconsistente (*Self Consistent Field; SCF*). Partiendo de la descripción de un sistema de N electrones mediante un determinante de Slater, este procedimiento conduce a la siguiente ecuación de valores propios

$$f(i)\chi(q_i) = \varepsilon\chi(q_i) \quad (2.10)$$

donde $f(i)$ es un operador efectivo monoeléctrico, denominado *operador de Fock*, que puede expresarse matemáticamente, para un sistema de M núcleos, como

$$f(i) = -\frac{1}{2} \nabla_i^2 - \sum_{A=1}^M \frac{Z_A}{r_{iA}} + v^{HF}(i) \quad (2.11)$$

En esta ecuación, la energía cinética del electrón i -ésimo queda recogida en el primer término, su interacción con los núcleos atómicos en el segundo y la repulsión electrónica queda descrita en base al potencial efectivo creado por el resto de electrones, indicado como $v^{HF}(i)$.

La solución para un sistema de muchos electrones se simplifica, pues, transformándose en un problema monoeléctrico. Ahora bien, el potencial electrónico efectivo para un electrón i ($v^{HF}(i)$) depende de los spinorbitales del resto de electrones, por lo que la ecuación 2.11 ha de resolverse de forma iterativa. El conjunto de *spinorbitales* Hartree-Fock ortonormales que minimizan la energía constituyen la aproximación variacional más precisa a la función de onda del estado fundamental del sistema.

En la práctica se recurre a la aproximación LCAO (*Linear Combination of Atomic Orbitals*) propuesta por Roothaan (Roothaan 1951), donde cada orbital molecular (MO) χ se expresa como una combinación lineal de orbitales atómicos (φ)

$$\chi_i = \sum_{\mu=1}^N c_{\mu i} \varphi_{\mu} \quad (2.12)$$

donde cada uno de los N orbitales atómicos (conjunto de base) contribuye al orbital molecular según el coeficiente $c_{\mu i}$ asociado.

La resolución del método MO-LCAO conduce a resolver la siguiente ecuación matricial, que se trata de una ecuación de autovalores en la que \mathbf{C} y $\boldsymbol{\varepsilon}$ se refieren a los coeficientes de los orbitales moleculares expresados en base a orbitales atómicos y energías orbitálicas, respectivamente, \mathbf{S} representa la matriz de solapamiento y \mathbf{F} es la matriz de Fock.

$$\mathbf{FC} = \mathbf{SC}\boldsymbol{\varepsilon} \quad (2.13)$$

Aumentando el tamaño del conjunto de funciones de base, la energía obtenida tiende al denominado *límite Hartree-Fock*. Por otra parte, cabe distinguir el formalismo RHF (*Restricted Hartree-Fock*), que considera idéntica la componente espacial de los *spinorbitales* de los dos miembros de un par electrónico, con independencia del spin, y el formalismo UHF (*Unrestricted Hartree-Fock*), donde la parte espacial varía dependiendo del spin.

2.1.1.2.4. Funciones de base

La identificación de funciones de base que proporcionen una descripción de la función de onda cercana al límite Hartree-Fock de manera eficiente ha supuesto un avance crucial en la aplicación de la mecánica cuántica a sistemas químicos. Inicialmente, se buscaron funciones similares a los orbitales atómicos del átomo de hidrógeno (funciones hidrogenoides), denominadas orbitales de Slater (STO; ec. 2.14) (Slater 1930),

$$\phi_i(\alpha, n, l, m; r, \theta, \varphi) = Nr^{n-1} e^{-\alpha r} \cdot Y_l^m(\theta, \varphi) \quad (2.14)$$

donde N es una constante de normalización, α designa el factor exponencial, r , θ y φ son coordenadas esféricas e Y_l^m es la parte referida al momento angular. Las letras n , l y m indican, respectivamente, los números cuánticos principal, momento angular y magnético. Estas funciones convergen rápidamente pero no resultan adecuadas para un cálculo eficiente de las integrales bielectrónicas requeridas en el método HF.

La sustitución de la función radial $e^{-\alpha r}$ por $e^{-\alpha r^2}$ (Boys 1950) conduce a orbitales con forma de funciones gaussianas (GTO; Gaussian-type Orbitals), que difieren de un STO en que la parte del momento angular se expresa como una función simple de las coordenadas cartesianas (x, y, z) modificadas por unos exponentes (l, m, n) , que ya no corresponden a números cuánticos, y el factor r^{n-1} se ha eliminado.

$$g(\alpha, l, m, n; x, y, z) = Ne^{-\alpha r^2} x^l y^m z^n \quad (2.15)$$

Las propiedades matemáticas de orbitales GTO simplifican enormemente el problema de cálculo de integrales planteado para los STOs. No obstante, dado que la forma de un GTO varía respecto a la de una función hidrogenoide en las zonas muy cercana y muy distante al núcleo atómico, se suelen emplear combinaciones lineales de varios GTOs (*primitivas*) para describir cada STO (*función de base contraída*). Los trabajos de Pople y colaboradores (Hehre et al. 1969) realizaron una optimización sistemática de los parámetros (exponentes y coeficientes) de cara a reproducir STOs mediante combinación lineal de GTOs para una gran variedad de átomos.

Siguiendo este procedimiento, la función de onda puede describirse empleando diferentes conjuntos de funciones de base. Si cada orbital atómico queda descrito por una única función de base, se trata de una *base mínima*, mientras que la consideración de dos o tres funciones de base para cada orbital atómico conduce a bases *doble (DZ)* y *triple (TZ)* zeta, respectivamente. Si se hace una distinción entre los orbitales de valencia, para los que se emplean dos o más funciones, y los de capas internas, que se describen con una única función, se obtienen bases *split-valence*. El desarrollo de estas bases se debe fundamentalmente a Pople y colaboradores (Hariharan and Pople 1973; Hehre et al. 1986), destacando las bases M-NG, donde M indica el número de primitivas utilizadas para cada función de base contraída de los orbitales internos y N es el análogo para los orbitales de valencia. Así, la base 6-31G emplea seis gaussianas para construir un único orbital en la descripción de las capas internas, mientras que representa la capa de valencia mediante dos orbitales, uno con tres y otro con una función gaussiana. Una mejor descripción se obtiene mediante funciones de polarización (de momento angular *p* para los átomos de hidrógeno, y *d* o *f* para átomos pesados) y difusas (que, al presentar exponentes muy bajos, permiten una deslocalización hacia zonas alejadas de los núcleos). A continuación se detallan las características de las principales bases empleadas en el presente trabajo:

- 6-31G(d) → a la base 6-31G se le añade una función de polarización de tipo *d*, para mejorar la descripción de los átomos pesados
- 6-31G(d,p) → se añade otra función de polarización, esta vez de tipo *p* sobre los hidrógenos
- 6-31+G(d,p) → añade un tipo de funciones difusas (Clark et al. 1983; Frisch et al. 1984) a la descripción de los átomos pesados

- 6-311++G(d,p) → para los átomos de la primera fila de la tabla periódica emplea una base 6-311G según la nomenclatura descrita anteriormente, y el doble signo “+” indica la adición de funciones difusas sobre los átomos pesados y los hidrógenos.
- aug-cc-pVDZ → se trata de las bases de Dunning (Dunning Jr. 1989; Kendall et al. 1992; Woon and Dunning Jr. 1993; Peterson et al. 1994), que incluyen funciones de polarización, empleando en este caso una descripción *DZ* para los electrones de valencia, y el prefijo “aug” indica la adición de funciones difusas.
- aug-cc-pVTZ → difiere de la base anterior en la descripción de la capa de valencia, que es de tipo *TZ*.

2.1.1.3. Introducción de la correlación electrónica

En un cálculo basado en la aproximación Hartree-Fock, la interacción entre electrones se considera únicamente de un modo promediado. Es necesario recurrir a los procedimientos *post-Hartree-Fock* para introducir la correlación electrónica, existiendo diversas estrategias.

2.1.1.3.1. Interacción de configuraciones (CI)

Se basa en la descripción de estados excitados del sistema mediante la construcción de nuevos determinantes de Slater obtenidos al sustituir uno o varios orbitales ocupados del determinante original por orbitales virtuales. Una combinación lineal de todos los determinantes de Slater de los diferentes estados excitados (Ψ_1, Ψ_2, \dots), conduce a la expresión de la función de onda correlacionada (Ψ).

$$\Psi = c_0\Psi_0 + c_1\Psi_1 + c_2\Psi_2 + \dots \quad (2.16)$$

La aplicación más exhaustiva de este método implica tener en cuenta todas las permutaciones de todos los electrones del sistema en todos los orbitales (*fullCI*). Dado que para la mayoría de sistemas habituales su coste computacional es prohibitivo, suele limitarse el número de excitaciones permitidas. De esta forma, cabe hablar de excitaciones simples (CIS) y procedimientos que incorporan sucesivamente excitaciones

dobles (CISD), triples (CISDT) y cuádruples (CISDTQ), así como la restricción de las excitaciones a los electrones de la capa de valencia (aproximación *frozen core*). Estrictamente, la metodología CI sólo permite la optimización de los coeficientes (c_0, c_1, c_2, \dots) asociados a cada determinante de Slater en la ecuación 2.16. Existen estrategias alternativas que, adicionalmente, contemplan la variación de los coeficientes de las funciones de base de cada determinante por separado, mejorando la descripción de la función de onda. Se trata de los métodos multiconfiguracionales (*MultiConfigurational Self-Consistent Field; MCSCF*).

Una limitación de los métodos CI es el hecho de que, salvo en el caso de una aplicación *fullCI*, no son consistentes de tamaño, lo cual motivó la elaboración de metodologías como *Quadratic Configuration Interaction* (Pople et al. 1987).

2.1.1.3.2. Teoría perturbacional de Møller-Plesset (MP)

Se trata de una aplicación de la teoría de perturbaciones a sistemas químicos polielectrónicos (Moller and Plesset 1934). En esta estrategia, el Hamiltoniano se divide en dos partes, una llamada *de referencia*, de la que conocemos su función de onda y valor propio, y otra llamada *de perturbación*, desconocida pero mucho menor que la anterior. En concreto, como parte de referencia (H_0) se escoge el Hamiltoniano HF y, para obtener el Hamiltoniano final, se le suma un término acoplado a un parámetro perturbacional (λ), que puede ser expandido en una serie de términos ($\lambda^1 V_1 + \lambda^2 V_2 + \lambda^3 V_3 + \dots$). De esta forma, se puede expresar la función de onda perturbada y su energía correspondiente como

$$\Psi = \Psi^{(0)} + \lambda\Psi^{(1)} + \lambda^2\Psi^{(2)} + \dots \quad (2.17)$$

$$E = E^{(0)} + \lambda E^{(1)} + \lambda^2 E^{(2)} + \dots \quad (2.18)$$

La suma de la energía procedente de los dos primeros términos ($E^{(0)} + \lambda E^{(1)}$) es la energía de referencia o del primer nivel. A partir de esta referencia, las correcciones para la correlación electrónica empiezan en el orden 2, dando lugar a cálculos MP2, MP3, MP4,...

La metodología *MP* no presenta el problema de la consistencia de tamaño, si bien su naturaleza perturbacional no permite asegurar que el valor de energía obtenido sea superior al valor real del estado fundamental.

2.1.1.3.3. Metodología de clusters acoplados (Coupled cluster; CC)

Partiendo de una filosofía similar a la de los métodos perturbativos (Cizek 1966; 1969), esta estrategia introduce la novedad de expresar la función de onda como una ecuación exponencial

$$\Psi_{CC} = \left(1 + T + \frac{1}{2!} T^2 + \frac{1}{3!} T^3 + \dots \right) \Psi_0 \quad (2.19)$$

en la que T es un sumatorio de n operadores de cluster (T_i).

$$T = \sum_{i=1}^n T_i \quad (2.20)$$

La aplicación de cada operador de cluster T_i sobre la función de onda de referencia genera todas las excitaciones de orden i . Por ello, análogamente a lo que sucedía para los métodos CI, la aplicación de un operador T de muchos términos no resulta computacionalmente viable salvo en sistemas muy pequeños, por lo que suele truncarse la función de onda CC (Ψ_{CC}) en un cierto número de términos o *nivel de excitación*. Pueden realizarse aplicaciones que incluyan excitaciones dobles (CCD) (Pople et al. 1978), y aditivamente simples y dobles (CCSD) (Purvis and Bartlett 1982) y considerar perturbativamente el efecto de triples (CCSD(T)) (Pople et al. 1987) .

La función de onda CCD contiene los mismos términos que la función de onda CI de mismo orden, diferenciándose de ésta en una serie de términos adicionales que hacen que la estrategia CCD sea consistente con el tamaño del sistema. Cabe resaltar, como diferencia respecto a los métodos MP, que en CCD se considera la suma de las dobles excitaciones a orden infinito, de forma que quedan implícitamente incluidas las excitaciones cuádruples, resultando en una determinación más precisa de la energía de

correlación que, adicionalmente, converge más rápidamente que la serie de cálculos MPn. La estrategia CC, no obstante, es computacionalmente muy cara.

2.1.2. Métodos basados en la teoría del funcional de la densidad

Provinientes de ideas ya formuladas por los padres de la física cuántica, estos métodos encuentran realce a partir de los teoremas de Hohenberg-Kohn (Hohenberg and Kohn 1964; Parr and Yang 1989), que establecen que la energía del estado fundamental de un sistema de N electrones es un funcional de su densidad electrónica.

La teoría de Kohn-Sham (Kohn and Sham 1965) calcula la energía electrónica como un sumatorio de cuatro términos

$$E = E^T + E^V + E^J + E^{XC} \quad (2.21)$$

que incluyen energía cinética (E^T), repulsiones-atracciones del conjunto de núcleos y electrones ($E^V + E^J$) y un término referido a la energía de intercambio-correlación (E^{XC}). Exceptuando la repulsión internuclear, todos los elementos incluidos en esta descripción dependen de la densidad electrónica (ρ) y pueden aproximarse mediante integrales que incluyen la densidad electrónica como función de las coordenadas espaciales ($\rho(r)$), distinguiendo la referida a spin α o β , o el gradiente de ésta $\nabla\rho(r)$. En esta notación, podría expresarse el término de intercambio-correlación como

$$E^{XC}(\rho) = \int f[\rho_\alpha(r), \rho_\beta(r), \nabla\rho_\alpha(r), \nabla\rho_\beta(r)] d^3r \quad (2.22)$$

Este término se desglosa para su cálculo en dos partes: una de intercambio y una de correlación, pudiendo expresarse cada una mediante funcionales distintos. Las aproximaciones más empleadas pueden clasificarse en dos tipos: la aproximación de densidad local (LDA) (Vosko et al. 1980) y la aproximación de densidad no-local (GDA) (Perdew and Wang 1992). La nomenclatura de los métodos DFT se asigna según los funcionales utilizados para cada una de estas partes.

La precisión de los cálculos DFT ha mejorado notablemente en los últimos años. Mediante la aplicación de funcionales desarrollados recientemente (Tao et al. 2003a; b; 2004a; b), pueden alcanzarse errores inferiores a 0.001 Å para distancias de enlace y del orden de 0.2 eV en la determinación de energías moleculares. Por ello, su calidad es comparable a cálculos MP2, si bien manteniendo un coste computacional reducido, por lo que su uso es muy frecuente en sistemas de mediana y gran magnitud.

El funcional empleado con más frecuencia en este trabajo ha sido B3LYP, que expresa la parte de intercambio con el funcional híbrido de Becke de tres parámetros (B3) (Becke 1993), y la parte de correlación mediante el funcional de Lee-Yang-Parr (LYP) (Lee et al. 1988).

2.1.3. Métodos semiempíricos

Estos métodos tratan de aumentar la eficiencia computacional del cálculo HF al reducir el número de integrales bielectrónicas mediante simplificaciones introducidas en la expresión del Hamiltoniano. No obstante, la pérdida de calidad intenta ser compensada por la inclusión de parámetros ajustados para reproducir propiedades experimentales (energías de ionización, geometrías moleculares,...) o resultados de cálculos *ab initio* de alto nivel (Stewart 1990; Zerner 1991).

Una característica básica es que no tratan explícitamente los electrones internos, que junto con el núcleo forman el *core atómico*. Adicionalmente, emplean un conjunto de base de tamaño mínimo, formado por orbitales STO tipo s, p y en algunos casos d, para describir los electrones de valencia. Finalmente, la aproximación ZDO (Zero Differential Overlap), que omite la consideración de solapamiento entre orbitales, se aplica en diferentes grados. El método CNDO (*Complete Neglect of Differential Overlap*) (Pople et al. 1965; Pople and Segal 1965) la aplica estrictamente a todos los productos de funciones de base. Por otro lado, la estrategia NDDO (*Neglect of Diatomic Differential Overlap*) anula únicamente el solapamiento entre funciones de base pertenecientes a átomos distintos. En este nivel cabe mencionar los métodos MNDO (*Modified Neglect of Diatomic Overlap*) (Dewar and Thiel 1977), que tiende a sobreestimar la repulsión entre átomos separados por una distancia cercana al mínimo

de una interacción de van der Waals, y AM1 (*Austin Model 1*) (Dewar et al. 1985), que fue desarrollado para corregir las deficiencias del primero. La automatización de los procesos de parametrización de la metodología AM1 condujo al método PM3 (*Parametrized Model 3*) (Stewart 1989), dando lugar a los modelos más empleados actualmente (Cramer 2002).

Los métodos semiempíricos, por ser de escaso coste computacional, permiten el tratamiento de sistemas de gran tamaño. Sin embargo, varias consideraciones deben tenerse en cuenta de cara a emplear estas metodologías con la debida cautela.

- Su utilización pasa por la disponibilidad de parámetros para todos los átomos del sistema, que no necesariamente se habrán desarrollado.
- Los errores van asociados a la parametrización, por lo que tienden a ser más sistemáticos que en los cálculos *ab initio*, dependiendo fuertemente del tipo de compuestos estudiados.
- El conjunto de moléculas empleado en la parametrización puede comprometer la aplicabilidad de los modelos a moléculas muy disimilares.

La utilización de los métodos semiempíricos en este trabajo no ha sido especialmente relevante, salvo usos puntuales en el proceso de preparación de sistemas para su modelado por mecánica clásica o para continuar con cálculos cuánticos de mayor nivel.

2.1.4. Detalles relacionados con las aplicaciones realizadas

Error de superposición de base (BSSE)

La aproximación de la supermolécula calcula la energía de interacción entre dos moléculas como la diferencia entre la energía del complejo (por ejemplo, un dímero) y la suma de las energías de cada monómero por separado. Este procedimiento, cuando se usa una base finita, produce resultados erróneos, ya que la función de onda obtenida para el dímero permite a los electrones una distribución más amplia que la que se les permite en el monómero. Este efecto es especialmente notable con bases pequeñas, o en cálculos donde se considera la correlación electrónica (Jensen 1996). Para solventar este

problema se emplea el método *counterpoise* (Boys and Bernardi 1970), que considera para cada monómero la diferencia entre la energía obtenida con las funciones de base del dímero y con las del monómero. Esto permite cuantificar el grado de estabilización generada por el dímero sobre cada monómero.

Potencial electrostático molecular (MEP), potencial de interacción molecular (MIP) y potencial de interacción generalizado con correcciones para la polarización (GMIP_p)

El potencial electrostático en un punto r (MEP) alrededor de un determinado sistema molecular se define como el trabajo necesario para situar en él la unidad de carga positiva desde distancia infinita. El producto de una carga por dicho potencial, $q \cdot V(r)$, corresponde a la energía electrostática de interacción entre ella y la molécula en ese punto (Scrocco and Tomasi 1973; Kaufman et al. 1981; Murray and Sen 1996). Esta magnitud, que recoge toda la información acerca de la distribución de carga de los núcleos y electrones de la molécula, resulta muy apropiada en la descripción cualitativa de su reactividad.

En el marco del formalismo *LCAO* puede obtenerse mediante la siguiente expresión

$$MEP = \sum_A \frac{Z_A}{|r_1 - R_A|} - \sum_{\mu} \sum_{\nu} P_{\mu\nu} \int \frac{\chi_{\mu}(r)\chi_{\nu}(r)}{|r_1 - r|} dr \quad (2.23)$$

donde Z_A y R_A corresponden respectivamente a la carga nuclear y la posición del átomo A , χ_{μ} y χ_{ν} se refieren a las funciones de base de los orbitales μ y ν , y $P_{\mu\nu}$ representa al elemento $\mu\nu$ de la matriz de densidad de primer orden.

Para su cálculo puede emplearse una función de onda proveniente de cálculos *ab initio*, semiempíricos o DFT. La obtenida a partir de un cálculo HF/6-31G(d) es particularmente adecuada para sistemas bioquímicos en base al balance entre calidad y coste computacional, y al hecho de que sobrestima la polaridad de la molécula en fase gas a una magnitud similar a la que produce el efecto del solvente.

La descripción anterior puede generalizarse a la interacción del sistema QM con cualquier tipo de partícula clásica (Orozco and Luque 1993; Alhambra et al. 1995; Orozco and Luque 1996; Luque and Orozco 1998). Así, en el formalismo GMIP (*Generalized Molecular Interaction Potential*) el potencial en un punto se obtiene por la adición, al valor de energía electrostática, de un término clásico de van der Waals,

$$GMIP = \sum E_{ele} + \sum E_{vdW} \quad (2.24)$$

donde la sumatoria se extiende a todas las partículas de la molécula clásica, y el primer término se calcula, para un conjunto de cargas discretas, mediante la expresión

$$E_{ele} = \sum_{s=1}^S q_s V(r_s) \quad (2.25)$$

donde $V(r_s)$ se refiere al potencial electrostático molecular en la posición de la carga s .

Una estrategia desarrollada en nuestro grupo (Luque and Orozco 1998) permite, mediante el empleo de teoría de perturbaciones, introducir la polarización inducida por la partícula clásica en la parte QM (E_{pol}) de una forma computacionalmente eficiente. Así, la energía de polarización se calcula mediante la expresión

$$E_{pol} = \sum_j^{vir} \sum_i^{occ} \frac{1}{\epsilon_i - \epsilon_j} \left\{ \sum_{\mu} \sum_{\nu} c_{\mu i} c_{\nu j} \left\langle \phi_{\mu} \left| \sum_S \frac{q_s}{r_s - r} \right| \phi_{\nu} \right\rangle \right\}^2 \quad (2.26)$$

en la que el primer y segundo sumatorios se realizan, respectivamente, sobre todos los orbitales moleculares virtuales y ocupados, ϵ se refiere a la energía del correspondiente orbital molecular y c designa a los coeficientes de la combinación lineal de orbitales atómicos (ϕ) que forman los orbitales moleculares.

La adición de E_{pol} a GMIP conduce al desarrollo del potencial generalizado de interacción molecular con correcciones para la polarización ($GMIP_p$), que permite una rápida y precisa evaluación de la interacción de un sistema clásico y uno cuántico, fácilmente implementable en algoritmos QM/MM de Monte Carlo (Cubero et al. 2003).

$$GMIP_p = GMIP + E_{pol} \quad (2.27)$$

Inclusión de los efectos del solvente en cálculos cuánticos

La estrategia más directa para incluir el solvente a nivel cuántico es la de incluir un conjunto reducido de moléculas de disolvente alrededor del soluto. Este método adolece de tres problemas: i) al ignorar efectos dinámicos, los clusters resultantes dan una imagen más cercana al hielo que al agua líquida, ii) el tamaño de los sistemas es siempre excesivamente pequeño para representar un medio solvatado real, y iii) el coste computacional puede ser enorme en cuanto el tamaño del sistema aumenta. Por ello, el efecto del solvente suele tratarse bien i) recurriendo a partículas clásicas que se muestrean en cálculos de Monte Carlo o dinámica molecular (métodos clásicos puros o QM/MM), o bien ii) utilizando representaciones macroscópicas del solvente (métodos continuos). En la presente tesis se mencionarán los principales rasgos del método continuo MST (*Miertus, Scrocco, Tomasi*), basado en el método PCM (*Polarizable Continuum Model*) (Miertus et al. 1981; Miertus and Tomasi 1982).

A fin de calcular la energía libre de solvatación (ΔG_{sol}), el método MST resume este proceso en dos etapas:

- a) Creación de la cavidad en el solvente para que pueda albergar al soluto.
- b) Inserción del soluto en la cavidad generada.

El primer paso lleva asociada la energía libre de cavitación (ΔG_{cav}), mientras que en el segundo paso se establecen las nuevas interacciones soluto-solvente, tanto electrostáticas (ΔG_{ele}) como de van der Waals (ΔG_{vdw}). Así pues, podemos expresar la energía libre de solvatación como una suma de estas tres contribuciones.

$$\Delta G_{solv} = \Delta G_{cav} + \Delta G_{vdw} + \Delta G_{ele} \quad (2.28)$$

Cálculo de la componente de cavitación

Partiendo de la teoría de partículas escaladas (Pierotti 1976), podemos expresar ΔG_{cav} como

$$\Delta G_{cav} = K_0 + K_1 R_{SX} + K_2 R_{SX}^2 + K_3 R_{SX}^3 \quad (2.29)$$

$$R_{SX} = R_S + R_X \quad (2.30)$$

donde R_S es el radio del soluto, R_X el de la molécula de solvente, y las diferentes K_n son coeficientes empíricos que dependen de la naturaleza del solvente y de las condiciones de presión y temperatura.

Claverie introdujo una mejora a este método adaptando las cavidades para que tengan forma molecular (Claverie 1978). Según este procedimiento, cada átomo por separado contribuye a la ΔG_{cav} final, que puede expresarse con la fórmula

$$\Delta G_{cav} = \sum_{i=1}^N \frac{S_i}{S_T} \Delta G_{cav,i}^P \quad (2.31)$$

donde S_i es la superficie expuesta al solvente de cada átomo, S_T la suma de estas superficies para todos los átomos del sistema y $\Delta G_{cav,i}^P$ la energía libre de cavitación para ese átomo según la fórmula de Pierotti.

Cálculo de la componente electrostática

Este término se determina mediante el formalismo SCRF (*Self-Consistent Reaction Field*), que considera que el solvente genera un campo de reacción como respuesta a la distribución de carga del soluto. En este marco, el operador hamiltoniano del soluto en vacío (H^0) queda modificado por la adición de una perturbación (V_R), que corresponde al campo de reacción del solvente, y se aplica en la siguiente pseudo-ecuación de Schrödinger

$$(H^0 + V_R)\Psi = E\Psi \quad (2.32)$$

La perturbación V_R se describe en base a la distribución de carga ($\sigma(s)$) inducida en la superficie de la cavidad, generada como resultado de la polarización que ejerce el soluto sobre el solvente. Para abordar en la práctica el cálculo de V_R , se divide el área de la cavidad en M elementos S_i cuyo tamaño sea tan reducido que permita considerar que la distribución de carga ($\sigma(S_i)$) permanece constante en su interior. De este modo, se puede hallar V_R en términos de cargas puntuales de polarización (q_i) distribuidas en la superficie de la cavidad y que pueden determinarse mediante la resolución de la ecuación de Laplace

$$V_R = \int_S \frac{\sigma(s)}{|r_0 - r|} dS = \sum_{i=1}^M \frac{\sigma(S_i)S_i}{|r_0 - r|} = \sum_{i=1}^M \frac{q_i}{|r_0 - r|} \quad (2.33)$$

en la que M se refiere al número de elementos de superficie i , siendo S_i el área de cada elemento, σ la densidad electrónica en esa zona y q_i la carga total que representa la respuesta del solvente al elemento de superficie i .

La componente electrostática de la energía libre de solvatación se obtendría mediante la ecuación

$$\Delta G_{ele} = \left\langle \Psi^{sol} \left| H^0 + \frac{1}{2} V_R \right| \Psi^{sol} \right\rangle - \left\langle \Psi^0 \left| H^0 \right| \Psi^0 \right\rangle \quad (2.34)$$

donde los superíndices 0 y sol indican respectivamente el estado de vacío y disolución.

Cálculo de la componente de Van der Waals

El término de van der Waals se determina mediante una relación lineal con la superficie expuesta de cada átomo al solvente. Se puede calcular fácilmente como

$$\Delta G_{vdW} = \sum_{i=1}^N \Delta G_{vdW,i} = \sum_{i=1}^N \xi_i S_i \quad (2.35)$$

donde S_i es la superficie accesible al solvente de cada átomo i , y el factor de proporcionalidad ζ_i es la tensión superficial atómica, obtenida a partir de un ajuste a datos empíricos de energías libres de solvatación (Luque et al. 1996a; Luque et al. 1996b; Curutchet et al. 2001).

2.2. Descripción de sistemas químicos basada en mecánica clásica

Aplicando únicamente el formalismo de la mecánica newtoniana sobre las partículas (núcleos) de un sistema químico, se pueden describir razonablemente muchos rasgos de su comportamiento, siempre que no impliquen cambios en el esqueleto covalente o en su distribución electrónica. El conjunto de ecuaciones y parámetros empíricos que posibilitan esta tarea se denomina campo de fuerza (*force-field*), que comprende:

- i) Interacciones entre átomos enlazados (átomos unidos en la estructura química por un máximo de tres enlaces covalentes)
- ii) Interacciones entre átomos no enlazados

Un ejemplo de expresión analítica estándar para un campo de fuerzas, donde se recoge esta división, lo encontramos en la siguiente ecuación

$$E_{TOTAL} = \sum_{\text{enlaces}} k_i (l_i - l_{i,0})^2 + \sum_{\text{ángulos}} k'_i (\theta_i - \theta_{i,0})^2 + \sum_{\text{torsiones}} \frac{V_n}{2} (1 + \cos(n\omega - \gamma))$$

$$+ \left(\sum_{i=1}^N \sum_{j=i+1}^N 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right) \quad (2.36)$$

Seguidamente se comentará cada término brevemente, si bien puede encontrarse una explicación más detallada en la literatura (van Gunsteren et al. 1996; Scott et al. 1999; Leach 2001; van der Spoel et al. 2004).

2.2.1 Interacciones entre átomos enlazados

En función del número de enlaces covalentes que separan dos átomos, distinguiremos tres tipos de interacciones: las que implican átomos separados por un enlace (términos 1-2 o de *stretching*), por dos enlaces (términos 1-3 o de *bending*), y por tres enlaces (términos 1-4 o de torsión).

Modelización del *stretching*

La relación energía-distancia entre dos átomos enlazados obedece una curva como la que se muestra en la figura 2.1.

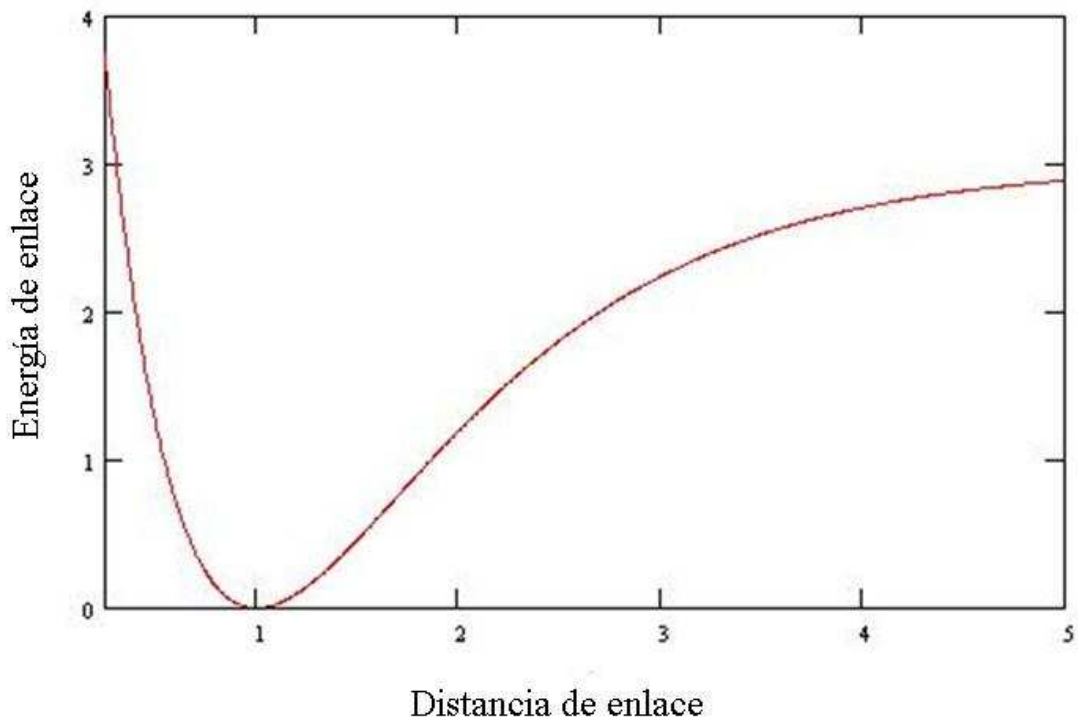


Figura 2.1. Evolución de la energía a medida que se modifica la distancia interatómica

El formalismo más comúnmente empleado para describir dicha curva de energía es la ley de Hooke, especialmente en la zona cercana al equilibrio, o bien la expresión de Morse si interesa incluir conformaciones más tensionadas.

$$E_{stretching} = \frac{k}{2} (l - l_0)^2 \quad (2.37)$$

En la aproximación armónica únicamente son necesarios dos parámetros: k , que es la constante de fuerza, y l_0 , que es la distancia de equilibrio. Estos parámetros se obtienen a partir de datos experimentales, tal como estructuras obtenidas por cristalografía de rayos X para las distancias de equilibrio, y frecuencias de vibración de enlaces obtenidas por espectroscopia sobre compuestos modelo para las constantes de fuerza (Weiner et al. 1984)). El aumento de la capacidad computacional permite que actualmente se tienda a suplementar este trabajo de parametrización mediante el ajuste a datos procedentes de cálculos *ab initio* de alto nivel.

Modelización del bending

Se suele emplear un formalismo armónico análogo al utilizado en el *stretching*, donde los parámetros implicados designan el ángulo de equilibrio (θ_0) y la constante de fuerza (K_{bend}) asociada a la distorsión de los ángulos de enlace.

$$E_{bending} = \frac{k_{bend}}{2} (\theta - \theta_0)^2 \quad (2.38)$$

Existen en la literatura tablas de valores experimentales de distancias de enlace y ángulos de equilibrio (Harmony et al. 1979; Allen et al. 1987), que suelen usarse en la parametrización de *stretchings* y *bendings*, combinados con datos de frecuencias obtenidos mediante experimentos de espectroscopia para la determinación de las constantes de fuerza. Se han propuesto algunos métodos de cara a la optimización computacional de este proceso (Hopfinger and Pearlstein 1985; Aleman et al. 1991; Cannon 1993).

Modelización del término de torsión

Las torsiones son los grados de libertad internos que presentan mayor variabilidad, siendo responsables de la flexibilidad de las moléculas. La variación de la energía asociada a la rotación de un diedro puede expresarse de diversas formas según el campo de fuerzas. En el *force field* AMBER (Case et al. 2004), se usa una expansión en serie de Fourier,

$$E_{torsión} = \sum_{tor} \sum_n \frac{V_n}{2} (1 + \cos(n\theta - \gamma)) \quad (2.39)$$

donde $V_n/2$ indica la barrera energética, n es la multiplicidad (el número de mínimos que presenta la función entre 0° y 360°), θ es el valor del diedro y γ representa el ángulo de fase.

En algunos casos, por ejemplo en AMBER, se usan términos genéricos de torsión, cuya forma viene marcada únicamente por la naturaleza de los átomos centrales independientemente de los extremos, lo cual facilita el proceso de parametrización.

La determinación de los parámetros de torsión se realiza mediante el ajuste de la energía cuántica a lo largo de la rotación completa del enlace, determinada generalmente a nivel HF/6-31G(d). No obstante, la última versión del campo de fuerzas AMBER (*ff3*) (Duan et al. 2003) emplea datos MP2/cc-pVTZ//HF/6-31G(d,p) para el ajuste de las torsiones peptídicas, a fin de mejorar el comportamiento de estos sistemas en simulaciones con solvente implícito.

Puede ser necesario, para ajustarse a la evidencia experimental, mantener la planaridad de algunos grupos. Para ello, se han propuesto diferentes formalismos. En AMBER, se utilizan torsiones impropias a cuatro átomos que no están secuencialmente enlazados y se aplica un potencial de torsión en serie de Fourier análogo a los anteriores.

Modelización de términos cruzados

Se trata de términos que intentan simular el acoplamiento entre movimientos internos de la molécula. Fueron muy utilizados en los primeros campos de fuerzas (orientados a la predicción de espectros vibracionales de moléculas pequeñas), pero no se usan en aquéllos destinados al estudio de macromoléculas.

2.2.2 Interacciones entre átomos no enlazados

Este tipo de interacciones dependen únicamente de la distancia entre las partículas implicadas y su naturaleza. La energía asociada a ellas suele expresarse analíticamente como una relación de proporcionalidad inversa con la distancia. Generalmente se distinguen dos términos: electrostático y van der Waals.

Interacciones electrostáticas

Se han descrito múltiples estrategias para modelar la interacción entre las distribuciones de carga de dos moléculas. La mayoría de ellas se basan en discretizar dichas distribuciones en cargas atómicas puntuales, cuya interacción obedece la ley de Coulomb.

$$E_{ele} = \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (2.40)$$

donde N_A y N_B representan el número de cargas distribuidas sobre las moléculas A y B, q es la magnitud de la carga puntual, r_{ij} la distancia entre dos cargas concretas y ϵ_0 la permitividad del vacío.

La metodología empleada de forma estándar en este trabajo para derivar cargas atómicas se basa en el potencial electrostático molecular (véase apartado 2.1.4), pues las cargas se ajustan de forma que reproduzcan el potencial electrostático en una región del espacio próxima a la molécula (Cox and Williams 1981). En particular, en AMBER se considera una distribución de puntos sobre superficies de van der Waals de tamaño

creciente (Singh and Kollman 1984). Una mejora sustancial provino del protocolo RESP (*Restrained Electrostatic Potential Fit*) (Bayly et al. 1993), donde se introducen restricciones hiperbólicas para evitar la localización de cargas elevadas sobre átomos pesados enterrados y disminuir la dependencia conformacional del conjunto de cargas. Dichas cargas generalmente se ajustan al potencial electrostático calculado a nivel HF/6-31G*, pues al exagerar ligeramente la polaridad de la molécula permite introducir implícitamente el efecto de polarización del solvente.

Interacciones de Van der Waals

El caso de los gases nobles es un ejemplo paradigmático de sistemas sujetos a *interacciones de van der Waals*. La energía asociada a una interacción de van der Waals entre dos átomos se muestra en la figura 2.2, donde se aprecia que la energía de interacción presenta un mínimo (ϵ) a una distancia característica (σ). Aumentar dicha distancia anula la interacción de forma asintótica y disminuirla conlleva un aumento de la energía a medida que los radios atómicos se solapan.

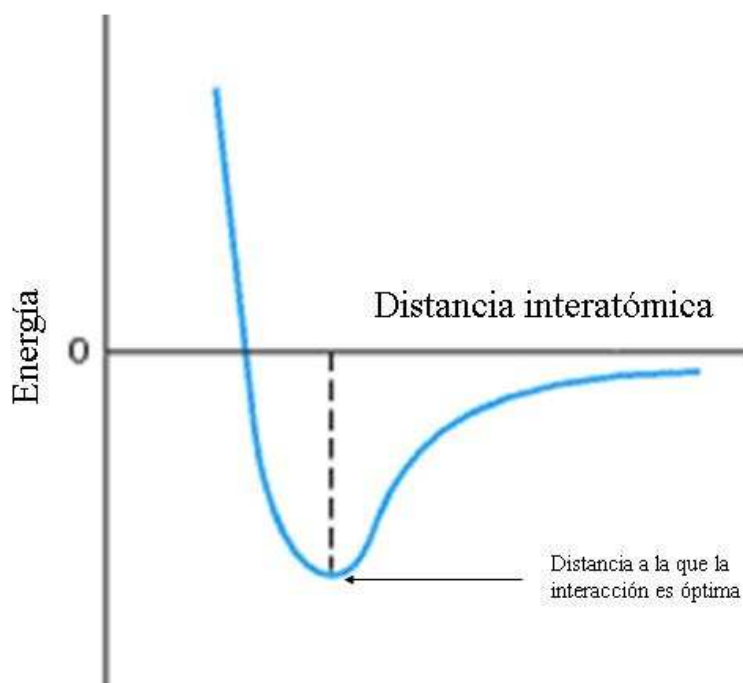


Figura 2.2. Energía asociada a la interacción de Van der Waals entre 2 átomos

Las fuerzas de van der Waals recogen una componente atractiva a distancias medias y otra repulsiva a cortas distancias. La parte atractiva depende de la sexta potencia de la distancia interatómica, mientras que el término repulsivo es típicamente considerado función de la doceava potencia.

$$E_{vdw} = \frac{A}{(r_{ij})^{12}} - \frac{B}{(r_{ij})^6} \quad (2.41)$$

La forma usualmente empleada en la obtención de los parámetros de van der Waals es la reproducción de datos experimentales estructurales. Conviene tener en cuenta que un ajuste de estas características no tiene porqué dar un conjunto único de parámetros, siendo así necesario validar la calidad del *force-field*.

2.2.3 Consideraciones prácticas en el manejo de *force-fields*

La creciente mejora de las técnicas de simulación basadas en dinámica molecular, unida al aumento de la capacidad computacional, ha conducido a la posibilidad de explorar el movimiento de sistemas de interés biológico de miles de átomos en la escala de tiempo del nanosegundo. Ahora bien, la validez de estas metodologías como herramientas de experimentación *in silico* en biología va asociada de forma crucial a la precisión que sean capaces de ofrecer, es decir, a su capacidad para reproducir propiedades observables experimentalmente.

Las mejoras algorítmicas como el tratamiento de las interacciones electrostáticas de largo alcance (Ewald 1921; Darden et al. 1993) y otros tratamientos del contorno (Beglov and Roux 1994; Bishop et al. 1997; Im et al. 2001), o la introducción de nuevos algoritmos de integración (Tuckerman et al. 1992; Feller et al. 1995; Barth and Schlick 1998), que permiten la simulación del colectivo termodinámico adecuado a cada problema, han contribuido notablemente a este incremento en la precisión. No obstante, la calidad del *force field* y su implementación correcta es siempre un factor crítico. Por ello, debe tenerse presente que:

- La parametrización es una de las tareas más delicadas y trascendentes del modelado molecular basado en mecánica clásica. No obstante, el objetivo concreto que se persigue puede modificar la precisión que debe alcanzarse en esta etapa. Así, la estimación de propiedades termodinámicas requiere un gran esfuerzo de parametrización, mientras que una menor calidad suele ser suficiente en la determinación de propiedades estructurales.
- Los *force-fields* se parametrizan generalmente para predecir cierto tipo de propiedades y garantizan una reproducción de la evidencia experimental concreta a la que se ajustan. Este carácter empírico exige un uso racional, ya que predicciones que queden fuera de este espectro no tienen porqué ser correctas.
- Un campo de fuerzas ha de considerarse siempre en su conjunto. No es estrictamente correcto el dividir la energía en contribuciones particulares o el intercambio de parámetros entre campos de fuerzas.

2.3. Simulación y análisis del movimiento de los sistemas químicos

Al estudiar biomacromoléculas, el concepto de *estructura de mínima energía* presenta muchas veces un interés parcial. Históricamente, se tiende a expresar una molécula como una estructura tridimensional fija, pero dicha representación omite el comportamiento dinámico, que involucra la existencia de diferentes estados conformacionales. Y quizás dicha *plasticidad* es precisamente el factor que justifica la selección de dicha macromolécula por los mecanismos de la evolución biológica. Resulta entonces de gran importancia conocer no sólo la estructura, sino también la flexibilidad de una molécula. Para ello se han desarrollado varias metodologías que persiguen simular con sentido físico el movimiento de sistemas de partículas. A continuación, se describirá el método de dinámica molecular (MD).

2.3.1 Dinámica molecular

Simular el comportamiento de un sistema por dinámica molecular (MD) consiste en generar configuraciones sucesivas del mismo mediante la aplicación repetida de la segunda ley de Newton sobre cada una de sus partículas. Hace casi medio siglo se realizó la primera MD de un sistema en fase condensada (Alder and Wainwright 1957). Innovaciones posteriores condujeron a la primera MD que utilizaba un potencial continuo (Rahman 1964), donde cualquier variación en la posición de una partícula implica un cambio en las fuerzas que actúan sobre ella, y a la ejecución de la primera simulación MD de un líquido molecular (Rahman and Stillinger 1971). Pocos años más tarde, la simulación del retinol como primera biomolécula (Warshel 1976), la primera proteína (McCammon 1976; McCammon et al. 1977) y el primer ácido nucleico (Levitt 1983; Tidor et al. 1983) constituyeron los primeros pasos en la creciente aplicación de esta técnica al estudio de los sistemas biológicos.

Algoritmo general y métodos de integración

Dado que los movimientos de todas las partículas están acoplados entre sí, la descripción de la evolución dinámica de un sistema precisa de métodos de integración numéricos. En el método de diferencias finitas, se fragmenta el tiempo de simulación en muchas etapas de duración muy corta y se integra la ecuación de Newton para cada paso. Así, dada una configuración del sistema, la fuerza en cada átomo se calcula por derivación numérica del *force-field*. Generadas las fuerzas, el sistema se deja evolucionar a merced de ellas durante un intervalo corto de tiempo (paso de integración o *timestep*). Sobre la nueva disposición espacial se recalculan las fuerzas y así sucesivamente, tal y como se esquematiza en la figura 2.3.

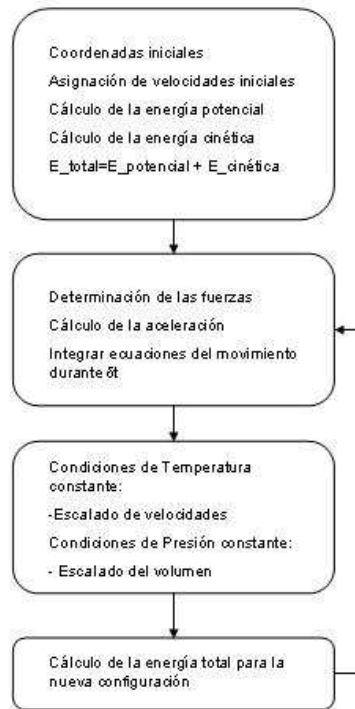


Figura 2.3. Pasos de un algoritmo estándar de dinámica molecular

Algoritmo de Verlet

Existen diferentes métodos de integración. Uno de los más utilizados es el algoritmo de Verlet (Verlet 1967), que emplea las posiciones y velocidades de un instante (t) y del inmediatamente anterior ($t - dt$) para calcular las nuevas posiciones ($t + dt$), según se indica en la ecuación 2.44 derivada de las dos anteriores (2.42-2.43).

$$\vec{r}(t + \delta t) = \vec{r}(t) + \delta t \vec{v}(t) + \frac{1}{2} \delta t^2 \vec{a}(t) \quad (2.42)$$

$$\vec{r}(t - \delta t) = \vec{r}(t) - \delta t \vec{v}(t) + \frac{1}{2} \delta t^2 \vec{a}(t) \quad (2.43)$$

$$r(t + \delta t) = 2r(t) - r(t - \delta t) + \delta t^2 a(t) \quad (2.44)$$

Algunos problemas, no obstante, acompañan a esta metodología. La expresión final contiene la adición de una cantidad muy pequeña, $\delta t^2 a(t)$, a la diferencia entre dos cantidades grandes, lo que puede introducir cierta imprecisión. Además, las velocidades han de obtenerse *a posteriori* una vez se conoce la posición final y, al depender de las posiciones anteriores, es difícil iniciar la simulación a tiempo cero.

Algoritmo de Leap-Frog

Este algoritmo (Hockney 1970) soluciona el problema del inicio a tiempo cero. Como se muestra en las ecuaciones 2.45 y 2.46, este método permite el avance de la MD a partir del conocimiento de las velocidades a tiempo $(t-1/2\delta t)$.

$$\vec{r}(t + \delta t) = \vec{r}(t) + \delta t \vec{v}\left(t + \frac{1}{2} \delta t\right) \quad (2.45)$$

$$\vec{v}\left(t + \frac{1}{2} \delta t\right) = \vec{v}\left(t - \frac{1}{2} \delta t\right) + \delta t \vec{a}(t) \quad (2.46)$$

En el momento inicial, dichas velocidades pueden ser asignadas al azar, restringiendo que globalmente se ajusten a una distribución de Boltzmann para una temperatura determinada, según la ecuación 2.47.

$$E_{\text{cinética}} = \sum_{i=1}^N \frac{|\vec{p}_i|^2}{2m_i} = \frac{K_B T}{2} (3N - N_C) \quad (2.47)$$

donde p_i se refiere al vector velocidad de la partícula i , m_i es la masa, K_B la constante de Boltzmann, N el número de partículas y N_C el número de grados de libertad congelados que presenta el sistema.

Algoritmo predictor-corrector de Gear

Es una metodología más precisa (Gear 1971) cuya ventaja principal es que permite un alargamiento de la etapa de integración. Consiste básicamente de cuatro pasos:

- generación de las nuevas posiciones, velocidades y aceleraciones a partir de un desarrollo en serie de Taylor
- obtención de fuerzas y aceleraciones para las nuevas posiciones
- comparación de estas aceleraciones con las predichas en el paso 1
- utilizar la diferencia entre aceleraciones para corregir las posiciones y velocidades

Su coste computacional, compensado con la posibilidad de alargamiento del *timestep*, es aún bastante superior al coste de un algoritmo de Verlet, por lo que su popularidad es pequeña en simulación de macromoléculas.

Paso de integración e introducción de restricciones

El tiempo de integración en una simulación por MD ha de escogerse más pequeño que la duración del movimiento más rápido del sistema, el cual corresponde a la vibración de *stretching*, que ocurre en el orden del femtosegundo. Por ello, sería necesario integrar las ecuaciones de Newton al menos cada 0.5 femtosegundos, lo que conlleva un inmenso número de integraciones para seguir una trayectoria, aún de corta duración temporal. Una estrategia utilizada para aumentar la longitud de la etapa de integración consiste en congelar algunos grados de libertad de alta frecuencia mediante la introducción de constricciones holonómicas. Un ejemplo es el método SHAKE (Ryckaert et al. 1977), que permite mantener estática la longitud de los enlaces covalentes. Dicho protocolo mejora notablemente la eficiencia computacional de la MD al hacer posible que se cuadruple la etapa de integración empleada.

El tratamiento de los límites del sistema

Los bordes de un sistema finito son una zona delicada donde la aplicación de las ecuaciones del *force-field* puede conducir a comportamientos artefactuales. Para evitar este efecto se emplean *condiciones periódicas de contorno* (PBC), que consiste en hacer múltiples réplicas del sistema en las 3 direcciones del espacio. Todo movimiento de cualquier partícula de la caja central se aplica sobre todas las réplicas y, en su conjunto, las réplicas establecen un macroentorno en torno a la celda central.

El empleo de PBC lleva asociado un enorme incremento en el número de interacciones a calcular, lo que obliga a aplicar ciertas restricciones. Una es la truncación de las interacciones cuya magnitud decae rápidamente con la distancia. Cada partícula únicamente verá a aquellas que estén dentro de un radio inferior a un límite (*cutoff*). Este procedimiento genera graves problemas si se aplica sobre interacciones que presentan un decaimiento muy lento respecto a la distancia, como son las electrostáticas. Una mejora al tratamiento de éstas puede venir del empleo de *funciones de switching* o *asignación de grupos cargados* (van der Spoel et al. 2004; Bhandarkar et al. 2005).

Un método que combina un tratamiento adecuado de las interacciones electrostáticas con una buena eficiencia computacional es *Particle Mesh Ewald (PME)* (Darden et al. 1993), derivado del método de Ewald (Ewald 1921) creado para representar sistemas cristalinos. Su implementación ha significado una mejora considerable en las simulaciones por MD, siendo su uso especialmente crucial en la simulación de ácidos nucleicos. El método originalmente propuesto por Ewald sustituye el sumatorio de todas las interacciones electrostáticas del sistema por una suma de dos términos que convergen más rápidamente.

$$E_{ele} = Ee_{dir} + Ee_{rec} + E_0 \quad (2.48)$$

El primer y segundo término se refieren, respectivamente, al *espacio directo* y *recíproco*, y el término E_0 es una constante. El término Ee_{dir} recoge las interacciones de corto alcance, calculadas mediante un potencial coulombico modificado. Con vistas a una convergencia más rápida, las cargas originales se representan en el espacio directo mediante una serie de funciones gaussianas. Estas funciones son contrarrestadas en el espacio recíproco por una serie de distribuciones gaussianas de signo opuesto. El método *PME* optimiza el cálculo del espacio recíproco transformando las cargas puntuales a mallas tridimensionales mediante interpolación *B-spline*, lo que permite emplear transformadas de Fourier en la resolución del segundo término de la ecuación 2.48, aumentando considerablemente la velocidad de cálculo.

Mantenimiento de la temperatura y presión constantes

Las simulaciones de MD pueden realizarse dentro de diferentes colectivos. Como colectivos más comunes encontramos NVE (microcanónico), NVT (canónico) y NPT (isotérmico-isobárico), según se mantenga constante el número de partículas (N), el volumen (V), la presión (P) y/o la temperatura (T).

Temperatura constante

Conociendo la energía cinética de todas las partículas, podemos estimar el valor de la temperatura según la expresión

$$\langle E_{\text{cinética}} \rangle_{NVT} = \frac{3}{2} NK_B T \quad (2.49)$$

donde K_B es la constante de Boltzmann.

Así, pues, un escalado de las velocidades (Woodcock 1971) soluciona el problema de modular la temperatura. El algoritmo de Berendsen o *weak-coupling algorithm* (Berendsen et al. 1984) permite al sistema intercambiar calor con un baño externo mantenido a la temperatura (T) deseada. En función de la diferencia de temperatura con el baño, el sistema reajusta su temperatura según

$$\frac{\delta T(t)}{\delta t} = \frac{1}{\tau} (T_{\text{baño}} - T(t)) \quad (2.50)$$

donde τ se refiere a un parámetro que regula la fuerza del acoplamiento térmico, adoptando generalmente valores entre 0.5-5.0 ps (valores inferiores forzarían un acoplamiento térmico más severo, disminuyendo la naturalidad de los movimientos del sistema), que a su vez determina el valor del factor de escalado λ .

$$\lambda^2 = 1 + \frac{\delta t}{\tau} \left(\frac{T_{\text{baño}}}{T(t)} - 1 \right) \quad (2.51)$$

Una desventaja de este método es que provoca un aumento artificial de las diferencias térmicas entre partes del sistema. Se ha descrito en este sentido el problema del enfriamiento relativo del soluto (*cold solute-hot solvent problem*) (Harvey et al. 1998), que se debe a una creación de gradiente de calor entre soluto y solvente. Una solución proviene de aplicar factores de acoplamiento diferentes a soluto y solvente.

Presión constante

Dado que un gran número de experimentos se realizan en condiciones de presión y temperatura constantes, la simulación MD en el colectivo isotérmico-isobárico es una herramienta indispensable. Las oscilaciones de la presión, que son mucho más grandes que las de la energía o la temperatura, se ajustan mediante la variación del volumen del sistema. Una expresión analítica empleada con frecuencia para cuantificar esta fluctuación se presenta en la ecuación 2.52.

$$\kappa = -\frac{1}{V} \left(\frac{\delta O}{\delta P} \right)_T \quad (2.52)$$

donde κ es la compresibilidad del sistema (generalmente del solvente) a T constante.

El método de regulación de la presión más utilizado es análogo al acoplamiento térmico de Berendsen (Berendsen et al. 1984), donde la variación de la presión se obtiene a partir de la siguiente fórmula

$$\frac{\delta P(t)}{\delta t} = \frac{1}{\tau_p} (P_{baño} - P(t)) \quad (2.53)$$

donde τ_p es el parámetro que regula la intensidad del acoplamiento de presión.

Así, el factor de escalado de posiciones se puede expresar según la ecuación 2.54, generando nuevas posiciones para cada partícula del sistema según la relación 2.55.

$$\lambda = 1 - \kappa \frac{\delta t}{\tau_P} (P - P_{baño}) \quad (2.54)$$

$$\vec{r}_i' = \lambda^{1/3} \vec{r}_i \quad (2.55)$$

donde el factor de escalado puede aplicarse de forma isotrópica o anisotrópica.

2.4. Cálculo de la energía libre

La energía libre es una magnitud termodinámica clave para el estudio de cualquier proceso químico o bioquímico, por lo que su determinación es el objeto de muchos estudios de modelado molecular. Si en el sistema de estudio no se produce variación de la temperatura, el volumen ni el número de partículas (colectivo canónico o NVT), la energía libre se expresa mediante la función de Helmholtz (A). Si nos encontramos en el colectivo isotérmico-isobárico (NPT), se expresa como la función de Gibbs (G). Una estimación rigurosa de la energía libre requiere un muestreo total del espacio de fase del sistema, lo que la hace inviable en la mayoría de sistemas de interés. Es preciso recurrir a técnicas que, sesgando el espacio de muestreo, permiten obtener estimados de diferencias de energía libre entre estados bien definidos.

2.4.1 Cálculo de variaciones de energía libre

Dos técnicas muy utilizadas para calcular cambios de energía libre asociados a la transición entre estados son la Perturbación de la Energía Libre (FEP) y la Integración Termodinámica (TI).

Perturbación de la energía libre (FEP).

Dados dos estados 0 y 1 (inicial y final) de un proceso, definidos cada uno por un hamiltoniano clásico, se transforma el estado 0 en el 1 mediante una combinación de

ambos hamiltonianos acoplada a la variación progresiva de un parámetro que denominamos λ y que va de 0 a 1. El sistema se simula con este hamiltoniano híbrido y la energía libre se calcula en intervalos discretos, tal como se indica en las ecuaciones siguientes

$$G_{\lambda(i+1)} - G_{\lambda(i)} = -RT \ln \left\langle \exp \left[\frac{-(V_{\lambda(i+1)} - V_{\lambda(i)})}{RT} \right] \right\rangle_{\lambda(i)} \quad (2.56)$$

$$\Delta G = G_1 - G_0 = \sum_i G_{\lambda(i+1)} - G_{\lambda(i)} \quad (2.57)$$

donde los subíndices 1 y 0 se refieren a los estados final e inicial, respectivamente, y $V_{\lambda(i)}$ es el hamiltoniano clásico híbrido definido para cada estado (Zwanzig 1954).

Integración termodinámica.

Un método alternativo es la integración termodinámica (Kirkwood 1935), donde la diferencia de energía libre entre los estados 0 y 1 se calcula como

$$G_1 - G_0 = \int_0^1 \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_{\lambda} d\lambda \quad (2.58)$$

En este caso, lo que se evalúa en cada paso es la derivada del hamiltoniano con respecto a λ . En la práctica, la técnica suele emplearse combinada con el uso de “ventanas” de simulación como en el caso de FEP.

2.4.2. El uso de ciclos termodinámicos

La energía libre de un proceso químico es función de estado, lo cual permite calcular diferencias de energía libre entre dos estados siguiendo cualquier ruta reversible entre ellos. Esto se aprovecha para, por ejemplo, obtener la diferencia de energía libre de un

proceso en disolución a partir de la energía libre de dicho proceso en fase gas y de las energías libres de solvatación correspondientes (véase figura 2.4).

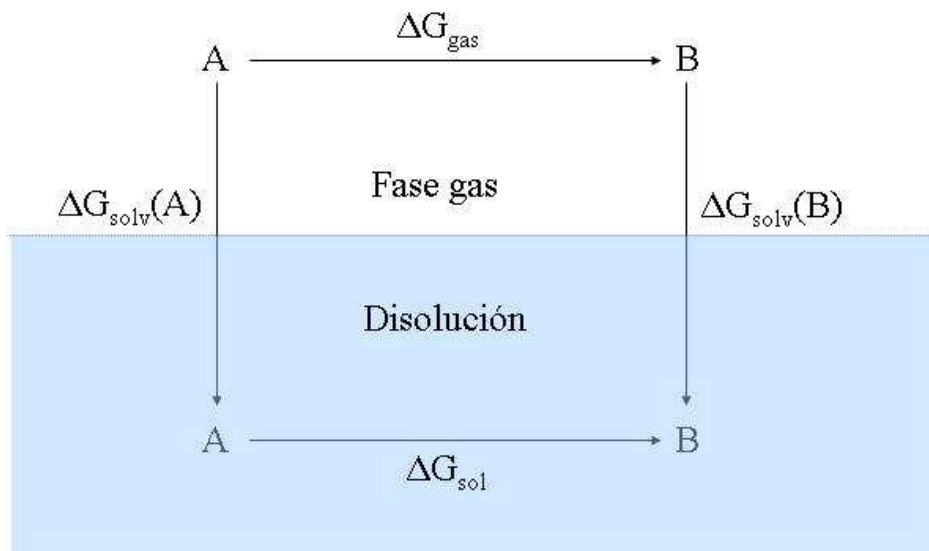


Figura 2.4. Ciclo termodinámico genérico, concretado para el proceso de solvatación de A y B.

La introducción del uso de ciclos termodinámicos acoplados a técnicas FEP/TI se debe a McCammon y Jorgensen (Tembe and McCammon 1984; Jorgensen and Ravimohan 1985). Esta aproximación computacional puede ser empleada en la determinación de las diferencias en energía libre de solvatación de dos estados o compuestos si obtenemos, mediante cálculos tipo FEP o TI, el coste en energía libre asociado a su interconversión en fase gas y en disolución. Nótese que el concepto “solvente” puede ser extendido a cualquier entorno discreto, lo que hace a esta metodología extremadamente útil en el estudio de las interacciones en sistemas bioquímicos.

2.5. Análisis de los sistemas simulados

La comprensión a nivel atómico de la relación entre la estructura y la función de sistemas biomacromoleculares implica no sólo disponer de simulaciones realistas, sino también realizar un análisis riguroso de las mismas. Las metodologías empleadas en dicho análisis son muy variadas y cambian en función del interés de estudio, que puede centrarse en propiedades estructurales, dinámicas, energéticas y de reactividad. Este capítulo citará sólo aquellas que han sido utilizadas en esta tesis.

2.5.1 Información estructural

El seguimiento de variables geométricas concretas puede ser útil en la determinación de detalles estructurales finos o de propiedades globales del sistema (amplitud de surcos en ácidos nucleicos, distancias ligando-receptor, estabilidad del patrón de puentes de hidrógeno, ángulos de torsión, población de las diversas esferas de hidratación,...). Por otro lado, los parámetros estructurales globales, diseñados para informar de la estructura conjunta de un grupo de átomos, como los parámetros de hélice en ácidos nucleicos (Lavery and Sklenar 1988; Lu and Olson 2003), el radio de giro y la orientación relativa de las piezas de un receptor orgánico, entre otros, constituyen una vía para estandarizar las observaciones sobre una molécula y permitir comparaciones globales del sistema.

En concreto, a partir del análisis de los parámetros de hélice del DNA, podemos expresar el sistema en un espacio de coordenadas helicoidales. Esta operación permite cuantificar el grado de curvatura de la fibra (*bending*) y sus principales deformaciones de estiramiento (*stretching*) y enrollamiento (*twisting*). Explorados desde una perspectiva dinámica, estos datos pueden emplearse para derivar las constantes de fuerza y energías elásticas asociadas a cada movimiento, aportando información sobre la deformabilidad isotrópica y anisotrópica de ácidos nucleicos.

2.5.2 Información dinámica

La desviación cuadrática media (*RMSD*) respecto a una estructura de referencia (ec. 2.59) informa cuantitativamente sobre el grado de similitud estructural entre la simulación y la estructura de referencia, resultando útil para evaluar la convergencia a lo largo de la trayectoria. Una variante del *RMSD*, la desviación promedio, puede relacionarse con la varianza del sistema (ver resultados (Perez et al. 2005)) y por lo tanto con la entropía del sistema.

$$RMSD_k = \left[\frac{1}{N} \sum_{l=1}^{3N} (x_{kl} - x_l)^2 \right]^{1/2} \quad (2.59)$$

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

Representando los sistemas mediante un conjunto de coordenadas globales (consideremos el caso de los ácidos nucleicos representados mediante un conjunto de parámetros helicoidales) y procesando las matrices de varianza-covarianza asociadas, se pueden obtener constantes de deformabilidad helicoidal. Estas constantes permiten cuantificar energéticamente (Lankas et al. 2000; Lankas et al. 2003; Perez et al. 2004; Perez et al. 2005) la facilidad/dificultad que la fibra de DNA encuentra para modificar los parámetros de hélice de forma global. Partiendo de la inversión de la matriz de varianza-covarianza en el sistema de coordenadas helicoidales (\mathbf{C}_h^{-1}), se obtiene una matriz $\mathbf{F} = kT \cdot \mathbf{C}_h^{-1}$, cuyos elementos diagonales (F_{ii}) representan la deformación de las variables helicoidales puras y los elementos no diagonales (F_{ij}) recogen la información del acoplamiento entre las diferentes variables helicoidales.

El análisis de componentes principales (PCA) aplicado sobre la fluctuación de las coordenadas cartesianas de un sistema molecular proporciona información sobre su flexibilidad intrínseca (García 1992; Amadei et al. 1993; Wlodeck et al. 1997). Este análisis parte de la construcción y posterior diagonalización de la matriz de varianza-covarianza de las fluctuaciones cartesianas, de donde se obtienen $3N-6$ vectores propios, con sus respectivos valores propios asociados, que corresponden a los grados de libertad del sistema. El estudio de estos vectores proporciona información sobre la dinámica esencial del sistema, que puede resumirse en los siguientes puntos:

- la magnitud del valor propio (λ_i) indica la contribución de un modo principal a la varianza total del sistema,
- puede obtenerse la constante de deformación del sistema a lo largo del vector propio i aplicando la siguiente fórmula

$$K_i = k_B T / \lambda_i \quad (2.60)$$

donde k_B es la constante de Boltzmann (en kcal/mol·K), T es la temperatura absoluta y λ_i es el valor propio (en \AA^2),

- considerando un modelo armónico, podemos derivar la energía asociada a este movimiento (ec. 2.61), y usando este valor de energía en el cálculo de factores de Boltzmann para cada movimiento, asignar un peso específico a cada vector propio en análisis de comparación de movimientos de sistemas diferentes.

$$E_i = \frac{K_i}{2} (\Delta X_i)^2 \quad (2.61)$$

donde ΔX_i refiere los cambios en la magnitud de la proyección de una configuración concreta sobre el vector propio.

- el recuento del número de valores propios de magnitud superior a 1\AA^2 nos informa sobre la dimensionalidad real del sistema simulado, es decir, del número de movimientos esenciales que contribuyen a la flexibilidad del sistema.
- el producto de los vectores propios que contribuyen a la dimensionalidad del sistema reporta el valor del volumen conformacional muestreado, que, en su forma generalizada (independiente del número de valores propios empleados), se expresa como

$$Vol_g = \prod_{i=1}^l (1 + \lambda_i)^{1/2} \quad (2.62)$$

- se pueden definir índices de similitud en las fluctuaciones de dos macromoléculas como producto de vectores propios (v) (Hess 2000; Hess 2002). usando la fórmula

$$\gamma_{AB} = \frac{1}{n} \sum_{j=1}^n \sum_{i=1}^n (v_i^A \bullet v_j^B)^2 \quad (2.63)$$

donde A y B se refieren a los dos sistemas comparados y n es el número total de vectores propios relevantes para definir los movimientos de ambos sistemas.

Este índice se puede calcular para comparar la dinámica de dos porciones independientes de una misma trayectoria, generando índices de autosimilitud. Esto resulta especialmente útil porque se pueden definir descriptores normalizados por autosimilitud, en los que el ruido debido a fluctuaciones dentro de una misma trayectoria queda eliminado. Además, a partir de la energía de deformación asociada a cada movimiento del sistema (ec. 2.61), pueden usarse los pesos de Boltzmann para ponderar la importancia de cada vector propio en el valor final del índice de similitud (Perez et al. 2005).

2.5.3 Información energética

El grado de estabilidad termodinámica de un sistema viene esencialmente determinado por su energía libre. Para sistemas moleculares de tamaño medio, simulados con solvente explícito, la exploración del espacio de fase es computacionalmente inviable, lo que impide la determinación exacta de esta magnitud. No obstante, podemos aproximarnos a este valor mediante la evaluación separada de los diferentes términos que contribuyen a la energía libre: i) energía interna del soluto, ii) efecto del solvente y iii) entropía del soluto.

2.5.3.1. Energía interna del soluto

La evaluación de la energía interna de los diferentes estados conformacionales de un sistema puede extraerse tal como viene determinada por el campo de fuerzas. Por otro

lado, cálculos de la energía de interacción entre fragmentos de una misma macromolécula o entre receptor y ligando, empleando un *force-field* clásico, permiten desglosar las contribuciones a la energía total. Esto resulta de particular utilidad, por ejemplo, en el desglose de las energías de interacción entre bases del DNA en contribuciones de apilamiento o *stacking*, de puente de hidrógeno, y de repulsión electrostática entre fosfatos, entre otras. Numerosos estudios demuestran que la estimación de energías de interacción con el campo de fuerzas AMBER presenta una notable concordancia con datos de cálculos cuánticos de alto nivel (Hobza et al. 1997).

2.5.3.2. Estudio de la solvatación

Otro análisis particularmente relevante es el estudio de la solvatación. Si se ha empleado una representación explícita del solvente en la simulación, todos los parámetros estructurales del soluto quedan modificados por el efecto del entorno. Adicionalmente, puede utilizarse la información almacenada en las coordenadas de las partículas de solvente para estudiar las preferencias de solvatación del soluto:

- elaboración de mapas de densidad de aguas
- cálculo de funciones de distribución radial
- recuento del número de moléculas de solvente en esferas de tamaño creciente alrededor del soluto
- la aplicación de la teoría de la respuesta lineal para estimar la energía libre de solvatación a partir de datos de MD (Morreale et al. 2004)

Otras metodologías permiten estimar la energía libre de solvatación basándose en modelos de solvente continuo, como GB/SA (*Generalized Born-Solvent Accessible*) y PB/SA (*Poisson-Boltzmann/Solvent Accessible*), observándose una notable similitud en los resultados obtenidos por ambas (Cubero et al. 2001; Wang et al. 2001).

2.5.3.3. Entropía del soluto

El parámetro termodinámico que cuantifica la flexibilidad conformacional es la entropía del sistema, que puede ser estimada a partir de sus fluctuaciones espaciales.

Genéricamente, la entropía puede conocerse para todo sistema en el que se definan una serie de microestados y se obtenga la probabilidad de cada uno de ellos de forma estadísticamente significativa, según la fórmula

$$S \propto \sum_k P_k \log_2 P_k \quad (2.64)$$

donde P_k es la probabilidad con la que el microestado k será visitado.

Dado que el sumatorio de la ecuación 2.64 debe extenderse a todo el espacio de fase, y que la arbitrariedad en la definición de los microestados afecta y hace inestable el valor final de la entropía, el empleo de esta fórmula es técnicamente problemático. Por ello, se emplean con más frecuencia métodos que asumen un comportamiento armónico de los movimientos del sistema, como el método de Schlitter (ec. 2.65) (Schlitter 1993) y el de Andricioaei-Karplus (ec. 2.66) (Andricioaei and Karplus 2001).

$$S \approx 0.5k \sum_i \ln \left(1 + \frac{e^2}{\alpha^2} \right) \quad (2.65)$$

$$S = k \sum_i \frac{\alpha_i}{e^{\alpha_i} - 1} - \ln(1 - e^{-\alpha_i}) \quad (2.66)$$

donde $\alpha_i = \hbar\omega_i / kT$ y ω es cada uno de los valores propios que se obtienen de diagonalizar la matriz de varianza-covarianza pesada por masas.

La entropía varía dependiendo de la longitud de la simulación, pero su valor a tiempos de simulación largos (S_∞) puede modelizarse a partir de un ajuste según la ecuación

$$S_\infty = S(t) + \frac{a}{t^b} \quad (2.67)$$

donde a y b son parámetros ajustados empíricamente (Harris et al. 2001; Perez et al. 2005).

2.5.4 Información sobre reactividad

La elaboración de mapas de potencial de interacción molecular (MIP) alrededor de las estructuras obtenidas por MD es una estrategia muy informativa acerca de sus preferencias de reactividad. La metodología CMIP (Gelpi et al. 2001) permite calcular con bajo coste computacional mapas de reactividad del soluto frente a cualquier molécula que se pueda construir con los parámetros de un *force-field* clásico. Se basa en el cálculo del potencial de interacción como una suma de un término estérico y uno electrostático. El término estérico se introduce mediante un potencial de Lennard-Jones (ver apartado 2.2.2), y el término electrostático se evalúa mediante diferentes estrategias, que permiten introducir el efecto apantallador del entorno químico.

El modelo del dieléctrico constante constituye la estrategia más sencilla. Puede expresarse mediante la ecuación

$$V_{ele}(r_i) = \sum_{j=1}^J \frac{Q_j}{\epsilon |r_j - r_i|} \quad (2.68)$$

donde la sumatoria se extiende sobre J cargas puntuales Q_j , que corresponden a las cargas puntuales del soluto, y ϵ es la constante dieléctrica.

Pueden emplearse modelos donde el dieléctrico varía con la distancia de forma lineal o sigmoideal (Mehler and Solmajer 1991), o puede utilizarse el potencial derivado de la resolución de la ecuación de Poisson, especialmente adecuado en sistemas que combinan de forma heterogénea medios dieléctricos muy diferentes, como es el caso de los sistemas bioquímicos. La ecuación de Poisson relaciona el desplazamiento eléctrico ($D(r)$) con la densidad de carga (ρ), mediante la expresión

$$\nabla D(r) = 4\pi\rho(r) \quad (2.69)$$

en la que $D(r)$ se define como

$$D(r) = -\varepsilon(r)\nabla\phi(r) \quad (2.70)$$

donde $\phi(r)$ y $\varepsilon(r)$ designan el potencial electrostático y la constante dieléctrica, respectivamente.

Modificando esta expresión para que sea sensible a la presencia de iones móviles en el sistema, se obtiene la ecuación de Poisson-Boltzmann

$$\nabla\varepsilon(r)\nabla\phi(r) - \varepsilon(k^2)\sinh[\phi(r)] = -4\pi\rho(r) \quad (2.71)$$

en la que k^2 puede calcularse como se indica

$$k^2 = \frac{8\pi N_A e^2 I}{1000\varepsilon k_B T} \quad (2.72)$$

siendo N_A el número de Avogadro, e la carga del electrón, I la fuerza iónica de la disolución, ε la constante dieléctrica, k_B la constante de Boltzmann y T la temperatura.

En el estudio de macromoléculas, dada su complejidad estructural, esta ecuación ha de resolverse de forma numérica mediante la teoría de diferencias finitas (Gilson and Honig 1987; Gilson and Honig 1988).

2.6. Referencias

Alder, B. J. and T. E. Wainwright (1957). J Chem Phys **27**: 1208.

Aleman, C., E. I. Canela, R. Franco and M. Orozco (1991). Journal of Computational Chemistry **12**(6): 664.

Alhambra, C., F. J. Luque and M. Orozco (1995). J. Phys. Chem. **99**: 3084.

Allen, F. H., O. Kennard, D. G. Watson, L. Brammer, A. G. Orpen and R. Taylor (1987). S1.

- Amadei, A., A. B. Linssen and H. J. Berendsen (1993). Proteins **17**(4): 412.
- Andricioaei, I. and M. Karplus (2001). Journal of Chemical Physics **115**(14): 6289.
- Barth, E. and T. Schlick (1998). J Chem Phys **109**: 1633.
- Bayly, C. I., P. Cieplak, W. D. Cornell and P. A. Kollman (1993). Journal of Physical Chemistry **97**: 10269.
- Becke, A. D. (1993). J Chem Phys **98**: 5648.
- Beglov, D. and B. Roux (1994). J Chem Phys **100**: 9050.
- Berendsen, H. J. C., J. P. M. Postma, W. F. Van Gunsteren, A. DiNola and J. R. Haak (1984). Journal of Chemical Physics **81**(8): 3684.
- Bhandarkar, M., R. Brunner, C. Chipot, A. Dalke, P. Dixit, P. Grayso, J. Gullinsgrud, A. Gurov, W. Humprey, D. Hurwitz, N. Krawetz, M. Nelson, J. Phillips, A. Shinozaki and F. Zhu (2005). NAMD User's Guide. Urbana-Champaign.
- Bishop, T. C., R. D. Skeel and K. Shulten (1997). J Comput Chem **18**: 1785.
- Born, M. and R. Oppenheimer (1927). Annalen der Physik (Berlin, Germany) **84**: 457.
- Boys, S. F. (1950). Proc R Soc (London) **200**: 542.
- Boys, S. F. and F. Bernardi (1970). Mol Phys **19**: 553.
- Cannon, J. F. (1993). J Comput Chem **14**: 995.
- Case, D. A., D. A. Pearlman, J. W. Caldwell, T. E. Cheatham, III, W. S. Ross, C. L. Simmerling, T. L. Darden, K. M. Marz, R. V. Stanton, A. L. Cheng, J. J. Vincent, M. Crowley, V. Tsui, R. J. Radmer, Y. Duan, J. Pitera, I. Massova, G. L. Seibel, U. C. Singh, P. K. Weiner and P. A. Kollman (2004). AMBER8. San Francisco, University of California.
- Cizek, J. (1966). J Chem Phys **45**: 4256.
- Cizek, J. (1969). Adv Chem Phys **14**: 35.

- Clark, T., J. Chandrasekhar, G. W. Spitznagel and P. v. R. Schleyer (1983). J Comput Chem **4**: 294.
- Claverie, P. (1978). Intermolecular Interactions: From Diatomics to Biomolecules. B. Pullman. Chichester, Wiley: 69.
- Cox, S. R. and R. E. Williams (1981). J Comput Chem **2**: 303.
- Cramer, C. J. C. (2002). Essentials of Computational Chemistry: Theories and Models. Chichester, Wiley.
- Cubero, E., F. J. Luque and M. Orozco (2001). J Am Chem Soc **123**(48): 12018.
- Cubero, E., F. J. Luque, M. Orozco and J. Gao (2003). J Phys Chem B **107**(7): 1664.
- Curutchet, C., M. Orozco and F. J. Luque (2001). J Comput Chem **22**: 1180.
- Darden, T. L., D. York and L. Pedersen (1993). Journal of Chemical Physics **98**: 10089.
- Dewar, M. and W. Thiel (1977). J Am Chem Soc **99**: 4899.
- Dewar, M. J. S., E. G. Zoebisch, E. F. Healy and J. J. P. Stewart (1985). J Am Chem Soc **107**: 3902.
- Duan, Y., C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo and T. Lee (2003). J Comput Chem **24**: 1999.
- Dunning Jr., T. H. (1989). J Chem Phys **90**: 1007.
- Ewald, P. P. (1921). Ann. Phys. **64**: 253.
- Feller, S. E., Y. Zhang, R. W. Pastor and B. R. Brooks (1995). J Chem Phys **103**: 4613.
- Frisch, M. J., J. A. Pople and J. S. Binkley (1984). J Chem Phys **80**: 3265.
- García, A. E. (1992). Physical Review Letters **68**: 2696.
- Gear, C. W. (1971). Numerical initial value problems in ordinary differential equations. Englewood Cliffs, NJ, Prentice Hall.
- Gelpi, J. L., S. G. Kalko, X. Barril, J. Cirera, X. de La Cruz, F. J. Luque and M. Orozco (2001). Proteins **45**(4): 428.

- Gilson, H. S. and P. H. Honig (1988). Proteins **4**: 7.
- Gilson, M. K. and P. H. Honig (1987). Nature **330**: 84.
- Hariharan, P. C. and J. A. Pople (1973). Theor Chim Acta **28**: 213.
- Harmony, M. D., R. W. Laurie, R. L. Kuczkowki, R. H. Schwendemann, D. A. Ramsay, F. J. Lovas, F. J. Lafferty and A. G. Maki (1979). J Phys Chem Ref Data **8**: 619.
- Harris, S. A., E. Gavathiotis, M. S. Searle, M. Orozco and C. A. Laughton (2001). J Am Chem Soc **123**(50): 12658.
- Harvey, S. C., T. R.K. and I. Cheatham, T.E. (1998). J Comput Chem **19**: 726.
- Hehre, W. J., L. Radom, P. Schleyer and J. A. Pople (1986). Ab Initio Molecular Orbital Theory. New York, Wiley.
- Hehre, W. J., R. F. Stewart and J. A. Pople (1969). J Chem Phys **51**: 2657.
- Hess, B. (2000). Phys Rev E Stat Nonlin Soft Matter Phys(6): 8438.
- Hess, B. (2002). Phys Rev E Stat Nonlin Soft Matter Phys **65**(3 Pt 1): 031910.
- Hobza, P., M. Kabelac, J. Sponer, P. Mejzlik and J. Vondrasek (1997). Journal of Computational Chemistry **18**(9): 1136.
- Hockney, R. W. (1970). Methods in Computational Physics **9**: 136.
- Hohenberg, P. and W. Kohn (1964). Phys Rev B **136**: 864.
- Hopfinger, A. J. and R. A. Pearlstein (1985). J Comput Chem **5**(486).
- Im, W., S. Bernéche and B. Roux (2001). J Chem Phys **114**: 2924.
- Jensen, F. (1996). Chem Phys Letters **261**: 633.
- Jorgensen, W. L. and C. Ravimohan (1985). J Chem Phys **83**: 3050.
- Kaufman, J. J., P. C. Hariharan and F. C. Tobin (1981). Chemical Applications of Atomic and Molecular Electrostatic Potential. New York, Plenum Press.
- Kendall, R. A., T. H. Dunning Jr. and R. J. Harrison (1992). J Chem Phys **96**: 6796.
- Kirkwood, J. (1935). **3**: 300.

- Kohn, W. and L. J. Sham (1965). Phys Rev A **140**: 1133.
- Lankas, F., J. Sponer, P. Hobza and J. Langowski (2000). J Mol Biol **299**(3): 695.
- Lankas, F., J. Sponer, J. Langowski and T. E. Cheatham, 3rd (2003). Biophys J **85**(5): 2872.
- Lavery, R. and H. Sklenar (1988). J Biomol Struct Dyn **6**: 63.
- Lavery, R. and H. Sklenar (1989). J Biomol Struct Dyn **6**(4): 655.
- Leach, A. R. (1996). Molecular Modelling. Principles and Applications. Singapore, Longman.
- Leach, A. R. (2001). Molecular Modelling: Principles and applications, 2nd edition. Essex, Pearson Education (Prentice Hall).
- Lee, C., W. Yang and R. G. Parr (1988). Physical Review B. Condensed Matter **37**(2): 785.
- Levitt, M. (1983). Cold Spring Harb Symp Quant Biol **47 Pt 1**: 251.
- Lu, X. J. and W. K. Olson (2003). Nucleic Acids Res **31**: 5108.
- Lu, X. J., Z. Shakked and W. K. Olson (2000). J Mol Biol **300**(4): 819.
- Luque, F. J., M. Bachs, C. Alemán and M. Orozco (1996a). J Comput Chem **17**: 806.
- Luque, F. J. and M. Orozco (1998). J Comput Chem **19**: 866.
- Luque, F. J., Y. Zhang, C. Alemán, M. Bachs, J. Gao and M. Orozco (1996b). J. Phys. Chem. **100**: 4269.
- McCammon, J. A. (1976). Models for Protein Dynamics. H. J. C. Berendsen. Orsay (France), CECAM - Universite de Paris IX: 137.
- McCammon, J. A., B. R. Gelin and M. Karplus (1977). Nature **267**(5612): 585.
- Mehler, E. L. and T. Solmajer (1991). Protein Eng **4**(8): 903.
- Miertus, S., E. Scrocco and J. Tomasi (1981). J Chem Phys **55**: 117.
- Miertus, S. and J. Tomasi (1982). J Chem Phys **65**: 239.

- Moller, C. and M. S. Plesset (1934). Phys Rev **46**: 618.
- Morreale, A., X. de la Cruz, T. Meyer, J. L. Gelpi, F. J. Luque and M. Orozco (2004). Proteins **57**(3): 458.
- Murray, J. S. and K. Sen (1996). Molecular Electrostatic Potential: Principles and Applications. Amsterdam, Elsevier.
- Orozco, M. and F. J. Luque (1993). J Comput Chem **14**: 587.
- Orozco, M. and F. J. Luque (1996). Theoretical and Computational Chemistry. New York, Elsevier Science. **3**.
- Parr, R. G. and W. Yang (1989). Density-Functional Theory of Atoms and Molecules. Oxford, Oxford University Press.
- Perdew, J. P. and Y. Wang (1992). Phys Rev B **45**: 13244.
- Perez, A., J. R. Blas, M. Rueda, J. M. López-Bes, X. de La Cruz, F. J. Luque and M. Orozco (2005). Journal of Chemical Theory and Computation **1**(5): 790.
- Perez, A., A. Noy, F. Lankas, F. J. Luque and M. Orozco (2004). Nucleic Acids Res **32**(20): 6144.
- Peterson, K. A., D. E. Woon and T. H. Dunning Jr. (1994). J Chem Phys **100**: 7410.
- Pierotti, R. A. (1976). Chem Rev **76**: 717.
- Pople, J. A., M. Head-Gordon and K. Raghavachari (1987). J Chem Phys **87**: 5968.
- Pople, J. A., R. Krishnan, H. B. Schlegel and J. S. Binkley (1978). Int J Quant Chem **XIV**: 545.
- Pople, J. A., D. P. Santry and G. A. Segal (1965). J Chem Phys **43**: S129.
- Pople, J. A. and G. A. Segal (1965). J Chem Phys **43**: S136.
- Purvis, G. D. and R. J. Bartlett (1982). J Chem Phys **76**: 1910.
- Rahman, A. (1964). Physical Review **136**(2A): 405.
- Rahman, A. and F. H. Stillinger (1971). J Chem Phys **55**: 3336.

- Roothaan, C. C. J. (1951). Rev Mod Phys **23**: 69.
- Ryckaert, J. P., G. Ciccotti and H. J. C. Berendsen (1977). Journal of Computational Physics **23**(3): 327.
- Schlitter, J. (1993). Chemical Physics Letters **215**(6): 617.
- Scott, W. R. P., P. H. Huenenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Krueger and W. F. van Gunsteren (1999). Journal of Physical Chemistry A **103**(19): 3596.
- Scrocco, E. and J. Tomasi (1973). Top. Curr. Chem. **42**: 95.
- Singh, U. C. and P. A. Kollman (1984). J Comput Chem **5**: 129.
- Slater, J. C. (1930). Phys Rev **36**: 57.
- Stewart, J. J. P. (1989). J Comput Chem **10**: 209.
- Stewart, J. J. P. (1990). Reviews in Computational Chemistry. New York, VCH. **1**: 45.
- Szabo, A. and N. S. Ostlund (1982). Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory. New York, McGraw-Hill.
- Tao, J., V. N. Perdew, V. N. Staroverov and G. E. Scuseria (2003a). Physical Review Letters **91**: 146401.
- Tao, J., V. N. Perdew, V. N. Staroverov and G. E. Scuseria (2003b). J Chem Phys **119**: 12129.
- Tao, J., V. N. Perdew, V. N. Staroverov and G. E. Scuseria (2004a). J Chem Phys **120**: 6898.
- Tao, J., V. N. Perdew, V. N. Staroverov and G. E. Scuseria (2004b). Physical Review B. Condensed Matter **69**: 075102.
- Tembe, B. L. and J. A. McCammon (1984). J Comput Chem **8**: 281.
- Tidor, B., K. K. Irikura, B. R. Brooks and M. Karplus (1983). J Biomol Struct Dyn **1**(1): 231.

- Tuckerman, M., B. J. Berne and G. J. Martyna (1992). J Chem Phys **97**: 1990.
- van der Spoel, D., E. Lindahl, B. Hess, A. R. van Buuren, E. Apol, P. J. Meulenhoff, D. P. Tieleman, A. Sijbers, K. A. Feenstra, R. van Drunen and H. J. Berendsen (2004). Gromacs User Manual version 3.2.
- van Gunsteren, W. F., S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, A. M. Scott and I. G. Tironi (1996). Biomolecular simulations: The GROMOS96 manual and user's guide. Hochschulverlag Ag an der ETH Zürich. Zurich.
- Verlet, L. (1967). Physical Review **159**: 98.
- Vosko, S. H., L. Wilk and M. Nusair (1980). Canadian Journal of Physics **58**: 1200.
- Wang, W., O. Donini, C. M. Reyes and P. A. Kollman (2001). Annu Rev Biophys Biomol Struct **30**: 211.
- Warshel, A. (1976). Nature **260**(5553): 679.
- Weiner, S. J., P. A. Kollman, D. A. Case, U. Chandra Singh, C. Ghio, G. Alagona, S. J. Profeta and P. Weiner (1984). J Am Chem Soc **106**: 765.
- Wlodeck, S. T., T. W. Clark, L. R. Scott and J. A. McCammon (1997). J Am Chem Soc **119**: 9513.
- Woodcock, L. V. (1971). Chemical Physics Letters **10**: 257.
- Woon, D. E. and T. H. Dunning Jr. (1993). J Chem Phys **98**: 1358.
- Zerner, M. C. (1991). Reviews in Computational Chemistry. New York, VCH. **2**: 313.
- Zwanzig, R. W. (1954). Journal of Chemical Physics **22**: 1420.