

La simulación computacional

de procesos genéticos a nivel molecular

Eduardo
González Jiménez
Valery
I. Poltev

A partir del descubrimiento de la estructura de la doble hélice del ácido desoxirribonucleico (ADN), realizado por Watson y Crick en 1953, se inició un intenso desarrollo de la biofísica molecular, que tiene repercusiones en genética, medicina y biotecnología. Este creciente interés por investigar problemas genéticos a nivel molecular tiene como punto de partida el estudio de la estructura y la dinámica de los ácidos nucleicos.

El ADN es el banco de la información genética y consta de dos cadenas de una sucesión de nucleótidos. Hay cuatro tipos de nucleótidos dependiendo de la base nitrogenada que contienen (Adenina, Guanina, Timina y Citosina) (Figura 1).

Las bases contienen átomos con características donadoras y aceptores de protones que pueden formar enlaces de hidrógeno (enlaces-H) entre sí o con otras moléculas. Sólo dos pares de bases se pueden formar para dar una doble hélice de ADN regular; éstos son los pares de Watson y Crick (WC), A:T y G:C (Figura 1).

La investigación de los ácidos nucleicos influyó en el desarrollo de los métodos experimentales y teóricos que lograron determinar diversas características de su estructura. En 1981, Dickerson y colaboradores lograron cristalizar pequeños fragmentos de ADN (oligonucleótidos) para investigar por el método de rayos X y obtener varias de sus propiedades. Hoy en día, han sido caracterizadas varias decenas de fragmentos, como octámeros, decámeros y dodecámeros.^{1,2}

Un hecho de actualidad que ocupó los encabezados de los medios de comunicación a nivel mundial fue la determinación del genoma humano, que consistió en conocer la secuencia nucleotídica e identificar los genes de que consta el ADN humano. Sin embargo, esto

es solo una pequeña parte del problema, pues la aplicación de los resultados aún no es plena y no lo será, mientras no se comprendan en detalle los mecanismos de expresión, reparación y transferencia de los genes para una manipulación dirigida y segura.

En la actualidad, comprendemos en términos generales que la secuencia de bases y la conformación estructural del ADN dirigen la construcción de las proteínas, y que este proceso se puede ver influenciado por factores que alteraran la información genética o afectan los mecanismos que aseguran la transmisión correcta de la información en la célula.

Muchos de los mecanismos físicos de estos procesos transcurren en el nivel de átomos o grupos atómicos y dependen de sus interacciones, así como de la influencia del medio, a través de la interacción con los constituyentes moleculares de éste. Por lo tanto, se precisa de métodos aplicables a nivel molecular, que puedan cuantificar energías, coordenadas y otras características asociadas a los componentes de los ácidos nucleicos.

LOS MÉTODOS EXPERIMENTALES

Los más usados en el estudio de los ácidos nucleicos son la técnica de rayos X en cristales de ADN y los métodos de Resonancia Magnética Nuclear (NMR) para ácidos nucleicos en solución. Los datos obtenidos por rayos X sobre la estructura de cientos de cristales de oligonucleótidos de doble espiral y sus complejos con proteínas, permitieron comprobar la dependencia de la conformación de los ácidos nucleicos respecto de la secuencia nucleotídica y los alrededores moleculares.

Otros datos experimentales muestran la importante influencia del agua en la formación de la estructura espacial del ADN y su relación e interacciones con proteínas y antibióticos. Por ejemplo, se encontró un arreglo ordenado de moléculas de agua (estructura de hidratación) en la forma cristalina del dodecámero d(CGCGAATTGCG) en la conformación B.¹ Este arreglo se forma en la hendidura menor de la doble hélice a lo largo de la secuencia central AATT, región que es importante porque en ella hay una unión preferente de antibióticos como el netropsin³ o el distamycin –pero este último para la secuencia d(GGCCAATTGG).⁴ De estos resultados se desprende la propuesta de que los átomos que sobresalen de las hendiduras de ciertas secuencias abundantes en bases A y T, forman parte de sitios promotores o de reconocimiento.

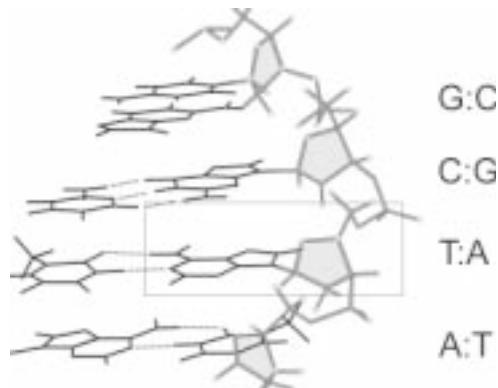


FIGURA 1. Estructura del ácido desoxirribonucleico en la conformación B, obtenida con el programa HiperChem v5.02; sólo se incluye una cadena azúcar-fosfato. El rectángulo indica un nucleótido (base, azúcar y grupo fosfato), la molécula de azúcar se ilumina de gris. Se observan los pares de bases unidos con enlaces de hidrógeno (líneas punteadas).

Por tal motivo, es importante tener una descripción detallada de la estructura espacial de los ácidos nucleicos y del solvente a nivel atómico.

LOS MÉTODOS TEÓRICOS

Aun con la utilización de los más sofisticados y modernos métodos experimentales, no es posible determinar los detalles de la estructura del ADN y sus complejos moleculares.

La investigación directa de los detalles atómicos con los métodos de rayos X es imposible, ya que la precisión de los resultados frecuentemente no llega al nivel atómico y la posición y orientación de las moléculas de agua quedan determinadas sólo aproximadamente. Además de los datos de la estructura de cristales de ADN, sólo es posible determinar una o algunas conformaciones de la molécula de ADN; sin embargo, es fundamental conocer todas las posibles conformaciones y los mecanismos de las transformaciones entre ellas. Esta tarea es asumida por los métodos teóricos, y su aplicación a los sistemas moleculares ha dependido del desarrollo de las supercomputadoras.

La descripción de un sistema molecular obedece a las leyes de la mecánica cuántica. Estas leyes físicas se resumen en la ecuación de Schrödinger propuesta en 1926. Para predecir el comportamiento del sistema se deben resolver directamente las ecuaciones de la mecánica cuántica.

En principio, las propiedades electrónicas de una molécula pueden ser calculadas de las interacciones entre todos los electrones y todos los núcleos de la misma. Fue en los años sesenta, con el desarrollo de las computadoras, cuando la ecuación de Schrödinger se resolvió para sistemas moleculares en su más rigurosa aproximación de primeros

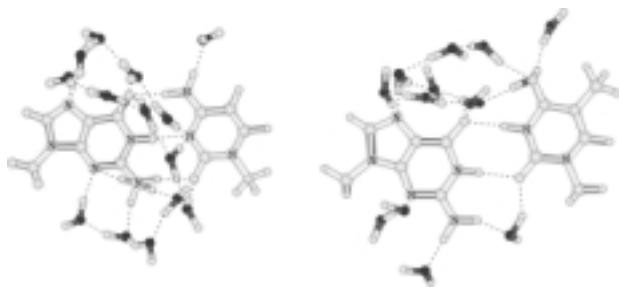


FIGURA 2. Simulación de la estructura de la capa de hidratación de los pares de bases G:C_{WC} (izquierda) y G:T (derecha). Los puentes de agua pueden estabilizar algún par incorrecto como es el caso del par G:T encontrado en estudios experimentales. La estructura mostrada se obtuvo de una simulación de Monte Carlo con 400 moléculas de agua a temperatura de 500K, las restantes moléculas de agua se eliminaron.

principios (*ab initio*), así se pudieron definir propiedades importantes de las moléculas. Sin embargo, es suficientemente pequeño el tamaño de las moléculas que la aproximación *ab initio* puede tratar con suficiente exactitud (relativa) para poder comparar con los datos experimentales.

Otra opción para la construcción de un cuadro completo y detallado de las posibles conformaciones del ADN, que consideren las interacciones con agua, iones y otras moléculas (como proteínas), es utilizar esquemas empíricos para la simulación computacional molecular. Éstos comprenden la aproximación de la mecánica molecular.

SIMULACIÓN COMPUTACIONAL: MECÁNICA MOLECULAR

Los métodos de la simulación molecular han llegado a utilizarse ampliamente en las últimas décadas y en diferentes disciplinas de las ciencias exactas; tienen su origen en modelos simples, como los modelos moleculares construidos con alambre, madera o de otros materiales. Estos modelos permitían cambios manuales en las posiciones mutuas de los átomos. A pesar de lo burdo de los modelos, contribuyeron al descubrimiento de la estructura de la doble espiral del ADN.

Hoy existe software con paquetes gráficos muy sofisticados que son estupendos para visualizar moléculas biológicas, pero no dicen mucho acerca de cómo cambian las moléculas cuando están sujetas a su acción mutua. Por tanto, se necesitan modelos físicos que reproduzcan confiablemente la estructura, las conformaciones y las rutas para obtener una conformación favorable bajo ciertas condiciones.

La versión moderna de los métodos de simulación para ácidos nucleicos es la mecánica molecular. Ella nos permite calcular la estructura espacial de una molécula con base en

la búsqueda de los mínimos locales a través del análisis de la energía de interacción no-valente.

La mecánica molecular se basa en la idea de que una molécula se puede representar por un conjunto de puntos con carga neta situados en los núcleos de los átomos y distribuidos en una superficie de energía potencial generada por ellos mismos. El cálculo de los valores de la energía se hace utilizando un potencial efectivo para interacciones átomo-átomo, el cual se construye basándose en datos experimentales y en cálculos mecánico cuánticos.

El método de Monte Carlo y el de Dinámica Molecular son los dos métodos de la mecánica molecular más utilizados para obtener propiedades promedio de los sistemas moleculares. El primero parte de una configuración inicial que cambia aleatoriamente generando un conjunto de "fotografías" instantáneas del sistema que se eligen de acuerdo con una regla (muestreo por importancia). Las propiedades de interés se obtienen promediando éstas sobre el conjunto de fotografías. El segundo resuelve las ecuaciones de Newton para los átomos del sistema, a los cuales se les asignan posiciones y velocidades iniciales. El movimiento de los átomos se simula calculando sucesivamente las posiciones y velocidades del sistema cuando transcurre el tiempo, de esta forma la conformación va cambiando hasta la conformación final. Ambos métodos pueden ser aplicados a sistemas con miles de átomos debido a que las funciones de energía y sus derivadas son relativamente fáciles de calcular.

Partiendo de la viabilidad de los métodos de cálculo, la diferencia en la obtención de resultados más acordes con la realidad (el experimento) lo constituyen las interacciones entre los átomos, expresadas a través de las funciones de potencial.

LAS FUNCIONES DE POTENCIAL

En la actualidad existen varios conjuntos de funciones de potencial. Algunos de ellos son bastante sofisticados y exigen grandes recursos de cómputo y complejos programas que abarcan interacciones en ácidos nucleicos, proteínas y diversas moléculas orgánicas, pero esta generalidad limita la exactitud. Respecto de los resultados experimentales, los modelos sofisticados no dan mejores valores, respecto de los modelos simples.⁵ Las funciones de potencial deben reproducir satisfactoriamente los datos experimentales disponibles; también deben estar acordes con los más estrictos resultados de los cálculos mecánico cuánticos.

El interés en la investigación detallada de las posibles conformaciones de fragmentos de ADN que expliquen las peculiaridades de las conformaciones de diferentes secuencias y ayuden a comprender el papel de la capa de solvente acuoso sobre la estructura y el funcionamiento del ADN, se logra al elegir potenciales exclusivos para ácidos nucleicos y elaborar programas de cómputo especiales para estos sistemas.

La formulación de uno de los primeros sistemas de funciones de potencial átomo-átomo para interacciones de moléculas orgánicas con miras a aplicaciones a ácidos nucleicos fue introducido por uno de los autores en 1967. En 1984 el potencial se ajustó y fue presentado como el potencial Poltev-Malenkov (PM).⁶ Casi simultáneamente otro campo de fuerzas hace su aparición presentado por Kollman y colaboradores y después optimizado en 1995.⁷ Adquirió gran popularidad al ser implementado en el software AMBER para simulación de ácidos nucleicos y proteínas. Pero para hidratación de bases nitrogenadas sus resultados no son mejores que los obtenidos con el potencial PM. Otros potenciales para estudiar ácidos nucleicos son el campo de fuerza en el programa CHARMM⁸ y el "potenciales optimizados para la simulación de líquidos" (OPLS) de Jorgensen,⁹ este último no obtiene buenos resultados para ácidos nucleicos.⁵

Una característica de los campos de fuerzas es que continuamente deben someterse a refinamientos sucesivos para una descripción cuantitativa más precisa. Recientemente se inició una optimización del potencial PM, pero sin cambiar su forma analítica.¹⁰ El potencial PM ha obtenido importantes resultados que han sido corroborados experimentalmente, contribuyendo a comprender los procesos genéticos a nivel molecular.

SIMULACIÓN DE LA ESTRUCTURA Y DINÁMICA DE ADN

Las simulación no puede sustituir al experimento, tal como el modelo no sustituye a la realidad, pero logra aproximaciones que sugieren afirmaciones que pueden ser corroboradas por el experimento o que le indican al experimentador en qué dirección trabajar. En este sentido mencionamos algunas conclusiones que fueron hechas con base en el análisis teórico de los cálculos de la simulación con el potencial PM, sobre la estructura y propiedades de los ácidos nucleicos, mucho antes de que aparecieran los datos experimentales que confirmaran dichas proposiciones.



Un proceso fundamental de la vida es la replicación de la información genética a partir de un molde o plantilla. Las copias de la información se realizan con una alta fidelidad durante la replicación del ADN en la célula, formando sólo un par incorrecto por cada $10^8\text{-}10^{11}$ pares correctos, a pesar de que, de consideraciones teóricas y datos experimentales sobre "sistemas modelo", se puede garantizar una exactitud no mayor que una base errónea por 10^3 bases correctas. ¿Cuáles son los mecanismos atómico-moleculares de la exactitud de la biosíntesis del ADN? El papel crucial de esta exactitud lo juegan las enzimas de reconocimiento, la ADN-polimerasa.

En 1974 se propuso por primera vez, por uno de los autores, un mecanismo de reconocimiento primario para la incorporación de nucleótidos por ADN-polimerasa.¹¹ Según este mecanismo, el centro de reconocimiento de la ADN-polimerasa interactúa con invariantes estructurales de los pares de WC, éstos son átomos que sobresalen del par de bases y que pueden formar enlaces de hidrógeno. Por su parte, el centro de reconocimiento de la ADN-polimerasa contiene dos grupos protón-donador, N-H u O-H situados de manera que corresponden a los átomos invariantes del par de Watson y Crick.

Utilizando las funciones de potencial PM, se calculó la energía de interacción de todos los posibles pares de bases co-planares con un modelo del sitio de reconocimiento de la ADN-polimerasa, permitiendo explicar los datos experimentales sobre la frecuencia relativa de mutaciones espontáneas. La conclusión que se obtuvo fue que los sitios de reconocimiento de la polimerasa actúan como una plantilla adicional que incrementa la probabilidad de incorporar un nucleótido correcto y la disminuye para incorporar uno incorrecto.

En 1994 aparecen los primeros datos experimentales de rayos X de complejos de varias ADN-polimerasas con fragmentos de plantillas de ADN;¹² estos datos comprueban la predicción que se obtuvo 30 años antes con los resultados de la simulación.

Usando el potencial PM se estudió la hidratación de los diferentes pares de bases (Figura 2); los resultados revelan que, respecto de la morfología de las bases, hay diferentes patrones de hidratación que dependen del tipo de base y de la conformación de ADN.¹³

Ambos pares de bases de Watson y Crick poseen un patrón de hidratación similar, no así los pares incorrectos. Ello está de acuerdo con el mecanismo de fidelidad descrito antes.



© Oscar Necoechea, de la serie *El cielo sobre nosotros*, 1989-2001.

Se observaron estructuras similares a las espinas de hidratación en secuencias específicas corroborando la propuesta de que los átomos de proteínas involucrados en la unión con ADN ocupan las posiciones donde normalmente se sitúan las moléculas de agua en ADN.

En los últimos años se ha incrementado la eficiencia de cálculo con el desarrollo de algoritmos que incrementan la exactitud de los resultados, aunado a esto, el aumento en la potencia de cálculo de las computadoras ha permitido que este desarrollo impacte en la habilidad para simular de manera más realista la estructura de los ácidos nucleicos y los procesos que involucran, contribuyendo a tener una visión más completa del funcionamiento biológico de los ácidos nucleicos y demás moléculas bio-orgánicas.

N O T A S

¹ Dickerson R.E. Drew D.R., Structure of a B-DNA Dodecamer. III. Geometry of Hydration, *J.Mol.Biol.*, Vol.151, 1981, pp.535-556.

² Ng H.L., Kopka M.L., Dickerson R.E., *Proc.Natl.Acad.Sci. USA*, Vol.97, 2000, pp.2035.

³ Kuroda R., Neidle S., Drug-Nucleic Acid Interactions at the Molecular Level. In: *Steric Aspects of Biomolecular Interactions*, CRC Press. , 1987, pp.213-233.

⁴ Vlieghe D., Sponer J., Van Meervelt L., The Crystal Structure of d(GGCCATTG) Complexed with DAPI Reveals Novel Binding Mode, *Biochemistry*, Vol.38, 1999, pp.16451-16459.

⁵ Poltev V.I., Malenkov G.G., Gonzalez E., Teplukhin A.V., Rein R., Shibata

M. and Miller J.H., Modeling DNA Hydration: Comparison of Calculated and Experimental Hydration Properties of Nucleic Acid Bases, *J.Biomol.Struc.Dyn.*, No.4, Vol.13, 1996, pp.717-725.

⁶ Poltev V.I., Grokhina J.I. and Malenkov G.G., Hydration of Nucleic Acids Bases Studied Using Novel Atom-Atom Potential Functions *J.Biomol.Struc.Dyn.*, No.2, Vol.2, 1984, pp.413-429.

⁷ Cornell W.D., Cieplak P., Bayly C.I., Gould I.R., Merz K.M., Ferguson D.M., Spellmeyer D.C., Fox T., Caldwell J.W. and Kollman P.A., A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules *J.Amer.Chem.Soc.*, Vol.117, 1995, pp.5179-5197.

⁸ MacKerell Jr. A.D., Wirkiewicz-Kuczera J., Karplus M., An all-Atom Empirical energy Function for the Simulation of Nucleic Acids, *J.Am.Chem.Soc.*, Vol.117, 1995, pp.11946-11975.

⁹ Jorgensen W.L., Tirado-Rives J., The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J.Am.Chem.Soc.*, Vol.110, 1988, pp.1657-1666.

¹⁰ González E., Cedeno F.I., Teplukhin A.V., Malenkov G.G., Poltev V.I. Refinamiento de la Metodología de la Simulación de los Hidratación de los Ácidos Nucleicos, *Rev.Mex. de Física*, S.2, Vol.46, 2000, pp.142-147.

¹¹ Bruskov V.I., Poltev V.I., Recognition by encimes of complementary pairs of nitrogen bases and intensification of the specificity of the interaction in processes of template synthesis, *Dokl.Akad.Nauk. SSSR*, Vol.219, 1974, pp.231-234.

¹² Pelletier H., Azuaya M.R., Kumar A., Wilson S.H., Kraut J., Structures of Ternary Complexes of Rat DNA Polymerase α , a DNA Template-Primer, and dCTP, *Science*, Vol.264, 1994, pp.1891-1903.

¹³ Gonzalez E., Deriabina A.S., Poltev V.I., Computer Simulation of Biomolecular Systems. Hydration of DNA Fragments., In: *Advances in Systems Theory, Mathematical Methods and Applications* 310p, ISBN 960 8052 610, 2002 (en prensa).

Eduardo González Jiménez y Valery I. Poltev son profesores de la Facultad de Ciencias Físico-Matemáticas de la BUAP. gonzalez@fcfm.buap.mx; poltev@fcfm.buap.mx.



© Oscar Necochea, de la serie *El cielo sobre nosotros*, 1989-2001.