

UNIVERSITAT DE BARCELONA
DEPARTAMENT DE BIOQUÍMICA I BIOLOGIA MOLECULAR

**ESTUDIO TEÓRICO DE NUCLEOBASES:
IMPLICACIONES ESTRUCTURALES EN
ÁCIDOS NUCLEICOS**

ELENA CUBERO I JORDÀ



UNIVERSITAT DE BARCELONA

U

B

UNIVERSITAT DE BARCELONA
DEPARTAMENT DE BIOQUÍMICA I BIOLOGIA MOLECULAR

**ESTUDIO TEÓRICO DE NUCLEOBASES:
IMPLICACIONES ESTRUCTURALES EN
ÁCIDOS NUCLEICOS**

ELENA CUBERO I JORDÀ

Programa de doctorado de Bioquímica i Biologia Molecular
Bienio 1997-1999

Estudio teórico de nucleobases:
Implicaciones estructurales en ácidos nucleicos

El trabajo experimental y la redacción de la presente memoria han sido realizados por la licenciada Elena Cubero i Jordà bajo la dirección del Dr. Modesto Orozco i López, profesor titular del Departament de Bioquímica i Biologia Molecular de la Universitat de Barcelona, y del Dr. Francisco Javier Luque i Garriga, profesor titular del Departament de Físico-química de la Universitat de Barcelona, para optar al título de doctora en Química.

Barcelona, septiembre de 2001

Dr. Modesto Orozco i López

Dr. F. Javier Luque i Garriga

A mi familia

Índice

<u>Abreviaturas y acrónimos.....</u>	1
<u>1 Introducción.....</u>	7
1.1 Estudio teórico de nucleobases: Implicaciones estructurales en ácidos nucleicos.....	12
1.2 Referencias bibliográficas.....	13
<u>2 Métodos teóricos.....</u>	15
2.1 Descripción del sistema.....	17
2.1.1 Métodos cuánticos (QM).....	18
2.1.1.1 <i>Métodos ab initio</i>	19
2.1.1.2 <i>Métodos semiempíricos</i>	22
2.1.1.3 <i>Métodos basados en el funcional de la densidad (DFT)</i> ..	23
2.1.1.4 <i>Estudio de propiedades moleculares</i>	25
2.1.1.5 <i>Consideraciones prácticas</i>	30
2.1.2 Métodos clásicos (MM).....	32
2.1.2.1 <i>Cálculo de la energía en un campo de fuerzas</i>	33
2.1.2.2 <i>Parametrización</i>	36
2.1.2.3 <i>Métodos basados en campos de fuerzas</i>	36
2.1.3 Métodos mixtos (QM/MM).....	44
2.2 Inclusión de los efectos de entorno.....	49

2.2.1	Naturaleza de la interacción soluto-solvente.....	49
2.2.2	Métodos teóricos de representación del solvente.....	50
2.2.2.1	<i>Métodos discretos</i>	51
2.2.2.2	<i>Métodos continuos</i>	51
2.2.2.3	<i>Métodos mixtos discreto-continuo o semicontinuos</i>	52
2.2.2.4	<i>Cálculo de la energía libre de solvatación</i>	53
2.2.3	Ciclos termodinámicos.....	58
2.3	Análisis de resultados.....	60
2.4	Referencias bibliográficas.....	64
3	<u>Estructura de los ácidos nucleicos</u>	73
3.1	Componentes estructurales de los ácidos nucleicos.....	75
3.1.1	Estructura y nomenclatura.....	75
3.1.2	Conformación.....	79
3.1.3	Propiedades fisicoquímicas y reactividad.....	85
3.1.4	Derivados sintéticos de ácidos nucleicos.....	89
3.2	Interacciones entre bases.....	90
3.2.1	Interacciones por puente de hidrógeno.....	90
3.2.2	Interacciones por apilamiento (<i>stacking</i>).....	94
3.2.3	Dinámica de interacciones entre bases.....	96
3.3	Estructura secundaria del ADN.....	97
3.3.1	Parámetros conformacionales para describir el ADN.....	98
3.3.2	Estructuras canónicas del ADN.....	98
3.3.3	Estructura no canónicas del ADN.....	105
3.3.4	Fuerzas estabilizadoras.....	113

3.4 Aspectos dinámicos de la flexibilidad conformacional del ADN.....	116
3.5 Interacción del ADN con otras moléculas.....	117
3.6 Estructura terciaria del ADN.....	119
3.7 Referencias bibliográficas.....	123
4 Resultados.....	133
4.1 Interacciones π -catión.....	135
4.1.1 Is polarization important in cation- π interactions?.....	137
4.1.2 A topological analysis of electron density in cation- π complexes.....	145
4.2 Interacciones por puente de hidrógeno impropio.....	155
4.2.1 Hydrogen bond versus anti-hydrogen bond: A comparative analysis based on the electron density topology.....	157
4.2.2 C-H \cdots O contacts in the adenine-uracil Watson-Crick and uracil-uracil nucleic acid base pairs: Nonempirical <i>ab initio</i> study with inclusion of electron correlation effects.....	167
4.3 Interacciones canónicas con bases modificadas.....	177
4.3.1 The effect of amino groups on the stability of DNA duplexes and triplexes based on purines derived from inosine.....	179
4.3.2 Observation of spontaneous base pair breathing events in the molecular dynamics simulation of a difluorotolueno-containing DNA oligonucleotide.....	195
4.3.3 Molecular dynamics study of oligonucleotides containing difluorotoluene.....	199
4.4 Interacciones no canónicas entre bases.....	211

4.4.1	Theoretical studies of d(A:T)-based parallel-stranded DNA duplexes.....	213
4.4.2	Parallel-stranded hairpins containing 8-aminopurines. Novel efficient probes for triple-helix formation.....	235
4.4.3	Hoogsteen-based parallel-stranded duplexes of DNA. The effect of 8-amino derivatives.....	241
4.5	Referencias bibliográficas.....	265
5	<u>Discusión de resultados.....</u>	267
5.1	La naturaleza de las interacciones π -catión.....	269
5.2	La naturaleza de los puentes de hidrógeno impropios.....	276
5.3	El efecto de las bases modificadas.....	284
5.4	Apareamientos no canónicos.....	291
5.5	Referencias bibliográficas.....	298
6	<u>Conclusiones.....</u>	303
	<u>Otros trabajos publicados.....</u>	309
	<u>Agradecimientos.....</u>	313

Abreviaturas y acrónimos

(Esta página está intencionadamente en blanco)

A	Adenina
AA (o A ^N)	AminoAdenina
AI (o I ^N)	AminoInosina
AIM	Atoms In Molecules theory
ADN (o DNA)	Ácido DesoxirriboNucleico
ADP	Adenosine DiPhosphate
AG (o G ^N)	AminoGuanina
AM1	Austin Model 1
AMBER	Assisted Model Building with Energy Refinement
aps (o APS)	antiparallel stranded
ARN (o RNA)	Ácido RiboNucleico
ATP	Adenosine TriPhosphate
B3LYP	Becke 3 parameter hybrids & Lee-Yang-Parr correlation
BCP	Bond Critical Point
<i>BSSE</i>	Basis Set Superposition Error
C	Citosina
cAMP	cyclic Adenosine MonoPhosphate
cav	cavitación
CC	Coupled Cluster method
CD	Circular Dichroism spectroscopy
CI	Configuration Interaction method
cMIP	classical Molecular Interaction Potential
DFT	Density Functional Theory
DZ	Double-Zeta basis set
<i>E</i>	Envelope conformer
E	Energía; East region
EDT	Electron Density Transfer
ele	electrostático/a
ESP	ElectroStatic Potential

ext	externo/a
F	Difluorotolueno
F	Fuerza
FAD	Flavine Adenine Dinucleotide
FEP	Free Energy Perturbation
FF	Force Field
G	Energía libre; Guanina
GGA	Generalized Gradient Approximation
GMIP	Generalized Molecular Interaction Potential
GMIPp	Generalized Molecular Interaction Potential with polarization
GTO	Gaussian Type Orbital
H	Operador Hamiltoniano
H	Entalpía; Hoogsteen base pair
hb (o H-bond)	hydrogen bonding
HF	Hartree-Fock method
HOMO	Highest Occupied Molecular Orbital
HPLC	High Performance Liquid Chromatography
I	Inosina
int	interacción; interno/a
IR	InfraRed spectroscopy
k	cinético/a
LCAO	Linear Combination Atomic Orbitals
LDA	Local Density Approximation
LUMO	Lowest Unoccupied Molecular Orbital
m	minor groove
M	Major groove
MBPT	Many Body Perturbation Theory
MC	Monte Carlo
MD	Molecular Dynamics
MEAD	Macroscopic Electrostatics with Atomic Detail
MEP	Molecular Electrostatic Potential
MIP	Molecular Interaction Potential

MIPp	Molecular Interaction Potential with polarization
mM	minor Major groove
MM	Molecular Mechanics; Major Major groove
MO	Molecular Orbital
MP	Møller-Plesset perturbation theory
MST	Miertus-Scrocco-Tomasi method
N	North region
NAD ⁺	Nicotinamide Adenine Dinucleotide
NADP ⁺	Nicotinamide Adenine Dinucleotide Phosphate
NBO	Natural Bond Orbital
NDDO	Neglect of Diatomic Differential Overlap
OA	Orbitales Atómicos
OPLS	Optimized Potentials for Liquid Simulations
P	Presión; Ángulo de pseudorotación
PB	Poisson-Boltzmann
PBC	Periodic Boundary Conditions
PCA	Principle Component Analysis
PCM	Polarized Continuum Model
PME	Particle Mesh Ewald method
PMF	Potential Mean Force
PNA	Polyamide Nucleic Acid
pol	polarización
pot	potencial
ps (o PS)	parallel-stranded
<i>pur</i>	purine
<i>pyr</i>	pyridine
QM	Quantum Mechanics
RESP	Restrained ElectroStatic Potential
rH	reverse Hoogsteen base pair
RHF	Restricted Hartree-Fock method
RMN (o NMR)	Resonancia Magnética Nuclear
RMSd (o r.m.s.d)	Root Mean Square deviation

rWC	reverse Watson-Crick base pair
S	Entropía; South region
SAS (o SA)	Solvent Accessible Surface
SASA	Solvent Accessible Surface Area
SCF	Self Consistent Field
SCRf	Self Consistent Reaction Field
SM	Statistical Mechanics
sol	solución
solv	solvatación
STO	Slater Type Orbital
<i>T</i>	Twist conformer
T	Temperatura; Timina
T_m (o T_m)	melting Temperature
TI	Thermodynamic Integration
TLC	Thin Layer Chromatography
TFO	Triplex Forming Oligonucleotide
tot	total
TZ	Triple-Zeta basis set
U	Uracilo
ua	unidades atómicas
UHF	Unrestricted Hartree-Fock method
UV	Ultraviolet/Visible spectroscopy
<i>V</i>	Perturbación
V	Volumen; Potencial
vdW	van der Waals
W	West region
WC	Watson-Crick base pair
ZDO	Zero-Differential Overlap
ZPE	Zero Point Energy

Introducción

(Esta página está intencionadamente en blanco)

Los nucleótidos, presentes en todo organismo vivo, son una familia de moléculas muy versátiles¹ que participan en casi todos los procesos bioquímicos desempeñando una gran variedad de funciones, tales como unidades monoméricas de ácidos nucleicos, especies implicadas en vías liberadoras de energía, componentes de intermediarios activados en procesos de biosíntesis, reguladores metabólicos, mediadores de señales hormonales o componentes de coenzimas. Todas estas funciones son vitales para asegurar el buen funcionamiento celular.

Tal vez la función más destacable de los nucleótidos es la de ser constituyentes de los ácidos nucleicos^{2,3}, los responsables de almacenar y transmitir la información genética. La composición de los ácidos nucleicos es relativamente simple, al estar formados por largas cadenas en las que intervienen unos pocos nucleótidos. De hecho, el código genético involucra tan sólo 5 nucleótidos, que se diferencian por la nucleobase unida al azúcar: adenina (A), timina (T), guanina (G), citosina (C) y uracilo (U). Además la modificación química de alguna de estas unidades monoméricas podría tener consecuencias letales para el organismo.

El descubrimiento de la naturaleza de las moléculas portadoras del material genético y de su estructura química y propiedades ha sido uno de los grandes hitos alcanzados en la investigación científica. El modelo tridimensional de doble hélice dextrógira para el ADN, propuesto por Watson y Crick⁴ en 1953, abrió el camino hacia la comprensión de cómo podía desempeñar esta molécula sus funciones, almacenando y transmitiendo la información genética. Desde entonces se ha registrado un gran avance en el conocimiento de la estructura y función de los ácidos nucleicos. No obstante, todavía quedan muchas incógnitas sobre su estructura y su variabilidad conformacional; el efecto de cambios en el medio; los mecanismos de expresión del código genético, de los procesos de replicación y transcripción, del reconocimiento del ADN con proteínas, fármacos u otras moléculas.

La funcionalidad de los ácidos nucleicos viene predeterminada por su estructura tridimensional y ésta se mantiene gracias a una serie de interacciones no covalentes. La estructura de un dúplex de ADN, en solución acuosa, está gobernada por un balance de fuerzas no covalentes⁵. Algunas fuerzas son virtualmente siempre estabilizantes, como los puentes de hidrógeno Watson-Crick y las interacciones de apilamiento entre bases. Otras son casi siempre desestabilizantes, como la repulsión electrostática de los fosfatos. La interrelación entre fuerzas es compleja y puede resultar difícil obtener un balance adecuado de las diferentes interacciones. Consecuentemente, el estudio y la comprensión de dichas interacciones y su papel en la estructura y estabilidad del ADN son clave para conocer mejor las funciones de los ácidos nucleicos.

Actualmente se considera que el ADN presenta *in vivo* una estructura de doble hélice dextrógira de cadenas antiparalelas, con una repetitividad de 10 pares de bases por vuelta (forma B), donde cada paso está formado por pares de bases A·T o G·C. No obstante, esta estructura no es única y experimentalmente se han descrito otras formas del ADN, alguna de las cuales puede tener gran importancia fisiológica². Así, el ADN puede adoptar diversas formas de doble hélice, como el A-ADN o el Z-ADN y otras estructuras tridimensionales inusuales, incluyendo hélices de 3 y 4 cadenas. Asimismo, se han determinado nuevos apareamientos no canónicos entre bases y se ha descubierto el *bending* natural de ADN e incluso el apareamiento espontáneo de algunos pasos. Todas estas observaciones revelan que el ADN posee una enorme variabilidad conformacional mucho más allá de la doble hélice propuesta por Watson y Crick.

Una creencia generalizada es que la flexibilidad conformacional del ADN es, en gran parte, responsable de su funcionalidad biológica. Así, son obvias las posibilidades terapéuticas que comportaría poder interferir en las preferencias conformacionales de los ácidos nucleicos, del ADN en particular. Esto ha abierto un nuevo campo de investigación como es la modificación química de los ácidos nucleicos⁵, que puede conducir a generar nuevas propiedades en los ácidos nucleicos útiles para sus aplicaciones en áreas de biotecnología o biomedicina.

Gran parte del progreso en el campo de los ácidos nucleicos procede del desarrollo de nuevos métodos y de nuevas aproximaciones al análisis de datos². En particular, la mejora de nuestro conocimiento estructural de los ácidos nucleicos se ha logrado fundamentalmente gracias a los estudios de difracción de cristales de alta resolución y a la espectroscopia de RMN. Los métodos de análisis computacional, así como el análisis de las bases de datos termodinámicas para la predicción de estructura secundaria, proporcionan información adicional. De hecho, el rápido progreso de las técnicas computacionales, debido tanto al desarrollo de *hardware* y *software*, así como al hecho de disponer de un marco formal mucho más maduro para estas técnicas, ha permitido su extensa aplicación a las biodisciplinas, incluyendo varios aspectos del estudio de los ácidos nucleicos y sus constituyentes⁶⁻⁸.

La química computacional puede ayudarnos en la comprensión de la estructura y de las interacciones, la energética y las propiedades reactivas de los ácidos nucleicos^{7,8}. Los estudios teóricos incluyen la aplicación de la mecánica cuántica, la mecánica y dinámica molecular, y la mecánica estadística como herramientas para racionalizar, comprender y predecir propiedades observables experimentalmente. Por otro lado, si los fundamentos teóricos y los parámetros están suficientemente bien establecidos, la química computacional también puede ser usada para calcular aspectos del comportamiento molecular que no son observables experimentalmente⁶.

En conjunto, pues, la comprensión de la función y la estructura de los ácidos nucleicos y de sus constituyentes requiere un profundo conocimiento de las interacciones no covalentes. Los experimentos en fase gas y en solución, los cálculos teóricos cuánticos y los estudios dinámicos en solución proporcionan un paso clave en esta dirección.

Estudio teórico de nucleobases: Implicaciones estructurales en ácidos nucleicos

Esta memoria recoge el estudio teórico de algunos de los más importantes tipos de interacciones débiles no covalentes en el ADN. Los sistemas de estudio van desde bases nitrogenadas y nucleósidos en fase gas hasta los ácidos nucleicos en solución.

Concretamente se tratarán las interacciones débiles π -catión y los contactos entre bases con el objetivo de profundizar en su naturaleza. Dentro de las interacciones entre bases podemos diferenciar entre: (i) las interacciones de puente de hidrógeno impropio, (ii) las interacciones con bases modificadas, como la 8-aminoinosina y el difluorotolueno, y finalmente (iii) las interacciones no canónicas: Hoogsteen y Watson-Crick reverso, que pueden dar lugar a nuevas estructuras secundarias del ADN.

Se utilizan todo un conjunto de técnicas teóricas que van desde la mecánica clásica hasta los más elevados niveles de mecánica cuántica, con la finalidad de aportar la máxima información estructural, reactividad y dinámica que complementen así los datos experimentales disponibles sobre las nucleobases y los ácidos nucleicos.

En la primera parte de la presente memoria se describirán los métodos teóricos empleados para caracterizar las interacciones no covalentes. En la segunda parte se detallarán brevemente la estructura y las propiedades de los ácidos nucleicos. A continuación, se presentarán los resultados de las interacciones estudiadas. Y finalmente, se procederá a la discusión de los resultados obtenidos y a las conclusiones.

Referencias bibliográficas

- 1) Voet, D. & Voet, J.G. (1992) En *Bioquímica*. Ediciones Omega, S.A., Barcelona.
- 2) Saenger, W. (1988) En *Principles of Nucleic Acid Structure*. Springer-Verlag, New York.
- 3) Bloomfield, V.A., Crothers, D.M., & Tinoco, Jr., I. (2000) En *Nucleic Acids: Structures, Properties, and Functions*. University Science Books, Sausalito, CA.
- 4) Watson, J.D. & Crick, F.H.C. (1953) *Nature*, **171**, 737.
- 5) Kool, E.T. (1997) *Chem. Rev.*, **97**, 1473.
- 6) Müller-Dethlefs, K. & Hobza, P. (2000) *Chem. Rev.*, **100**, 143.
- 7) Hobza, P. & Šponer, J. (1999) *Chem. Rev.*, **99**, 3247.
- 8) Cheatham III, T.E. & Kollman, P.A. (2000) *Annu. Rev. Phys. Chem.*, **51**, 435.

(Esta página está intencionadamente en blanco)

2 Métodos teóricos

(Esta página está intencionadamente en blanco)

El conjunto de métodos teóricos abarca un amplio abanico de metodologías computacionales^{1,2}, las cuales difieren entre ellas básicamente en el grado de rigurosidad en que se lleva a cabo la aplicación del formalismo teórico. Este nivel de teoría se refleja especialmente en el funcional utilizado para evaluar las interacciones entre partículas del sistema. Como regla general, a medida que aumenta el nivel teórico, aumenta la calidad de los resultados que se derivan y el coste computacional del mismo, lo cual disminuye el tamaño del sistema susceptible de estudio.

Una descripción detallada de todos ellos no entra dentro de los objetivos de esta memoria, si bien en este capítulo se revisan brevemente aquellos métodos que se han empleado en el estudio teórico de las interacciones débiles de los ácidos nucleicos.

2.1 Descripción del sistema

El estudio teórico molecular de cualquier sistema químico implica tres etapas. Primero, es necesario decidir el nivel de detalle microscópico que se desea alcanzar a la hora de describir las especies que constituyen el sistema, lo cual a su vez depende de las propiedades que son objeto de interés. Posteriormente, debe escogerse el formalismo teórico empleado para evaluar las interacciones entre las partículas del sistema, lo que dependerá de los siguientes factores: la naturaleza de las interacciones que se desean estudiar, el tamaño del sistema y los recursos computacionales disponibles. Finalmente, se realiza el cálculo que permite obtener la energía del sistema y determinadas propiedades de éste.

A continuación se presentan los fundamentos de las estrategias computacionales más empleados para describir a nivel atómico los ácidos nucleicos.

2.1.1 Métodos cuánticos (QM)¹⁻⁵

El objeto de la metodología cuántica (QM) es la interpretación y la predicción de la estructura molecular y de la reactividad química a partir de la formulación cuántica de la física molecular. Estas técnicas se desarrollaron inicialmente para abordar el estudio de pequeños sistemas en fase gas. Su extensión a sistemas de impacto bioquímico ha precisado del desarrollo de aproximaciones y métodos computacionales que permitan abordar cada vez con mayor exactitud cuantitativa sistemas de complejidad creciente.

Dentro del formalismo de la mecánica cuántica (QM), un sistema queda descrito por su función de onda, ya que ésta incluye toda la información del sistema. La función de onda puede ser obtenida a partir de la resolución de la correspondiente ecuación de Schrödinger dependiente del tiempo o dentro del nivel de la aproximación usual de la ecuación independiente del tiempo (ecuación 2.1), es decir, encontrando las funciones propias del operador hamiltoniano (H).

$$H\Psi = E\Psi \quad (2.1)$$

donde H es el operador hamiltoniano que incluye la energía cinética y potencial de núcleos y electrones, E es la energía del sistema y Ψ la función de onda.

Para la gran mayoría de sistemas químicos de interés, las dificultades que comporta la resolución de esta ecuación (2.1) hace necesaria la aplicación de toda una serie de aproximaciones.

La primera de ellas es la aproximación de Born-Oppenheimer, fundamentada en el hecho que los electrones se mueven mucho más rápido que los núcleos. Basándose en esta aproximación se consideran por separado el movimiento de los núcleos y el de los electrones.

El resto de aproximaciones que pueden aplicarse (sobre los términos que componen el hamiltoniano, sobre el tipo de función de onda y sobre el proceso de resolución de la ecuación) da lugar a las diferentes metodologías cuánticas que existen, como los métodos *ab initio*, los métodos semiempíricos o los métodos basados en el funcional de la densidad (DFT).

2.1.1.1 Métodos *ab initio*

Los métodos *ab initio*⁶ se caracterizan por tratar la resolución del sistema de interés de la forma más rigurosa posible. Es decir, no se recurre a ninguna información adicional de tipo empírico para la resolución de la ecuación de Schrödinger (2.1).

Entre estos métodos, destacan aquellos basados en la aplicación del principio variacional. Este principio establece que la energía de una función de onda aproximada siempre será superior a la energía exacta. Consecuentemente, la mejor función de onda se obtiene cuando la energía es mínima.

- Método Hartree-Fock (HF)

En la metodología Hartree-Fock (HF)³ se emplea un operador para el que se busca la función propia, pero cuya expresión depende de esta misma función propia. Por tanto, se lleva a cabo un procedimiento de tipo iterativo, en virtud del cual, partiendo de unas funciones prueba iniciales, se van generando sucesivamente operadores y funciones hasta llegar a la convergencia. En este momento se ha encontrado la solución HF y se dice que se ha llegado a la autoconsistencia (SCF, *self-consistent field*).

La función analítica que se utiliza en el método HF es un determinante de espín-orbitales monoelectrónicos. Dentro de la aproximación MO-LCAO (Orbitales Moleculares-Combinación Lineal de Orbitales Atómicos), estos espín-orbitales (denominados entonces orbitales atómicos, OA) quedan expresados a su vez como combinaciones lineales de funciones de base centradas en los núcleos atómicos. En la

práctica, lo que se optimiza no es la función analítica en sí, sino los diferentes coeficientes de las combinaciones lineales.

El tipo de funciones de base, así como su número, afectará de manera clave a la calidad del resultado que se obtenga. Una posibilidad es utilizar funciones con la forma de las soluciones de la ecuación de Schrödinger que se hallan para un átomo hidrogenoide, las denominadas funciones de Slater (STO). En la práctica, las funciones de base se construyen, por razones de eficiencia computacional, como combinaciones lineales de funciones gaussianas (GTO) que den una descripción correcta de la forma de las funciones STO.

De esta manera la función de onda puede ser descrita utilizando diferentes tipos de base. Así, si se asigna una sola función de base (como combinación lineal de funciones gaussianas) a cada orbital atómico, se tiene una base mínima. Las funciones denominadas doble Z (DZ) y triple Z (TZ) son aquéllas que incluyen dos o tres funciones de base por cada OA, respectivamente. En las bases denominadas *split-valence*, los orbitales correspondientes a los electrones internos quedan descritos con una sola función de base y los de valencia con dos o más. Habitualmente se incluyen funciones adicionales (las denominadas funciones de polarización y funciones difusas) para mejorar la descripción de la densidad electrónica en las regiones más alejadas de los núcleos.

Las bases de tipo *split-valence* de uso más extendido corresponden a la siguiente notación: K-M1G, desarrolladas por el grupo de Pople^{1,6,7}. En esta notación, K designa al número de funciones gaussianas contraídas para representar las funciones de base de los electrones de las capas internas, mientras que M indica el número de primitivas contraídas para generar las funciones de base de los electrones de la capa de valencia.

Este método puede aplicarse tanto en sistemas de capas cerradas (RHF) como en sistemas de capas abiertas (UHF).

Dentro del método HF, una aproximación importante es la asunción de que cada electrón interactúa con el resto de forma promediada. Es decir, no se incluye la correlación electrónica. Así pues, aunque en general la metodología HF da buenos resultados para algunas propiedades, aquéllas en que la correlación electrónica pueda tener cierta importancia hacen necesaria la aplicación de métodos más sofisticados. Debido a esto, se ha desarrollado metodología específica para la inclusión de la correlación electrónica.

- Métodos post-Hartree-Fock

Los métodos que incluyen correlación electrónica pertenecen a tres categorías: los de interacción de configuraciones (CI), los métodos perturbativos (MBPT, *many body perturbation theory*) y los de *coupled cluster* (CC). Muchos de ellos parten del formalismo HF, al que le añaden una corrección de la energía obtenida en el proceso SCF.

Los métodos post-HF más empleados son los derivados del tratamiento de Møller y Plesset⁸. Dicho método está basado en la teoría de perturbaciones de Rayleigh y Schrödinger⁹, según la cual el hamiltoniano de un sistema (H) se describe como la suma de dos términos: el hamiltoniano sin perturbar (H^0) y una perturbación (V) (ecuación 2.2).

$$H = H^0 + V \quad (2.2)$$

En el método Møller-Plesset, el hamiltoniano sin perturbar corresponde al utilizado en el método HF, mientras que dentro de la perturbación se incluyen los efectos de correlación electrónica. Dentro del desarrollo de la teoría perturbativa, se introduce un hamiltoniano generalizado (H_λ) (ecuación 2.3), donde λ es un parámetro adimensional que varía entre 0 y 1.

$$H_\lambda = H^0 + \lambda V \quad (2.3)$$

La función de onda (Ψ_λ) y la energía (E_λ) correspondientes al sistema descrito por el hamiltoniano H_λ pueden expresarse por una expansión en serie según las ecuaciones (2.4) y (2.5), respectivamente.

$$\Psi_\lambda = \Psi^{(0)} + \lambda\Psi^{(1)} + \lambda^2\Psi^{(2)} + \dots \quad (2.4)$$

$$E_\lambda = E^{(0)} + \lambda E^{(1)} + \lambda^2 E^{(2)} + \dots \quad (2.5)$$

Los diferentes términos $E^{(n)}$ y $\Psi^{(n)}$ corresponden a las correcciones de orden-n de la energía y la función de onda.

La energía HF es la suma de los términos $E^{(0)}$ y $E^{(1)}$, de manera que las correcciones para la correlación electrónica empiezan en el segundo orden (MP2). El grado de correlación que se considera da lugar a diferentes órdenes en que se valora la energía de correlación: MP2, MP3, MP4, etc.

2.1.1.2 Métodos semiempíricos

Los métodos semiempíricos^{10,11} fueron desarrollados fundamentalmente para reducir el coste computacional de los cálculos cuánticos y permitir su aplicación a sistemas químicos de mayor tamaño. Todos ellos parten de la teoría HF e introducen diversas simplificaciones en la expresión del hamiltoniano, que son compensadas por la inclusión de una serie de parámetros. Estos parámetros son ajustados para poder reproducir datos experimentales o cálculos cuánticos *ab initio*, de manera que muy a menudo incluyen de forma indirecta efectos de correlación electrónica. Entre las simplificaciones se encuentran el uso de una base mínima para los electrones de valencia, la definición del núcleo y los electrones internos en una única entidad denominada *core*, y la aplicación de la aproximación ZDO (*Zero-differential overlap*), si bien existen diferencias en cuanto al grado de aplicación de dicha aproximación.

De entre los métodos semiempíricos de última generación, el método semiempírico AM1 (*Austin model 1*)¹² es posiblemente el más popular. Se trata de un método basado en la aproximación NDDO (*Neglect of diatomic differential overlap*), que se desarrolló con la finalidad de solventar las principales limitaciones de anteriores métodos semiempíricos. Entre estas limitaciones, se encontraba la tendencia a sobreestimar las repulsiones entre átomos a distancias cercanas a las de van der Waals, y la pobre descripción de los puentes de hidrógeno.

La forma en que el método AM1 aborda estos problemas consiste en la introducción de un conjunto de funciones gaussianas adicionales a cada átomo. Estas funciones gaussianas permiten modular los efectos de las interacciones a largo alcance y mejorar la descripción de las interacciones no covalentes, sin que haya un aumento significativo del coste computacional. De esta manera, el método AM1 constituye una herramienta muy útil para un estudio preliminar de sistemas de tamaño relativamente grande, en particular para la descripción de geometrías moleculares y energías de sistemas estabilizados por puentes de hidrógeno.

2.1.1.3 Métodos basados en el funcional de la densidad (DFT)

La teoría del funcional de la densidad (DFT)^{13,14} es una alternativa a la metodología *ab initio* convencional para introducir los efectos de la correlación electrónica en la resolución de la ecuación de Schrödinger electrónica.

De acuerdo con la DFT, la energía del estado fundamental de un sistema polieletrónico puede expresarse a partir de su densidad electrónica (Teorema de Hohenberg y Konh¹⁵). De hecho el uso de la densidad electrónica en lugar de la función de onda para el cálculo de la energía constituye la base fundamental de la DFT. Por desgracia, la formula matemática exacta que relaciona la energía con la densidad electrónica no es conocida, y es necesario recurrir a expresiones aproximadas.

En la literatura se dispone de una gran cantidad de estudios que muestran diferentes desarrollos matemáticos de este método. De entre todas las formulaciones DFT, la más habitual es la autoconsistente DFT (SCF-DFT) de Kohn y Sham¹⁶, en la cual se resuelven unas ecuaciones muy parecidas a las de HF y se obtienen también unos orbitales, aunque éstos no tienen el mismo sentido físico que los orbitales HF.

Según esta formulación, se puede expresar la energía del sistema en función de la densidad electrónica, como suma de diferentes términos (ecuación 2.6).

$$E(\rho) = E_{KE}(\rho) + V(\rho) + E_{XC}(\rho) \quad (2.6)$$

donde $E(\rho)$ es la energía total, $E_{KE}(\rho)$ es la energía cinética, $V(\rho)$ es la interacción con el potencial externo y $E_{XC}(\rho)$ es la energía de correlación-intercambio.

Un aspecto básico dentro de la DFT es disponer de un buen potencial de correlación-intercambio. En general, $E_{XC}(\rho)$ se divide en una parte de intercambio (E_X) y una parte de correlación (E_C), como se muestra en la ecuación (2.7).

$$E_{XC}(\rho) = E_X(\rho) + E_C(\rho) \quad (2.7)$$

Ambas partes se pueden calcular con funcionales distintos que pueden o no incluir términos dependientes del gradiente de la densidad. Se han propuesto dos tipos de aproximaciones para calcular el funcional de correlación-intercambio: la aproximación de la densidad local (LDA)¹⁷ y la aproximación de la densidad no local (GGA)¹⁸. El acoplamiento de un funcional de intercambio con un funcional de correlación da nombre a cada método DFT. Así el método B3LYP usa el funcional híbrido de Becke¹⁹ de 3 parámetros para el término de intercambio y el funcional de Lee-Yang-Parr²⁰ para el término de correlación.

El principal inconveniente de estos métodos se encuentra en el hecho de que para sistemas en que la DFT da resultados erróneos no hay forma sistemática de mejorarlos. Aún así, en los últimos años las aplicaciones DFT han crecido espectacularmente, ya

que para muchos sistemas proporciona buenos resultados (calidad comparable a la que ofrece el método MP2) a un coste reducido. Actualmente es ampliamente usado para el estudio de los sistemas de mediano y gran tamaño.

2.1.1.4 *Estudio de propiedades moleculares*

La mecánica cuántica puede ser usada para calcular un amplio rango de propiedades estructurales y termodinámicas. La mecánica cuántica permite también derivar todas aquellas propiedades que dependen de la distribución electrónica, como los momentos dipolares y cuadrupolares, los orbitales moleculares frontera (HOMO y LUMO), la densidad de carga, el potencial electrostático, etc. Estas propiedades a menudo no pueden ser determinadas por ningún otro método teórico ni experimental.

i) Propiedades estructurales y termodinámicas

Entre las diferentes aplicaciones de los métodos cuánticos destaca la localización de puntos singulares (mínimos y estados de transición) en la superficie de energía potencial lo que permite determinar estructuras de reactivos, productos, intermediarios y estados de transición. Para obtener estas geometrías se han desarrollado diferentes algoritmos, que difieren según si lo que se busca es un mínimo o un estado de transición.

Para la obtención de mínimos, en general se emplean métodos en los cuales se evalúa el gradiente y el hessiano de la energía del sistema. Según la expresión utilizada para el hessiano y el criterio de optimización de la dirección de búsqueda, se tienen diferentes algoritmos. Los resultados de estos cálculos de optimización pueden ser comparados con estructuras obtenidas de experimentos en fase gas. La concordancia entre la teoría y los experimentos, para cálculos *ab initio*, es normalmente muy buena y generalmente mejora cuando se aumenta el tamaño de la base y se incluye correlación electrónica.

Cada vez que se localiza un punto singular se tiene que verificar su condición como tal, mediante un análisis de frecuencias²¹. Dicho análisis permite, asimismo, estimar correcciones térmicas y entrópicas de la energía del sistema, y derivar magnitudes termodinámicas. Dichas correcciones térmicas corresponden a la energía vibracional del punto cero (*ZPE*), debida al valor no nulo de la energía vibracional a una temperatura de 0 K, y a las correcciones térmicas (ΔE^T) fruto de que la temperatura (*T*) del sistema químico es diferente de 0 K.

La termodinámica en vacío²¹ se calcula incluyendo la energía del punto cero (*ZPE*) y las correcciones térmicas (ΔE^T) a la energía del sistema que incluye la contribución electrónica (E_{ele}) y nuclear (E_N), obteniendo la entalpía (*H*) (ecuación 2.9). Posteriormente, la energía libre (*G*) se estima con la adición de la contribución entrópica (ecuación 2.10).

$$E_{\text{sist}} = E_{\text{ele}} + E_N \quad (2.8)$$

$$H = E_{\text{sist}} + ZPE + \Delta E^T \quad (2.9)$$

$$G = H - TS \quad (2.10)$$

ii) Análisis de población

Los métodos de análisis de población realizan una partición de la densidad electrónica entre los núcleos, es decir, cada núcleo tiene un número de electrones asociado a él. Cabe destacar que no existe ningún operador cuántico para la carga atómica, por lo tanto cualquier partición debe ser arbitraria y como tal debe ser interpretada. Si bien el método de Mulliken²² es ampliamente utilizado, el tratamiento más riguroso para evaluar la carga atómica es el análisis de población basado en la teoría de Bader²³.

La aproximación de Bader es sin duda una de las más cercanas a la intuición química, y está considerada como la más potente para la descripción de la topología

electrónica de las moléculas. Esta aproximación está basada en el concepto de *gradient vector path*, definido a partir del gradiente de la densidad electrónica. Un conjunto de *gradient paths* se muestran en la figura 2.1a para la formamida. Como puede verse, algunos de los *gradient paths* terminan en el núcleo atómico. Otros son atraídos a puntos (denominados puntos críticos) que no están localizados en el núcleo, siendo de particular relevancia los puntos críticos de enlace (*bond critical points*) que se localizan entre átomos unidos. Existen otros tipos de puntos críticos, por ejemplo, los puntos críticos de anillo (*ring critical point*), que se encuentran en el centro de un anillo, o los puntos críticos de celda (*cage critical point*), que se encuentran en el interior de una molécula con superficies aromáticas.

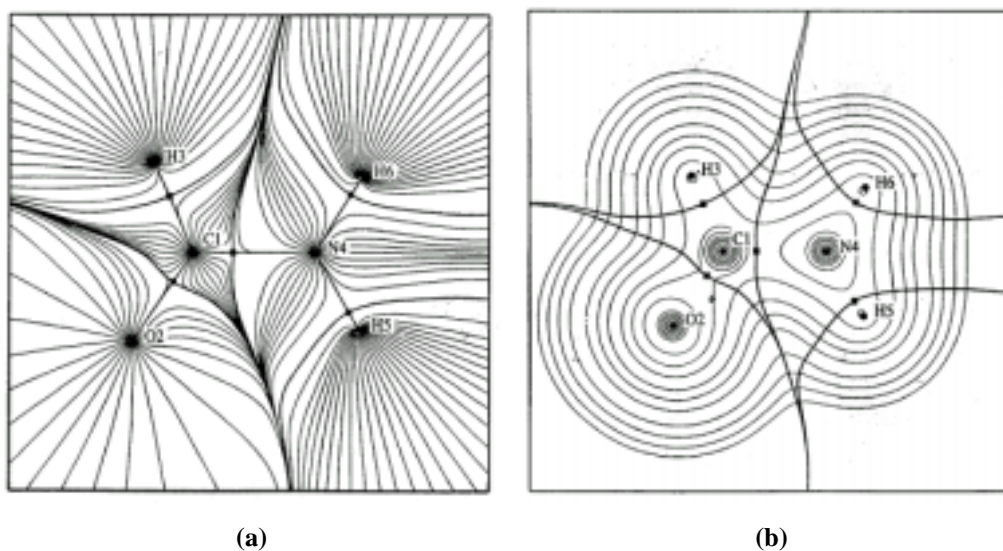


Figura 2.1 (a) *Gradient vector paths* alrededor de la formamida. Las líneas terminan en los átomos o en los puntos críticos de enlace (indicados con cuadrados); (b) Partición de la densidad electrónica en la formamida. (Fuente: véase referencia 1, p. 117-118)

Los puntos críticos de enlace son puntos de mínima densidad de carga electrónica a lo largo del *gradient path* que une dos átomos. Dichos puntos permiten fraccionar la densidad electrónica, tal como se muestra en la figura 2.1b para la formamida. Este procedimiento puede llevarse a cabo en cada enlace, resultando una partición

tridimensional de la densidad electrónica. La población electrónica, que se asigna a cada átomo, se calcula integrando numéricamente la densidad de carga dentro de la región que rodea a cada átomo.

Las cargas que se obtienen con esta teoría son relativamente invariantes a las bases empleadas. Además de cargas, la teoría de Bader^{24,25} también permite examinar en detalle las propiedades topológicas de la densidad de carga electrónica. Esta información puede utilizarse para caracterizar la formación de interacciones débiles, como los puentes de hidrógeno.

iii) Potencial electrostático molecular (MEP)

El potencial electrostático (ESP) en un punto \mathbf{r} , $V(\mathbf{r})$, se define como el trabajo realizado para traer la unidad de carga positiva desde el infinito hasta este punto. La energía electrostática de interacción entre una carga q situada en \mathbf{r} y la molécula es igual a $q \cdot V(\mathbf{r})$, y se conoce con el nombre de potencial electrostático molecular (MEP)²⁶⁻²⁸.

El MEP contiene toda la información sobre las distribuciones de carga nuclear y electrónica de la molécula (figura 2.2). Esta magnitud tiene gran utilidad para dar una idea cualitativa de la reactividad del compuesto. También tiene aplicación en la parametrización de campos de fuerzas y para la definición del operador de perturbación en métodos SCRF (*Self-consistent reaction field*).

El MEP puede ser definido como el componente electrostático de la energía de interacción entre una molécula y una carga positiva, o también como el valor esperado de operador r^{-1} . Su cálculo en una posición \mathbf{r}_1 , dentro del formalismo MO-LCAO, puede realizarse a partir de la ecuación (2.11),

$$\text{MEP} = V(\mathbf{r}_1) = \sum_A \frac{Z_A}{|\mathbf{r}_1 - \mathbf{R}_A|} - \sum_{\mu} \sum_{\nu} P_{\mu\nu} \int \frac{\chi_{\mu}(\mathbf{r}) \chi_{\nu}(\mathbf{r})}{|\mathbf{r}_1 - \mathbf{r}|} d\mathbf{r} \quad (2.11)$$

donde Z_A se refiere a la carga nuclear del átomo A, \mathbf{R}_A es su posición, χ_μ y χ_ν son las funciones de base de los orbitales atómicos μ y ν , y $P_{\mu\nu}$ representa al elemento $\mu\nu$ de la matriz de densidad de primer orden.

Este cálculo puede realizarse con una función de onda *ab initio*, semiempírica o derivada de un cálculo DFT. En sistemas bioquímicos se suele emplear la función de onda HF/6-31G(d) que proporciona un balance adecuado de calidad y coste computacional.

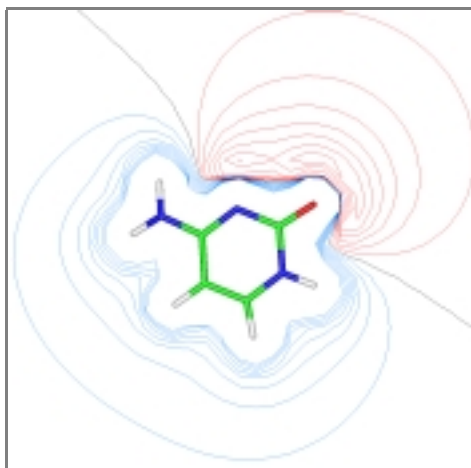


Figura 2.2 Mapa de potencial electrostático en el plano molecular de la citosina. El contorno negativo es el rojo (interacción favorable con un protón), el contorno positivo es el azul (interacción desfavorable) y el contorno cero es la línea negra. Las líneas de isocontorno van de ± 80 kcal/mol y se muestra cada 10 kcal/mol. Se observan mínimos cercanos al N3 y al O2.

iv) Cargas atómicas

Las cargas atómicas son utilizadas para determinar las interacciones electrostáticas dentro de los métodos clásicos. Además, también han sido consideradas como índices de reactividad química. Se ha constatado que las cargas obtenidas mediante un análisis de población de Mulliken²² dan una pobre representación de las características

electrostáticas del sistema. Mejores representaciones se obtienen con cargas derivadas del método de Bader, pero sin duda las cargas más útiles para determinar la energía electrostática de interacción son las cargas derivadas del potencial electrostático molecular (ESP).

La estrategia ESP^{29,30} aplicada se basa en ajustar el MEP calculado por mecánica cuántica (ecuación 2.11) con el obtenido a partir de una distribución de cargas puntuales (ecuación 2.12) en una nube de puntos (\mathbf{r}_1) alrededor de la molécula de interés.

$$V(\mathbf{r}_1) = \sum_i \frac{q_i}{|\mathbf{r}_1 - \mathbf{R}_i|} \quad (2.12)$$

donde q_i y \mathbf{R}_i son la carga atómica y la posición del núcleo i respectivamente.

Se ha comprobado que este método tiene una gran eficacia y que es capaz de reproducir de forma aceptable mapas de potencial electrostático, momentos dipolares experimentales e interacciones electrostáticas entre moléculas²⁹⁻³³. Recientemente se han aplicado variantes de esta técnica que intenta mejorar la representación de carga de los átomos internos.

2.1.1.5 Consideraciones prácticas

Extensión de la base^{1,6,7}

El principio variacional nos asegura que para cálculos no perturbativos un conjunto más extenso de funciones de base rendirá una energía más baja y por lo tanto más cercana al valor real. El uso de una base más extendida no sólo repercute en la energía, sino que también permite definir una función de onda más flexible y, por tanto, una mejor descripción de la molécula.

Desafortunadamente, el aumento de la base implica un aumento del coste computacional. Puede que sea necesario asumir dicho aumento en el cálculo de alguna de las propiedades, pero no en otras, como por ejemplo la geometría de equilibrio, que se ha visto que queda bien representada a un nivel medio de teoría. Por ello es una práctica común realizar optimizaciones de geometría a niveles medios de teoría (como HF/6-31G(d) o B3LYP/6-31G(d)) para realizar después cálculos puntuales de energía a un nivel superior con una base más extendida.

Error de superposición de base (BSSE)^{1,34}

Este error siempre se pone de manifiesto en el cálculo de interacciones intermoleculares con bases finitas. Se podría esperar que la energía de interacción corresponda con la diferencia de la energía de la supermolécula “dímero” menos la energía de los monómeros a distancia infinita. Sin embargo, la energía calculada de esta manera sobreestimaré el valor real debido a que en el dímero, las funciones de base de un monómero son utilizadas por el monómero contrario para describir mejor su distribución electrónica, produciendo un efecto de pseudoextensión de base en el dímero.

Esta claro que cuanto más pobre sea la base más importante será este efecto, debido a que la distribución electrónica del monómero en cuestión buscará mayor flexibilidad y empleará las funciones de base del otro monómero³⁵. También será importante cuando en la función de onda se introduzcan efectos de correlación electrónica dado que los electrones tenderán a relajar su repulsión alojando densidad de carga en los orbitales atómicos del monómero adyacente.

La forma más usual de corregir este efecto es utilizar el método *counterpoise* de Boys y Bernardi³⁶, desarrollado para dimerizaciones atómicas o de monómeros rígidos. Éste se basa en calcular la energía de un monómero en presencia de las funciones de base de todo el dímero y luego restarle la energía del monómero calculada sólo con sus funciones de base. Luego se repite este cálculo para el segundo monómero. Esto permite cuantificar la estabilización artificial en el dímero. La inclusión de los efectos de

distorsión de geometría no es obvia, ya que es necesario realizar hasta 7 cálculos (3 optimizaciones de geometría y 4 cálculos puntuales de energía) para la evaluación de la energía de dimerización.

Para un dímero A-B la energía de interacción no corregida por el *BSSE* viene dada por la ecuación (2.13). En las energías el subíndice indica geometría, el superíndice las funciones de base usadas y entre paréntesis el sistema objeto del cálculo.

$$E_{\text{int}} = E_{\text{AB}}^{\text{AB}}(\text{AB}) - E_{\text{A}}^{\text{A}}(\text{A}) - E_{\text{B}}^{\text{B}}(\text{B}) \quad (2.13)$$

El error de superposición de base viene dado por la ecuación (2.14).

$$BSSE = E_{\text{AB}}^{\text{AB}}(\text{A}) - E_{\text{AB}}^{\text{A}}(\text{A}) + E_{\text{AB}}^{\text{AB}}(\text{B}) - E_{\text{AB}}^{\text{B}}(\text{B}) \quad (2.14)$$

Finalmente, la energía corregida por el *BSSE* viene dada por la ecuación (2.15).

$$E_{\text{int}}^{\text{CP}} = E_{\text{int}} - BSSE \quad (2.15)$$

2.1.2 Métodos clásicos (MM)^{1,2,37,38}

Los métodos clásicos son una alternativa a las metodologías cuánticas para abordar el estudio de sistemas químicos de gran tamaño en los que no se producen alteraciones en enlaces covalentes, ni cambios drásticos en la distribución electrónica.

Dentro de las técnicas clásicas, la energía de un sistema químico se expresa como una función únicamente de las posiciones de los núcleos y se evalúa a través de las expresiones analíticas clásicas y de parámetros ajustados que incorporan de forma promediada la información electrónica. El conjunto de expresiones y parámetros utilizados constituye lo que se denomina campo de fuerzas (FF, *force-field*).

2.1.2.1 Cálculo de la energía en un campo de fuerzas

Los diversos campos de fuerzas pueden utilizar diferentes expresiones para evaluar la energía de un sistema de una manera más o menos rigurosa. En general, la energía se puede expresar básicamente de la forma que muestra la siguiente ecuación (2.16),

$$E_{\text{total}} = E_{\text{términos enlazantes}} + E_{\text{términos no enlazantes}} \quad (2.16)$$

donde el primer término hace referencia a la interacción entre pares de átomos enlazados entre sí, y el segundo incluye la interacción entre átomos no directamente enlazados.

- Términos enlazantes

- i) Términos de *stretching* y *bending*

Estos términos ($E_{\text{stretching}}$ y E_{bending}) están relacionados con el coste energético que conlleva la distorsión de las distancias y ángulos de enlace respecto a unos valores de equilibrio. Generalmente se calculan considerando un comportamiento de tipo armónico (ecuaciones 2.17 y 2.18),

$$E_{\text{stretching}} = K_s (d - d_0)^2 \quad (2.17)$$

$$E_{\text{bending}} = K_b (\theta - \theta_0)^2 \quad (2.18)$$

donde d y θ son las distancias y ángulos de enlace, d_0 y θ_0 los valores de equilibrio y K_s y K_b las constantes de fuerza. La aproximación armónica sólo resulta válida para distorsiones pequeñas respecto a los valores de equilibrio.

Existen campos de fuerzas que además de considerar estos dos términos añaden términos cruzados que incluyen los efectos asociados al acoplamiento de las

distorsiones de las distancias y ángulos de enlace. Otros emplean expresiones polinomiales más complejas, o incluso términos exponenciales.

ii) Términos de torsión

Estos términos ($E_{torsion}$) describen las barreras de rotación respecto a los enlaces químicos. Estas rotaciones presentan una periodicidad en el valor de la energía con diferentes máximos y mínimos, que los campos de fuerza representan mediante series de Fourier (ecuación 2.19). El número de términos de la serie dependerá del tipo de torsión a considerar, y en general no es superior a tres.

$$E_{torsion} = \sum_i \frac{V_i}{2} (1 + \cos(n\chi - \gamma_i)) \quad (2.19)$$

En esta ecuación, V_i representa la constante de fuerza de cada término de la función, n la periodicidad del término, χ el ángulo diedro y γ_i el ángulo de fase.

Un caso particular de torsión es aquélla que está asociada al movimiento de un átomo respecto de un plano. Este tipo de movimiento recibe el nombre de torsión impropia y está asociado a los fenómenos de conjugación de enlaces. Dentro de los campos de fuerzas, la energía se calcula mediante diferentes tipos de expresiones, desde unas similares a las de *bending* hasta otras análogas a las empleadas para una torsión propia de periodicidad dos.

- Términos no enlazantes

i) Término electrostático

Este término representa las interacciones entre las distribuciones de carga de las moléculas. La evaluación de esta contribución puede realizarse mediante diferentes tipos de expresiones. La estrategia más habitual consiste en asignar cargas puntuales a

los diferentes átomos del sistema y calcular las energías de interacción (E_{ele}) con la fórmula de Coulomb (ecuación 2.20),

$$E_{ele} = \sum_{i,j>i} \frac{q_i q_j}{\epsilon r_{ij}} \quad (2.20)$$

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

ii) Término de van der Waals

Esta contribución (E_{vdw}) incluye diversas interacciones de naturaleza no electrostática entre átomos no directamente enlazados. Entre estas interacciones se encuentran la energía de dispersión y la repulsión entre núcleos.

Para describir esta interacción existen diferentes tipos de expresiones, la forma general de todas ellas se reduce a la suma de dos fuerzas: una repulsiva para las distancias cortas entre los átomos, y otra atractiva para distancias mayores. Las dos tienden asintóticamente a cero a medida que aumenta la distancia.

Entre estas expresiones, una muy utilizada es la denominada 12-6 (o potenciales de Lennard-Jones) (ecuación 2.21),

$$E_{vdw} = \sum_{i,j>i} \left(\frac{A_{ij}}{r_{ij}^{12}} \right) - \left(\frac{B_{ij}}{r_{ij}^6} \right) \quad (2.21)$$

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

Una expresión alternativa muy empleada es la denominada exponencial-6, indicada en la ecuación (2.22),

$$E_{\text{vdW}} = \sum \left(\frac{D_{ij}}{\exp(-Cr_{ij})} \right) \left(\frac{B_{ij}}{r_{ij}^6} \right) \quad (2.22)$$

donde D_{ij} y C vuelven a ser constantes que dependen de los parámetros de van der Waals de cada átomo.

2.1.2.2 Parametrización

La determinación de los parámetros del campo de fuerzas es la etapa clave en el desarrollo de un *force-field*. Los campos de fuerza actuales han sido cuidadosamente parametrizados usando datos experimentales y cálculos cuánticos en sistemas modelo utilizados como referencia. Ello garantiza la calidad de los cálculos clásicos posteriores y su capacidad de reproducir los valores experimentales. Existen muchos campos de fuerzas desarrollados con marcos de aplicación diferentes³⁹⁻⁴⁴, es decir, que han sido diseñados para tratar sistemas moleculares diferentes. Se pueden encontrar campos de fuerza para estudiar polipéptidos y proteínas, otros para estudiar ácidos nucleicos, otros para estudiar compuestos orgánicos de pequeño tamaño, etc.

Frecuentemente, en el diseño de nuevas moléculas se estudian estructuras con grupos funcionales no contemplados en la base de datos *force-field*. En estos casos es necesario realizar una parametrización *in situ* que, ante la inexistencia en general de datos experimentales para dicha molécula, debe realizarse a partir de cálculos mecanocuánticos.

2.1.2.3 Métodos basados en campos de fuerzas

El cálculo de la energía de un sistema a partir de las expresiones de la mecánica clásica permite aplicar diferentes tipos de técnicas teóricas. Las principales son: la

mecánica molecular (MM), la dinámica molecular (MD), los métodos de Monte Carlo (MC) y la mecánica estadística (SM).

- Técnicas de mecánica molecular (MM): minimización de la energía⁴⁵

La minimización de la energía es la técnica clásica más sencilla. Básicamente, este método trata de encontrar una o más configuraciones estables de una molécula, a partir de una optimización de la energía del sistema (figura 2.3).

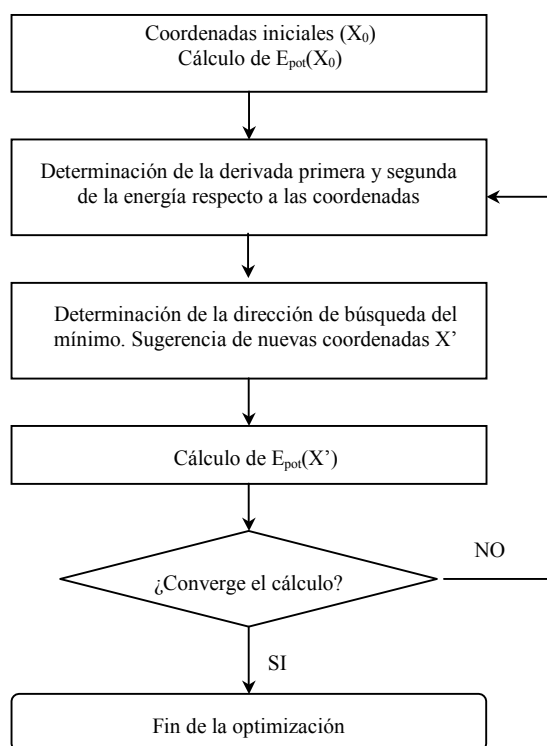


Figura 2.3 Algoritmo básico de los métodos de mecánica molecular.

La minimización suele utilizarse con tres finalidades principales: (i) la obtención de la geometría más estable de un sistema; (ii) el cálculo de la energía relativa entre las estructuras de equilibrio de diferentes especies; y (iii) la obtención de una estructura de partida que no presente interacciones fuertemente desfavorables para una simulación de dinámica molecular.

Esta metodología presenta diferentes limitaciones. Una de ellas es que proporciona una visión estática del sistema, mientras que lo que se observa experimentalmente es en realidad el resultado promedio de toda una serie de estructuras dinámicas. Adicionalmente, cuando el sistema presenta un tamaño considerable no resulta viable considerar el efecto de la temperatura en la conformación del mismo. Tampoco se introducen adecuadamente los términos entrópicos del sistema. Una importante limitación proviene de la aparición de múltiples mínimos locales alejados del mínimo absoluto, fruto del elevado número de grados de libertad que suelen tener las moléculas consideradas que puede llevar a considerar como geometrías óptimas conformaciones muy alejadas del mínimo absoluto.

- Técnicas de dinámica molecular (MD)⁴⁵⁻⁴⁹

La dinámica molecular proporciona una descripción del comportamiento del sistema a lo largo del tiempo aplicando las leyes de mecánica clásica. La evolución temporal se simula mediante la integración de las ecuaciones de Newton para cada átomo (ecuación 2.23).

$$\frac{\partial^2 \mathbf{x}_i}{\partial^2 t} = \frac{1}{m_i} \mathbf{F}_i \quad (2.23)$$

donde \mathbf{x}_i corresponde a las coordenadas de cada átomo, m_i a su masa, \mathbf{F}_i a las fuerzas que actúan sobre el mismo y t es el tiempo.

La fuerza \mathbf{F}_i que actúa sobre cada partícula del sistema en un instante de tiempo t se determina a partir de la primera derivada de la energía potencial. De acuerdo con la ecuación 2.23 dicha fuerza da directamente la aceleración de la partícula. A partir de ésta, integrando las ecuaciones del movimiento, se puede obtener la velocidad cuya integral a su vez proporciona el cambio de posición del átomo. Con las nuevas posiciones y velocidades de cada partícula, se obtienen las energías potencial y cinética del sistema. Aplicando sucesivamente este procedimiento se obtiene lo que se denomina

una trayectoria, el conjunto de posiciones y velocidades de cada partícula a lo largo del tiempo. El esquema de esta metodología se muestra en la figura 2.4.

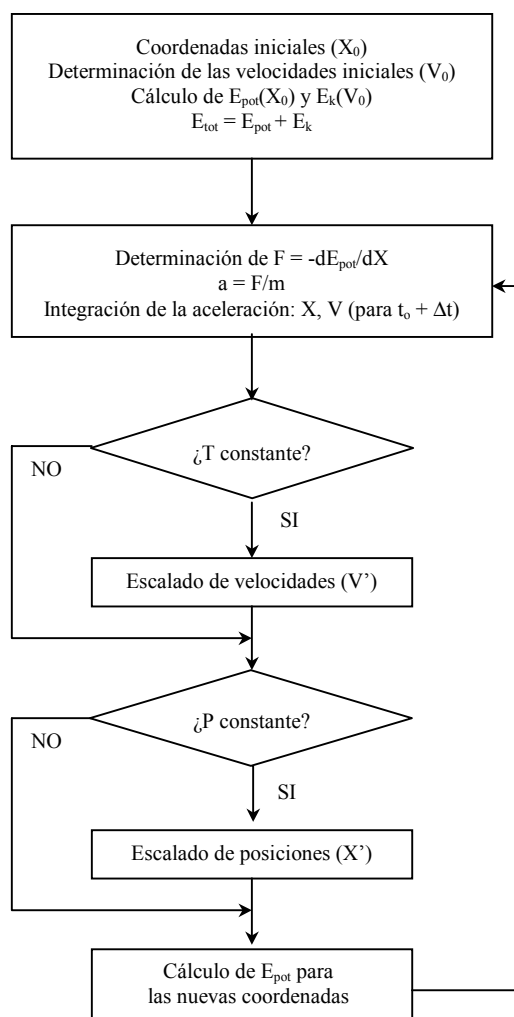


Figura 2.4 Algoritmo básico de los métodos de dinámica molecular.

La integración de las ecuaciones se realiza mediante métodos numéricos en los cuales se recurre a un paso de integración lo suficientemente pequeño para poder simular los movimientos de más alta frecuencia del sistema, que suelen ser las vibraciones de los enlaces. Esto fuerza a emplear pasos de integración en el orden del femtosegundo. Ello da lugar a uno de los principales inconvenientes de esta técnica, ya

que la mayoría de procesos tienen lugar en escalas de tiempo del orden del nanosegundo o superiores.

La dinámica molecular puede ser aplicada a sistemas en vacío o dentro de un determinado entorno. Además, es posible fijar las condiciones de la simulación, como la energía total del sistema (E), el número de partículas (N), el volumen (V), la temperatura (T) o la presión (P). Según que condiciones se fijen, se tienen diferentes formalismos, como el microcanónico (NVE), el isobárico-isotérmico (NPT) y el canónico (NVT).

Generalmente, la estructura de partida que se utiliza para iniciar una dinámica molecular proviene de una minimización de energía. La velocidad de cada partícula se asigna de forma aleatoria considerando una distribución de Maxwell-Boltzmann. Los primeros pasos de la simulación constituyen lo que se denomina periodo de equilibrado, durante el cual las propiedades del sistema no se mantienen constantes. Las energías, coordenadas y velocidades generadas durante el equilibrado no son utilizadas en el análisis posterior del sistema.

La duración del equilibrado es variable y depende del sistema. Generalmente para sistemas bioquímicos es del orden de 100 picosegundos o algo más. El momento a partir del cual finaliza el equilibrado puede establecerse a partir de la evolución temporal de una determinada variable (por ejemplo, la energía potencial), considerándose que ha acabado cuando el promedio de la variable no cambia. Tras el equilibrado, diversas propiedades se mantienen constantes, a no ser que tenga lugar un cambio conformacional importante.

La visión dinámica del sistema que proporciona esta metodología hace que pueda ser utilizada para describir procesos que tienen lugar en escalas de tiempo asequibles por los recursos computacionales disponibles, y también para realizar exploraciones del espacio conformacional en sistemas moleculares complejos. Debido a ello, constituye uno de los métodos más utilizados en las simulaciones de sistemas poliméricos, como los ácidos nucleicos.

- Técnicas de Monte Carlo (MC)⁵⁰

Este tipo de técnicas tienen como objetivo primordial, al igual que la dinámica molecular (MD), obtener una visión promediada del sistema. La principal diferencia entre la MD y MC es que el método Monte Carlo no incluye el tiempo como variable.

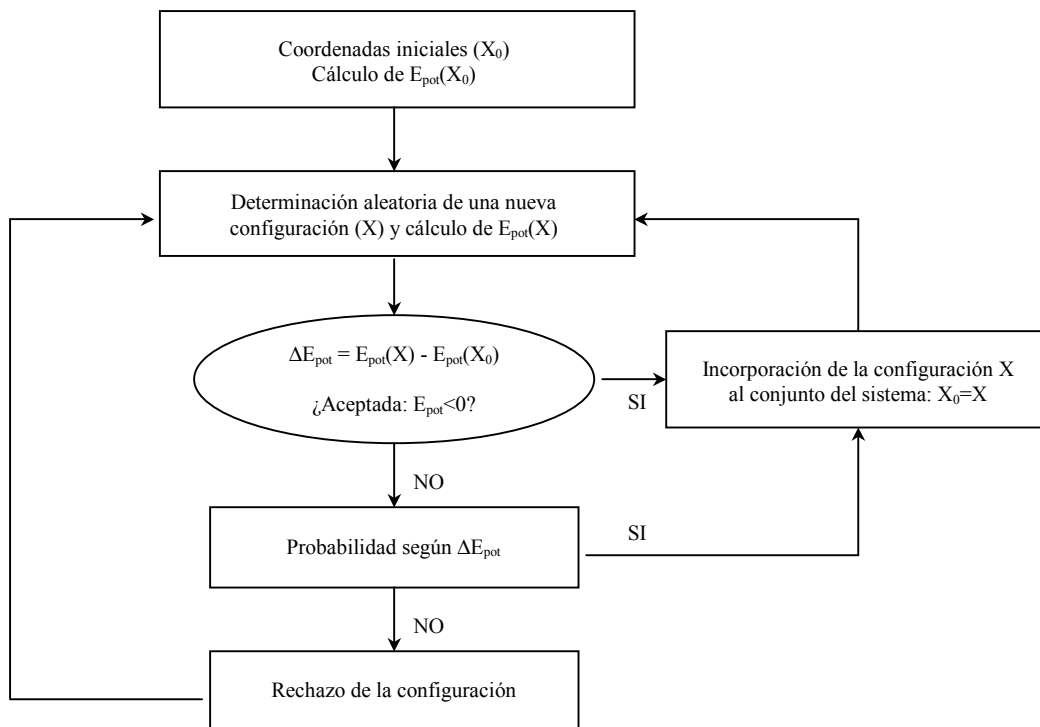


Figura 2.5 Algoritmo básico de los métodos de Monte Carlo.

La filosofía de las técnicas Monte Carlo está ligada a los formalismos de la mecánica estadística. De acuerdo con ellos, un sistema debe cumplir la ley de Boltzmann, por la cual las configuraciones más estables estarán ocupadas por un mayor número de moléculas. Los métodos de Monte Carlo más usuales usan el algoritmo de Metrópolis (figura 2.5). De acuerdo con éste se parte de una configuración inicial, a partir de la cual se genera otra al azar, cuya energía es evaluada con un *force-field*. Si ésta es menor que la energía de la configuración original, entonces es aceptada, pero en caso contrario aún tendrá una probabilidad de ser aceptada, tanto menor cuanto mayor

sea la diferencia energética con respecto a la configuración inicial. Después de varios millones de configuraciones generadas de esta manera se consigue un muestreo del sistema similar al obtenido por dinámica molecular para sistemas ergódicos.

Los métodos Monte Carlo son especialmente útiles para analizar sistemas pequeños y de rigidez elevada en fases condensadas, mientras que los sistemas moleculares más grandes son estudiados preferiblemente por dinámica molecular (MD).

- Técnicas de mecánica estadística (SM)⁵¹

Las técnicas de mecánica estadística permiten procesar la información de un sistema obtenida por cálculos de dinámica molecular o Monte Carlo, a fin de obtener información sobre magnitudes termodinámicas. Un tipo de métodos de especial utilidad son los cálculos basados en las diferencias de energía libre entre diferentes estados de un sistema. Estos métodos, entre los que cabe citar la perturbación de energía libre (FEP) y la integración termodinámica (TI), permiten estimar con sorprendente precisión cambios de energía libre en procesos bioquímicos. Paralelamente, el uso de ciclos termodinámicos junto con estas técnicas incrementa enormemente su aplicación.

Como ya se han mencionado, hay diversas técnicas que evalúan la diferencia de energía libre (ΔG) entre dos estados “0” y “1” (ecuación 2.24), todas ellas basadas en la mecánica estadística.

$$\Delta G = G_1 - G_0 \quad (2.24)$$

Cada uno de estos estados “0” y “1” se encuentra representado por una función de energía potencial (V_i). La diferencia de energía libre es calculada en una serie de pasos incrementales que conectan los estados físicos “0” y “1” a través de una serie de intermedios no necesariamente físicos. El carácter del sistema en cada uno de estos pasos intermedios se relaciona con un parámetro λ . Este parámetro λ controla el cambio entre las funciones de energía potencial de los 2 estados “0” y “1” (ecuación 2.25).

$$V_\lambda = \lambda V_1 + (1-\lambda) V_0, \quad \lambda \in [0,1] \quad (2.25)$$

donde λ puede variar de 0 ($V_\lambda=V_0$) a 1 ($V_\lambda=V_1$) y V_λ es la función de energía potencial efectiva.

Perturbación de la energía libre (FEP)⁵²

En el método FEP, la energía libre es calculada por intervalos discretos, empleando las siguientes ecuaciones 2.26 y 2.27:

$$G_{\lambda(i+1)} - G_{\lambda(i)} = -RT \ln \left\langle \exp \left[\frac{-(V_{\lambda(i+1)} - V_{\lambda(i)})}{RT} \right] \right\rangle_{\lambda(i)} \quad (2.26)$$

$$\Delta G = G_1 - G_0 = \sum_i G_{\lambda(i+1)} - G_{\lambda(i)} \quad (2.27)$$

donde G_0 y G_1 son las energías libres de los estados “0” y “1” respectivamente, $V_{\lambda(i)}$ es la función de energía potencial representativa de cada estado $\lambda(i)$ y $\langle \rangle_{\lambda(i)}$ es el promedio del conjunto de valores de energía obtenidos de la trayectoria de MD representativo de cada estado $\lambda(i)$.

Integración termodinámica (TI)⁵²

En el método TI, en lugar de las ecuaciones (2.26) y (2.27) se emplea la fórmula (2.28) para calcular la diferencia de energía libre.

$$G_1 - G_0 = \int_0^1 \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_\lambda d\lambda \quad (2.28)$$

La aplicación de esta ecuación (2.28) requiere evaluar el promedio del conjunto de valores de la derivada de la función de energía potencial con respecto a λ , $\langle \partial V / \partial \lambda \rangle_\lambda$ a

varios valores de λ . En la práctica, la integral es aproximada por una sumatoria sobre los intervalos finitos de λ .

Estas metodologías son susceptibles a cometer errores cuando la simulación no es suficientemente larga para obtener valores convergidos de las cantidades promedio que van a servir para calcular la diferencia de energía libre o cuando los dos estados entre los que se une el sistema son demasiado disimilares.

2.1.3 Métodos mixtos (QM/MM)^{53,54}

La mecánica molecular no está adaptada al estudio de las reacciones químicas, ya que la creación o ruptura de los enlaces químicos es difícil de reproducir por un potencial efectivo. Por otro lado, la mecánica cuántica no puede estudiar en la práctica en su totalidad sistemas de cierto tamaño, como proteínas, ácidos nucleicos o sistemas en solución. Afortunadamente, los efectos cuánticos (formación/ruptura de enlaces, transferencia de carga, polarización, etc.) son habitualmente locales y es posible estudiar el problema de reactividad en entornos complejos mediante una aproximación mixta que combina la mecánica cuántica (QM) y la mecánica molecular (MM).

La idea central de estos métodos mixtos, denominados QM/MM, es dividir el sistema en dos partes: una cuántica y una clásica. Una se modela a partir de métodos cuánticos, mientras que el resto del sistema se simula por técnicas clásicas. En consecuencia, la energía total del sistema queda dividida en tres contribuciones: la correspondiente al subsistema cuántico, la correspondiente al subsistema clásico y la de interacción entre ambos subsistemas cuántico-clásico.

El hamiltoniano efectivo del sistema se describe como:

$$H_{\text{ef}} = H_{\text{MM}} + H_{\text{QM}} + H_{\text{QM/MM}} \quad (2.29)$$

donde H_{QM} y H_{MM} son los hamiltonianos del subsistema QM y MM aislados, y $H_{QM/MM}$ es el hamiltoniano de interacción entre los subsistemas QM y MM.

Debido al elevado costo de los cálculos cuánticos, la mayoría de los métodos QM/MM usan hamiltonianos semiempíricos para describir la parte QM. Sin embargo, también se han usados métodos basados en el funcional de la densidad y métodos *ab initio* HF con bases pequeñas^{55,56}. Todos estos métodos consideran que los cambios que se producen por el efecto del sistema clásico sobre el cuántico resultan únicamente del acoplamiento electrostático, despreciando los cambios debidos a la dispersión.

Los cálculos electrónicos se efectúan solamente sobre la parte QM teniendo en cuenta las interacciones con la parte MM. El subsistema MM se trata con la ayuda de potenciales de la mecánica clásica, mediante un conjunto de cargas y parámetros de van der Waals.

Para una configuración nuclear determinada la energía del sistema se obtiene resolviendo la correspondiente pseudo-ecuación de Schrödinger (2.30), donde el hamiltoniano efectivo incluye el hamiltoniano QM del soluto y el término de acoplamiento QM/MM. En este cálculo la distribución de carga de la parte QM se adapta a la distribución de cargas discretas que se distribuyen a su alrededor que actúan como un elemento perturbador de la distribución electrónica del sistema QM.

$$H_{ef} \Psi = E \Psi \quad (2.30)$$

Utilizando este formalismo, no sólo se puede calcular la energía, sino que es perfectamente posible calcular las fuerzas sobre los átomos cuánticos y clásicos. Debe remarcarse que en el formalismo QM/MM cada movimiento nuclear requiere un cálculo SCF (resolución de la ecuación 2.30) y esto supone un esfuerzo computacional muy elevado.

Existen diferentes variantes de técnica mixtas que parten de formalismos diferentes y que poseen marcos de aplicación también distintos.

- Método de generalización del potencial electrostático molecular (MEP)⁵⁷⁻⁶⁰

Estos métodos explotan la descripción de la interacción electrostática obtenida a partir del MEP. La idea principal es extender el MEP de manera que permita no sólo considerar la interacción con un protón, sino con cualquier partícula clásica. Para ello emplean un formalismo denominado GMIP (*Generalized molecular interaction potential*) que incluye un término electrostático y un término de van der Waals clásico parametrizado a partir de cálculos SCF para representar las interacciones no electrostáticas.

$$\text{GMIP} = E_{\text{ele}} + E_{\text{vdW}} \quad (2.31)$$

En nuestro grupo⁶⁰, se ha desarrollado una estrategia que permite introducir la polarización en la parte QM por la parte MM mediante el uso de teoría de perturbaciones. Esto evita la resolución de la ecuación (2.30) para cada configuración de la parte MM, pero sin renunciar a la introducción del efecto de la polarización. Según este método, la función de onda se halla para la parte QM aislada y luego se usa para evaluar: (i) la energía de interacción electrostática de la parte QM con las cargas discretas de su entorno (E_{ele}) y (ii) la energía de polarización que las cargas discretas provocan en la distribución de carga QM (E_{pol}). La energía de interacción electrostática se calcula rigurosamente mediante la generalización del MEP (véase el apartado 2.1.1.4) para un conjunto de cargas, según la ecuación (2.32).

$$E_{\text{ele}} = \sum_{s=1}^S q_s V(\mathbf{r}_s) \quad (2.32)$$

donde s representa las cargas clásicas de la parte MM y $V(\mathbf{r}_s)$ el MEP en la posición de cada carga.

La energía de polarización se evalúa perturbativamente según la ecuación (2.33),

$$E_{\text{pol}} = \sum_j^{\text{vir}} \sum_i^{\text{occ}} \frac{1}{\epsilon_i - \epsilon_j} \left\{ \sum_{\mu} \sum_{\nu} c_{\mu i} c_{\nu j} \left\langle \phi_{\mu} \left| \sum_S \frac{q_s}{|\mathbf{r}_s - \mathbf{r}|} \right| \phi_{\nu} \right\rangle \right\}^2 \quad (2.33)$$

donde el primer sumatorio es sobre todos los orbitales moleculares virtuales, el segundo sobre todos los ocupados, ϵ son las energías de los orbitales molecular, ϕ son orbitales atómicos y c los coeficientes de la combinación lineal de orbitales atómicos para formar los orbitales moleculares.

En este marco, a partir del E_{ele} y E_{pol} se han definido un potencial que permite el cálculo de energías de interacción entre una parte QM y una MM a un coste computacional reducido: el GMIPp (*Generalized molecular interaction potential with polarization correction*).

El GMIP se calcula como la interacción electrostática más una interacción de Lennard-Jones (ecuación 2.31), mientras que el GMIPp incluye, además, la polarización provocada por las cargas clásicas (ecuación 2.34).

$$\text{GMIPp} = \text{GMIP} + E_{\text{pol}} \quad (2.34)$$

- Métodos SCRF (*Self-consistent reaction field*)⁶¹⁻⁶³

Su objetivo es estudiar sistemas en solución, permitiendo obtener con gran precisión variables importantes en sistemas químicos como la energía libre de solvatación. Utilizan una representación cuántica del soluto, mientras que el solvente es representado como un dieléctrico continuo. Una explicación más detallada de este método se encuentra en el apartado 2.2.2.2.

- Métodos QM/MM (*Quantum mechanics/molecular mechanics*)^{64,65}

En este caso se emplea una descripción a nivel semiempírico, del funcional de la densidad o *ab initio* con bases pequeñas de la parte cuántica y una descripción de

dinámica molecular o Monte Carlo de la parte clásica. El acoplamiento cuántico-clásico permite capturar efectos de polarización, ausentes en los métodos clásicos convencionales. Estos métodos son muy poco eficientes desde el punto de vista computacional, al precisar miles (MD) o millones (MC) de cálculos SCF. Por ello, su uso actual se ha circunscrito al estudio de pequeñas moléculas en solución y al estudio de reactividad enzimática.

2.2 Inclusión de los efectos de entorno

La mayoría de las metodologías descritas hasta este punto permiten obtener una descripción suficientemente precisa del comportamiento y de las propiedades del sistema químico. Estas son particularmente aplicables al estudio de procesos químicos que tienen lugar en fase gas. Pero en muy pocas ocasiones estas condiciones son reales, ya que la mayor parte de los procesos químicos tienen lugar en fase condensada, donde el entorno químico juega un papel fundamental. La representación de los efectos de entorno es de vital importancia a la hora de profundizar en el estudio de los sistemas y procesos químicos. Este hecho ha dado origen al desarrollo de toda una serie de estrategias computacionales que tratan de simular el efecto del solvente, tanto implícita como explícitamente.

2.2.1 Naturaleza de la interacción soluto-solvente

Para cualquier proceso químico el paso de fase gas a fase líquida provoca toda una serie de cambios en sus propiedades termodinámicas y cinéticas. Estas variaciones se reflejan en los cambios que experimentan las especies moleculares involucradas en el proceso para adaptarse a las condiciones que impone el nuevo medio que las envuelve⁶⁶.

Un líquido formado por moléculas polares carece de un campo estático permanente, pero es capaz de generarlo como reacción a un campo externo aplicado. Este efecto viene caracterizado por la denominada constante dieléctrica ϵ ⁶⁷. Desde un punto de vista microscópico, cuando el soluto se sumerge en un solvente se genera de forma instantánea una distorsión en la estructura local del líquido, y esto se traduce en la creación de un campo eléctrico en el solvente. Este campo eléctrico reacciona con el soluto dando lugar a una interacción soluto-solvente, que es el principal responsable de la energía de solvatación en solventes polares.

La solvatación del soluto produce una relajación de su densidad electrónica desde las condiciones de vacío a las de solución, así como una relajación de la conformación nuclear, de manera que se optimizan las interacciones moleculares soluto-solvente⁶⁸. Los cambios electrónicos y estructurales se pueden monitorizar a través de la variación de diferentes propiedades respecto al valor correspondiente en vacío, como el momento dipolar, el volumen molecular, el potencial electrostático o el desplazamiento del espectro⁶⁹. El efecto de entorno afecta asimismo a la magnitud de las interacciones intermoleculares y la reactividad en fase condensada.

2.2.2 Métodos teóricos de representación del solvente

Existen diversos métodos para el estudio de la influencia del entorno, consiguiendo un mayor o menor éxito en función del modelo que se utiliza para representar el solvente y del tipo de tratamiento para representar las interacciones entre los componentes del sistema (soluto-soluto, soluto-solvente y solvente-solvente).

Una primera clasificación diferenciaría aquellos métodos que consideran explícitamente las moléculas de solvente de los que consideran el solvente como un medio continuo caracterizado por la constante dieléctrica ϵ . Los primeros se denominan métodos *discretos* mientras que los segundos se denominan métodos *continuos*. En ambos casos, sin embargo, el solvente se describe clásicamente: microscópicamente en los primeros y macroscópicamente en los segundos. Una segunda clasificación consideraría el nivel de teoría con el que se describe soluto: clásico o cuántico. Según estas distinciones, los métodos teóricos para cálculos en solución se pueden clasificar en cuatro grupos⁶⁹⁻⁷⁴: continuos-clásicos, discretos-clásicos, continuos-cuánticos y discretos-cuánticos.

En esta memoria se han considerado los efectos de la solvatación a través de métodos discretos, en las simulaciones de dinámica molecular, y de métodos continuos para la determinación de la energía libre de solvatación.

2.2.2.1 *Métodos discretos*

Los métodos discretos, más rigurosos, se basan en la inclusión explícita de las moléculas de solvente dentro del sistema estudiado. De esta manera, se da una descripción de la estructura microscópica del solvente y de las interacciones específicas soluto-solvente. Dentro del marco de la mecánica cuántica, razones de coste computacional limitan la aplicabilidad de los métodos discretos a un número de microestructuras de soluto rodeado por unas pocas moléculas de solvente. Los métodos basados en campos de fuerza, en cambio, permiten simular un elevado número de microestructuras con muchas moléculas de solvente. Sin embargo, estos métodos presentan el inconveniente de no representar adecuadamente los cambios inducidos por el solvente en la estructura electrónica del soluto.

Para poder solventar las limitaciones de los métodos discretos, se han desarrollado los métodos híbridos cuántico/clásico (QM/MM). En ellos, se realiza una descripción clásica de los términos correspondientes al solvente y a las interacciones soluto-solvente, mientras que el soluto se trata de forma mecano-cuántica. Diferentes estudios han demostrado que estos métodos proporcionan muy buenos resultados en la descripción de sistemas en solución⁵³⁻⁵⁶. (Véase el apartado 2.1.3)

2.2.2.2 *Métodos continuos*⁷²

A diferencia de los métodos discretos, los métodos continuos sólo tratan de forma explícita el soluto, mientras que el solvente se representa como un medio dieléctrico continuo, las propiedades del cual se representan de forma promediada. Estos métodos consideran que el solvente reacciona frente a la distribución de carga del soluto generando un campo de reacción (SCRF), que a su vez introduce una perturbación al soluto.

Estos métodos fueron diseñados para determinar la energía libre de solvatación, que suele incluir 3 contribuciones: término de cavitación, término de van der Waals y

término electrostático. Las principales diferencias entre los diversos métodos SCRF estriban en (i) la forma de la interfase soluto-solvente (cavidad), (ii) la descripción de la distribución de cargas del soluto, (iii) la definición del campo de reacción, (iv) el formalismo empleado para calcular el término electrostático y (v) el cálculo de las contribuciones no electrostáticas de la energía libre de solvatación.

2.2.2.3 *Métodos mixtos discreto-continuo o semicontinuos*⁷³

Tanto en los métodos discretos como en los métodos continuos, hay ciertos efectos que no son tomados en consideración. Así, los métodos discretos-cuánticos están limitados a sistemas de reducido tamaño y a una representación muy limitada de las configuraciones soluto-solvente, mientras que los clásicos omiten los efectos de polarización. Por otro lado, los métodos continuos no tratan de forma explícita los puentes de hidrógeno específicos soluto-solvente ni consideran detalles microscópicos de la primera capa de solvatación. Estas limitaciones quedan solventadas con los métodos combinados discreto-continuos, que aprovechan las ventajas que proporcionan las diferentes metodologías. Trabajos recientes han puesto de manifiesto que las contribuciones omitidas tanto en los modelos discretos como los continuos pueden tener una gran relevancia a la hora de dar una buena descripción del sistema^{74,75}, resultando conveniente recurrir a un método combinado.

En estos métodos se describe de forma discreta moléculas de solvente de la primera o primeras capas de solvatación del soluto, mientras que el resto de moléculas de solvente se representan por un continuo dieléctrico. Así las interacciones específicas que se presentan en la zona próxima al soluto son tratadas a un nivel microscópico y el promedio estadístico del resto del solvente se trata por medio de una representación macroscópica.

2.2.2.4 Cálculo de la energía libre de solvatación

La energía libre de solvatación (ΔG_{solv}) se define como el trabajo necesario para llevar un mol de soluto de la fase gas (1M) a solución (1M). La energía libre de solvatación se puede descomponer en tres componentes (ecuación 2.35).

$$\Delta G_{\text{solv}} = \Delta G_{\text{ele}} + \Delta G_{\text{vdW}} + \Delta G_{\text{cav}} \quad (2.35)$$

donde ΔG_{ele} es la contribución electrostática, ΔG_{vdW} es el término de van der Waals y ΔG_{cav} es el término de cavitación. Estos dos últimos términos constituyen la contribución estérica.

La contribución electrostática representa la reacción del soluto frente al solvente y puede ser obtenida de varias formas en función de la naturaleza de la densidad de carga del soluto y del tipo de reacción del solvente. El término de van der Waals describe las interacciones de repulsión-dispersión entre el soluto y el solvente, mientras que el término de cavitación tiene en cuenta el trabajo necesario para crear la cavidad en el seno del solvente en ausencia de soluto. Ambos términos se suelen calcular a partir de la distribución nuclear del soluto, ignorando los detalles de su distribución electrónica.

A continuación se describen los dos métodos empleados en la presente memoria para evaluar estas contribuciones: (i) método del continuo polarizable y (ii) método basado en diferencias finitas.

- Método del continuo polarizable (PCM/MST)

Es un método continuo-cuántico, que consiste en una versión del método desarrollado por Miertus, Scrocco y Tomasi (MST) adaptada a un hamiltoniano cuántico.

En la metodología MST, el soluto se trata tanto a nivel *ab initio* HF/6-31G(d)^{62,77,78} como semiempírico AM1^{63,79,80} y también DFT⁸¹, lo que conlleva a un tratamiento soluto-solvente riguroso que incorpora de forma explícita la polarización mutua. El campo de reacción del solvente se representa mediante un conjunto de cargas ficticias de polarización distribuidas sobre la superficie que delimita la interfase soluto-solvente. Dichas cargas contrarrestan el campo creado por la distribución de carga del soluto. La forma de la cavidad se adapta a la forma molecular del soluto, a partir de los radios de van der Waals atómicos, convenientemente escalados, centrados en cada uno de los átomos que componen la molécula de soluto⁸².

La interacción electrostática entre el soluto y el solvente se determina a través del formalismo SCRF mencionado anteriormente. Dentro de este formalismo, al hamiltoniano del soluto en el vacío (H^0) se le añade una perturbación (V_R).

$$(H^0 + V_R)\Psi = E\Psi \quad (2.36)$$

Este operador de perturbación corresponde al campo de reacción generado por una distribución de carga ($\sigma(s)$) en la superficie de la cavidad, que es fruto del efecto de polarización del soluto sobre el solvente (ecuación 2.36). Para su cálculo, la superficie del solvente se divide en M elementos S_i suficientemente pequeños como para considerar que la distribución de carga ($\sigma(S_i)$) en su interior es constante. Esto hace posible encontrar V_R en términos de cargas puntuales de polarización (q_i) en la superficie de la cavidad (ecuación 2.37), y que se encuentran resolviendo la correspondiente ecuación de Laplace.

$$V_R = \int_S \frac{\sigma(s)}{|\mathbf{r}_0 - \mathbf{r}|} dS = \sum_{i=1}^M \frac{\sigma(S_i)S_i}{|\mathbf{r}_0 - \mathbf{r}|} = \sum_{i=1}^M \frac{q_i}{|\mathbf{r}_0 - \mathbf{r}|} \quad (2.37)$$

El término electrostático se obtiene mediante la ecuación (2.38), donde los índices '0' y 'sol' representan el entorno fase gas y solución respectivamente.

$$\Delta G_{\text{ele}} = \langle \Psi^{\text{sol}} | \mathbf{H}^0 + \frac{1}{2} \mathbf{V}_R | \Psi^{\text{sol}} \rangle - \langle \Psi^0 | \mathbf{H}^0 | \Psi^0 \rangle \quad (2.38)$$

Por otro lado, el término de cavitación se obtiene aplicando el modelo *scaled particle theory* (teoría de la partícula escalada) de Pierotti⁸³ o versiones más sofisticadas como el modelo de Claverie-Pierotti⁸⁴ adaptado a cada tipo de solvente.

La contribución de van der Waals se calcula a partir de una relación lineal con el área de la superficie molecular (ecuación 2.39).

$$\Delta G_{\text{vdW}} = \sum_i \xi_i S_i \quad (2.39)$$

donde ξ_i es el parámetro de tensión del átomo i y S_i la porción de superficie molecular correspondiente al átomo i .

Los parámetros de tensión de los diferentes átomos en agua, cloroformo, tetracloruro de carbono y octanol se representan mediante valores previamente determinados por ajuste con datos experimentales⁸⁵⁻⁸⁷.

- Método basado en diferencias finitas (PB/SA)

Es un método continuo-clásico, conceptualmente similar a los continuo-cuánticos. Se diferencian en el nivel de teoría usado para describir la distribución de la carga del soluto, que se trata a nivel clásico. Esta metodología es muy popular para el estudio de la energía libre de solvatación de grandes sistemas moleculares, donde el uso de métodos QM-SCRF es menos adecuado.

Bajo este modelo el soluto se representa asignando una serie de radios de van der Waals, con los cuales se construye la superficie molecular y un conjunto de cargas (q_i) centradas en cada uno de sus átomos. La interfase soluto-solvente se determina

utilizando los parámetros estándar de van der Waals en conjunción con los radios de exclusión del solvente.

El término electrostático de la energía libre de solvatación se obtiene de dos cálculos independientes⁸⁸, uno con la constante dieléctrica de valor ϵ_{int} dentro y fuera de la cavidad (ϕ^0) y otro con la constante dieléctrica interior de valor ϵ_{int} y de valor ϵ_{ext} en el exterior.

$$\Delta G_{\text{ele}} = \frac{1}{2} \sum_i q_i (\phi_i^{\text{sol}} - \phi_i^0) \quad (2.40)$$

donde los índices '0' y 'sol' representan el entorno fase gas y solución respectivamente y q_i son las cargas puntuales clásicas del soluto.

El punto de partida para determinar el potencial electrostático en cualquier punto ϕ_i es la ecuación de Poisson (2.41)⁸⁹, que relaciona el desplazamiento eléctrico con la densidad de carga.

$$\nabla \{ \epsilon(\mathbf{r}) \nabla \phi(\mathbf{r}) \} = - 4\pi \rho(\mathbf{r}) \quad (2.41)$$

donde ρ es la densidad de carga, ϕ es el potencial electrostático y ϵ es la constante dieléctrica.

La densidad de carga total incluye la distribución de carga del soluto dentro de la cavidad (ρ_{int}) y la densidad de carga generada por la atmósfera de iones fuera de la cavidad (ρ_{ext}) en solución.

$$\rho = \rho_{\text{int}} + \rho_{\text{ext}} \quad (2.42)$$

La densidad de carga debido a la atmósfera iónica puede ser aproximada al equilibrio por la distribución de Boltzmann.

$$\rho_{\text{ext}} = -\varepsilon(\kappa^2)\sinh[\phi(\mathbf{r})] \quad (2.43)$$

donde κ es la inversa de la distancia Debye-Hückle, que se define en función del número de Avogadro, de la carga del electrón, de la fuerza iónica, de la constante de Boltzmann y de la temperatura.

La ecuación de Poisson-Boltzmann puede expresarse de 3 maneras diferentes dependiendo de la fuerza iónica: (i) en ausencia de atmósfera iónica (ecuación de Poisson (2.44)); (ii) para las fuerzas iónicas pequeñas, la función $\sinh[\phi(\mathbf{r})]$ se reemplaza por el primer término de la expansión de la serie de Taylor originando la expresión lineal (ecuación 2.45); y (iii) para fuerzas iónicas, donde se requiere la expresión no lineal (ecuación 2.46).

$$\nabla \{ \varepsilon(\mathbf{r}) \nabla \phi(\mathbf{r}) \} = -4\pi\rho_{\text{int}}(\mathbf{r}) \quad (2.44)$$

$$\nabla \{ \varepsilon(\mathbf{r}) \nabla \phi(\mathbf{r}) \} = \kappa'_0{}^2 \phi(\mathbf{r}) - 4\pi\rho_{\text{int}}(\mathbf{r}) \quad (2.45)$$

$$\nabla \{ \varepsilon(\mathbf{r}) \nabla \phi(\mathbf{r}) \} = \kappa'_0{}^2 \sinh[\phi(\mathbf{r})] - 4\pi\rho_{\text{int}}(\mathbf{r}) \quad (2.46)$$

donde $\kappa'_0{}^2$ es 0 en el interior de la cavidad y κ^2 en el exterior de la cavidad.

La resolución de la ecuación de Poisson-Boltzmann permite, en principio, obtener el potencial electrostático en cualquier punto ϕ_i . Desafortunadamente, la solución analítica no es factible, siendo necesario el uso de diferentes métodos como el de las diferencias finitas. Esta aproximación se lleva a cabo utilizando una malla finita cúbica que engloba al soluto y un cierto volumen de solvente continuo a su alrededor. El potencial electrostático en cada punto se determina como función de los potenciales en los puntos contiguos en la malla, repitiéndose este proceso hasta la convergencia⁹⁰. Los resultados son muy sensibles al tamaño de la malla usada, siendo a menudo necesario emplear

mallas muy finas, lo que aumenta el coste computacional del cálculo. Estrategias, como el enfoque (*focussing*)⁹⁰, han sido diseñadas para minimizar este coste.

Una vez conocido el potencial electrostático, el cálculo del componente electrostático de la energía libre de solvatación es determinado a partir de la ecuación 2.40. El componente estérico se determina a partir de la superficie accesible al solvente de cada uno de los átomos mediante una relación lineal empírica.

$$\Delta G_{\text{estérico}} = \sum_{k=1}^N \sigma_k (\text{SAS})_k \quad (2.47)$$

donde σ_k es el parámetro de solvatación del átomo k por unidad de área y $(\text{SAS})_k$ es la superficie accesible al solvente.

2.2.3 Ciclos termodinámicos

El uso de ciclos termodinámicos es una estrategia muy común en el estudio de procesos químicos en solución. Debido al carácter de función de estado de la energía libre, su valor alrededor de un ciclo termodinámico debe ser cero. Aprovechando esta característica es posible calcular la ΔG de un proceso en solución de forma indirecta, a partir de la ΔG del proceso en fase gas y de las energías libres de solvatación de los compuestos implicados. La figura 2.6 muestra un ciclo termodinámico estándar para el cálculo de un proceso en solución.

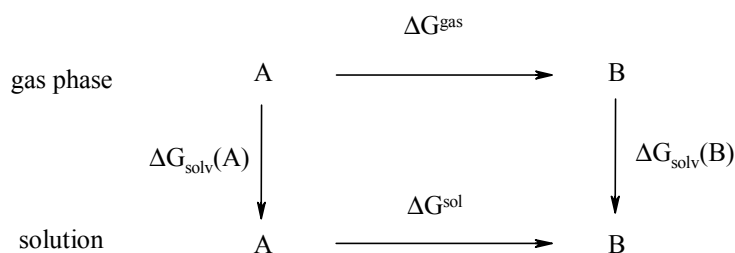


Figura 2.6 Ciclo termodinámico para el cálculo de la energía libre de procesos en solución.

El cálculo de la reacción $A \rightarrow B$ en solución (ΔG^{sol}) puede llevarse a cabo mediante el citado ciclo termodinámico y la ecuación (2.48).

$$\Delta G^{\text{sol}} = \Delta G^{\text{gas}} + (\Delta G_{\text{solv}}(\text{B}) - \Delta G_{\text{solv}}(\text{A})) = \Delta G^{\text{gas}} + (\Delta \Delta G_{\text{solv}}(\text{B-A})) \quad (2.48)$$

donde $\Delta G_{\text{solv}}(\text{A})$ y $\Delta G_{\text{solv}}(\text{B})$ son las energías libres de solvatación de A y B respectivamente y ΔG^{gas} es la energía libre del proceso $A \rightarrow B$ calculado en fase gas.

Similarmente, se puede usar un ciclo termodinámico como el de la figura 2.7 y la ecuación 2.49 para calcular las energías libres de solvatación (ΔG_{solv}) a partir de cálculos discretos tipo FEP o TI.

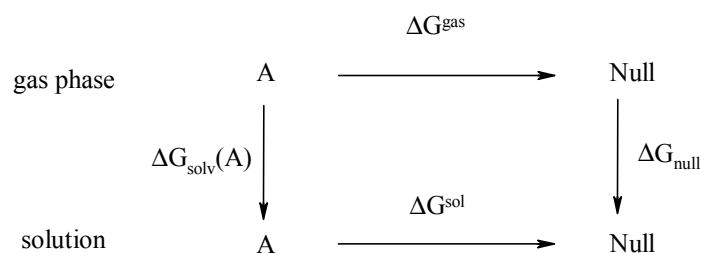


Figura 2.7 Ciclo termodinámico para el cálculo de la energía libre de solvatación.

$$\Delta G_{\text{solv}}(\text{A}) = \Delta G^{\text{gas}} - \Delta G^{\text{sol}} \quad (2.49)$$

La identidad de las dos fases (gas y solución) en ambos ciclos termodinámicos puede generalizarse a otras fases (solvente 1 y solvente 2).

2.3 Análisis de resultados^{91,92}

Una vez finalizada una simulación de dinámica molecular se procede a analizar la trayectoria. Generalmente se realizan una gran variedad de análisis tanto estáticos (de la estructura promedio) como dinámicos (a lo largo de toda la trayectoria). En el caso de los ácidos nucleicos, de entre todos ellos cabe destacar los análisis siguientes:

- (i) Características estructurales: parámetros de hélice, tamaño de los surcos, conformaciones del esqueleto, etc.
- (ii) Hidratación (densidad de agua aparente)
- (iii) Potencial de interacción molecular (MIP)
- (iv) Componentes principales (PCA)
- (v) Energías de interacción: por puente de hidrógeno, por apilamiento, entre fosfatos, etc.
- (vi) Energías de solvatación
- (vii) Cálculos entrópicos
- (viii) Otros cálculos dependiendo del sistema concreto

Algunos de estos análisis precisan una metodología específica que se describe a continuación.

- Análisis del potencial de interacción molecular (MIP)⁹¹⁻⁹³

En este tipo de análisis se calcula el potencial clásico asociado a la interacción de un átomo prueba O^+ con la estructura promedio de un ácido nucleico, empleando el programa cMIP⁹³. El potencial electrostático molecular (MIP) se obtiene a partir de la suma de un término electrostático más un término de van der Waals de tipo Lennard-Jones. El término electrostático se calcula utilizando la ecuación de Poisson-Boltzmann

(PB)⁷¹, que se resuelve numéricamente empleando el formalismo no lineal para un medio con una fuerza iónica de 0.145 M, una constante dieléctrica de 2 para el ácido nucleico y de 80 para el agua, y un espaciado de malla de 0.5 Å con la estructura centrada en el centro de geometría. La contribución de Lennard-Jones se determina usando los parámetros estándar (típicamente TIP3P para el átomo prueba y OPLS o AMBER-94/98 para el ácido nucleico).

Este análisis determina la capacidad de las estructuras de interactuar con grupos cargados positivamente. Así, las regiones de potencial negativo definen las áreas donde la interacción con iones o ligandos cargados positivamente sería favorable. Las regiones con un elevado potencial, tanto positivo como negativo, definen áreas donde la interacción con ligandos polares sería favorable.

- Análisis de componentes principales (PCA)^{94,95}

El análisis de la flexibilidad molecular de las estructuras de los ácidos nucleicos se realiza empleando el método PCA.

En este tipo de análisis se calcula la matriz de covarianza posicional de una porción equilibrada de trayectoria. La diagonalización de dicha matriz proporciona un conjunto de vectores propios ortogonales en conjunción con sus correspondientes valores propios. Los vectores propios proporcionan una representación vectorial de cada modo de deformación estructural, y el valor propio del modo indica la contribución relativa que este modo tiene al movimiento dentro de la trayectoria. Es decir, que cada vector propio obtenido representa una fracción del movimiento total de la estructura.

Posteriormente se realiza una medida cuantitativa de la variación de la contribución de cada vector propio sobre toda la trayectoria. Para ello se rescribe la trayectoria sobre los mayores vectores propios, y estas proyecciones de la trayectoria se analizan por su comportamiento dependiente del tiempo y su distribución de probabilidades. Para facilitar la interpretación de las deformaciones asociadas a cada componente principal, se generan artificialmente breves trayectorias de MD. Estas nuevas trayectorias

contienen estructuras donde las coordenadas varían linealmente entre el valor mínimo y el máximo observado de un determinado vector propio. El conjunto de trayectorias resultantes se visualizan y se analizan con la finalidad de identificar el movimiento involucrado más notable.

En general, la flexibilidad de una estructura de un ácido nucleico proviene de un complejo y amplio rango de movimientos de alta y baja frecuencia. Los primeros modos, que corresponden a movimientos de baja frecuencia, explican una parte muy significativa de la varianza estructural de la trayectoria. La inspección de los vectores propios (modos normales de vibración) asociados con los valores propios (frecuencias de vibración) muestra la dinámica esencial de la estructura, es decir, permite determinar los movimientos conformacionales más importantes de la estructura.

- Método de solvatación

La energía libre de solvatación de las estructuras de los ácidos nucleicos se determina como la adición de dos contribuciones: la electrostática y la estérica. Siguiendo la metodología PB/SA (véase el apartado 2.2.2.4), el componente electrostático se determina resolviendo la ecuación de Poisson-Boltzmann (PB)⁷¹ - empleando el programa MEAD^{96,97} - a una temperatura de 300 K. La malla empleada para el cálculo se centra en el centro geométrico del sistema con un espaciado inicial de 1 Å y un espaciado final de 0.4 Å (*focussing*). Para simular un entorno acuoso se emplea una constante dieléctrica externa de 80 y una fuerza iónica de 0.145 M, mientras que el interior del ácido nucleico presenta una constante dieléctrica de 2. La interfase soluto-solvente se determina empleando los parámetros estándar de van der Waals^{96,97} en conjunción con los radios de exclusión de 1.4 (agua) y 2.0 Å (iones). El componente estérico de la solvatación se determina escalando la superficie accesible al solvente por un factor de 0.005 kcal/mol Å²⁸⁸.

- Cálculo de entropía^{98,99}

El cálculo de la entropía de las estructuras de los ácidos nucleicos se realiza empleando la fórmula de Schlitter.

De acuerdo con esta aproximación se estiman los valores de la entropía S' a partir de la diagonalización de la matriz covarianza cartesiana corregida por masa de la trayectoria de una simulación. La formula heurística generalizada para un sistema viene dada por:

$$S < S' = \frac{1}{2} k_B \ln \det \left[1 + \frac{k_B T e^2}{\hbar^2} M \sigma \right] \quad (2.50)$$

donde k_B es la constante de Boltzmann, T es la temperatura, e es el número de Euler, \hbar es la constante de Planck dividida por 2π , M es la matriz de masas y σ es la matriz de covarianza de la fluctuación posicional de los átomos.

En esta generalización los grados de libertad congelados, como el centro de masas fijo o la orientación de la molécula, no contribuyen al valor de la entropía S' .

2.4 Referencias bibliográficas

- 1) Leach, A.R. (1996) En *Molecular Modelling. Principles and Applications*. Longman, Singapore.
- 2) Andrés, J. & Beltrán, J. Eds. (2000) En *Química Teórica y Computacional*. Publicacions de la Universitat Jaume I, Castelló de la Plana.
- 3) Szabo, A. & Ostlund, N.S. (1982) En *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. McGraw-Hill, New York.
- 4) Pople, J.A. & Beveridge, D.L. (1970) En *Approximate Molecular Orbital Theory*. McGraw-Hill, New York.
- 5) Foresman, J.B. & Frisch, A. (1996) En *Exploring Chemistry with Electronic Structure Methods*. Gaussian, Inc., Pittsburgh.
- 6) Hehre, W.J., Radom, L., Schleyer, P. vP. & Pople, J.A. (1986) En *Ab initio Molecular Orbital Theory*. Wiley, New York.
- 7) Hariharan, P.C. & Pople, J.A. (1973) *Theor. Chim. Acta*, **28**, 213.
- 8) Møller, C. & Plesset, M.S. (1934) *Phys. Rev.*, **46**, 618.
- 9) Bates, D.R. (1961) En *Quantum Theory*. Academic Press, New York.
- 10) Stewart, J.J.P. (1990) En *Reviews in Computational Chemistry*; Volume 1. VCH Publishers, Inc., New York.
- 11) Zerner, M.C. (1991) En *Reviews in Computational Chemistry*; Volume 2. VCH Publishers, Inc., New York.

-
- 12) Dewar, M.J.S., Zoebisch, E.G., Healy, E.F. & Stewart, J.J.P. (1985) *J. Am. Chem. Soc.*, **107**, 3902.
- 13) Parr, R.G. & Yang, W. (1995) *Annu. Rev. Phys. Chem.*, **46**, 701.
- 14) Parr, R.G. & Yang, W. (1989) En *Density Functional Theory of Atoms and Molecules*. Oxford University Press, Oxford.
- 15) Hohenberg, P. & Kohn, W. (1964) *Phys. Rev. B*, **136**, 864.
- 16) Kohn, W. & Sham, L.J. (1965) *Phys. Rev. A*, **140**, 1133.
- 17) Vosko, S.H., Wilk L. & Nusair, M. (1980) *Can. J. Phys.*, **58**, 1200.
- 18) Perdew, J.P. & Wang, Y. (1992) *Phys. Rev. B*, **45**, 13244.
- 19) Becke, A.D. (1988) *Phys. Rev. A*, **38**, 3098.
- 20) Lee, C., Yang, W. & Parr, R.G. (1988) *Phys. Rev. B*, **37**, 785.
- 21) Daudel, R., Leroy, G., Peeters, D. & Sana, M. (1983) En *Quantum Chemistry*. John Wiley, Chichester.
- 22) Mulliken, R.S. (1955) *J. Chem. Phys.*, **23**, 1833.
- 23) Bader, R.F.W. (1985) *Acc. Chem. Res.*, **18**, 9.
- 24) Bader, R.F.W. (1990) En *Atoms in Molecules – A Quantum Theory*. Oxford University Press, Oxford.
- 25) Bader, R.F.W. (1991) *Chem. Rev.*, **91**, 893.

- 26) Scrocco, E. & Tomasi, J. (1973) *Top. Curr. Chem.*, **42**, 95.
- 27) Kaufman, J.J., Harharan, P.C. & Tobin, F.C. (1981) En *Chemical Applications of Atomic and Molecular Electrostatic Potential*. Plenum Press, New York.
- 28) Murray, J.S. & Sen, K. (1996) En *Molecular Electrostatic Potential: Concepts and Applications*. Elsevier, Amsterdam.
- 29) Bayly, C.I., Cieplak, P., Cornell, W.D. & Kollman, P.A. (1993) *J. Phys. Chem.*, **97**, 10269.
- 30) Cieplak, P., Cornell, W.D., Bayly, C. I. & Kollman P.A. (1995) *J. Comp. Chem.*, **16**, 1357.
- 31) Alhambra, C., Luque, F.J. & Orozco, M. (1994) *J. Comp. Chem.*, **15**, 12.
- 32) Beck, B., Clark, T. & Glen, R.C. (1997) *J. Comp. Chem.*, **18**, 744.
- 33) Wang, J.M., Cieplak, P. & Kollman P.A. (2000) *J. Comp. Chem.*, **21**, 1049.
- 34) Davidson, E.R. & Chakravory, S.J. (1994) *Chem. Phys. Lett.*, **217**, 49.
- 35) Jensen, F. (1996) *Chem. Phys. Lett.*, **261**, 633.
- 36) Boys, S.F. & Bernardi, F. (1970) *Molecular Physics*, **19**, 553.
- 37) Burkett, O. & Allinger, N.L. (1982) En *Molecular Mechanics*. ACS Monograph Series. Washington.
- 38) Orozco, M., Alemán, C. & Luque, F.J. (1993) *Acta Chim. Hung. Models in Chemistry*, **130**, 695.

-
- 39) Allinger, N.L. (1977) *J. Am. Chem. Soc.*, **99**, 8127.
- 40) Allinger, N.L., Yuh, Y.H. & Lii, J.H. (1989) *J. Am. Chem. Soc.*, **111**, 8551.
- 41) Brooks, B.R., Bruccoleri, R.E. Olaf-son, B.D., States, D.J., Swaminathan, S. & Karplus, M. (1983) *J. Comp. Chem.*, **4**, 187.
- 42) van Gunsteren, W.F. & Berendsen, H.J.C. (1993) En *GRONingen Molecular Simulations (GROMOS)*. ETH, Zurich.
- 43) Weiner, S.J., Kollman, P.A., Nguyen, D.T. & Case, D.A. (1986) *J. Comp. Chem.*, **7**, 230.
- 44) Jorgensen, W.L. & Tirado-Rives, J. (1988) *J. Am. Chem. Soc.*, **110**, 1657.
- 45) van Holde, K.E., Jonson, W.C. & Ho, P.S. (1998) En *Principles of Physical Biochemistry*. Prentice Hall, New Jersey.
- 46) McCammon, J.A. & Karplus, M (1980) *Annu. Rev. Phys. Chem.*, **31**, 29.
- 47) McCammon, J.A. & Harvey, S. (1987) En *Dynamics of Proteins and Nucleic Acids*. Cambridge University Press, Cambridge.
- 48) Cheatham III, T.E. & Kollman, P.A. (2000) *Annu. Rev. Phys. Chem.*, **51**, 435.
- 49) van Gunsteren, W.F. & Mark, A.E. (1998) *J. Chem. Phys.*, **108**, 6109.
- 50) Frenkel, D. & Smith, B. (1996) En *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press, Inc., San Diego.
- 51) McQuarri, D.A. (1976) En *Statistical Mechanics*. Harper & Row, New York.

- 52) Kollman, P.A. (1993) *Chem. Rev.*, **93**, 2395.
- 53) Field, M.J., Bash, P.A. & Karplus, M. (1990) *J. Comp. Chem.*, **11**, 700.
- 54) Gao, J. (1994) *J. Am. Chem. Soc.*, **116**, 9324.
- 55) Maseras, F. & Morokuma, K. (1995) *J. Comp. Chem.*, **16**, 1170.
- 56) Lyne, P.D., Hodoscek, M. & Karplus, M. (1999) *J. Phys. Chem. A*, **103**, 3462.
- 57) Orozco, M. & Luque, F.J. (1993) *J. Comp. Chem.*, **14**, 587.
- 58) Alhambra, C., Luque, F.J. & Orozco, M. (1995) *J. Phys. Chem.*, **99**, 3084.
- 59) Orozco, M. & Luque, F.J. (1996) En *Theoretical and Computational Chemistry*, Volume 3. Elsevier Science, New York.
- 60) Luque, F.J. & Orozco, M. (1998) *J. Comp. Chem.*, **19**, 866.
- 61) Cramer, C.J. & Truhlar, D.G. (1992) *Science*, **256**, 213.
- 62) Bachs, M., Luque, F.J. & Orozco, M. (1994) *J. Comp. Chem.*, **15**, 446.
- 63) Luque, F.J., Bachs, M., & Orozco, M. (1994) *J. Comp. Chem.*, **15**, 847.
- 64) Gao, J. & Xia, X. (1992) *Science*, **258**, 631.
- 65) Luzhkov, V. & Warshel, A. (1992) *J. Comp. Chem.*, **13**, 199.
- 66) Reichardt, C. (1990) En *Solvent and Solvent Effects in Organic Chemistry*. VCH, Weinheim.

- 67) Jackson, J.D. (1975) En *Classical Electrodynamics*. John Wiley & Sons, New York.
- 68) Beveridge, D.L. & Schnuell, G.W. (1974) *J. Phys. Chem.*, **78**, 2064.
- 69) Orozco, M., Alhambra, C., Barril, X., López, J.M. Busquets, M.A. & Luque, F.J. (1996) *J. Mol. Model.*, **2**, 1.
- 70) Luque, F.J., Bofill, J.M. & Orozco, M. (1995) *J. Chem. Phys.*, **103**, 10183.
- 71) Orozco, M. & Luque, F.J. (2000) *Chem. Rev.*, **100**, 4187.
- 72) Tomasi, J. & Persico, M. (1994) *Chem. Rev.*, **94**, 2027.
- 73) Cramer, C.J. & Truhlar, D.G. (1999) *Chem. Rev.*, **99**, 2161.
- 74) Contador, J.C., Aguilar, M.A. & Olivares del Valle, F.J. (1997) *Chem. Phys.*, **214**, 113.
- 75) Alemán, C. & Galembeck, S.E. (1998) *Chem. Phys.*, **232**, 151.
- 76) Rivail, J.L., Rinaldi, D. (1996) En *Computational Chemistry. Review of Current Trends*. World Scientific Publishing, Singapore.
- 77) Miertus, S., Scrocco, E. & Tomasi, J. (1981) *Chem. Phys.*, **55**, 117.
- 78) Miertus, S. & Tomasi, J. (1982) *Chem. Phys.*, **65**, 239.
- 79) Chudinov, G.E., Napolov, D.V. & Basilevsky, M.V. (1998) *Chem. Phys.*, **160**, 41.
- 80) Orozco, M., Bachs, M., & Luque, F.J (1995) *J. Comp. Chem.*, **16**, 563.

- 81) Fortunelli, A. & Tomasi, J. (1994) *Chem. Phys. Lett.*, **231**, 34.
- 82) Pascual-Ahuir, J.L., Silla, E., Tomasi, J. & Bonaccorsi, R. (1987) *J. Comp. Chem.*, **7**, 778.
- 83) Pierotti, R.A. (1976) *Chem. Rev.*, **76**, 717.
- 84) Claverie, P. (1978) En *Intermolecular Interactions: from Diatomics to Biomolecules*. John Wiley, Chichester.
- 85) Luque, F.J., Zhang, Y., Alemán, C., Bachs, M., Gao, J. & Orozco, M. (1996) *J. Phys. Chem.*, **100**, 4269.
- 86) Luque, F.J., Bachs, M., Alemán, C. & Orozco, M. (1996) *J. Comp. Chem.*, **17**, 806.
- 87) Curutchet, C., Orozco, M. & Luque, F.J. (2001) *J. Comp. Chem.*, **22**, 1180.
- 88) Sitkoff, D., Sharp, K.A. & Honig, B. (1994) *J. Phys. Chem.*, **98**, 1978.
- 89) Sharp, K.A. & Honig, B. (1990) *Annu. Rev. Biophys. Chem.*, **19**, 301.
- 90) Nicholls, A. & Honig, B. (1991) *J. Comp. Chem.*, **12**, 435.
- 91) Shields, G.C., Laughton, C.A. & Orozco, M. (1997) *J. Am. Chem. Soc.*, **119**, 7463.
- 92) Shields, G.C., Laughton, C.A. & Orozco, M. (1998) *J. Am. Chem. Soc.*, **120**, 5895.
- 93) Gelpí, J.L., Kalko, S., Barril, X., Cirera, J., de la Cruz, X., Luque, F.J. & Orozco, M. (2001) Aceptado en *Proteins*.

- 94) Wlodek, S.T., Clark, T.W., Scott, L.R. & McCammon, J.A. (1997) *J. Am. Chem. Soc.*, **119**, 9513.
- 95) Sherer, E.C., Harris, S.A., Soliva, R., Orozco, M. & Laughton, C.A. (1999) *J. Am. Chem. Soc.*, **121**, 5981.
- 96) Bashford, D. & Gerwert, K. (1992) *J. Mol. Biol.*, **224**, 473.
- 97) Bashford, D. (1997) En *Scientific Computing in Object-Oriented parallel environments*. Springer, Berlin.
- 98) Schlitter, J. (1993) *Chem. Phys. Lett.*, **215**, 617.
- 99) Schäfer, H., Daura, X., Mark, A.E. & van Gunsteren, W.F. (2001) *Proteins*, **43**, 45.

(Esta página está intencionadamente en blanco)

3 Estructura de los ácidos nucleicos

(Esta página está intencionadamente en blanco)

Existen dos tipos fundamentales de ácidos nucleicos, el ácido ribonucleico (ARN) y el ácido desoxirribonucleico (ADN). Éstos desempeñan un papel crucial en el almacenaje y la transmisión de la información genética y en la biosíntesis de las proteínas. Por todo ello, los ácidos nucleicos son unas de las biomoléculas más relevantes.

En este capítulo se describen brevemente algunos de los conceptos esenciales para la comprensión de la estructura y de las propiedades de los ácidos nucleicos, haciendo especial énfasis en el ADN.

3.1 Componentes estructurales de los ácidos nucleicos¹⁻⁵

Los ácidos nucleicos están constituidos por 3 unidades elementales: un azúcar, un fosfato y una base heterocíclica. La unión del azúcar y la base se denomina *nucleósido* y la unión del fosfato, del azúcar y de la base se denomina *nucleótido*.

3.1.1 Estructura y nomenclatura^{6,7}

Las bases nitrogenadas son estructuras heterocíclicas aromáticas y planas. Se clasifican en función de su estructura en bases púricas (derivadas de la purina) y bases pirimidínicas (derivadas de la pirimidina). Las bases nitrogenadas más comunes son adenina (A), guanina (G), citosina (C), uracilo (U) y timina (T) (figura 3.1). Las tres primeras bases (A, G y C) se encuentran en todos los ácidos nucleicos, mientras que la timina sólo se localiza en el ADN y es substituida por el uracilo en el ARN.

En la nomenclatura⁷ de las bases nitrogenadas sólo se enumeran los átomos del anillo aromático y los substituyentes adoptan el mismo número que el átomo al que están unidos.

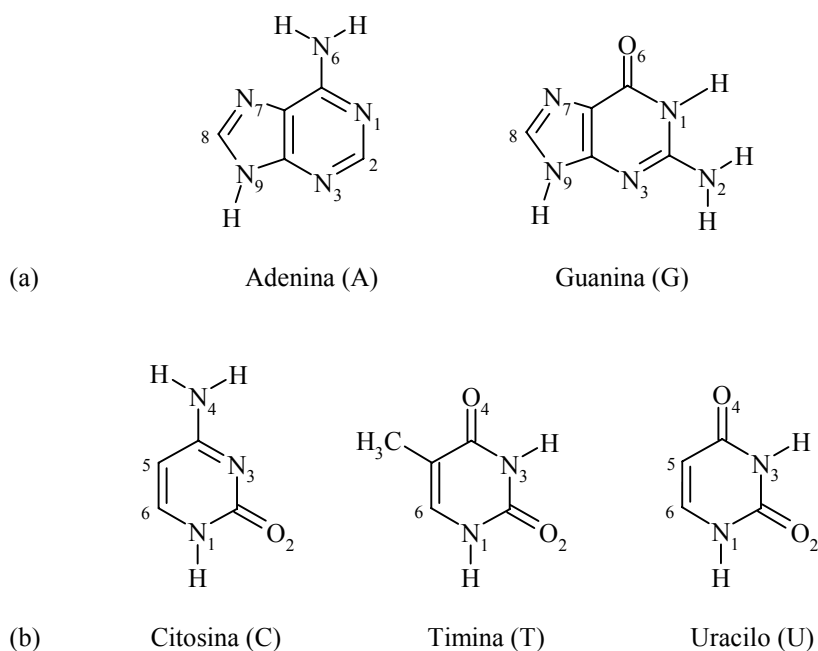


Figura 3.1 Estructura química de las bases nitrogenadas (a) púricas y (b) pirimidínicas.

Es posible localizar otros derivados de estas bases en los ácidos nucleicos, especialmente en el ARN de transferencia⁸. Estos derivados se obtienen por modificaciones enzimáticas específicas de las bases o bien por mutaciones espontáneas. En la mayoría de los casos, las bases nitrogenadas presentan sólo ligeras modificaciones, como: 2-tiocitosina, N6-metiladenina, hipoxantina, 5,6-dihidrouracil, etc. No obstante, en algunos casos se pueden encontrar modificaciones más complejas, como N6-(2-isopentenil) adenina o N6-treoninocarbonil adenina.

Los nucleósidos están formados por la unión de una base nitrogenada y una pentosa (figura 3.2). Las bases pirimidínicas se enlazan al azúcar a través del N1 y las bases púricas a través del N9. Esta unión está en la misma cara del anillo de azúcar que el grupo $-CH_2OH$ y se conoce como un enlace β -glucosídico. Las pentosas que aparecen en los ácidos nucleicos son la D-ribosa en el ARN y la 2'-desoxi-D-ribosa en el ADN.

Los mencionados azúcares son anillos furanósidos que adoptan empaquetamientos (*puckering*) en los cuales 1 o 2 átomos están fuera del plano formado por el resto. Los azúcares siempre se nombran con “prima”. La nomenclatura⁷ de la ribosa/desoxiribosa comienza por el carbono unido a la base, que se denomina 1' y sigue de tal manera que el carbono unido al grupo -CH₂OH es el 4'. El carbono del sustituyente -CH₂OH es el 5'. El oxígeno furanósico se nombra O4' u O1' y el resto de los sustituyentes adoptan el mismo número que el átomo al que están unidos.

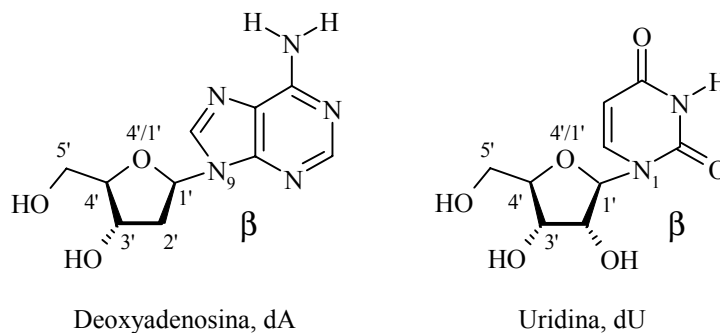


Figura 3.2 Estructura química de dos nucleósidos.

Un nucleótido es el resultado de la esterificación de la pentosa de un nucleósido con un fosfato (figura 3.3). La unión se produce entre cualquiera de los tres grupos hidroxilo libres del azúcar, aunque generalmente se da entre los que ocupan la posición 3' ó 5'. Algunos de los nucleótidos están unidos a dos o más grupos fosfato.

El grupo fosfato se nombra por la posición del azúcar a la que está unido y por el número de fosfatos esterificados: mono-, di- o trifosfato.

Algunos nucleótidos y derivados de éstos poseen interés biológico de *per se*, como son los fosfatos de adenosina. Entre ellos destacan: adenosina 5'-difosfato (ADP) y adenosina 5'-trifosfato (ATP) como almacén energético; adenosina 3',5'-monofosfato cíclica (cAMP) como mediador en procesos hormonales; y los que actúan como

coenzimas flavina adenina dinucleótido (FAD), nicotinamida adenina dinucleótido (NAD^+) y nicotinamida adenina dinucleótido fosfato (NADP^+).

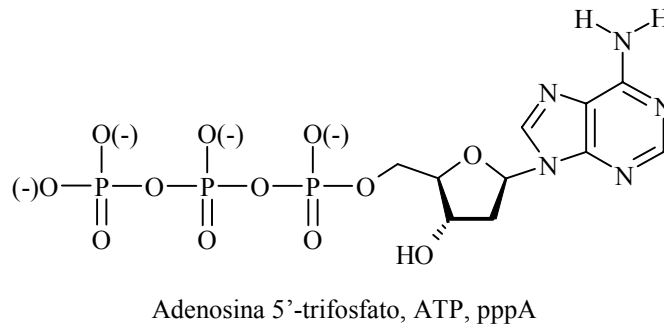


Figure 3.3 Estructura química de un nucleótido.

Los ácidos nucleicos son polímeros de nucleótidos, unidos a través de los fosfatos, que por una parte se enlazan con el C3' de la pentosa de un nucleósido y por otra con el C5' de la pentosa de otro nucleósido. Al enlace que se forma se le denomina enlace fosfodiéster. Los polímeros forman largas cadenas, con un sector idéntico en todas ellas (fosfato-pentosa-fosfato) y otro variable, las bases unidas a las pentosas (figura 3.4).

Se clasifican según la pentosa que poseen. Si la pentosa es desoxirribosa, tenemos el ADN y si se trata de la ribosa, tenemos el ARN. Tanto el ARN como el ADN pueden estar formados por una única cadena, por dos cadenas, e incluso, por tres y cuatro cadenas. Asimismo, el ADN puede ser circular o lineal.

Las cadenas se nombran a nivel de estructura primaria por la composición de las bases⁷. Siempre se nombra del extremo 5' al 3', donde 5' es el extremo que tiene el grupo fosfato libre unido al carbono C5'.

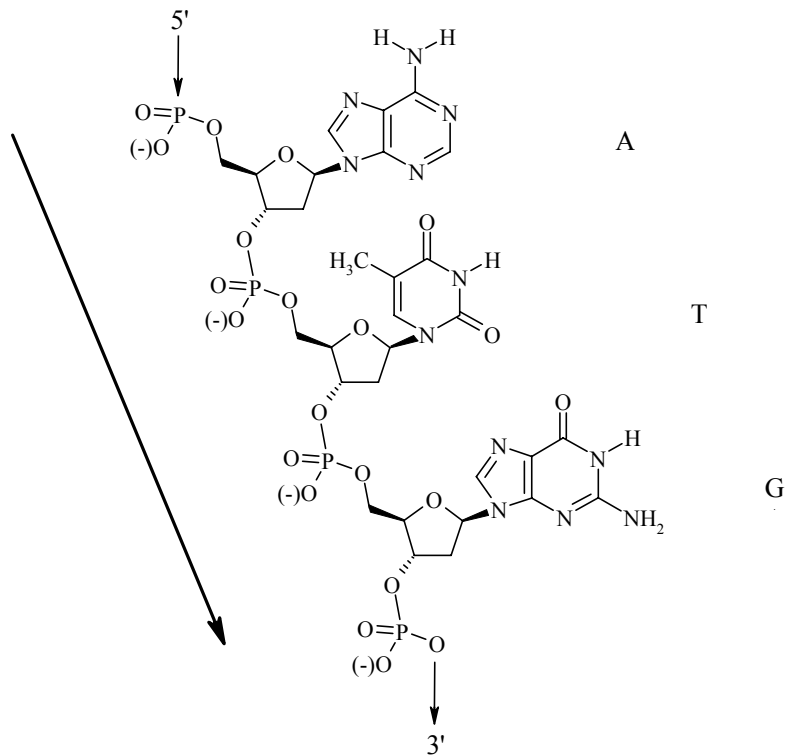


Figura 3.4 Fragmento de un ácido desoxirribonucleico (ADN) con la secuencia adenosina, timidina y guanósina unido por enlaces 3', 5'-fosfodiéster. La dirección de la cadena viene mostrada por la flecha.

3.1.2 Conformación

La flexibilidad conformacional de los ácidos nucleicos viene dada por los movimientos del anillo furanósico y por los movimientos de rotación respecto al enlace glucosídico (base-azúcar), al enlace exocíclico C4'-C5' y los enlaces del grupo fosfato.

Los movimientos de rotación respecto a los diferentes enlaces se nombran por un ángulo diedro o de torsión. Éste marca la rotación respecto al enlace en cuestión, siempre en el sentido de las agujas del reloj. Hay 7 ángulos diedros por nucleótido que tienen que ser especificados para caracterizar la estructura secundaria de los ácidos

nucleicos⁹ y todos ellos están correlacionados a excepción de γ (figura 3.5). Esto significa que los cambios estructurales siguen movimientos concertados.

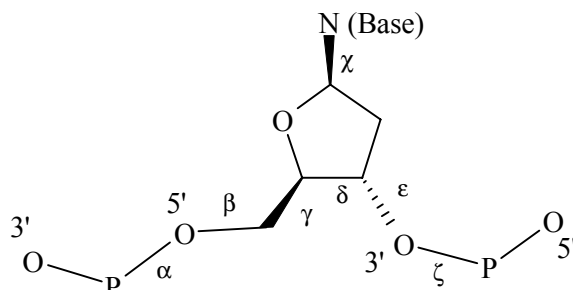


Figura 3.5 Definición de los ángulos diedros de una cadena polinucleótida.

i) Conformación del anillo furanósico

La conformación del anillo furanósico es compleja al existir una restricción en los movimientos de los ángulos por tratarse de un ciclo (movimientos concertados).

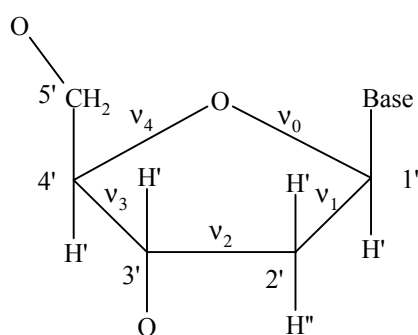


Figura 3.6 Ángulos diedros internos de un anillo de cinco miembros.

Estos anillos de cinco miembros siempre adoptan conformaciones fuera del plano, básicamente del tipo *envelope* (*E*) y del tipo *twist* (*T*)¹⁰ (figura 3.7), porque éstas son

más estables y favorables energéticamente que la conformación coplanar. Por otro lado, los confórmeros tipo *E* suelen ser más estables que los *T*. También se emplea la nomenclatura *endo*, si el átomo que está fuera del plano está en el mismo lado, o *exo*, si dicho átomo se encuentra en el lado contrario a la posición de C5' ¹¹ (figura 3.7).

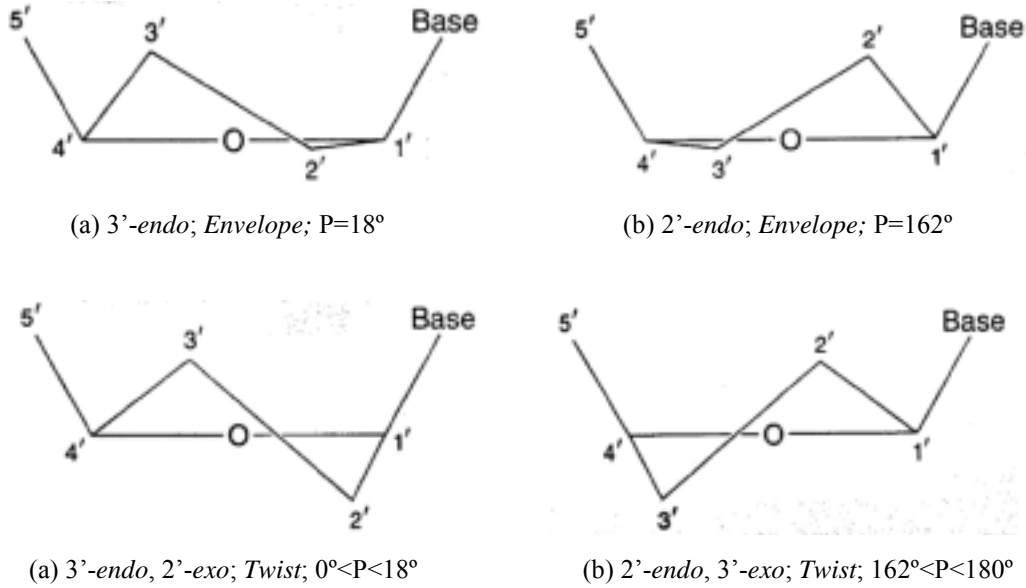


Figura 3.7 Conformaciones del azúcar encontradas en los nucleótidos. (a) Confórmeros Norte que se encuentran en la doble hélice de la forma A y (b) confórmeros Sur que se encuentran en la doble hélice de la forma B. (Fuente: véase referencia 5, p. 19)

La conformación de un anillo de cinco miembros puede ser descrita por sólo dos parámetros gracias a las restricciones del anillo: el ángulo de pseudorrotación o ángulo de fase (*P*) y la amplitud de puckering (v_m)^{12,13}, de acuerdo con las siguientes expresiones (3.1) y (3.2):

$$\tan P = \frac{(v_4 + v_1) - (v_3 + v_0)}{2 v_2 (\sin 36^\circ + \sin 72^\circ)} \quad (3.1)$$

$$v_m = \frac{v_2}{\cos P} \quad (3.2)$$

El primer parámetro da información sobre el átomo (o los átomos) que está fuera del plano formado por el resto y el segundo indica el grado de distorsión de la planaridad.

La conformación de la ribosa viene dada según su posición en el ciclo pseudorrotacional¹³ (figura 3.8). En éste se pueden distinguir cuatro zonas: Norte (N, $P \approx 0^\circ$), Sur (S, $P \approx 180^\circ$), Este (E, $P \approx 90^\circ$) y Oeste (W, $P \approx 270^\circ$). Dentro de esta zona es posible encontrar toda una gama de conformaciones, que se nombran por su ángulo de fase.

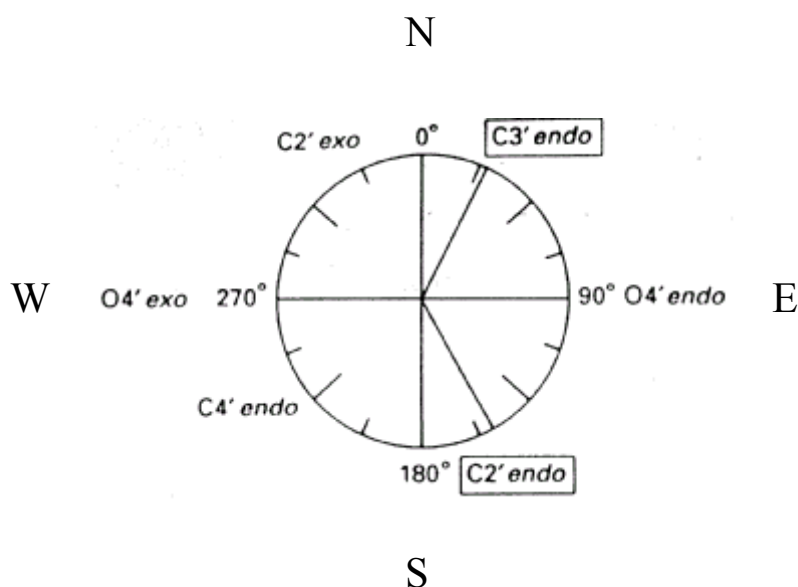


Figura 3.8 Ciclo de pseudorrotación. (Fuente: véase referencia 4, p. 20)

Se ha visto que los cambios de conformación ocurren por variación del ángulo de fase P , mientras que la amplitud de puckering se mantiene constante alrededor de $38-40^\circ$. Las conformaciones más favorecidas son la N y la S¹⁴. Las transiciones N-S tienen lugar siempre por la zona E y la zona W corresponde a una conformación inestable. La conformación en zona S suele ser C2'-endo con unos ángulos de fase de $160-164^\circ$, la conformación en zona N es C3'-endo con un ángulo de fase de $15-20^\circ$.

ii) Rotación respecto al enlace glucosídico (χ)

Es la que marca la posición del azúcar respecto a la base nitrogenada. Se han asignado dos rangos para este ángulo: *syn* y *anti*^{15,16} (figura 3.9).

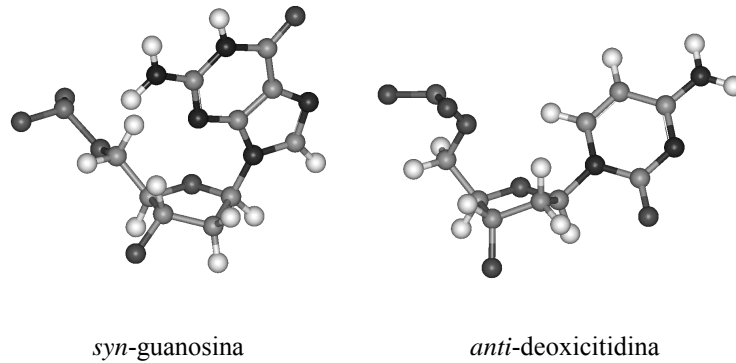


Figura 3.9 Conformaciones *syn* y *anti* de dos nucleósidos.

Las conformaciones *syn* son aquellas en las que la ribosa está eclipsando la base y las *anti* son aquellas en las que el azúcar está en posición opuesta a la base (figura 3.10). Las posiciones intermedias se conocen como zonas *high syn* y *high anti*.

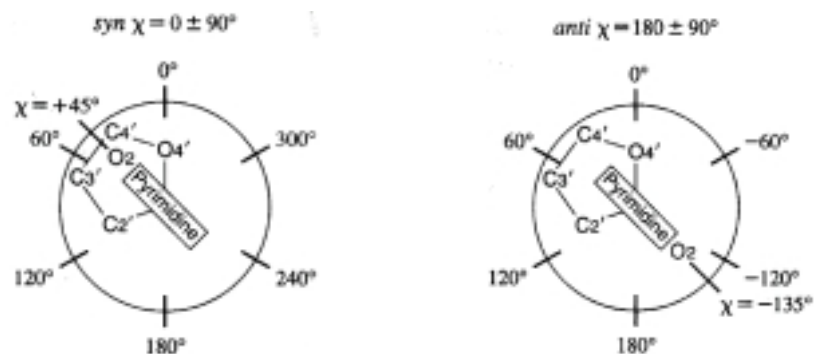


Figura 3.10 Definición de los rangos conformacionales *syn* y *anti* para un nucleótido pirimidínico. (Fuente: véase referencia 5, p. 24)

Las conformaciones *anti* presentan un ángulo de torsión glucosídico en torno a 210° y las *syn* entorno a 49° .

La conformación más estable es la *anti*¹⁷ (la única base con una población notable de conformaciones *syn*¹⁸ es la guanosina), al presentar una menor repulsión estérica entre la base y la ribosa. No obstante, algunos factores pueden influir en el equilibrio conformacional como la posible formación de puentes de hidrógeno (fuerza estabilizadora externa) entre bases y el azúcar, la existencia de un grupo voluminoso en posición 8, etc. En solución se puede dar una rápida interconversión *syn/anti*, ya que la diferencia energética entre *syn* y *anti* es pequeña (alrededor de 1-2 kcal/mol)¹⁹.

iii) Rotaciones respecto al enlace exocíclico C4'-C5' (γ)

El enlace C4'-C5' se encuentra mayoritariamente en las 3 conformaciones alternadas. En una proyección de Newman -se mira la posición del O5' respecto al O4'- si el O5' está entre O4' y C3' es *gauche gauche* (gg), si está en *trans* al C3' y al lado del O4' es *gauche trans* (gt) y si está en *trans* al O4' y al lado de C3' es *trans gauche* (tg) (figura 3.11).

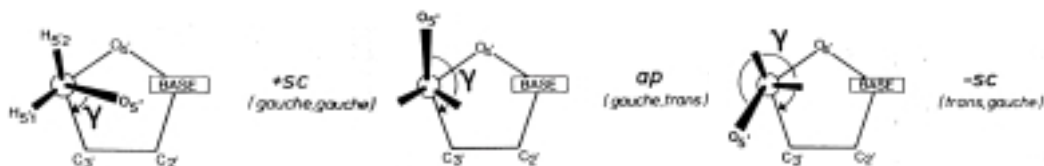


Figura 3.11 Definición de los rangos del ángulo diedro del enlace C4'-C5'. (Fuente: véase referencia 1, p.23)

En general la conformación *gauche gauche* es la más abundante tanto en purinas como en pirimidinas²⁰. Generalmente, ésta se encuentra estabilizada por puentes de hidrógeno entre la base nitrogenada y el O5'.

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

Las rotaciones respecto a los enlaces C-O y P-O están restringidas a determinados rangos, es decir, que no están libres de adoptar cualquier valor entre 0° y 360° . A pesar de esta limitación, le confieren un alto grado de flexibilidad a los ácidos nucleicos.

El enlace $C5'-O5'$ se define por el ángulo diedro β y suelen estar en el rango $180^\circ \pm 40^\circ$, mientras que el enlace $C3'-O3'$ se define por ϵ y suele estar alrededor de 220° (en el rango $180^\circ-270^\circ$)²¹.

El grupo fosfato presenta una configuración tetraédrica con distancias de enlace P-O variables ($P-O^-$; $P-O_{\text{éster}}$) debido a la diferente contribución de los enlaces π y con unos ángulos O-P-O que oscilan alrededor de $105-120^\circ$. Respecto a los enlaces P-O son dos los diedros que los definen: $\zeta(PO3')$ y $\alpha(PO5')$. El ángulo diedro $\zeta(PO3')$ varía alrededor de dos regiones concretas (50° ó 330°) y a su vez, el ángulo diedro $\alpha(PO5')$ también se mueve cerca de dos zonas determinadas (80° ó 350°)²².

3.1.3 Propiedades fisicoquímicas y reactividad

En los ácidos nucleicos se puede hablar de dos comportamientos diferenciados. El azúcar y los fosfatos tienen un papel básicamente estructural, mientras que las bases tienen un papel más funcional al modular específicamente las interacciones en los ácidos nucleicos. Este apartado se centra en las propiedades fisicoquímicas y en la reactividad de las bases.

- Propiedades fisicoquímicas

i) Ionización

El comportamiento ácido-base de los nucleótidos es una de las características más importantes. Éste determina su carga, su estructura tautomérica y su habilidad de actuar

como dador/aceptor de puentes de hidrógeno.

En el rango fisiológico de pH ($5 < \text{pH} < 9$), las bases nitrogenadas son neutras²³. Lo mismo ocurre con las pentosas, que sólo pueden perder el protón del grupo hidroxilo en medios extremadamente básicos ($\text{pH} > 12$)²⁴. El grupo fosfato, en cambio, se encuentra con carga negativa (carga -1). Cuando el grupo fosfato se une a dos nucleósidos, éste sólo tiene un único hidroxilo ionizable con un $\text{pK}_a \approx 1$ y tan solo cuando se trata de un monoéster, puede perder un segundo protón a pH 6-7²⁵. Así que a pH fisiológico, los ácidos nucleicos son polianiones.

ii) Tautomería

Un fenómeno que afecta a las bases es la tautomería y resulta ser una de las claves para entender la funcionalidad de los ácidos nucleicos.

El equilibrio tautomérico prototrópico²⁶ implica estructuras alternativas que sólo difieren en la localización de los átomos de hidrógeno. Las bases nitrogenadas de los ácidos nucleicos disponen de un equilibrio ceto-enol y de un equilibrio amino-imino (figura 3.12).

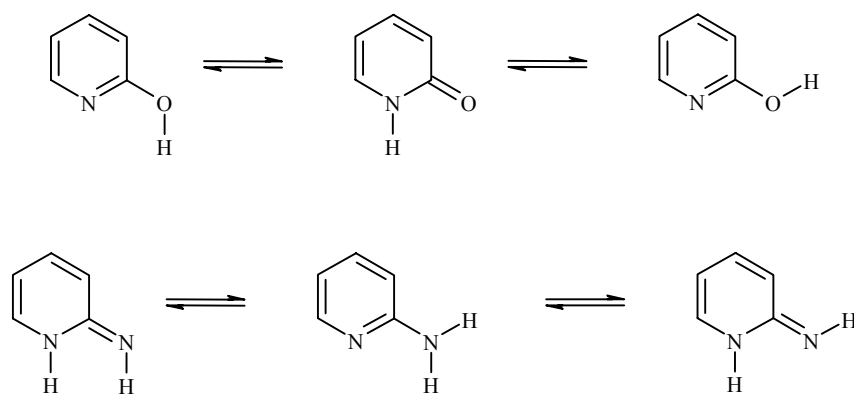


Figura 3.12 Modelo de la tautomería ceto-enol para la 2-piridona: 2-hidroxipiridina y de la tautomería amino-imino para 2-aminopiridina.

Las bases más comunes existen mayoritariamente en la forma tautomérica ceto-amino²⁷. Se ha establecido, mediante técnicas espectroscópicas²⁸, que estos tautómeros son más estables que los tautómeros enol-imino.

Cabe destacar, que muchas mutaciones²⁹ espontáneas tienen su origen en cambios temporales de la preferencia ceto-amino, ya que el cambio tautomérico altera totalmente el patrón de reconocimiento invirtiendo cualquier tipo de afinidad de las bases, afectando la reactividad y la funcionalidad de las mismas.

iii) Distribución de carga

Las bases nitrogenadas son estructuras polares que presentan una distribución de carga muy localizada en determinados átomos³⁰, presentando momentos dipolares elevados³¹. Esta distribución de carga justifica las interacciones no covalentes entre bases, tan importantes en el reconocimiento molecular, como son los puentes de hidrógeno. Concretamente, los grupos $-NH$ de las bases nitrogenadas -donde se concentra la carga positiva- son buenos dadores de puentes de hidrógeno y los pares libres del oxígeno de los grupos $-C=O$ y de los nitrógenos del anillo -donde se concentra la carga negativa- son buenos aceptores de puentes de hidrógenos (figura 3.13).

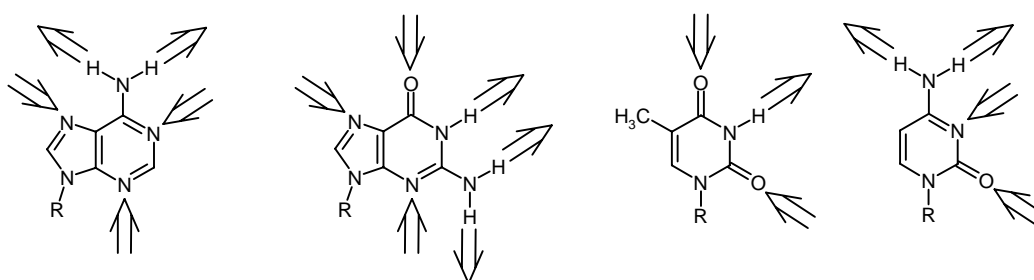


Figura 3.13 Patrón de centros dadores/aceptores de puentes de hidrógeno en las bases nitrogenadas. Los centros dadores se muestran con una flecha apuntando hacia fuera; los centros aceptores se muestran con una flecha apuntando hacia dentro.

Por otro lado, los nucleósidos y nucleótidos también aportan más oportunidad de formar puentes de hidrógeno a través del oxígeno de grupos hidroxilo y fosfato. Los puentes de hidrógeno son los responsables de la mayoría de interacciones de reconocimiento del ADN.

- Reactividad

- i) Protonación

Las reacciones de protonación/desprotonación en las bases nitrogenadas son las que tienen mayor transcendencia funcional. Primero porque este tipo de reacciones es básico en numerosas reacciones enzimáticas. Por otro lado, esta protonación tiene un papel importante para explicar determinados tipos de estructuras de ácidos nucleicos, como la triple hélice. Además, se piensa que este tipo de reacción puede estar detrás de determinados procesos mutagénicos.

- ii) Hidratación

Las bases tienden a agregarse en agua mediante interacciones de apilamiento, como se verá más adelante. Pero en baja concentración presentan una muy buena capacidad de hidratación. Por otra parte, la reacción de hidratación es importante en la vida celular, ya que a menudo resulta una de las primeras etapas del proceso de degradación de nucleósidos no deseados.

- iii) Sustitución nucleófila

El ataque de grupos nucleófilos, generalmente con carga negativa, a las bases puede dar lugar a dos reacciones mayoritarias que compiten: la abstracción de un protón (normalmente unido a un átomo de N) o la sustitución nucleófila. La primera es una reacción que tiene lugar en el plano molecular, mientras que la segunda es una reacción donde el nucleófilo ataca en el plano perpendicular a la molécula.

Las modificaciones químicas de las bases nitrogenadas o del esqueleto de los ácidos nucleicos pueden influir cambiando drásticamente las propiedades fisicoquímicas y la reactividad aquí descritas.

3.1.4 Derivados sintéticos de ácidos nucleicos

Existe una gran cantidad de posibles modificaciones en la estructura de los ácidos nucleicos³². En los últimos 10 años se han llevado a cabo numerosos estudios de estos derivados con finalidades muy diversas. Actualmente, el mayor interés de estas modificaciones radica en las nuevas propiedades que estas confieren a la estructura de los ácidos nucleicos y a sus futuras aplicaciones en el campo de la biotecnología y la biomedicina^{33,34}.

Por todo ello se han desarrollado distintas estrategias³⁵ que permiten obtener ácidos nucleicos sintéticos en los que se mejoran las interacciones no covalentes que los estabilizan, las propiedades termodinámicas y/o cinéticas, las propiedades de reconocimiento como la afinidad y/o selectividad, etc.

3.2 Interacciones entre bases^{1-5,36}

De todas las interacciones que afectan a los nucleótidos, las de asociación entre bases son las de mayor importancia, tanto funcional como estructuralmente³⁷. Las bases efectúan dos interacciones básicas: las de formación de puentes de hidrógeno en el plano de las bases (horizontal) y las de apilamiento (*stacking*) perpendiculares al plano de las bases.

3.2.1 Interacciones por puente de hidrógeno

Los puentes de hidrógeno son unas interacciones principalmente electrostáticas que se efectúan entre un grupo electronegativo poseedor de pares libres (aceptor) y otro que tiene unido un hidrógeno (dador). El sistema de tres átomos X-H...Y suele ser lineal con una desviación no superior a los 20° y la distancia entre heteroátomos está establecida entre 2.6 y 3.5 Å.

El reconocimiento entre bases se da por enlaces de hidrógeno. *A priori* existen muchos patrones de formación de puentes de hidrógeno entre bases³⁸, pero en el ADN biológico el modo más común de reconocimiento es el propuesto por Watson y Crick (WC)^{39,40} (figura 3.14).

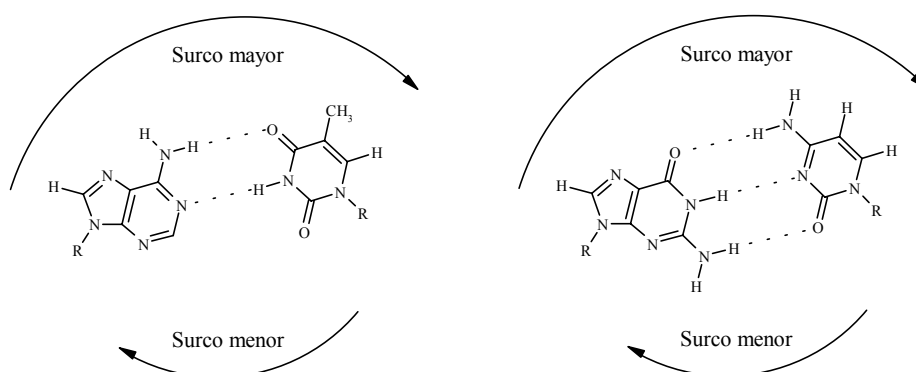


Figura 3.14 Pares de bases Watson-Crick encontrados en las dobles hélices de ADN y de ARN. Se muestra el surco menor (m) y el surco mayor (M) de la hélice.

Las características del apareamiento canónico WC⁴¹ son: (i) la preferencia por la coplanaridad de las dos bases, (ii) la formación de 3 puentes de hidrógeno para el apareamiento C·G y la formación de 2 puentes de hidrógeno para el apareamiento T(U)·A, y (iii) la distancia del par de bases G·C y A·T es prácticamente la misma y el ángulo de inclinación de los 2 pares de bases muy similar (geometría isomorfa). Esto permite la formación de una doble hélice regular.

Existen otros modos de reconocimiento de las bases que no son el propuesto por Watson y Crick. De hecho, es posible encontrar 31 posibles modos de aparear bases neutras con alta eficiencia^{15,42,43}, como se muestran en la figura 3.15 (véase la página siguiente). Resulta interesante ver que son posibles los apareamientos purina:purina y pirimidina:pirimidina, no solo purina:pirimidina. Si además se incluyen las posibilidades que proporciona el cambio en el estado de ionización, las formas tautoméricas minoritarias y los apareamientos con un solo puente de hidrógeno las posibilidades de dimerización son enormes.

Un modo de unión importante, que se ha detectado experimentalmente en ciertos tipos de ADN, es el apareamiento Hoogsteen (H)⁴⁴. Este tipo de apareamiento sólo se puede formar entre bases A·T en condiciones de electroneutralidad, pero si la citosina se protona se puede formar también en pares (G·C)⁺⁴⁵ (figura 3.16).

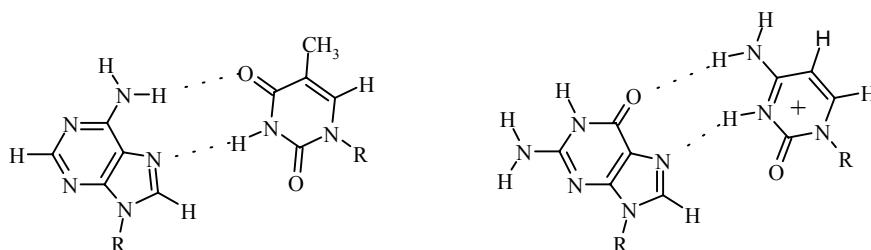


Figura 3.16 Apareamientos tipo Hoogsteen entre A·T y G·C protonada.

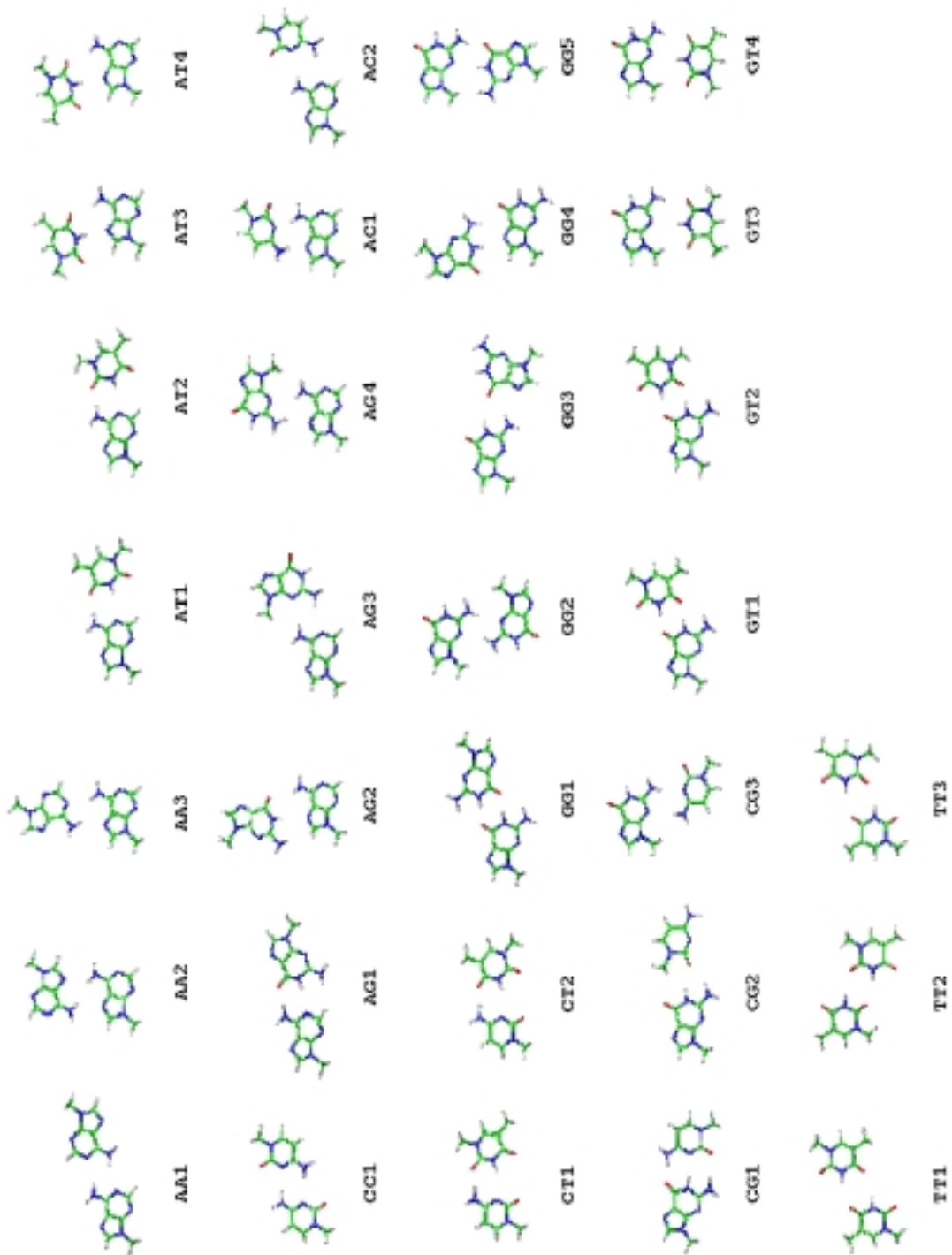


Figura 3.15 Pares de bases que pueden formarse a partir de las bases clásicas neutras en la forma tautomérica más estable y con al menos dos puentes de hidrógeno. En el esquema el par AT1 corresponde a A·T (WC), el par AT2 a A·T (rWC), el par AT3 a A·T (H), el par AT3 a A·T (rH), el par CG1 a G·C (WC) y el par CG2 a G·C (rWC). (Fuente: véase referencia 43, p.221)

Los pares Hoogsteen no son isomorfos a los pares Watson-Crick, ya que tienen un ángulo de 80° entre los enlaces glucosídicos (vs. $68 \pm 2^\circ$ del par WC) y una separación de 8.6 \AA entre los carbonos anoméricos (vs. $10.6 \pm 0.15 \text{ \AA}$ del par WC). También se han detectado otros modos de unión como los apareamientos Watson-Crick reverso (rWC)⁴⁶ y Hoogsteen reverso (rH) (figura 3.17), donde una de las bases del par está rotada 180° respecto a la otra.

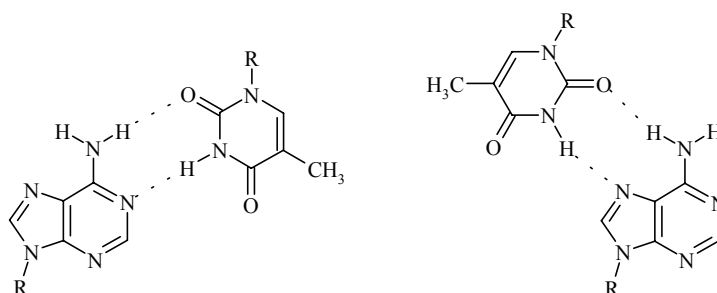


Figura 3.17 Apareamientos entre A·T de tipo rWC y de tipo rH.

Otro modo de unión detectado es el apareamiento denominado ‘wobble’⁴⁷ (figura 3.18). En este apareamiento hay un desplazamiento de una de las bases con respecto a la otra, en relación con la posición de la geometría WC. Como resultado de este desplazamiento, generalmente se pierde uno de los puentes de hidrógeno.

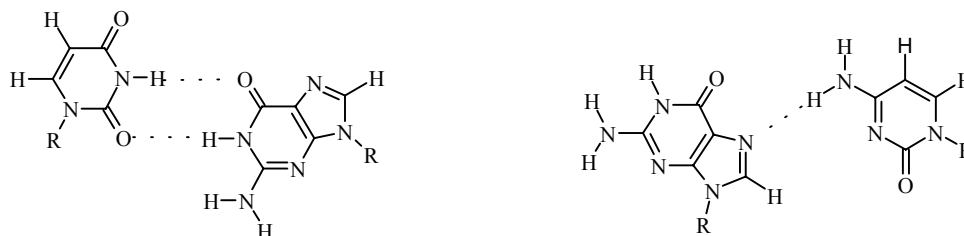


Figura 3.18 Apareamiento tipo ‘wobble’ entre U·G y G·C (Hoogsteen).

Estos tipos de apareamientos que permiten el reconocimiento no canónico de las bases, originan degeneraciones en el código genético, lo que puede conducir a mutaciones en el ADN.

Resulta difícil experimentalmente determinar la contribución de los puentes de hidrógeno a la estabilidad de los ácidos nucleicos, porque es difícil separar estos apareamientos de las interacciones por apilamiento (*stacking*). Esto hace que las estimaciones de la entalpía de dimerización de las bases por puente de hidrógeno se realicen con nucleósidos o nucleótidos aislados. Sin embargo, en estos estudios no se puede asegurar que las bases sólo se aparean de la forma canónica WC, así que se deben considerar todos los demás apareamientos potencialmente en equilibrio.

Se han realizado estudios experimentales en fase gas mediante espectrometría de masas⁴⁸ o en solventes orgánicos, donde se espera que las interacciones de *stacking* sean poco significativas, mediante técnicas de RMN o de espectroscopia de IR^{49,50}. De estos estudios sólo se obtiene una información cualitativa, pues la precisión de los resultados no está clara. Así, que la mayoría de información disponible se obtiene de cálculos teóricos^{51,52} que han permitido determinar con buena precisión todos los posibles modos de reconocimiento. De estos estudios se establece que el par G·C(WC) es el más favorable, que el par A·T(H) es ligeramente más estable que el A·T(WC) y que algunos apareamientos erróneos, como el G·G o el G·A, son sorprendentemente estables.

3.2.2 Interacciones por apilamiento (*Stacking*)

Las bases también interaccionan entre ellas por interacciones de apilamiento o *stacking* (figura 3.19). En este tipo de interacciones dos bases se disponen aproximadamente paralelas una sobre la otra a una distancia cercana a 3.4 Å, evitando siempre contactos electrostáticos no deseados^{53,54}. La interacción de *stacking* es básicamente una interacción de van der Waals relacionada con el solapamiento de orbitales π ocupados y virtuales de los anillos aromáticos. Actualmente, se considera la interacción más importante para mantener la estructura del ADN.

La termodinámica de las interacciones de *stacking* ha sido determinada experimentalmente en medios acuosos, donde las interacciones por puente de hidrógeno entre bases son muy débiles o inexistentes. Se han empleado técnicas muy diversas⁵⁵⁻⁵⁸ como la osmometría, la ultracentrifugación, la calorimetría, la desnaturalización térmica y la RMN. Los estudios teóricos^{52,59-62} confirman los resultados obtenidos por RMN. Los datos disponibles muestran que aquellos pares de bases que involucran G suelen ser los más estables, mientras los que involucran T son los menos estables. La estabilidad del apilamiento es en general la siguiente: purina:purina > pirimidina:purina > pirimidina:pirimidina.

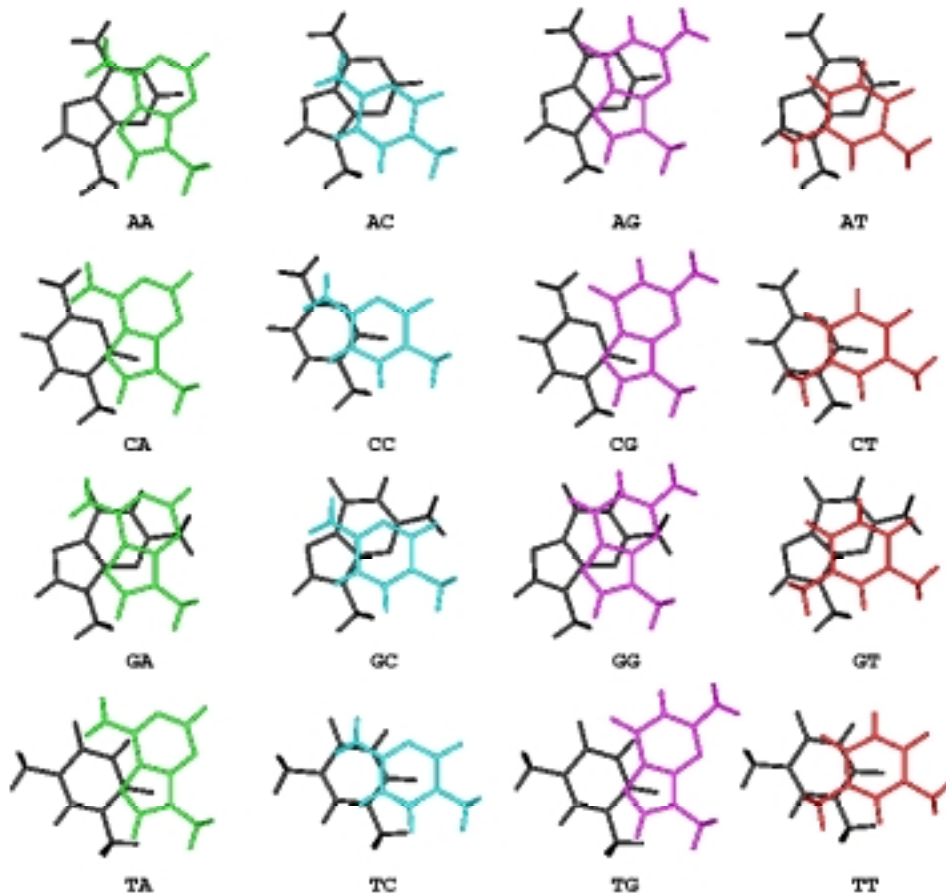


Figura 3.19 Posibles apareamientos por apilamiento entre las bases canónicas en una estructura B-ADN estándar.

En general, en fase gas la energética del proceso de apilamiento es favorable, pero no tanto como lo es la formación de puentes de hidrógeno. No obstante, en medios acuosos la situación se invierte haciendo que el *stacking* sea mucho más importante que la formación de puentes de hidrógeno para la estabilización de la doble hélice de ADN. (Véase el apartado siguiente 3.2.3)⁶³

3.2.3 Dinámica de interacciones entre bases^{43,64}

Experimentalmente se ha demostrado que las bases tienden a agregarse en solución acuosa. Este hecho indica que la interacción entre bases en agua es favorable termodinámicamente. Se conoce que la interacción de bases libres por puentes de hidrógeno en agua es desfavorable, porque la formación de puentes de hidrógeno entre las bases compite con la formación de puentes de hidrógeno con las moléculas de agua. Es decir, al formarse los puentes de hidrógeno entre bases se pierden muchas interacciones de puente de hidrógeno con el agua que son las que estabilizan a las bases en solución. En cambio, el apilamiento no cambia excesivamente por la presencia del agua, ya que las partes de las bases que se solapan son las más apolares.

Asimismo, la flexibilidad del ADN tiene que buscarse en parte en el papel que juega el agua en las interacciones entre bases⁶⁵. Los estudios demuestran que las interacciones entre bases son extremadamente fuertes en fase gas. Este hecho implicaría que la estructura del ADN sería muy estable pero excesivamente rígida, y por tanto incapaz de realizar eficazmente su función biológica.

3.3 Estructura secundaria del ADN^{1-5,36}

Las estructuras secundarias del ADN son todas helicoidales. Una simple cadena de ácido nucleico puede formar una hélice sin ningún apareamiento con otras bases si las bases del polinucleótido son capaces de efectuar un buen apilamiento. Son posibles hélices intra- o intercatenarias de 2, 3 y hasta 4 cadenas estabilizadas por puentes de hidrógeno e interacciones de apilamiento.

En un dúplex de ADN canónico (figura 3.20), G se aparea con C y A se aparea con T (pares de bases Watson-Crick³⁹). La secuencia de una de las cadenas define automáticamente la secuencia de la cadena complementaria, que se dispone de manera antiparalela a la primera cadena.

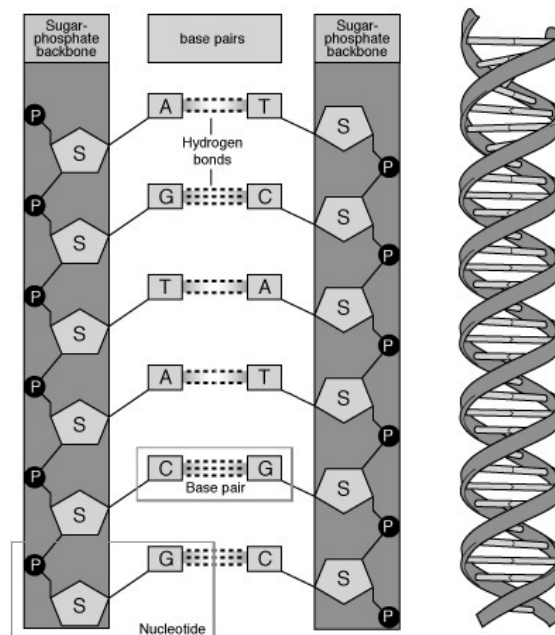


Figura 3.20 Las cuatro bases nitrogenadas del ADN se disponen a lo largo del esqueleto de azúcar-fosfato en un determinado orden: la secuencia de ADN, codificando todas las instrucciones genéticas para el organismo. Ambas cadenas de ADN se mantienen unidas por enlaces débiles entre bases. (Fuente: véase <http://www.accessexcellence.org/AB/GG/dna2.html>)

3.3.1 Parámetros conformacionales para describir el ADN

La estructura secundaria del ADN precisa de una nomenclatura especial⁶⁶. Esta nomenclatura distingue entre movimientos de translación y movimientos de rotación (figura 3.21). Dentro de cada uno de ellos diferencia entre movimientos de pares de bases respecto al eje de la hélice, movimientos de un par de bases respecto a su vecino y movimientos de bases por separado (tabla 3.1). Los sistemas de referencia son siempre el eje de la hélice y los surcos ancho y estrecho (*major/minor grooves*). La nomenclatura de los surcos es intrínseca a los pares de bases. Así, el surco mayor no tiene porqué ser el más ancho o el más profundo.

	Movimientos de translación	Movimientos de rotación
Par de bases respecto al eje	Desplazamiento Y (<i>y-displacement</i>) Desplazamiento X (<i>x-displacement</i>)	Ladeo (<i>Tip</i>) Inclinación (<i>Inclination</i>)
Par de bases respecto a otro par	Ascenso (<i>Rise</i>) Deslizamiento (<i>Slide</i>) Corrimiento (<i>Shift</i>)	Giro (<i>Twist</i>) Tambaleo (<i>Roll</i>) Ángulo basculante (<i>Tilt</i>)
Una base respecto a otra	Sesgo (<i>Stagger</i>) Estiramiento (<i>Stretch</i>) Cizalladura (<i>Shear</i>)	Apertura (<i>Opening</i>) Alabeo (<i>Propeller twist</i>) Plegamiento (<i>Buckle</i>)

Tabla 3.1 Parámetros locales de hélice.

3.3.2 Estructuras canónicas del ADN

Los primeros estudios de difracción de fibras de ADN parecían indicar que éste existía en unas formas fijas. Posteriormente, estudios de difracción de rayos X en cristales, estudios de RMN y especialmente estudios de dinámica molecular⁶⁷ (MD) han permitido concluir que esto no es así, sino que el ADN se encuentra en un amplio abanico de familias conformacionales.

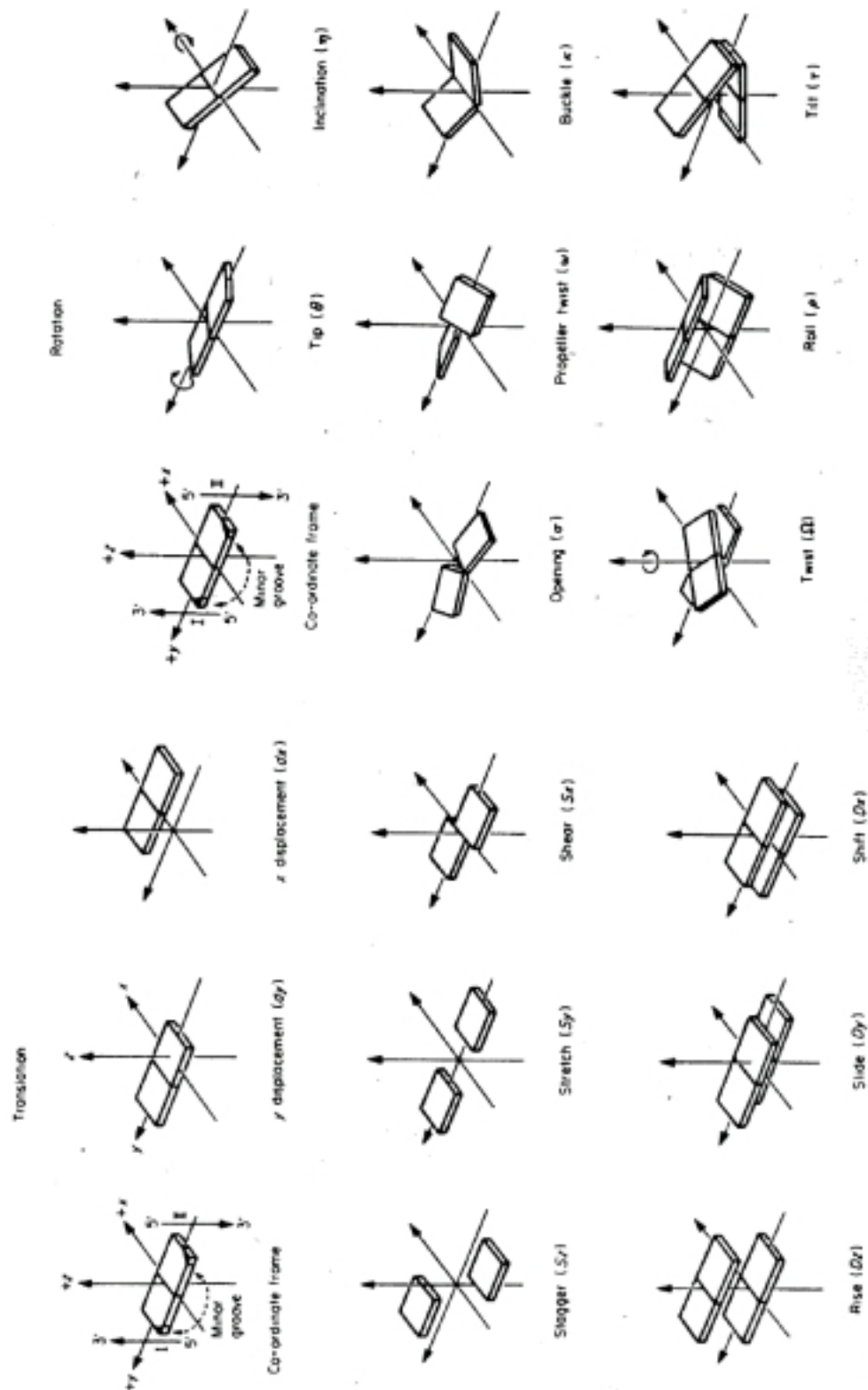


Figura 3.21 Definición de los parámetros conformacionales de translación y de rotación para describir el ADN. (Fuente: véase referencia 66, p. 789-790)

El ADN tiene una elevada flexibilidad conformacional⁶⁸. Por lo tanto, la estructura secundaria que conocemos es una información promedio. El ADN se mueve mucho con movimientos locales de las bases en la escala del picosegundo y movimientos globales de doblez (*bending*) en la escala del nanosegundo. Es decir, el ADN parece ser una estructura estable con una gran facilidad de cambiar de conformación. Estos cambios pueden verse afectados por una serie de factores externos como la secuencia, la fuerza iónica, la humedad relativa, las modificaciones de las bases o la presencia de fármacos.

Tradicionalmente se clasifica el ADN doble hélice en dos grupos, tres familias y 6 tipos. Los grupos son: dextrógiro y levógiro. Las familias son A, B (dextrógiras) y Z (levógira)⁶⁹. Todas las familias tienen un único tipo, a excepción de la B que cuenta con 4 tipos: B, C, D y T.

- Formas dextrógiras

Se caracterizan en que la doble hélice gira en sentido dextrógiro. El aspecto de un ADN de la familia B y uno de la familia A es totalmente diferente (figura 3.22). Las principales características⁷⁰ de estas familias se resumen en la tabla 3.2.

El tipo A-ADN da lugar a una hélice más ancha al contar con más residuos por vuelta de hélice. Esto mismo se refleja en el *twist*, que pasa a ser de 36-38° en la familia B a ser de 33° en la familia A, y en el *rise*, que demuestra que las bases en la familia A se encuentran más apiladas. Otra característica diferencial es el *puckering* de los azúcares, que en la familia B es C2'-*endo* (C3'-*exo*), mientras que en la familia A es C3'-*endo*. También son destacables las características de los surcos. En la familia B el surco mayor (M, *major groove*) es más ancho y un poco más profundo que el surco estrecho (m, *minor groove*). En cambio en la familia A, el surco ancho es mucho más profundo que el surco menor, pero también más estrecho. La familia B se distingue por los pares de bases perpendiculares al eje de la hélice, la forma estilizada de la hélice y por lo compacta que es ésta en su interior. Por el contrario, el A-ADN aparece mucho menos estilizado, siendo visible un hueco existente en el interior de la hélice y la inclinación de los pares de bases con respecto al eje de la hélice. La familia B es mucho

más flexible que la familia A y la conformación concreta depende mucho de la secuencia, al contrario del A-ADN que parece ser una conformación más rígida.

Familia B	Familia A
8-10 nucleótidos por vuelta	11 nucleótidos por vuelta
Bases perpendiculares al eje de la hélice con poca inclinación	Bases inclinadas 20° con respecto al eje de la hélice
Surco mayor más ancho y un poco más profundo que el surco menor	Surco mayor mas estrecho y más profundo que el surco menor
<i>Rise</i> de 3 a 3.3Å y <i>Twist</i> de 33°	<i>Rise</i> de 2.6 Å y <i>Twist</i> de 36-38°
Hélice compacta en su interior	Existe un hueco en el interior de la hélice; <i>x-displacement</i> 4.7 Å hacia el surco mayor
<i>Puckering</i> del azúcar en zona S (C2'-endo/ C3'-exo)	<i>Puckering</i> del azúcar en zona N (C3'-endo)
Conformación flexible y dependiente de secuencia	Conformación rígida

Tabla 3.2 Características de las formas dextrógiras.

i) Familia B^{71,72}

La familia B es la forma más habitual del ADN. Esta familia incluye los ADNs de los tipos B, C, D y T.

Las características prototípicas de la familia B las presenta el tipo B-ADN^{70,71} que es el más habitual *in vivo* y el que se considera biológicamente activo. A parte del B-ADN hay otros tipos de ADN dentro de la familia B, todos ellos muy similares. Las características diferenciales son que el C-ADN es ligeramente más estilizado y con surcos más estrechos y que las formas D y T del ADN presentan un surco menor profundo pero muy estrecho. Estas formas aparecen en ADN sintéticos, en ADN

naturales en diversas condiciones de humedad relativa (C-ADN) y en determinados fagos (T-ADN).

ii) Familia A⁷³

La familia A cuenta con un solo tipo: el A-ADN. Es la segunda forma más detectada de ADN, y ha sido determinada tanto en cristales como en fibras. Presenta secuencias ricas en C·G en condiciones de baja humedad.

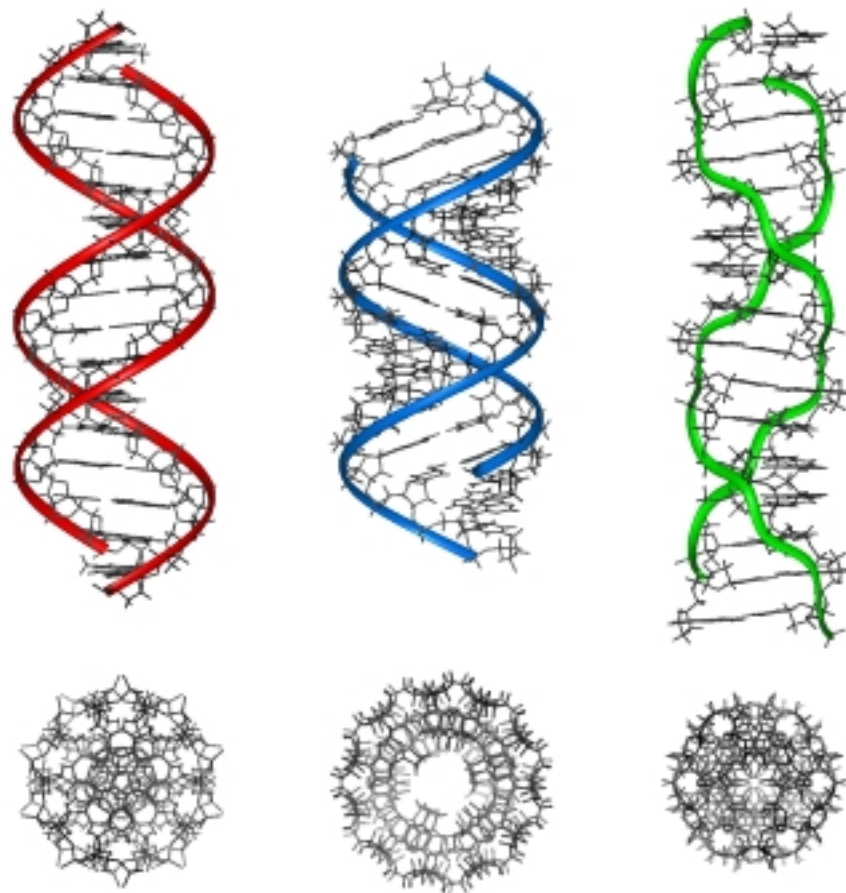


Figura 3.22 Modelos de doble hélice del ácido desoxirribonucleico (ADN). Se observa claramente el surco mayor (M) y el surco menor (m) formados por el esqueleto de azúcares unidos a través de fosfatos. Las bases nitrogenadas aparecen en el interior de la hélice. El B-ADN se muestra en color rojo, el A-ADN en color azul y el Z-ADN en color verde.

- Forma Levógira⁷⁴⁻⁷⁶

La característica más relevante de este ADN es que la doble hélice gira en sentido levógiro. Presenta una estructura más larga y delgada que las anteriores. Su aspecto es inconfundible por la forma de zig-zag del esqueleto de fosfatos (figura 3.22). Las características principales se muestran en la tabla 3.3.

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

Tabla 3.3 Características de la forma levógira.

El Z-ADN se ha detectado *in vivo* en secuencias purina-pirimidina alternantes G, C o en secuencias en las que alguna base se ha modificado por metilación o bromación, especialmente citosinas.

La forma Z del ADN se puede interconvertir espontáneamente en formas B-ADN en presencia de fármacos o cambios del medio. Por ejemplo, un aumento de la fuerza iónica produce un cambio reversible de B a Z-ADN.

- Estructuras irregulares del ADN^{77,78}

El ADN real en solución es una larguísima fibra que adopta conformaciones macroscópicas diferentes a las formas canónicas descritas anteriormente.

Estudios de RMN, de dinámica molecular (MD) y los cristales han demostrado que dependiendo de la secuencia la conformación del ADN puede variar mucho, es decir, cada secuencia de bases imprime sus propios rasgos en la hélice de ADN. Las secuencias tienen influencia en la amplitud de los surcos, en el *twist* de la hélice, en la curvatura, en la rigidez mecánica y en la resistencia a doblarse (*bending*). Por ejemplo, las regiones G·C y regiones mixtas tienen el surco menor más ancho que las regiones A·T, tanto homo- como heteropoliméricas.

Las irregularidades locales a nivel de par de bases de la estructura normal del ADN que pueden observarse son los movimientos de *tip*, *buckle* y *propeller twist*. Otra irregularidad es la doblez (*bending*) de ADN. Actualmente se conoce la tendencia de fragmentos A·T a formar *bending* y se ha determinado que éste existe en zonas de unión entre A y B-ADN. Por último, los apareamientos incorrectos también generan irregularidades locales. (Véase el apartado 3.3.3)

Las distorsiones más generales de la estructura del ADN son:

- (i) ADN curvado. Éste se presenta en zonas con sucesivas dobleces, generalmente ricas en poli(A·T).
- (ii) Estructuras de ADN *slipped* (lazos). Éstas se encuentran en zonas de *direct repeat*. Su estructura descrita es consistente con el patrón de escisión de las nucleasas de una única cadena, pero todavía no ha sido bien caracterizada.
- (iii) ADN anisomórfico. Son las conformaciones asociadas con el *direct repeat* que presentan propiedades químicas y físicas inusuales. Las dos cadenas complementarias tienen diferentes estructuras, pero se desconoce como son.
- (iv) ADN cruciforme. Se encuentra en zonas de *inverse repeat*, especialmente palíndromos. Da lugar a formas muy raras que se asemejan a cruces.

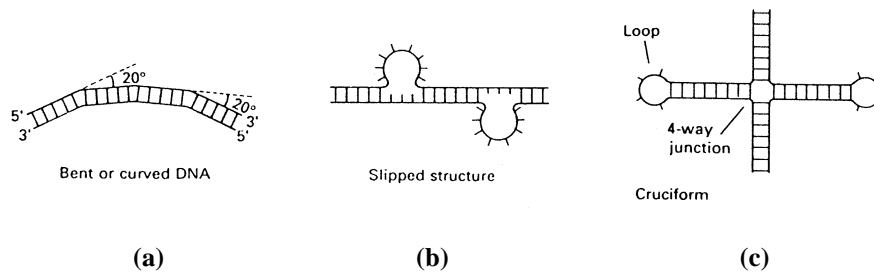


Figura 3.23 Algunas de las estructuras inusuales del ADN. (a) ADN curvado, mostrando *bendings* de 20° , (b) ADN *slipped*, presenta recodos en cadenas alternadas para compensar y (c) ADN cruciforme. (Fuente: véase referencia 2, p. 45 y 47)

3.3.3 Estructura no canónicas del ADN

En los últimos años se ha incrementado la evidencia que los ácidos nucleicos poseen una gran variabilidad conformacional mucho más allá de la doble hélice antiparalela de tipo B propuesta por Watson y Crick y ampliamente analizada desde entonces.

Un número de estructuras alternativas de ADN han sido descritas en la literatura, en las cuales se observan apareamientos incorrectos, interacciones entre bases no estándar y estructuras de 2, 3 y 4 cadenas, y cambios en la polaridad estándar de las cadenas.

- Dúplexes con apareamientos incorrectos^{29,79}

La fidelidad en la transmisión del código genético recae en la especificidad de los reconocimientos A·T y G·C. Se dan ciertas circunstancias donde aparecen errores en el reconocimiento entre bases (*mismatching*). Se puede hablar de dos tipos apareamientos erróneos entre bases normales: *transition*, donde una purina se aparea con una pirimidina errónea y *transversion*, donde los apareamientos se dan entre purina:purina o pirimidina:pirimidina.

Otro tipo de apareamiento incorrecto resulta de la presencia de una base extra en una de las cadenas (*insertion-deletion mispair*). En este caso, el residuo extra puede acomodarse entre las cadenas, en el interior de la hélice, o bien, ser forzada a una posición fuera de la hélice.

Los apareamientos no canónicos⁸⁰, como G·A, G·T, A·A, G·G, T·T, C·C, T·C, o A·C, adoptan una gran variedad de esquemas de interacción por puente de hidrógeno (figura 3.15) y algunos de estos apareamientos son bastante estables. Los cambios estructurales por la formación de un apareamiento erróneo o la presencia de una base extra son muy locales y la conformación global de los dúplexes de la forma A, B y Z se mantiene inalterada.

Estas mutaciones pueden estar ligadas a cambios en el estado de ionización de las bases fruto de variaciones locales de pH, a cambios en las preferencias tautoméricas, a distorsiones por factores estéricos del ADN que favorecen la formación de pares erróneos, o bien a un error de los enzimas encargados de la reparación de ADN que detectan la base incorrecta pero eliminan la base correcta.

- Doble hélice de ADN paralelos

Una característica común en las tres familias A, B y Z-ADN de dúplexes es la orientación antiparalela de las cadenas complementarias fruto del exclusivo apareamiento canónico WC. No obstante, la posibilidad que el ADN forme una estructura de doble hélice paralela ha sido demostrada experimentalmente -absorción UV, RMN, dicroísmo circular- desde hace dos décadas. Este tipo de estructuras, también conocidas como *parallel-stranded DNA*, ps-ADN o H-ADN, se forman bajo condiciones fisiológicas y en ausencia de alteraciones de las bases o del esqueleto dadas las secuencias apropiadas⁸¹. Ésta puede formarse a partir de una única cadena (ps-ADN intramolecular) o a partir de dos cadenas polinucleótidas independientes (ps-ADN intermolecular).

La disposición paralela también puede obtenerse en condiciones de pH bajo, como consecuencia de modificaciones químicas (sustituyentes voluminosos que bloquean los apareamientos WC; uniones fosfotriéster; dúplexes anoméricos α/β ; enlaces 2'-5'), o bien, puede venir mediada por interacciones con ligandos⁸².

Las características estructurales de ps-ADN, originalmente propuestas en un estudio teórico por Pattabiraman⁸³ para una secuencia d(A·T)₆, son un apareamiento Watson-Crick reverso (rWC)⁴⁶ con los enlaces glucosídicos en orientación *trans* y los surcos mayor y menor indistinguibles al tener una anchura similar (figura 3.24). La hélice paralela es dextrógira y con un *twist* comparable al del B-ADN. Desde un punto de vista estructural, cada una de las cadenas que forman el ps-ADN es esencialmente del tipo B-ADN.

La estabilidad de los dúplexes paralelos tipo Watson-Crick reverso decrece en el siguiente orden: secuencias homopoliméricas A·T > secuencias homopoliméricas y alternancias A·T > secuencias alternantes A·T. Se ha visto que el par G·C puede ser incorporado en secuencias A·T, pero éste desestabiliza los dúplexes⁸⁴.

En los últimos años han aparecido nuevas estructuras que demuestran que el rWC no es el único tipo de apareamiento posible para la obtención de ps-ADN. Los apareamientos que definen este tipo de estructuras son muy diversos. Por un lado, tenemos apareamientos simétricos entre homobases como C·C⁺, T·T, A·A, G·G (figura 3.15) y por otro lado, apareamientos entre heterobases de tipo rWC, donde las bases interactúan a través de formas tautoméricas minoritarias, o de tipo 'wobble'. También se ha propuesto el apareamiento de tipo Hoogsteen⁴⁴ (A·T^H y G·C⁺^H, donde la citosina está protonada para establecer dos puentes de hidrógeno con la guanina), para tratar de explicar la transición: hélice paralela a bajo pH → hélice antiparalela imperfecta a pH más elevados. Los datos estructurales disponibles^{85,86} para este tipo de hélices paralelas Hoogsteen (figura 3.24) parecen indicar que son muy similares a una triple hélice de ADN y bastante diferentes de una doble hélice WC. (Véase el siguiente punto)

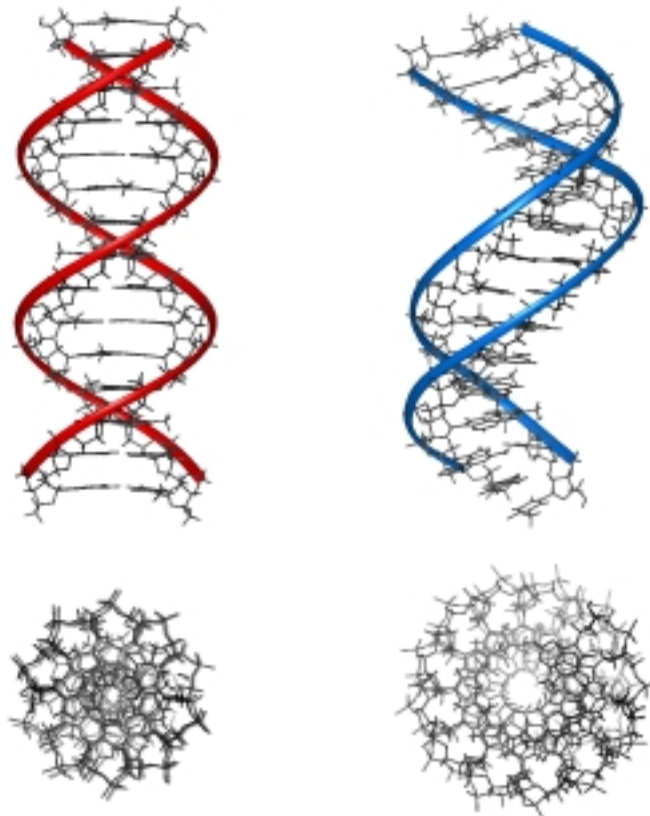


Figura 3.24 Modelos de doble hélice paralela del ácido desoxirribonucleico (ADN). Se observa que no existen diferencias entre el surco mayor y el surco menor de un ps-ADN Watson-Crick reverso (rojo) . El ps-ADN Hoogsteen se muestra en color azul.

El estudio de los componentes estructurales de las cadenas paralelas resulta útil para la comprensión de las hélices de ADN de 3 y 4 cadenas, donde también aparecen estos componentes.

- Triple hélice de ADN¹⁸⁷⁻⁸⁹

La habilidad del ADN para formar triples hélices se conoce desde los años 50-60, pero en los últimos años un gran número de estudios se han centrado de nuevo en el análisis de estas estructuras. El resurgido interés por este tipo de estructuras es debido a

su posible implicación en procesos biológicos y a sus potenciales aplicaciones biotecnológicas⁹⁰.

Una triple hélice se forma cuando una cadena polinucleótida se coloca en el surco mayor de una doble hélice de ADN y sus bases interaccionan por puentes de hidrógeno Hoogsteen con las purinas del apareamiento Watson-Crick (figura 3.25). Ésta puede formarse a partir de una única cadena polimérica (tríplex intramolecular) o a partir de diferentes cadenas polinucleótidas (tríplex intermolecular). El esqueleto de las cadenas puede provenir de ARN o ADN, de la combinación de éstos o de una gran variedad de esqueletos no naturales (p.e. PNA).

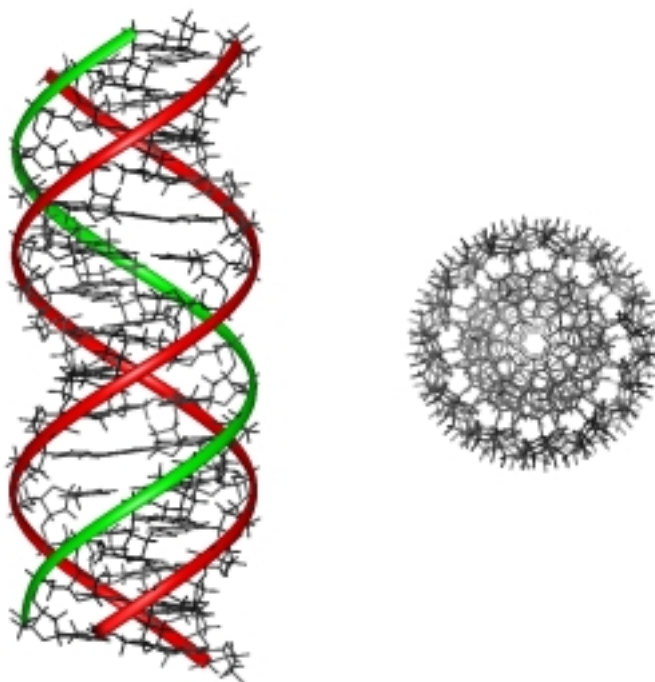


Figura 3.25 Modelo de triple hélice del ácido desoxirribonucleico (ADN). La estructura del ADN triple hélice está constituida por 2 cadenas en estructura WC formando algo similar a la estructura de un dúplex (rojo), pero con una cadena extra que se dispone a lo largo del surco mayor de la doble hélice (verde). La tercera cadena se une a una de las cadenas del dúplex por medio de puentes de hidrógeno de tipo Hoogsteen.

La incorporación de la tercera cadena, también denominada TFO (*triplex forming oligonucleotides*), en el surco mayor de la doble hélice hace que éste se ensanche y se divida en dos surcos asimétricos, denominados el surco menor mayor (mM) y el surco mayor mayor (MM), mientras que el surco menor (m) se mantiene con sólo pequeñas alteraciones (figura 3.25).

Las triples hélices en los ácidos nucleicos se dividen en dos familias dependiendo de la identidad de las bases (purina o pirimidina) de la tercera cadena. Las dos familias de tríplexes se distinguen por la siguiente nomenclatura: $(pyr \cdot pur) \cdot pyr$, en que la tercera cadena es rica en residuos pirimidina y $(pyr \cdot pyr) \cdot pur$, en que la tercera cadena es rica en residuos purina. En ambas familias, el dúplex involucrado en la formación del tríplex debe tener una secuencia homopurina en una de las cadenas.

En la familia $(pyr \cdot pur) \cdot pyr$, la tercera cadena es paralela a la cadena de purinas del dúplex e interacciona con ésta por puentes de hidrógeno Hoogsteen para formar las tríadas canónicas $(C \cdot G) \cdot C^+$ y $(T \cdot A) \cdot T$ que se muestran en la figura 3.26. La formación del triplete requiere la protonación del N3 de la citosina de la tercera cadena. En esta familia la secuencia de la tercera cadena viene determinada por la secuencia del dúplex.

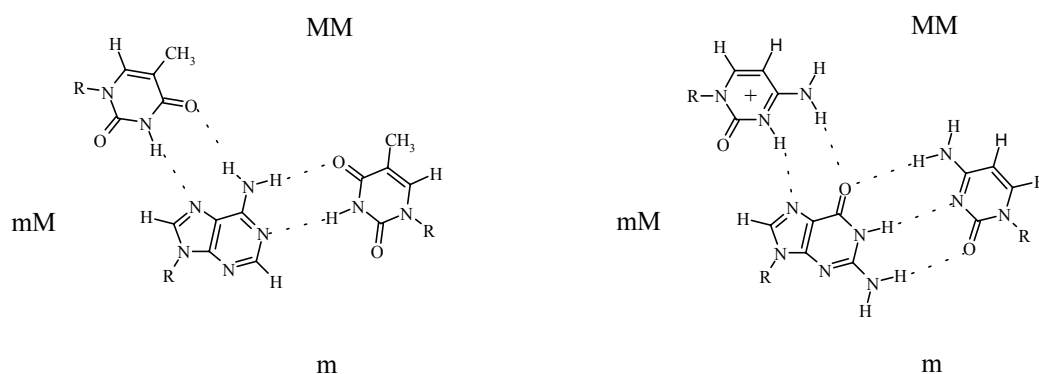


Figura 3.26 Triplete de bases canónicas para la familia $(pyr \cdot pur) \cdot pyr$. El remarcable isomorfismo de ambas tríadas canónicas hace posible la formación de una triple hélice regular. Se muestran los surcos de la hélice: el menor (m), el menor mayor (mM) y el mayor mayor (MM).

En la familia $(pyr \cdot pur) \cdot pur$, la tercera cadena es antiparalela a la cadena de purinas del dúplex e interacciona con ésta por puentes de hidrógeno Hoogsteen reverso para formar las tríadas $(C \cdot G) \cdot G$, $(T \cdot A) \cdot A$ y $(T \cdot A) \cdot T$. La estabilidad de esta familia depende dramáticamente de la presencia de iones bivalentes.

A falta de datos cristal de alta resolución se han sugerido 3 modelos distintos a partir de datos experimentales.

- (i) El primer modelo estructural de una triple hélice se obtuvo por difracción de rayos X de fibras⁹¹, donde se concluyó que la triple hélice adopta una conformación del tipo A-ADN. En esta estructura las ribosas están en una conformación $C3' \text{-endo}$, los pares de bases están inclinados alrededor de 8.5° respecto al eje de la hélice y el par de bases WC muestra un gran desplazamiento respecto al eje de la hélice (-3.2 \AA). El *rise* se encuentra alrededor de 3.3 \AA y el *twist* próximo a 30° , dejando la periodicidad helicoidal en 12 pares de bases por vuelta.
- (ii) Posteriores estudios espectroscópicos de RMN e IR⁹²⁻⁹⁴ han indicado que la triple hélice adopta una estructura de tipo B-ADN. En este modelo las ribosas son $C2' \text{-endo}$, los pares de bases son casi perpendiculares al eje de la hélice, el desplazamiento del par de bases WC es menor que el del modelo de la forma A (-2.0 \AA), el *rise* alrededor de 3.4 \AA y el *twist* entre 27° - 32° , dejando de nuevo la periodicidad helicoidal en 12 pares de bases por vuelta.
- (iii) Por último la determinación de la estructura de una triple hélice PNA·ADN·PNA por difracción de rayos X de alta resolución⁹⁵ ha añadido un nuevo modelo estructural. Esta nueva hélice de tipo P es completamente diferente a las hélices de tipo A y B: esta estructura muestra las ribosas en $C3' \text{-endo}$, un gran desplazamiento del par de bases WC respecto al eje de la hélice, de un *rise* promedio de 3.4 \AA y un *twist* de 23° dejando la periodicidad helicoidal en 15 pares de bases por vuelta.

Por otro lado, estudios de dinámica molecular^{96,97} también sugieren que los tríplexes son más similares a la familia B. Estas estructuras de triple hélice adoptan una configuración tipo B, a pesar de tener sutiles diferencias como que el *twist* sea significativamente menor al de la forma B canónica.

- ADN tétraplex⁹⁸

Es conocido que los oligo- o polinucleótidos ricos en guaninas pueden agregarse juntos y formar una hélice de cuatro cadenas en presencia de un catión monovalente. Estas estructuras se conocen como G-ADN (figura 3.27).

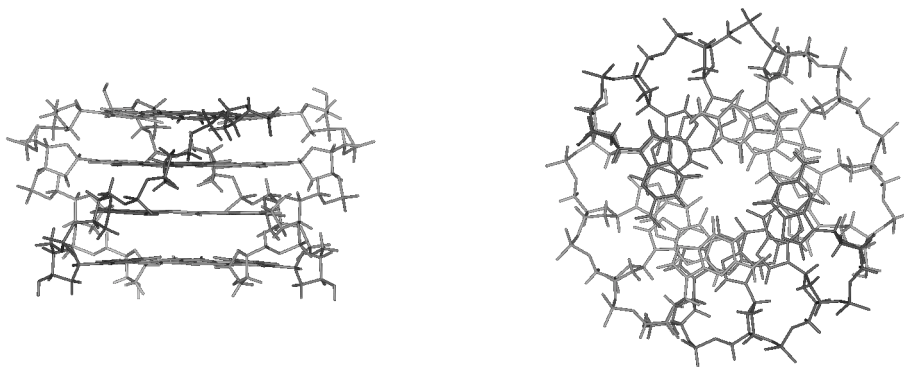


Figura 3.27 Modelo de la hélice de 4 cadenas del ácido desoxirribonucleico (ADN). La estructura del G-ADN es un tétraplex que está estabilizada por un ión Na^+ o K^+ en el interior.

El G-ADN es una estructura inusual de ADN que ocurre en las regiones teloméricas al final de los cromosomas⁹⁹. Durante el proceso de replicación es necesario que exista una porción al final del ADN extra en una de las cadenas (extremo 3') que actúe de cebo para la síntesis de la cadena complementaria. Estas regiones extra son ricas en guaninas y su presencia es muy importante en la replicación. El fragmento extra de poliG se repliega sobre el mismo y las guaninas se reconocen mutuamente por puentes de hidrógeno tipo Hoogsteen (figura 3.28).

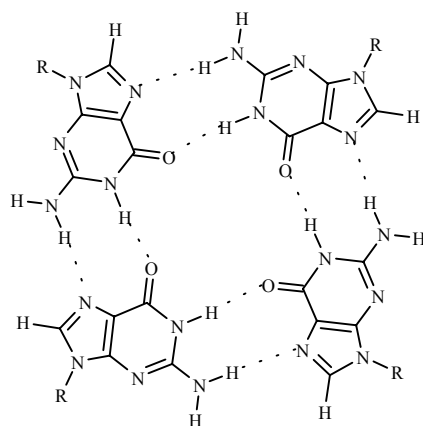


Figura 3.28 Tétrada de guaninas en un *G-quartets*. Las cuatro guaninas se aparean en una tetrada G·G mediante puentes de hidrógeno tipo Hoogsteen.

Las cuatro guaninas, conocidas como *G-quartets*, se encuentran apiladas unas encima de las otras a 3.4 Å de separación^{100,101}. Este tipo de estructura puede formarse a partir de una única cadena (cuádruplex intramolecular), a partir de dos cadenas o a partir de cuatro cadenas polinucleótidas (cuádruplex intermolecular). Generalmente, estas estructuras presentan un patrón alternado *syn/anti* de los ángulos glucosídicos a lo largo de la cadena, pero también es posible encontrar un patrón todo *anti*. La disposición de las cadenas puede ser paralela o antiparalela, dependiendo de la naturaleza de la secuencia que conecta las guaninas o de los cationes que estabilizan la estructura.

3.3.4 Fuerzas estabilizadoras

El ADN es una fibra larguísima con una enorme flexibilidad conformacional y globalmente estable. Las razones de su estabilidad son (i) la disposición conformacional favorable de sus unidades elementales, (ii) la solubilidad proporcionada por los fosfatos, (iii) las interacciones por puente de hidrógeno entre bases que unen las diversas cadenas y (iv) las interacciones por apilamiento entre bases, tanto intra- como intercatenarias.

Además, las interacciones con el agua y con los iones también son vitales para mantener la estructura, la estabilidad y la funcionalidad del ADN.

- El agua^{102,103}

El agua es el componente mayoritario y el solvente de los ácidos nucleicos en todas las situaciones biológicas relevantes. Se ha visto que el porcentaje de agua influye en la estructura que adopta la hélice de ADN, aunque la composición de las bases y la concentración de las sales también interviene a la hora de modular la estructura.

Las características más relevantes del agua son su capacidad para establecer puentes de hidrógeno y su elevado momento dipolar¹⁰⁴. Ambas características favorecen la estabilización de los ácidos nucleicos. El agua interacciona mediante puentes de hidrógeno con determinados grupos de los ácidos nucleicos. Experimentalmente se han determinado zonas de hidratación específicas en los ácidos nucleicos. Concretamente, en este tipo de estructuras se observa que el agua se dispone de manera muy ordenada en torno al surco estrecho del B-ADN paralela a la cadena de fosfatos, dando lugar a la espina de hidratación^{105,106}. La espina de hidratación parece ser un elemento clave para explicar la interacción de muchos fármacos con el ADN y unas de las razones para justificar la estabilidad del B-ADN frente a otros ADNs que no la poseen.

- Iones metálicos

Los ácidos nucleicos son polianiones que interaccionan fuertemente con iones¹⁰⁷, particularmente cationes, en el entorno celular. Estos iones son necesarios para equilibrar la repulsión negativa entre los grupos fosfatos cargados. Se conoce la importancia de la unión de iones metálicos a ácidos nucleicos y a enzimas implicados en los procesos relacionados con la replicación y transcripción del ADN, la traducción de ARN mensajero. También se sabe que la estructura y estabilidad de los ácidos se encuentra modulada significativamente por iones metálicos¹⁰⁸.

Los ácidos nucleicos contienen diferentes sitios de unión con cationes¹⁰⁹: (i) los átomos de oxígeno cargados negativamente del grupo fosfato, (ii) los grupos hidroxilo de las ribosas, (iii) los nitrógenos del anillo de las bases, (iv) los grupos carbonilo exocíclicos de las bases y (v) los anillos aromáticos de las bases¹¹⁰.

Los iones metálicos que interaccionan con los ácidos nucleicos se clasifican en metales alcalinos, metales alcalinotérreos, metales de transición (divalentes) y otros metales. Dependiendo del tipo de catión, éstos pueden unirse específicamente a diversos sitios de unión, o bien, inespecíficamente a cualquiera de los sitios de unión, e incluso algunos iones pueden unirse a más de un sitio simultáneamente (formación de quelatos). La influencia de los cationes es muy específica.

Se ha determinado que los cationes estabilizan tríplexes y otras estructuras no canónicas como G-tétrplex, también estabilizan tautómeros raros y apareamientos erróneos, etc^{111,112}.

3.4 Aspectos dinámicos de la flexibilidad conformacional del ADN^{1-5, 113}

El ADN es una molécula muy flexible capaz de sufrir cambios conformacionales pequeños a nivel de par de bases, como cambios en las características macroscópicas de fibra. Los enlaces fosfodiéster y el *puckering* de la ribosa son los puntos más flexibles del ADN, pero también son posibles cambios más costosos como las rotaciones del enlace glucosídico *syn/anti*.

Entre los movimientos a corta escala hay que destacar la respiración del ADN (*breathing*)¹¹⁴. Este movimiento local y reversible implica la abertura de las bases del ADN hacia el exterior durante un breve periodo de tiempo. Las bases se abren unos 15° cada 10 milisegundos. Esto genera un hueco momentáneo que puede ser aprovechado por moléculas que se intercalan, o bien por proteínas específicas que reconocen bases desplazadas (*flipped bases*).

Entre los movimientos a larga escala se encuentran las transiciones *helix-coil*. Se trata de la disociación total de las hélices en solución para dar cadenas individuales de ADN al aumentar la temperatura. Esta transición depende de la composición de las bases¹¹⁵, de la secuencia, de la longitud de las cadenas, de la concentración de sales, del pH del medio y especialmente de la temperatura. El punto medio de esta transición térmica se conoce como la T_m (*melting temperature*). El proceso inverso es la renaturalización de cadenas complementarias separadas para formar una hélice correctamente apareada. En la práctica, la desnaturalización sólo es reversible en oligómeros relativamente cortos y homopolímeros. En cadenas largas, este proceso intermolecular es muy lento pudiendo competir con la rápida formación de horquillas (*hairpin loops*) intramoleculares.

La curvatura del ADN⁷⁸ es uno de los fenómenos macroscópicos más importantes y explica el empaquetamiento de la fibra de ADN en niveles de estructuración de orden superior.

3.5 Interacción del ADN con otras moléculas¹⁻⁵

Las moléculas pueden interactuar con el ADN a través de interacciones covalentes o no covalentes. Las primeras conducen a una modificación del esqueleto covalente del ADN a través de reacciones de hidrólisis, redox, con nucleófilos, con electrófilos, con metales, con fármacos, etc.

Dentro de los modos no covalentes podemos hablar de las interacciones electrostáticas externas como son la unión a los surcos y la intercalación. La intercalación¹¹⁶ es una interacción que conduce a la apertura del ADN y la disposición de la molécula total o parcialmente entre un par de bases del ADN. Las moléculas intercaladoras se introducen entre 2 pares de bases estabilizándose por medio de interacciones de apilamiento. La estructura de estas moléculas contiene anillos aromáticos y de éstos puede colgar cadenas laterales que se ubican a lo largo de los surcos. Normalmente se encuentran cargados positivamente para interactuar con un polianión como el ADN.

La unión al surco, generalmente el surco menor¹¹⁷, es el posicionamiento de la molécula en el interior del surco desplazando total o parcialmente las moléculas de agua que se encontraban en este surco hidratando el ADN. La interacción molécula-ADN viene determinada por interacciones electrostáticas fuertes, típicamente un catión y los grupos fosfatos, por interacciones de van der Waals y por los puentes de hidrógeno entre los grupos dadores y aceptores (de puentes de hidrógeno) de las bases orientados al surco estrecho y la molécula.

Las interacciones entre ADN y las proteínas están guiadas por fuerzas electrostáticas, como los puentes de hidrógeno y los puentes salinos entre Lys⁺ y Arg⁺ y los grupos fosfato del ADN¹¹⁸. También son importantes los contactos de van der Waals. Podemos clasificar las interacciones en 2 tipos: aquellas poco o nada específicas, que dependen poco de la secuencia del ADN. (p.e. ADN polimerasas) y aquellas específicas, que reconocen fragmentos del ADN. (p.e. enzimas de restricción). La

mayoría de proteínas, especialmente aquellas que reconocen el ADN de manera específica interactúan por el surco mayor, aunque en muchos casos una pequeña parte de la proteína también establece contactos por el surco estrecho.

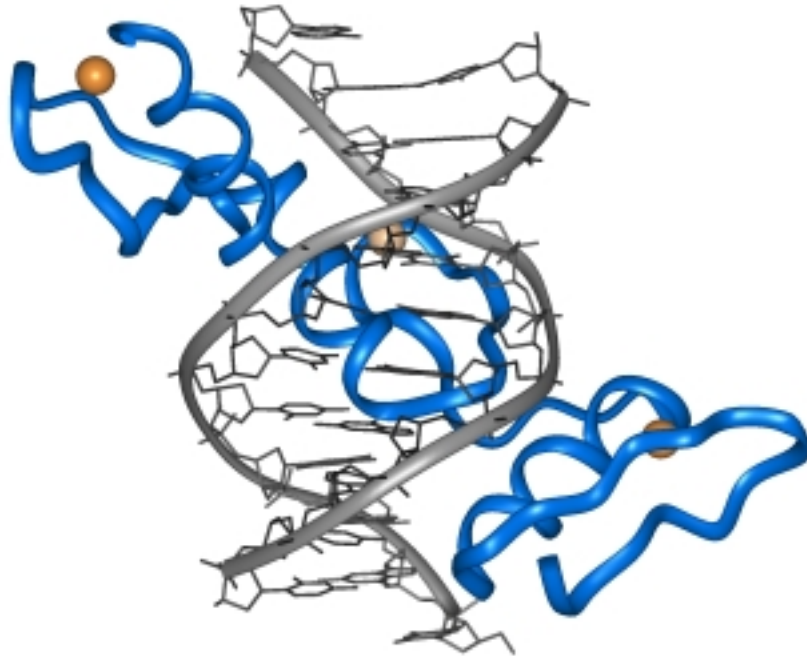


Figura 3.29 Interacción del ADN con una proteína a través del surco ancho. El dúplex de ADN se muestra en color gris, el fragmento de proteína *zinc finger* en color azul y los cationes Zn^{+2} en color naranja. (Fuente: basado en 1a1f.pdb; véase referencia 119)

3.6 Estructura terciaria del ADN¹⁻⁵

El ADN adopta niveles de estructuración mayor como consecuencia de la necesidad de compactarse eficientemente en el interior del núcleo celular. Estos niveles de compactación no son fijos a lo largo de la vida celular, ya que el ADN debe cambiar de nivel de estructuración en función del momento del ciclo en el que se encuentre la célula.

El eje de la hélice de ADN puede retorcerse para formar una superhélice, lo que provoca un mayor grado de empaquetamiento. Este enrollamiento lleva a la doble hélice de ADN a organizarse en cromosomas mediante un proceso por etapas. La primera etapa es la estructuración del ADN para formar los nucleosomas -asociaciones del ADN con octámeros de histonas-, la segunda etapa consiste en la condensación de los nucleosomas en fibras de cromatina y finalmente la cromatina se empaqueta hasta las estructuras que denominamos cromosomas¹²⁰⁻¹²².

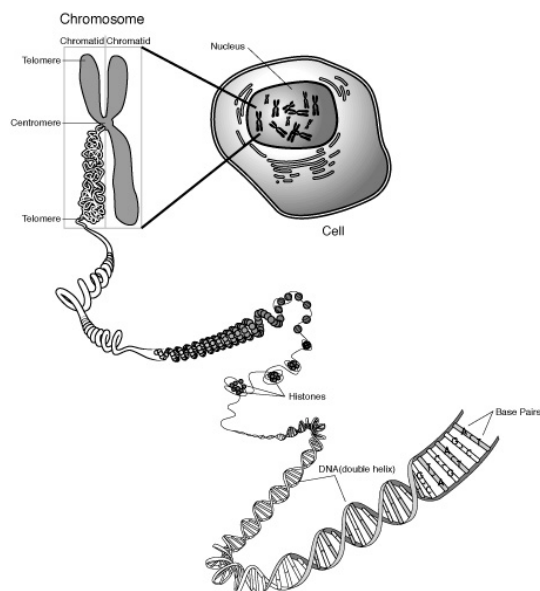


Figura 3.30 Niveles de empaquetamiento de un cromosoma. (Fuente: véase

<http://www.accessexcellence.org/AB/GG/chromosome.html>)

- Estructura de un nucleosoma

La primera etapa en la condensación del ADN es el nucleosoma, donde el ADN cromosómico es empaquetado en una estructura compacta con la ayuda de unas proteínas especializadas llamadas histonas.

Un nucleosoma (figura 3.31) se forma cuando un fragmento de doble hélice de ADN se envuelve fuertemente alrededor de un núcleo central de 8 proteínas histonas (2 copias de 4 histonas diferentes: H2A, H2B, H3 y H4). Este proceso permite obtener una superhélice levógira de ADN de 1.75 vueltas, consiguiendo un ratio de empaquetamiento de 7. El número de pares de bases involucrados en las estructuras nucleosómicas varía dependiendo de la especie, siendo 200 ± 40 pares de bases el valor promedio en los genomas eucarióticos. Asimismo, otra histona (H1) liga externamente el ADN al núcleo nucleosómico, estabilizando el complejo nucleoproteína.

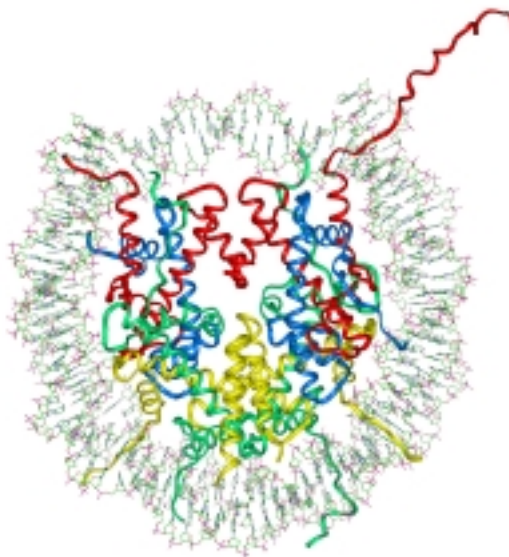


Figura 3.31 Estructura de un nucleosoma, unidad fundamental de empaquetamiento. Las 8 proteínas, 2 copias de cada histona, se encuentran en el centro, mientras que el ADN se repliega alrededor de las proteínas. (Fuente: basado en 1aoi.pdb; véase referencia 123, p.252)

Las unidades nucleosómicas se enlazan entre sí a través de un fragmento de ADN de unos 30-40 pares de bases (*linker DNA*).

El plegamiento del ADN en nucleosomas es el nivel de empaquetado mejor conocido, ya que éste ha sido cristalizado y analizado recientemente por difracción de rayos X¹²³.

- Estructuras de orden superior

Las estructuras correspondientes a órdenes superiores de empaquetamiento cromosómico (figura 3.32) son más especulativas, puesto que no se han podido realizar estudios a alta resolución dado el gran tamaño y la heterogeneidad de la cromatina.

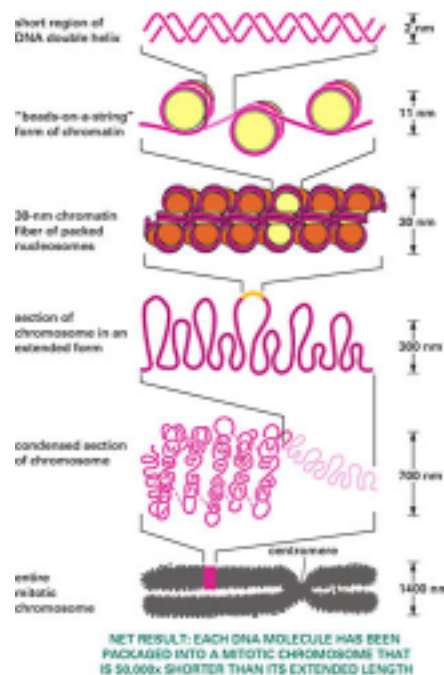


Figura 3.32 Niveles de empaquetamiento del DNA. Éstos órdenes superiores de empaquetamiento proporcionan al cromosoma una estructura más compacta. (Fuente: véase http://www.accessexcellence.org/AB/GG/chroma_packg.html)

El ADN en la cromatina se organiza en agrupaciones de nucleosomas. Los nucleosomas normalmente se empaquetan juntos, con la ayuda de la histona H1 y de un fragmento de ADN (*linker DNA*), formando fibras de 10-11 nm de diámetro que adoptan una apariencia de cuentas de un collar. Modelos de estructuras de orden superior presentan dichas fibras de nucleosomas enrolladas formando una nueva fibra de 30nm de diámetro, denominada solenoide. La hélice que se obtiene, que puede ser tanto levógira como dextrógira, permite un empaquetamiento compacto de nucleosomas con un ratio de 40.

En etapas posteriores de condensación del ADN, uno de los modelos propuestos¹²⁴ sugiere la formación de lazos a partir de estas fibras de 30nm. Cada lazo contendría alrededor de 50 vueltas de solenoides y posiblemente se empaquetarían en una superespiral, uniéndose a un núcleo central de una proteína soporte desde la cual se difunden hacia fuera. Una hélice continua de lazos podría constituir el cromosoma.

3.7 Referencias bibliográficas

- 1) Saenger, W. (1988) En *Principles of Nucleic Acid Structure*. Springer-Verlag, New York.
- 2) Blackburn, G.M. & Gait, M.J. Eds. (1990) En *Nucleic Acids in Chemistry and Biology*. Oxford University Press, Oxford.
- 3) Sinden, R.R. (1994) En *DNA Structure and Function*. Academic Press, San Diego.
- 4) Neidle, S. (1995) En *DNA Structure and Reactivity*. Oxford University Press, London.
- 5) Bloomfield, V.A., Crothers, D.M., & Tinoco, Jr., I. (2000) En *Nucleic Acids: Structures, Properties, and Functions*. University Science Books, Sausalito, CA.
- 6) Taylor, R. & Kennard, O. (1982) *J. Mol. Struct.*, **78**, 1.
- 7) IUPAC-IUB Joint Commission on Biochemical Nomenclature (1983) *Eur. J. Biochem.*, **131**, 9.
- 8) Kline, L.K. & Söll, D. (1982) En *The Enzymes*. Academic Press, New York.
- 9) Seeman, N.C., Rosenberg, J.M., Suddath, F.L., Park Kim, J.J. & Rich, A. (1976) *J. Mol. Biol.*, **104**, 142.
- 10) Hall, L.D. (1963) *Chem. Ind. (London)*, 950.
- 11) Jardetzky, C.D. (1960) *J. Am. Chem. Soc.*, **82**, 229.

- 12) Altona, C., Geise, H.J. & Romers, C. (1968) *Tetrahedron*, **24**, 13.
- 13) Altona, C. & Sundaralingam, M. (1972) *J. Am. Chem. Soc.*, **94**, 8205.
- 14) de Leeuw, H.P.M., Haasnoot, C.A.G. & Altona, C. (1980) *Isr. J. Chem.*, **20**, 108.
- 15) Donohue, J. & Trueblood, K.N. (1960) *J. Mol. Biol.*, **2**, 363.
- 16) Haschemeyer, A.E.V. & Rich, A. (1967) *J. Mol. Biol.*, **27**, 369.
- 17) Sundaralingam, M. (1976) *Annu. N. Y. Acad. Sci.*, **255**, 3.
- 18) Young, D.W., Tollin, P. & Wilson, H.R. (1974) *Nature*, **248**, 513.
- 19) Saran, A., Perahia, D. & Pullman, B. (1973) *Theor. Chim. Acta (Berlin)*, **30**, 31.
- 20) Hruska, F.E., Smith, A.A. & Dalton, J.G. (1971) *J. Am. Chem. Soc.*, **93**, 4334.
- 21) Saenger, W. (1973) *Angew. Chem. Int. Ed. Engl.*, **12**, 591.
- 22) Danyluk, S.S. (1979) En *Nucleoside Analogues. Chemistry, Biology, and Medical Application*. Plenum Press, New York.
- 23) Fasman, G.D. Ed. (1975) En *Handbook of Biochemistry and Molecular Biology*, Volume I, *Nucleic Acids*. Chem. Rubber Co., Cleveland.
- 24) Birnbaum, G.I., Giziewicz, J., Huber, C.P. & Shugar, D. (1976) *J. Am. Chem. Soc.*, **98**, 4640.
- 25) Philips, R., Eisenberg, P., George, P. & Rutman, R.J. (1965) *J. Mol. Biol.*, **240**, 4393.

- 26) Pauling, L. (1978). En *The Nature of the Chemical Bond*. Cornell University Press, Ithaca, New York.
- 27) Wolfenden, R.V. (1969) *J. Mol. Biol.*, **40**, 307.
- 28) Nowak, M.J., Szcsepaniak, K., Barski, A. & Shugar, D. (1978) *Z. Naturforsch. C.*, **33**, 876.
- 29) Topal, M.D. & Fresco, J.R. (1976) *Nature*, **263**, 285.
- 30) Renugopalakrishnan, V., Lakshminarayanan, A.V. & Sasisekharan, V. (1971) *Biopolymers*, **10**, 1159.
- 31) Pullman, B. & Pullman, A. (1971) *Adv. Heterocyc. Chem.*, **13**, 77.
- 32) Hall, R.H. (1971) En *The Modified Nucleosides in Nucleic Acids*. Columbia University Press, New York.
- 33) Uhlmann, E. & Peyman, D.T. (1990) *Chem.Rev.*, **90**, 543.
- 34) Cotton, R.G. (1993) *Mutat. Res.*, **285**, 125.
- 35) Kool, E.T. (1997) *Chem. Rev.*, **97**, 1473.
- 36) van Holde, K.E., Jonson, W.C. & Ho, P.S. (1998) En *Principles of Physical Biochemistry*. Prentice Hall, New Jersey.
- 37) DeVoe, H. & Tinoco, Jr., I. (1962) *J. Mol. Biol.*, **4**, 500.
- 38) Voet, D. & Rich, A. (1970) *Prog. Nucleic Acid Res. Mol. Biol.*, **10**, 183.
- 39) Watson, J.D. & Crick, F.H.C. (1953) *Nature*, **171**, 737.

- 40) Crick, F.H.D. & Watson, J.D. (1954) *Proc. Roy. Soc. (London) Ser. A*, **223**, 80.
- 41) Seeman, N.C., Rosenberg, J.M., Suddath, F.L., Kim, J.J.P. & Rich, A. (1976) *J. Mol. Biol.*, **104**, 109.
- 42) Donohue, J. (1956) *Proc. Nat. Acad. Sci. USA*, **42**, 60.
- 43) Orozco, M, Cubero, E., Hernández, B., López, J.M. & Luque, F.J. (1999) En *Computational Chemistry: Reviews of Current Trends*. Volume 4. World Scientific Publishing, Singapore.
- 44) Hoogsteen, K. (1963) *Acta. Crystallogr.*, **16**, 907.
- 45) Colominas, C., Luque, F.J. & Orozco, M. (1996) *J. Am. Chem. Soc.*, **118**, 6811.
- 46) Donohue, J. (1956) *Proc. Natl. Acad. Sci. USA*, **42**, 60.
- 47) Crick, F.H.C. (1966) *J. Mol. Biol.*, **19**, 548.
- 48) Yason, I.K., Teplitsky, A.B. & Sukhodub, L.F. (1979) *Biopolymers*, **18**, 1149.
- 49) Newmark, R.A. & Cantor, C.R. (1968) *J. Am. Chem. Soc.*, **90**, 5010.
- 50) Kyogoku y., Lord, R.C. & Rich, A. (1967) *J. Am. Chem. Soc.*, **89**, 496.
- 51) Pranata, J. & Jorgensen, W.L. (1991) *J. Am. Chem. Soc.*, **113**, 2810.
- 52) Hobza, P. & Šponer, J. (1999) *Chem. Rev.*, **99**, 3247.
- 53) Hanlon, S. (1966) *Biochem. Biophys. Res. Commun.*, **23**, 861.

- 54) Luo, R., Gilson, H.S.R., Potter, M.J. & Gilson, M.K. (2001) *Biophys. J.*, **80**, 140.
- 55) Ts'o, P.O.P. (1974) En *Basic Principles in Nucleic Acid Chemistry*. Volume 1. Academic, New York.
- 56) Solie, T.N. & Shellman, J.A. (1968) *J. Mol. Biol.*, **33**, 61.
- 57) Tribolet, R. & Sigel, H. (1987) *Biophys. Chem.*, **27**, 119.
- 58) Guckian, K.M., Schweitzer, B.A., Ren, R.X.F., Sheils, C.J., Paris, P.L., Tahmassebi, D.C. & Kool, E.T. (1996) *J. Am. Chem. Soc.*, **118**, 8182.
- 59) Ornstein, R.L., Rein, R., Breen, D.L. & MacElroy, R.D. (1978) *Biopolymers*, **17**, 2341.
- 60) Cieplak, P. & Kollman, P.A. (1988) *J. Am. Chem. Soc.*, **119**, 3734.
- 61) Friedman, R.A. & Honig, B. (1995) *Biophys. J.*, **69**, 1528.
- 62) Alhambra, C., Luque, F.J., Gago, F. & Orozco, M. (1997) *J. Phys. Chem. B*, **101**, 3846.
- 63) Nelson, J.C., Saven, J.G., Moore, J.S. & Wolynes, P.G. (1997) *Science*, **277**, 1793.
- 64) Orozco, M, Cubero, E., Barril, X., Colominas, C. & Luque, F.J. (1999) En *Computational Molecular Biology. Theoretical Computational Chemistry*. Volume 8. Elsevier Science, Amsterdam.
- 65) Jeffrey, G.A. (1994) *J. Mol. Struct.*, **322**, 21.

- 66) Dickerson, R.E., Bansal, M., Calladine, C.R., Diekmann, S., Hunter, W.N., Kennard, O., Lavery, R., Nelson, H.C.M., Olson, W.K., Saenger, W., Shakked, Z., Sklenar, H., Soumpasia, D.M., Tung, C.S., von Kitzing, E., Wang, A.H.J. & Zhurkin, V.B. (1989) *J. Mol. Biol.*, **205**, 787.
- 67) Cheatham III, T.E. & Kollman, P.A. (2000) *Annu. Rev. Phys. Chem.*, **51**, 435.
- 68) Leslie, A.G.W., Arnott, S., Chandrasekaran, R. & Ratliff, R.L. (1980) *J. Mol. Biol.*, **143**, 49.
- 69) Dickerson, R.E., Drew, H.R., Connor, B.N., Wing, R.M., Fratini, A.V. & Kapka, M.L. (1982) *Science*, **216**, 475.
- 70) Arnott, S. & Hukins, D.W.L. (1972) *Biochem. Biophys. Res. Commun.*, **47**, 1504.
- 71) Dickerson, R.E. & Drew, H.R. (1981) *J. Mol. Biol.*, **149**, 761.
- 72) Drew, H.R., Wing, R.M., Takano, T., Broka, C., Tanaka, S., Itakura, K. & Dickerson, R.E. (1981) *Proc. Nat. Acad. Sci. USA*, **78**, 2179.
- 73) Fuller, W., Wilkins, M.H.F., Wilson, H.R., Hamilton, L.D. & Arnott, S. (1965) *J. Mol. Biol.*, **12**, 60.
- 74) Arnott, S., Chandrasekaran, R., Birdsall, D.L. & Leslie, A.G.W. (1980) *Nature*, **283**, 743.
- 75) Wang, A.H.J., Quigley, G.J., Kolpak, F.J., Crawford, J.L. van Boom, J.H., van der Marel, G. & Rich, A. (1979) *Nature*, **282**, 680.
- 76) Wittig, B, Dorbic, T. & Rich, A. (1991) *Proc. Natl. Acad. Sci. USA.*, **88**, 2259.

- 77) Wells, R.D. (1988) *J. Biol. Chem.*, **263**, 1095.
- 78) Calladine, C.R., Drew, H.R., & McCall, M.J. (1988) *J. Mol. Biol.*, **210**, 127.
- 79) Drake, J.W., Allen, E.F., Forsberg, S.A., Preparata, R.M. & Greening, E.O. (1969) *Nature*, **221**, 1128.
- 80) Hunter W.N. (1992) *Methods in Enzymology*, **211**, 221.
- 81) Rippe, K. & Jovin, T.M. (1992) *Methods in Enzymology*, **211**, 199.
- 82) Germann, M.W., Zhou, N., van de Sande, J.H. & Vogel, H.J. (1995) *Methods in Enzymology*, **261**, 207.
- 83) Pattabiraman, N. (1986) *Biopolymers*, **25**, 1603.
- 84) van de Sande, J.H., Ramsing, N.B., Germann, M.W., Elhorst, W., Kalisch, B.W., Kitzing, E.v., Pon, R.T., Clegg, R.C. & Jovin, T.M. (1998) *Science*, **241**, 551.
- 85) Liu, K., Miles, H.T., Frazier, J. & Sasisekharan, V. (1993) *Biochemistry*, **32**, 11802.
- 86) Hashem, G.M., Wen, J.D., Do, Q. & Gray, D.M. (1999) *Nucl. Acids Res.*, **27**, 3371.
- 87) Frank-Kamenetskii, M.D. & Mirkin, S.M. (1995) *Annu. Rev. Biochem.*, **64**, 65.
- 88) Soyfer, V.N. & Potaman, V.N. (1996) *En Triple-Helical Nucleic Acids*. Springer-Verlag, New York.
- 89) Plum, G.E. (1997) *Biopolymers*, **44**, 241.

- 90) Sun, J.S., Garestier, T. & Hélène, C. (1996) *Curr. Opin. Struct. Biol.*, **6**, 327.
- 91) Arnott, S., Bond, P.J., Selsing, E. & Smith, P.J.C. (1976) *Nucl. Acids Res.*, **11**, 4141.
- 92) Macaya, R., Schultze, P. & Feigon, J. (1992) *J. Am. Chem. Soc.*, **114**, 781.
- 93) Howard, F.B., Miles, H.T., Liu, K., Frazier, J., Raghunathan, G. & Sasisekharan, V. (1992) *Biochemistry*, **31**, 10671.
- 94) Raghunathan, G., Miles, H.T. & Sasisekharan, V. (1993) *Biochemistry*, **32**, 455.
- 95) Betts, L., Josey, J.A., Veal, J.M. & Jordan, S.R. (1995) *Science*, **270**, 1838.
- 96) Shields, G.C., Laughton, C.A. & Orozco, M. (1997) *J. Am. Chem. Soc.*, **119**, 7463.
- 97) Soliva, R., Laughton, C.A., Luque, F.J. & Orozco, M. (1998) *J. Am. Chem. Soc.*, **120**, 11226.
- 98) Borman, S. (1998) *Chemical & Engineering*, October 5, 42.
- 99) Williamson, J.R., Raghuram, M.K. & Cech, T.R. (1989) *Cell*, **59**, 871.
- 100) Laughlan, G.; Murchie, A.I.H., Norman, D.G., Moore, M.H., Moody, P.C.E., Lilley, D.M.J. & Luisi, B. (1994) *Science*, **265**, 520.
- 101) Špačková, N., Berger, I. & Šponer, J. (2001) *J. Am. Chem. Soc.*, **123**, 3295.
- 102) Saenger, W. (1987) *Annu. Rev. Biophys. Biophys. Chem.*, **16**, 93.

- 103) Westhof, E. (1993) En *Water and Biological Macromolecules*. CRC Press, Boca Raton, FL.
- 104) Eisenberg, D. & Kauzmann, W. (1969) En *The Structure and Properties of Water*. University Press, Oxford.
- 105) Drew, H.R. & Dickerson, R.E. (1981) *J. Mol. Biol.*, **151**, 535.
- 106) Kubinec, M.G. & Wemmer, D.E. (1992) *J. Am. Chem. Soc.*, **114**, 8739.
- 107) Sigel, H. (1993) *Chem. Soc. Rev.*, **22**, 255.
- 108) Eichhorn, G.L. (1981) En *Advances in Inorganic Chemistry*. Volume 3. Elsevier/North Holland, The Netherlands.
- 109) Izatt, R.M., Christensen, J.J. & Rytting, J.H. (1971) *Chem. Rev.*, **71**, 439.
- 110) McFail-Isom, L., Shui, X. & Williams, L.D. (1998) *Biochemistry*, **37**, 17105.
- 111) Krakauer, H. & Sturtevant, J.M. (1968) *Biopolymers*, **6**, 491.
- 112) Hardin, C.C., Herdenson, E., Watson, T. & Prosser, J.K. (1991) *Biochemistry*, **30**, 4460.
- 113) McCammon, J.A. & Harvey, S.C. (1987) En *Dynamics of Proteins and Nucleic Acids*. Cambridge University Press, Cambridge.
- 114) Timsit, Y. & Moras, D. (1996) *Quart. Rev. Biophys.*, **29**, 279.
- 115) Marmur, J. & Doty, P. (1959) *Nature*, **183**, 1427.
- 116) Berman, H.M. & Young, P.R. (1981) *Annu. Rev. Biophys. Bioeng.*, **10**, 87.

- 117) Zimmer, C. & Wähnert, U. (1986) *Prog. Biophys. Mol. Biol.*, **47**, 31.
- 118) Steitz, T.A. (1990) *Quart. Rev. Biophys.*, **23**, 205.
- 119) Elrod-Erickson, M., Benson, T.E. & Pabo, C.O. (1998) *Structure*, **6**, 451.
- 120) Smith, G.R. (1981) *Cell*, **24**, 599.
- 121) Widom, J. (1989) *Annu. Rev. Biophys. Biophys. Chem.*, **18**, 365.
- 122) Adolph, K.W. Ed. (1988) En *Chromosomes and Chromatin*. CRC Press, Boca Raton, FL.
- 123) Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F. & Richmond, T.J. (1997) *Nature*, **389**, 251.
- 124) Rindt, K.P. & Nover, L. (1980) *Biol. Zentralblat.*, **99**, 641.