

# Claves teóricas de la mutación espontánea en el ADN

José Pedro Cerón-Carrasco, Denis Jacquemin, José Zúñiga y Alberto Requena

**Resumen:** El ADN es la molécula biológica que mayor atención ha atraído desde su descubrimiento como molécula portadora del código genético. Su estructura y propiedades permiten descifrar la información genética de los seres vivos y entender la forma en la que esta información se transmite de una generación a otra. Desafortunadamente, siguen sin conocerse completamente los mecanismos que producen alteraciones en la estructura del ADN, cuya consecuencia más grave es la aparición de mutaciones que dan lugar a enfermedades genéticas. En este trabajo presentamos los estudios teóricos recientes que hemos llevado a cabo para caracterizar uno de los posibles mecanismos responsables de la mutación espontánea, el equilibrio tautomérico en los pares de bases del ADN.

**Palabras clave:** ADN, mutación, tautómeros, transferencia protónica, química computacional.

**Abstract:** DNA has been the most widely studied biological molecule since it was discovered as the carrier of the genetic code. Its structure and properties contain the genetic information and allow us to unravel the process through which this information is transferred through generations. Unfortunately, even nowadays, the mechanism of the mutations of DNA, which might lead to important genetic diseases, has not been completely understood. In this work we review our recent theoretical studies focused on the possible spontaneous mutation mechanisms, specifically the process related to the tautomeric equilibria in DNA base pairs.

**Keywords:** DNA, mutation, tautomers, proton transfer, computational chemistry.

## Introducción

El ADN es la molécula responsable de codificar y transmitir la información genética necesaria para que las células funcionen. A pesar de estar protegido por el medio celular, el ADN sufre constantes ataques por parte de diversos agentes tanto físicos como químicos. Dentro del primer tipo se encuentra, por ejemplo, la radiación ultravioleta, cuya energía es suficiente para provocar la excitación electrónica de las bases del ADN.<sup>1</sup> Los cambios en la estructura electrónica pueden producir dímeros ciclobutánicos que afectan a la replicación del ADN, y que están en el origen del cáncer de piel.<sup>2</sup> El ADN sufre además el ataque de otras moléculas presentes en el medio celular como los radicales libres. Estas especies químicas son intermedios necesarios en un gran número de reacciones biológicas, pero en altas concentraciones pueden causar daños relacionados con la aparición de enfermedades degenerativas.<sup>3</sup> Existe además un tercer mecanismo de alteración del código genético en el que no participa ningún agente externo. Su origen está en la propia estructura

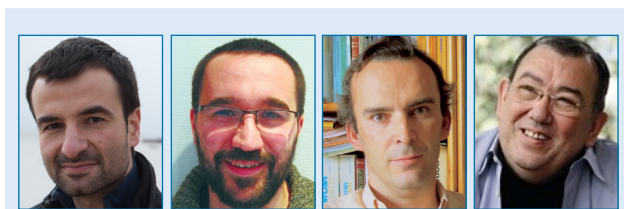
del ADN, y da lugar a lo que se conoce como mutaciones espontáneas.<sup>4</sup> Dado que las técnicas experimentales actuales no permiten identificar los mecanismos que provocan las mutaciones espontáneas en el ADN, la química teórica surge como una alternativa especialmente útil en el estudio de este importante fenómeno.

En este trabajo exponemos los conceptos básicos que rigen las propiedades y la estabilidad del ADN, y presentamos los resultados teóricos con los que hemos contribuido al esclarecimiento del origen de la mutación espontánea. En primer lugar repasamos brevemente los principales hallazgos que permitieron determinar la estructura molecular del ADN y los procesos que pueden dar lugar a mutaciones en su secuencia. Posteriormente, describimos los nuevos métodos de la química cuántica que pueden utilizarse para el estudio de las mutaciones; y, finalmente, presentamos los resultados que hemos obtenido en nuestras simulaciones teóricas y discutimos las implicaciones biológicas de los mismos.

## Estructura del ADN

### La molécula de la herencia

La primera teoría sobre la herencia, bastante alejada del concepto científico de la misma, proviene de Aristóteles, según el cual ésta se transmite a través de la sangre, un vínculo que todavía hoy seguimos utilizando de forma metafórica. Pero es Gregor Mendel quien, en el año 1886, enunció las leyes básicas de la herencia al presentar sus “Experimentos sobre híbridos de plantas”, en los que analizó los cruzamientos entre cepas de guisantes.<sup>5</sup> Las consecuencias de sus trabajos de jardín fueron, no obstante, ignoradas hasta quince años más tarde, cuando el botánico holandés Hugo de Vries descubrió que, incluso en la descendencia de cepas puras, aparece un número reducido de individuos con ciertas alteraciones. De Vries publicó en 1901 su obra “*La teoría de la mutación*”, en la que introdujo el término de mutación haciendo referencia a cualquier alteración en el material hereditario, si bien continuó sin conocerse el verdadero vehículo de la herencia.<sup>6</sup>



J.P. Cerón-Carrasco<sup>1</sup>

D. Jacquemin<sup>1</sup>

J. Zúñiga<sup>2</sup>

A. Requena<sup>2</sup>

<sup>1</sup> CEISAM, Université de Nantes  
Rue de la Houssinière, 44322 Nantes. Francia.  
C-e: [jose.ceron@univ-nantes.fr](mailto:jose.ceron@univ-nantes.fr)

<sup>2</sup> Departamento de Química Física. Universidad de Murcia  
Campus de Espinardo, 30100 Murcia, España.

**Recibido: 15/03/2012. Aceptado: 09/05/2012.**

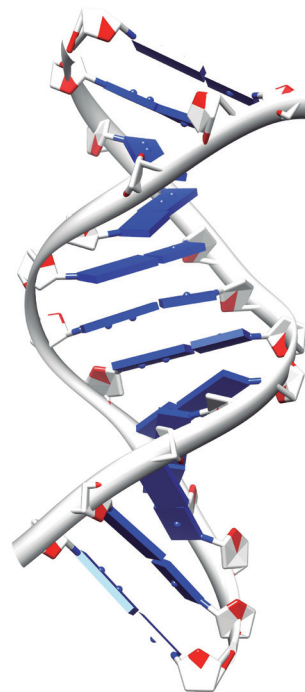
Un año después, en 1902, las investigaciones de Hermann Emil Fischer y Franz Hofmeister pusieron de manifiesto que las proteínas son secuencias de aminoácidos unidos por enlaces peptídicos, y a comienzos del siglo XX todo apuntaba a que estas macromoléculas eran, de alguna manera, las encargadas de codificar la información hereditaria. Hay que esperar, sin embargo, hasta el año 1952 en el que Alfred Hershey y su estudiante Martha Chase mostraron, mediante marcaje isotópico, que el ácido desoxirribonucleico (ADN), y no las proteínas, es el verdadero responsable de transmitir la información genética de una generación a la siguiente.<sup>7</sup> Después, y como contaron los propios protagonistas,<sup>8</sup> se aceleró la carrera para determinar la estructura del misterioso ADN en la están involucrados principalmente dos laboratorios ingleses, el Cavendish en Cambridge, donde trabajaron Francis Watson y James Crick, y el del King's College en Londres, donde Maurice Wilkins y Rosalind Franklin realizaron estudios de cristalografía por difracción de rayos X de la molécula de ADN. Forma ya parte de la historia la célebre fotografía de difracción número 51 que Wilkins mostró a Watson sin el conocimiento de Franklin y que permitió a Watson y Crick deducir la estructura de doble hélice.<sup>9</sup> El 25 de abril de 1953, la revista *Nature* publicó tres artículos consecutivos: el de Watson y Crick,<sup>9</sup> otro del equipo de Wilkins,<sup>10</sup> y un tercero de Franklin con sus datos de difracción de rayos X<sup>11</sup> que, evidentemente, respaldan la estructura del ADN propuesta en el primero. En el año 1962 Watson, Crick y Wilkins recibieron el Premio Nobel por estas investigaciones. Rosalind Franklin, cuya contribución, no exenta de numerosos obstáculos tanto científicos como humanos, resultó decisiva para el esclarecimiento de la estructura del ADN, había fallecido cuatro años antes, y hubiese sido tan merecedora del reconocimiento como los tres galardonados.

### La doble hélice

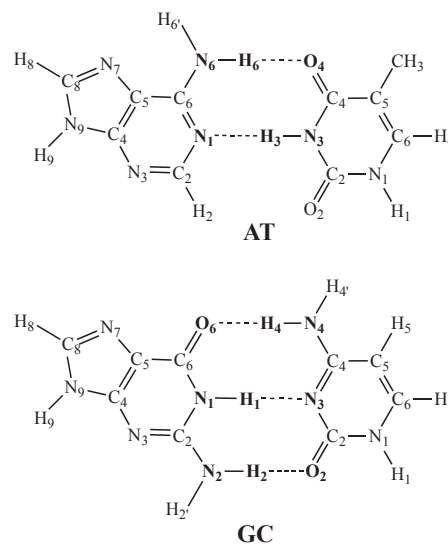
“Queremos proponer una estructura para la sal del ácido desoxirribonucleico (ADN). Esta estructura tiene características nuevas de considerable interés biológico”.<sup>9</sup> De esta manera comenzaban Francis Crick y James Watson su ya famoso artículo de 1953, en el que describían la estructura molecular del ADN en forma de doble hélice (Figura 1). Un artículo de tan sólo una página con el que pasar a la historia de la ciencia. La estructura del ADN consiste, pues, en dos cadenas de nucleótidos enrolladas alrededor de un mismo eje, de tal manera que los dos tipos de bases que contiene, las bases purínicas, adenina (A) y guanina (G), y las bases pirimidínicas, citosina (C) y timina (T), quedan enlazadas siempre bajo un mismo patrón. La doble hélice se mantiene unida mediante los enlaces de hidrógeno entre las bases adenina–timina (AT) y guanina–citosina (GC). Éstos son los dos únicos tipos de pares de bases presentes en el modelo de Watson y Crick, denominados, por tanto, pares de Watson–Crick, o simplemente pares de bases canónicos (Figura 2).

A, T, G y C; en estas cuatro letras se encuentra el secreto del código genético. Por un lado, el hecho de que la doble hélice esté unida mediante enlaces de hidrógeno permite que las dos cadenas puedan separarse con cierta facilidad, manteniendo intactas sus estructuras individuales para replicarse posteriormente. Y por otro lado, la secuencia en la que se disponen las bases en las cadenas de ADN es la que codifica

toda la información genética. En palabras de Watson y Crick “el emparejamiento específico que hemos postulado hace pensar al instante que se trata de un posible mecanismo de copia del material genético”. Esta afirmación constituye *per se* la explicación del mecanismo molecular de la herencia, ya que a partir de ella resulta fácil entender cómo se reproduce un cromosoma durante el proceso de mitosis mediante la duplicación de una doble hélice.



**Figura 1.** Representación esquemática de un fragmento de ADN compuesto por una secuencia de 12 pares de bases. Los pares de base que mantienen unida a la doble hélice aparecen representados en azul.



**Figura 2.** Estructura química de los pares de base adenina–timina (AT) y guanina–citosina (GC) en su forma canónica.

## Mutación espontánea

La estructura en forma de doble hélice del ADN le permite, por tanto, replicarse y transmitir la información genética a la siguiente generación. Podemos reinterpretar así la definición dada por Hugo de Vries del término mutación como cualquier modificación en la estructura química del ADN que se produzca bien por la acción de agentes externos, o bien de forma espontánea. ¿Cuál es entonces el origen de la mutación espontánea?

Para intentar responder esta pregunta conviene analizar previamente las estructuras químicas de los pares de bases del ADN, que se muestran en la Figura 2. Como vemos, en el par AT se forman dos enlaces de hidrógeno, mientras que en el par GC se forman tres. Estos enlaces de hidrógeno son, como hemos dicho, los puntos de anclaje de las dos cadenas de ADN, y los enlaces que se rompen, por tanto, durante el proceso de replicación. En su forma canónica, las cuatro bases tienen una estructura amino-cetónica, que es la más estable y, por tanto, la mayoritaria. Pero que haya una estructura más estable no significa que sea la única.<sup>12</sup> En realidad, los dos pares de bases del ADN se encuentran en equilibrio imino-enólico con varias formas tautoméricas como consecuencia de un proceso de intercambio de protones (véase Figura 3).<sup>13</sup> Estos tautómeros son mucho menos estables que las formas canónicas, y por ello es más difícil encontrarlos en el ADN de una célula, motivo por el cual se denominan tautómeros raros. El ADN humano está formado, sin embargo, por aproximadamente tres mil millones de pares de bases, seis mil millones de bases en total. Aunque la constante de equilibrio para la formación de tautómeros raros sea muy baja, con un número tan elevado de bases puede producirse una cantidad significativa de ellos durante la replicación. Las pequeñas, aunque importantes, variaciones en las estructuras de los tautómeros raros con respecto a las formas canónicas hace, además, muy difícil tanto su detección como la reparación de los mismos mediante

los mecanismos con los que la evolución ha dotado al ADN. Los tautómeros raros pueden acabar, por tanto, provocando un error en la transmisión de la secuencia.

La hipótesis anterior es la propuesta por Löwdin hace casi 50 años para entender la aparición de mutaciones espontáneas sin la implicación de ningún agente externo al propio ADN.<sup>14</sup> Desafortunadamente, no disponemos a día de hoy de métodos experimentales capaces de detectar la presencia de tautómeros raros en el ADN. Esto se debe, primero, a la dificultad de determinar las posiciones de los hidrógenos mediante las técnicas habituales como la cristalografía de rayos X o la de RMN, en las que se localizan primero los átomos pesados y, *a posteriori*, se utilizan herramientas informáticas para añadir los hidrógenos en función de las distancias obtenidas para tener una estructura química coherente. Y en segundo lugar, a la propia baja cantidad de tautómeros raros presentes en el ADN, que cae por debajo de los límites de sensibilidad a los que podemos llegar.<sup>15</sup> En contrapartida, el creciente desarrollo de los métodos de la química teórica y computacional permite explorar con un grado de precisión cada vez mayor las reacciones en las que están involucradas las moléculas de la vida.<sup>16</sup>

Es en este contexto en el que nuestros grupos de investigación han intentado contribuir al esclarecimiento del papel que desempeñan los tautómeros raros en la mutación espontánea del ADN empleando los métodos teóricos más avanzados para describir este tipo de procesos.

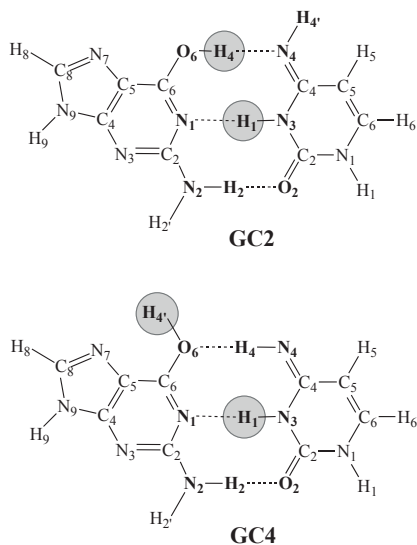
## Descripción cuántica

El primer paso para estudiar teóricamente los equilibrios tautoméricos de las bases del ADN consiste en diseñar un modelo químico adecuado y seleccionar un método de cálculo que proporcione el grado de precisión necesario con un coste computacional aceptable. Expondremos, en primer lugar, los modelos de ADN que hemos usado y a continuación discutiremos los niveles de teoría empleados en nuestras investigaciones.

## Modelos químicos

Podemos encontrar en la bibliografía excelentes trabajos, como los del grupo de Bertrán, Oliva y Sodupe, en los que se estudian sistemas biológicos mediante simulaciones teóricas.<sup>17</sup> Para el caso concreto de la mutación espontánea no es fácil, sin embargo, extrapolar los resultados de las simulaciones a la realidad *in vivo* debido a la complejidad del entorno biológico.<sup>16b</sup> Así, en el ADN existen dos tipos de interacciones básicas que deben estar adecuadamente representadas en los modelos químicos, las interacciones que provoca el apilamiento de las bases en la doble hélice, y las que genera el entorno eminentemente acuoso del medio celular. Estas interacciones afectan a los enlaces de hidrógeno que se establecen entre las bases,<sup>17</sup> y presumiblemente influyen de alguna manera en los equilibrios que mantienen las formas canónicas de las bases con sus tautómeros raros.

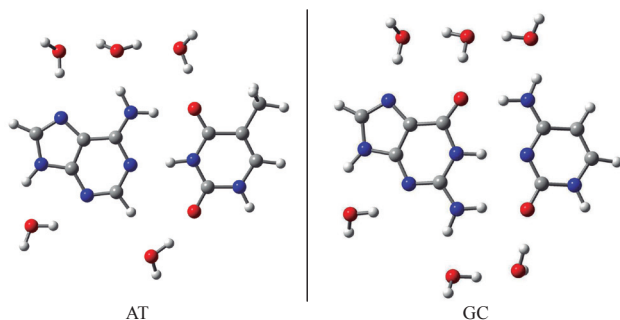
Con el fin de dilucidar el impacto del entorno biológico en la formación de los tautómeros raros y, en consecuencia, en la mutación espontánea, hemos investigado las estabildades relativas de los tautómeros raros en tres entornos diferentes: (i) en fase gaseosa, donde los pares AT y GC están completamente aislados tal y como aparecen en la Figura 2; (ii) en fase



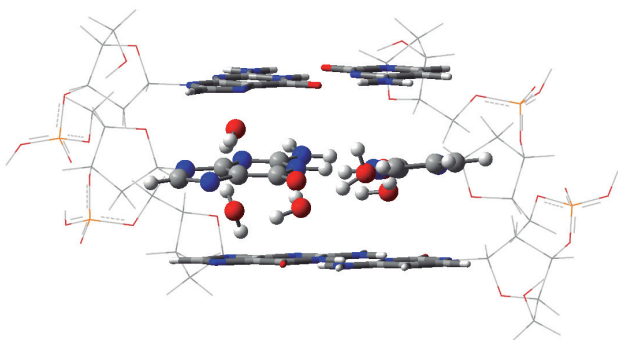
**Figura 3.** Estructura química de dos tautómeros raros del par GC donde se resaltan los protones transferidos. Aunque estrictamente hablando el término tautómero es solo aplicable a moléculas individuales, es frecuente encontrarlo en la bibliografía aplicado a los pares de bases AT y GC como hacemos en este trabajo.

solvatada,<sup>18</sup> incluyendo la primera capa de hidratación, según se muestra en la Figura 4; y (iii) en un entorno híbrido<sup>19</sup> como el representado en la Figura 5, que combina la secuencia completa de un fragmento de ADN formado por tres pares de bases con la primera capa de hidratación. Estos modelos están diseñados para dar cuenta de la hidratación de los heteroátomos de las bases, y para añadir a continuación los efectos de apilamiento de bases.

Cuando se tratan sistemas microhidratados hay dos parámetros claves a especificar, el número de moléculas de agua y la posición inicial de las mismas. En nuestro caso hemos usado los datos experimentales de rayos X,<sup>20a</sup> para extraer el número de moléculas de agua presentes en la primera capa de hidratación y las posiciones iniciales de sus átomos de oxígeno. Sin embargo, como ya hemos indicado, los rayos X no permiten localizar los átomos de hidrógeno, por lo que no es posible conocer de este modo la orientación relativa de las moléculas de agua. Esta información la obtenemos de simulaciones de Dinámica Molecular (MD),<sup>20b</sup> compatibles con las observaciones experimentales. Siguiendo el protocolo publicado recientemente por Kumar y Sevilla,<sup>20c</sup> combinamos de este modo los datos de rayos X y MD para reproducir adecuadamente el entorno acuoso de los pares de bases AT y GC.



**Figura 4.** Modelos empleados para simular la hidratación de los pares de bases AT (arriba) y GC (abajo). Los átomos de carbono se muestran en gris, nitrógeno en azul, oxígeno en rojo e hidrógeno en blanco. Por simplicidad, todos los enlaces se representan en forma de varillas.



**Figura 5.** Modelo de fragmento de ADN (dG:dG)<sub>3</sub> con el par de bases GC central solvatado por cinco moléculas de agua. El tratamiento teórico de este modelo combina tres niveles divididos en capa alta (átomos representados mediante esferas), capa media (átomos en varillas) y baja (átomos en alambre). Reimpreso desde *Phys. Chem. Chem. Phys.* **2011**, 13, 14584-14589, con el permiso de The Royal Society of Chemistry.

## Métodos teóricos

La química teórica es, sin lugar a dudas, una de las ramas de la química que más se ha desarrollado en los últimos años. En ella desempeña un papel fundamental la ecuación de Schrödinger, que se escribe abreviadamente de la forma,

$$H\Psi = E\Psi$$

donde  $H$ ,  $E$ ,  $\Psi$  son el operador Hamiltoniano, la energía electrónica del sistema y la función de onda, respectivamente. Resolviendo esta ecuación, es decir, determinando la energía y la función de onda del sistema, es posible conocer el resto de las propiedades del mismo. La aparente simplicidad de la ecuación de Schrödinger encierra, en realidad, un problema extraordinariamente complejo, ya que la función de onda de un sistema con  $n$ -electrones depende de  $4n$  coordenadas:  $3n$  espaciales y  $n$  de espín. La parte del operador Hamiltoniano que contiene los términos que dependen de las posiciones de los electrones, el Hamiltoniano electrónico ( $H_{elec}$ ), que se expresa por

$$H_{elec} = -\sum_i^n \frac{\nabla_i^2}{2} - \sum_i^n \sum_A^N \frac{Z_A}{r_{iA}} + \sum_i^n \sum_{j>i}^n \frac{1}{r_{ij}}$$

constituye de por sí un indicativo de la complejidad de la ecuación de Schrödinger cuya resolución se realiza a partir de la denominada aproximación del campo autoconsistente o de Hartree-Fock.<sup>21</sup>

Existe, sin embargo, una formulación alternativa denominada teoría del funcional de la densidad (DFT), en la que el elemento central no es la función de onda  $\Psi$ , sino la densidad electrónica  $\rho(r)$  dependiente únicamente de las coordenadas espaciales  $x$ ,  $y$ ,  $z$ . Esta simplificación hace que los cálculos DFT sean viables en sistemas complejos como los biológicos. En los métodos DFT, la densidad electrónica  $\rho(r)$  se calcula usando las denominadas ecuaciones de Kohn y Sham,<sup>22</sup> la energía depende de la densidad, es decir, es un funcional de la densidad,  $E[\rho]$ , escrito de la forma

$$E[\rho] = T_s[\rho] + \int \rho(r)v(r)dr + \frac{1}{2} \iint \frac{\rho(r_1)\rho(r_2)}{|r_1 - r_2|} dr_1 dr_2 + E_{xc}[\rho]$$

Sin entrar en detalles, los tres primeros términos en esta expresión están relacionados con la energía cinética de los electrones y con sus interacciones electrostáticas clásicas, mientras que el último término  $E_{xc}[\rho]$ , denominado energía de correlación e intercambio, recoge las interacciones electrónicas no clásicas. Todos los términos son conocidos, salvo el de correlación e intercambio,  $E_{xc}[\rho]$ , cuyas expresiones aproximadas dan lugar a los distintos tipos de funcionales que podemos utilizar. Ninguno de ellos puede considerarse, por tanto, como un funcional exacto, lo que obliga a realizar una elección cuidadosa del mismo. Sin lugar a dudas, el funcional más conocido y usado es el funcional híbrido B3LYP, que proporciona buenos resultados para las geometrías, las frecuencias vibracionales y las energías de un buen número de sistemas moleculares. Este funcional se define como híbrido porque incluye parte de la energía Hartree-Fock, cuyo coeficiente de participación está parametrizado. El funcional



B3LYP no es, sin embargo, la mejor alternativa para el estudio de las interacciones de apilamiento entre las bases del ADN ya que predice interacciones repulsivas entre las bases apiladas en contra de las evidencias experimentales.<sup>23</sup>

En nuestros estudios sobre la mutación espontánea en ADN hemos empleado los métodos de cálculo DFT. La primera etapa ha consistido en buscar y seleccionar el funcional que mejor describe el sistema, comparando entre sí y con los datos experimentales los resultados obtenidos empleando diferentes funcionales para describir las interacciones entre los pares de bases AT y GC.<sup>24-26</sup>

Una vez seleccionado el método pasamos al estudio de la transferencia de protones para identificar los tautómeros responsables de una posible mutación. Dicha transferencia está caracterizada, al igual que cualquier otra reacción química, por las energías de los reactivos y de los productos, y por la energía del estado de transición que los conecta. Matemáticamente, los reactivos y los productos corresponden a mínimos de la superficie de energía potencial que controla los movimientos nucleares del sistema completo, mientras que los estados de transición corresponden a puntos de silla de dicha superficie que son mínimos en todas las direcciones salvo en la que conecta a los reactivos y a los productos, que es la denominada coordenada de reacción. Para localizar todas estas especies se optimizan sus geometrías realizando un cálculo vibracional que proporciona las frecuencias de las mismas. Las estructuras que tienen todas sus frecuencias reales pueden identificarse como mínimos, y por tanto, como especies estables, mientras que las estructuras que poseen sólo una frecuencia imaginaria asociada a algún modo normal de vibración, son estados de transición que conectan entre sí las especies estables. El cálculo vibracional proporciona además las correcciones termodinámicas necesarias para calcular las variaciones de entalpía ( $\Delta H$ ) y de entropía ( $\Delta S$ ) del sistema, a partir de las cuales se obtiene la variación de energía libre de Gibbs ( $\Delta G$ ) de acuerdo con la ecuación básica de la termodinámica,

$$\Delta G = \Delta H - T\Delta S$$

La variación de energía libre proporciona, a su vez, la constante de equilibrio,

$$K_{eq} = e^{-\Delta G/RT}$$

donde  $R$  es la constante de los gases ideales y  $T$  es la temperatura absoluta.

Aunque los métodos DFT permiten abordar sistemas con un gran número de átomos, su potencia a día de hoy no alcanza para tratar de forma integral un sistema como el mostrado en la Figura 5 formado por aproximadamente 200 átomos. Para superar este tipo de limitaciones han aparecido recientemente los métodos ONIOM, desarrollados por Morokuma y colaboradores,<sup>27</sup> en los que se combinan diferentes niveles de descripción, o de teoría, en un mismo sistema. La región química de mayor interés, donde tiene lugar la reacción, queda descrita con un método de alto nivel, y para el resto del sistema se utilizan métodos teóricos de nivel más bajo, lo que da lugar a una división del sistema en diferentes "capas". En nuestro caso, dividimos el fragmento de ADN en tres capas. En la capa alta (método de cálculo más preciso) están los átomos del par de bases central y las moléculas de agua que lo rodean. Para describir esta parte

del modelo hemos elegido la serie de funcionales propuestos recientemente por el grupo de Truhlar en Minnesota, M05-2X y M06-2X, que están especialmente diseñados para describir enlaces no covalentes,<sup>28</sup> lo que muestra la estrecha relación existente entre el método de cálculo y el modelo químico.<sup>29</sup> Los dos pares de bases que confinan al par GC interior forman la capa media, bien descrita por el funcional B97-D que da cuenta de las interacciones débiles características del apilamiento de bases.<sup>29,30</sup> La capa baja contiene las cadenas laterales (grupos azúcar y fosfato), que se tratan a un nivel semiempírico PM3. Finalmente, se completa la descripción del disolvente, más allá de la primera capa de solvatación, usando el denominado modelo del continuo polarizable (PCM). Todos los cálculos se han realizado usando uno de los programas de cálculo de estructura electrónica más extendidos como es Gaussian09.<sup>31</sup>

### Resultados de las simulaciones

Nuestros resultados muestran que el funcional BP86 es el más adecuado para describir la geometría y la estabilidad de los pares de bases AT y GC en el vacío y en disolución (Figura 4).<sup>24-26</sup> En la Tabla 1 damos algunas de las distancias de enlace de estas especies moleculares calculadas teóricamente que, como vemos, concuerdan bastante bien con los valores de referencia experimentales.<sup>32</sup> Los resultados de la Tabla 1 ofrecen además una información valiosa sobre el efecto del medio acuoso en la interacción entre las bases. Concretamente, la hidratación del par AT provoca un alargamiento del enlace de hidrógeno N6-O4, dejando la distancia N1-N3 inalterada, mientras que en el par GC, este alargamiento se observa en el enlace O6-N4, y es menor en los enlaces N2-O2 y N1-N3. A pesar de que estos cambios son relativamente pequeños ( $< 0,1 \text{ \AA}$ ), pueden influir decisivamente en el equilibrio tautomérico de las bases.

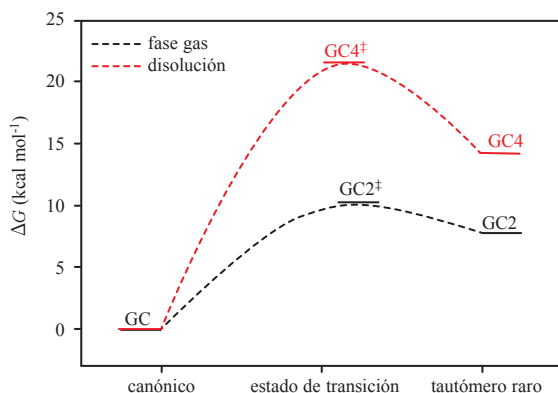
Veamos qué ocurre entonces con los equilibrios tautoméricos. Nuestros cálculos indican que el único par AT estable, tanto en fase gaseosa como en disolución, es el canónico, lo que descarta de partida los tautómeros del par de bases AT como responsables de la mutación espontánea en el ADN. La situación es muy diferente en el par de bases GC. En este caso existe en fase gas, además del par canónico, un tautómero raro termodinámicamente estable, el GC2 (Figura 3). La estabilidad se mide calculando la variación de energía libre  $\Delta G$  desde

**Tabla 1.** Geometrías optimizadas para las formas canónicas AT y GC en fase gaseosa y en disolución. Las distancias aparecen expresadas en angstroms [a].

Enlace	Gas		Disolución	
	Referencia	Calculado	Referencia	Calculado
<b>AT</b>				
N6-O4	2,86	2,89	2,95	2,95
N1-N3	2,83	2,83	2,82	2,82
<b>GC</b>				
O6-N4	2,75	2,75	2,91	2,81
N1-N3	2,90	2,91	2,95	2,95
N2-O2	2,89	2,90	2,86	2,94

[a] Referencias 24 y 25.

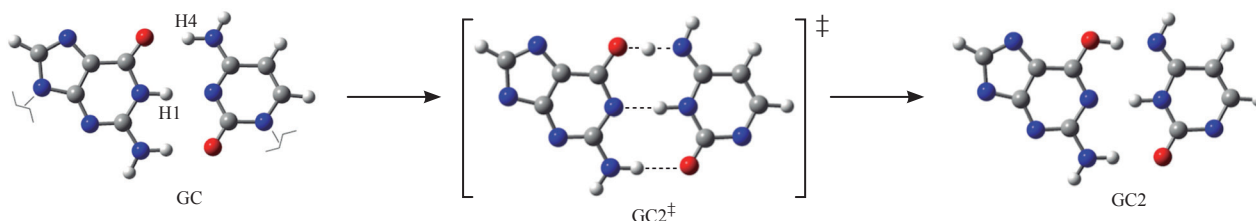
la especie molecular más estable (estructura canónica) usada como referencia, lo que proporciona un valor de 8,28 kcal/mol. En la Figura 6 representamos como cambia  $\Delta G$  al pasar del par canónico GC al tautómero estable CG2, y en la Figura 7 mostramos como evolucionan las estructuras moleculares en este equilibrio tautomérico. Como vemos en esta última figura, la forma canónica GC y el tautómero GC2 están conectados por un único estado de transición  $GC2^\ddagger$  caracterizado por una frecuencia imaginaria asociada a la tensión del enlace H4–N4. La estructura de  $GC2^\ddagger$  indica que un primer paso la guanina cede su protón H1 a la citosina, ya que como vemos en el estado de transición el protón H1 se encuentra enlazado a la base citosi-



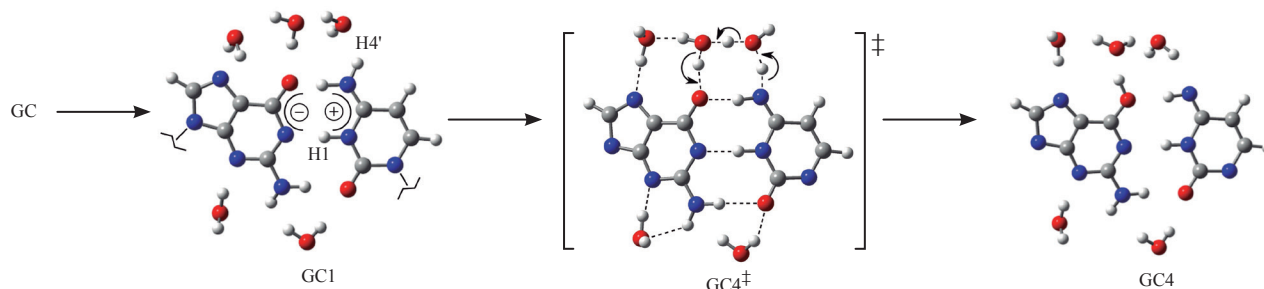
**Figura 6.** Perfil de energías libres de Gibbs para el equilibrio tautomérico en el par GC, calculadas tomando como valor de referencia la forma canónica. Los superíndices  $\ddagger$  indican el estado de transición que conecta la forma canónica con el correspondiente tautómero raro.

na. La neutralidad eléctrica puede recuperarse bien mediante la transferencia inversa del protón H1 desde la citosina a la guanina, volviendo a la especie canónica de partida, o bien mediante la transferencia del protón H4, también de la citosina, a la guanina, lo que da lugar al tautómero GC2. ¿Cuáles son las condiciones favorables para que ocurra esta segunda alternativa? Según los datos de la Tabla 1, la clave está en la menor distancia del enlace de hidrógeno O6–N4 (2,75 Å), que facilitaría la transferencia del protón H4 a lo largo del enlace O6–H4–N4. En fase gaseosa, por tanto, el par GC promueve la formación del tautómero raro GC2 mediante un equilibrio de transferencia de protones asincrónico concertado.

Consideremos a continuación las simulaciones realizadas en disolución. Estas indican para empezar que en este medio el tautómero GC2 deja de ser estable. Como hemos comentado antes, el enlace O6–N4, cuya fortaleza parece clave para promover la doble transferencia protónica, resulta ser precisamente el más afectado por las moléculas de agua del disolvente. La solvatación de GC debilita, pues, este enlace impidiendo la formación del tautómero GC2. Encontramos ahora, sin embargo, que en disolución el tautómero GC4 pasa a ser una especie estable ( $\Delta G = 14,22$  kcal/mol), que aparece como consecuencia de un mecanismo catalizado por las moléculas de agua que rodean al par. Este es uno de los resultados más destacables de nuestro trabajo, y merece un análisis más detallado. A diferencia de lo que sucede en fase gaseosa, en el medio acuoso la solvatación de los pares de bases debilita los enlaces de hidrógeno externos, lo que impide la transferencia directa de los protones, pero estas mismas moléculas de agua catalizan a su vez el equilibrio tautomérico alternativo. Como se observa en la Figura 8, la reacción evoluciona a través del



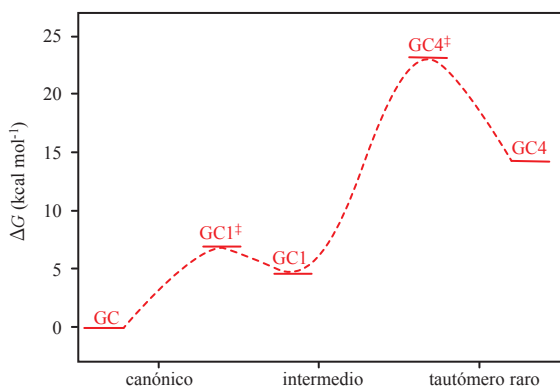
**Figura 7.** Mecanismo de transferencia directa de protones en el par de bases GC en gas. H1 y H4 son los protones transferidos. La forma canónica GC y el tautómero raro GC2 están conectados por un único estado de transición  $GC2^\ddagger$ . Para el estado de transición  $GC2^\ddagger$  se muestran los enlaces de hidrógeno mediante líneas discontinuas.



**Figura 8.** Mecanismo de transferencia de protones asistido por las moléculas de agua. H1 y H4' son los protones transferidos. Por simplicidad solo mostramos los átomos de la capa alta ONIOM. Para el estado de transición  $GC4^\ddagger$  se muestran los enlaces de hidrógeno mediante líneas discontinuas y el sentido de la transferencia protónica mediante flechas. Adaptado desde *Phys. Chem. Chem. Phys.* **2011**, 13, 14584-14589, con el permiso de The Royal Society of Chemistry.

estado de transición  $GC4^\ddagger$  que desemboca en la formación del tautómero GC4. Analizando los enlaces de hidrógeno existentes en el estado de transición  $GC4^\ddagger$  comprobamos que las moléculas de agua participan activamente como catalizadoras del equilibrio tautomérico aceptando y cediendo protones.

Vemos, pues, que la hidratación influye de forma decisiva en el equilibrio tautomérico, pero para simular más fielmente la realidad biológica hay que seguir mejorando el modelo químico. Con este fin utilizamos el fragmento de ADN de la Figura 5, que combina los efectos de hidratación y de apilamiento de los pares de bases. Los resultados de nuestras simulaciones indican que en estas condiciones son estables, además de la estructura canónica GC, los tautómeros GC1, GC2 y GC4. La estabilización del intermedio GC1 hace que la transferencia transcurra por etapas. Así, dentro del fragmento de ADN pueden tener lugar, en principio, los dos equilibrios discutidos anteriormente, la transferencia directa de los protones H1 y H4 mediante el mecanismo asincrónico que acaba en el tautómero GC2 (Figura 7), y la transferencia del protón H4' asistida por las moléculas de agua, que conduce al tautómero GC4 (Figura 8). Esta segunda transferencia asistida, cuyo perfil de reacción se muestra en la Figura 9, resulta ser el único camino de reacción termodinámicamente viable. Las variaciones de energía calculadas para el estado de transición  $GC4^\ddagger$  ( $\Delta G^\ddagger = 22,69$  kcal/mol) y para el tautómero GC4 ( $\Delta G = 14,10$  kcal/mol) indican además que este proceso está claramente desplazado hacia la forma canónica.



**Figura 9.** Perfil de las energías libres de Gibbs para el equilibrio tautomérico en el modelo hidratado de ADN. Las energías relativas están calculadas tomando como valor de referencia la forma canónica.

Afortunadamente para nuestra supervivencia el ADN es una molécula muy estable. Experimentalmente Topal y Fresco<sup>33</sup> han estimado que la frecuencia de mutación espontánea es del orden de  $10^{-8}$ – $10^{-10}$ , lo que indica que el equilibrio tautomérico se encuentra claramente desplazado hacia la forma canónica. Dado que los tautómeros raros son una de las fuentes de error en el ADN, pero no la única, su peso final en la aparición de mutaciones dependerá de la constante de equilibrio que controle su formación, y que debe estar forzosamente por debajo de los valores umbrales  $10^{-8}$ – $10^{-10}$ . Hasta ahora los estudios teóricos fallaban en la estimación de las variaciones de energía libre de los tautómeros raros, y predecían constantes de formación varios órdenes de magnitud superiores a los umbrales debido al empleo de modelos

demasiado simplificados. Por ejemplo, cuando se consideran los pares de bases aislados, el equilibrio tautomérico predicho es del orden de  $10^{-6}$ , que excede los valores reales de mutación espontánea en varios órdenes de magnitud.<sup>16b</sup> Con el modelo más elaborado que incorpora los efectos de solvatación y apilamiento (Figura 5), nuestras simulaciones predicen una constante de formación del tautómero GC4 de  $4,54 \times 10^{-11}$ . Este valor es, pues, coherente con el observado *in vivo* de  $10^{-8}$ – $10^{-10}$ , y proporciona la primera estimación cuantitativa del impacto real del equilibrio tautomérico en la mutación.<sup>29</sup> El tautómero GC4 se sitúa, como hemos visto, 14,10 kcal/mol por encima de la forma canónica. A pesar de esta gran diferencia de energía, la mutación puede acabar teniendo un impacto importante debido, por un lado, a que el tautómero GC4 es una mutación difícilmente detectable por los mecanismos de reparación del ADN ya que implica sólo un pequeño cambio respecto a la base canónica (posiciones de los protones H1 y H4'), y por otro lado a que este tautómero cumple con los requisitos energéticos propuestos por Florián y Leszczyński para inducir un error permanente en la secuencia genética.<sup>13</sup> El primero de estos requisitos es que la barrera energética que separa al tautómero raro de la forma canónica sea menos de 28 kcal/mol, lo que se cumple para el tautómero GC4 puesto que su barrera es de 22,69 kcal/mol. El segundo requisito es que la base permanezca en su forma no canónica durante el tiempo que tarda los doble hélice en replicarse. El tiempo de supervivencia del tautómero GC4 dependerá de la altura de la barrera para la transferencia inversa dada por la diferencia de energía entre  $GC4^\ddagger$  y GC4 (véase Figura 9). Según Florián y Leszczyński esta barrera debe de ser mayor de 3 kcal/mol, lo que también se cumple para el tautómero GC4 con un valor de 8,59 kcal/mol. Estos datos confirman pues nuestra hipótesis de que el tautómero GC4 tiene una implicación clave en la mutación espontánea del ADN.<sup>29</sup>

## Conclusiones y perspectivas

El estudio teórico de los procesos de transferencia de protones entre los pares de bases adenina-timina (AT) y guanina-citosina (GC) nos ha permitido desvelar el papel que desempeñan los llamados tautómeros raros en la aparición de mutaciones espontáneas en el ADN. Los resultados de nuestras simulaciones indican que el entorno biológico juega un papel activo y decisivo en la estabilidad de la forma canónica: el medio acuoso cataliza la transferencia de protones y el apilamiento de bases modifica el perfil de la reacción. Entre todas las formas no canónicas, hemos identificado al tautómero GC4 como uno de los posibles responsables de la aparición de mutaciones espontáneas en el ADN, destacando el papel tan importante que juegan las bases que rodean al par de base donde tiene lugar la transferencia. Finalmente conviene resaltar que el equilibrio tautomérico puede variar dependiendo de cuales sean las bases adyacentes, y ser alterado por la presencia de metales o la acción de radiaciones ionizantes,<sup>17b,33</sup> lo que abre la puerta a futuras investigaciones sobre el impacto de la secuencia genética en la aparición de mutaciones espontáneas. El trabajo presentado demuestra la importancia de este proceso biológico y la necesidad de continuar explorando los detalles moleculares de su origen.

## Agradecimientos

J. P. Cerón-Carrasco agradece la beca concedida por la Fundación Séneca, Agencia de Ciencia y Tecnología de la Región de Murcia, a través del Programa de Formación Posdoctoral de Personal Investigador. D. Jacquemin agradece la financiación de European Research Council (ERC) y *Région des Pays de la Loire* a través de los programas Starting Grant (Marches–278845) y *recrutement sur poste stratégique*, respectivamente. Este trabajo está financiado en parte por el Ministerio de Ciencia e Innovación, proyectos CTQ2011-25872 y CONSOLIDER CSD2009-00038 y la Fundación Séneca, Agencia de Ciencia y Tecnología de la Región de Murcia, proyecto 08735/PI/08.

## Bibliografía

1. a) I. González-Ramírez, D. Roca-Sanjuán, T. Climent, J. Serrano-Pérez, M. Merchán, L. Serrano-Andrés, *Theor. Chem. Acc.* **2011**, *128*, 705–711. b) D. Roca-Sanjuán, G. Olaso-González, I. González-Ramírez, L. Serrano-Andrés, M. Merchán, *J. Am. Chem. Soc.* **2008**, *130*, 10768–10779. c) A. Kumar, M. D. Sevilla, *Chem. Rev.* **2010**, *12*, 7002–7023. d) M. Boggio-Pasqua, G. Groenhof, L. V. Schäfer, H. Grubmüller, M. A. Robb, *J. Am. Chem. Soc.* **2007**, *129*, 10996–10997.
2. G. P. Pfeifer, *Photochem. Photobiol.* **1997**, *65*, 270–283.
3. J. M. Matxain, M. Ristilä, A. Strid, L. A. Eriksson, *J. Phys. Chem. A* **2006**, *110*, 13068–13072.
4. J. D. Watson, F. H. C. Crick, *Cold Spring Harbor Symposia on Quant. Biol.* **1953**, *18*, 123–131.
5. T. H. Noel Ellis, J. M. I. Hofer, G.M. Timmerman-Vaughan, C. J. Coyne, R.P. Hellens, *Trends Plant Sci.* **2011**, *16*, 1360–1385.
6. M. Nei, M. Nozawa, *Genome Biol. Evol.* **2011**, *3*, 812–829.
7. A. D. Hershey, M. Chase, *J. Gen. Physiol.* **1952**, *36*, 39–56.
8. a) J. D. Watson, *La doble hélice*, Salvat Editores, **2000**. b) F. H. C. Crick, *Qué loco propósito*, Tusquets Editores, **2011**. c) M. H. F. Wilkins, *The third man of the double helix: the autobiography of Maurice Wilkins*, Oxford University Press, **2005**. d) B. Maddox, *Rosalind Franklin: the dark lady of DNA*, Harper Perennial, **2003**.
9. J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 737–738.
10. M. H. F. Wilkins, A. R. Stocker, H. R. Wilson, *Nature* **1953**, *171*, 738–740.
11. R. E. Franklin, R. Gosling, *Nature* **1953**, *171*, 740–741.
12. C. Alemán, *Chem. Phys.* **2000**, *253*, 13–19.
13. a) J. Florián, V. Hroudá, P. Hobza, *J. Am. Chem. Soc.* **1994**, *116*, 1457–1460. b) J. Florián, J. Leszczynski, *J. Am. Chem. Soc.* **1996**, *118*, 3010–3017.
14. P. O. Löwdin, *Rev. Mod. Phys.* **1963**, *35*, 724–732.
15. G. Villani, *Chem. Phys.* **2005**, *316*, 1–8.
16. a) V. Gullar, A. Douhal, M. Moreno, J. M. Lluch, *J. Phys. Chem. A* **1999**, *103*, 6251–6256. b) L. Gorb, Y. Podolyan, P. Dziekonski, W. A. Sokalski, J. Leszczynski, *J. Am. Chem. Soc.* **2004**, *126*, 10119–10129. c) A. Pérez, M. E. Tuckerman, H. P. Hjalmarson, O. A. von Lilienfeld, *J. Am. Chem. Soc.* **2010**, *132*, 11510–11515. d) A. Kumar, V. Pottiboyina, M. D. Sevilla, *J. Phys. Chem. B*, **2011**, *115*, 15129–15137.
17. a) J. Bertrán, A. Oliva, L. Rodríguez-Santiago, M. Sodupe, *J. Am. Chem. Soc.* **1998**, *120*, 8159–8176. b) M. Noguera, J. Bertrán, M. Sodupe, *J. Phys. Chem. B* **2008**, *112*, 4817–4825. c) M. Noguera, R. Ríos-Font, L. Rodríguez-Santiago, X. Solans-Monfort, A. Oliva, J. Bertrán, M. Sodupe, *Theor. Chem. Acc.* **2009**, *123*, 105–111. d) A. Gil, V. Branchadell, J. Bertrán, A. Oliva, *J. Phys. Chem. B*, **2009**, *113*, 4907–4914. e) C. Acosta-Silva, V. Branchadell, J. Bertrán, A. Oliva, *J. Phys. Chem. B* **2010**, *114*, 10217–10227.
18. a) C. Fonseca-Guerra, F. Bickelhaupt, J. Snijders, E. Baerends, *J. Am. Chem. Soc.* **2000**, *122*, 4117–4128. b) T. van der Wijst, C. Fonseca-Guerra, M. Swart, F. Bickelhaupt, *Chem. Phys. Lett.* **2006**, *426*, 415–421.
19. H.-Y. Chen, C.-L. Kao, S. C. N. Hsu, *J. Am. Chem. Soc.* **2009**, *131*, 15930–15938.
20. a) B. Schneider, H. M. Berman, *Biophys. J.* **1995**, *69*, 2661–2669. b) P. Auffinger, E. Westhof, *J. Mol. Biol.* **2000**, *300*, 1113–1131. c) A. Kumar, M. D. Sevilla, S. Suhai, *J. Phys. Chem. B* **2008**, *112*, 5189–5198.
21. V. Fock, *Z. Physik.* **1930**, *61*, 126–148.
22. W. Kohn, L.J. Sham, *Phys. Rev. A* **1965**, *140*, 1133–1138.
23. Y. Zhao, N. E. Schultz, D. G. Truhlar, *J. Chem. Theory Comput.* **2006**, *2*, 364–382.
24. J. P. Cerón-Carrasco, A. Requena, C. Michaux, E. A. Perpète, D. Jacquemin, *J. Phys. Chem. A* **2009**, *113*, 7892–7898.
25. J. P. Cerón-Carrasco, A. Requena, J. Zúñiga, C. Michaux, E. A. Perpète, D. Jacquemin, *J. Phys. Chem. A* **2009**, *113*, 10549–10556.
26. J. P. Cerón-Carrasco, A. Requena, E.A. Perpète, C. Michaux, D. Jacquemin, *Chem. Phys. Lett.* **2009**, *484*, 164–168.
27. a) F. Maseras, K. Morokuma, *J. Comput. Chem.* **1995**, *16*, 1170–1179. b) M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber, K. Morokuma, *J. Phys. Chem.* **1996**, *100*, 19357–19363.
28. Y. Zhao, D. G. Truhlar, *Acc. Chem. Res.* **2008**, *41*, 157–167.
29. J. P. Cerón-Carrasco, J. Zúñiga, A. Requena, E. A. Perpète, C. Michaux, D. Jacquemin, *Phys. Chem. Chem. Phys.* **2011**, *13*, 14584–14589.
30. J. P. Cerón-Carrasco, A. Requena, D. Jacquemin, *Theor. Chem. Acc.* **2012**, *131*, 1188–1196.
31. M. J. Frisch *et al.*, Gaussian 03, Revision C.02; Gaussian 09, Revision A.02. GAUSSIAN, Inc., Wallingford, CT. **2009**.
32. a) J. Spöner, P. Jurecka, P. Hobza, *J. Am. Chem. Soc.* **2004**, *126*, 10142–10151. b) P. Jurecka, J. Spöner, J. Cerny, P. Hobza, *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993. c) N. Seeman, J. Rosenberg, F. Suddath, J. Kim, A. Rich, *J. Mol. Biol.* **1976**, *104*, 109–144. d) J. Rosenberg, N. Seeman, R. Day, A. Rich, *J. Mol. Biol.* **1976**, *104*, 145–167.
33. M. D. Topal, J. R. Fresco, *Nature* **1976**, *263*, 285–289.
34. a) J. P. Cerón-Carrasco, A. Requena, E. A. Perpète, C. Michaux, D. Jacquemin, *J. Phys. Chem. B* **2010**, *114*, 13439–13445. b) J. P. Cerón-Carrasco, D. Jacquemin, *Chem. Phys. Chem.* **2011**, *12*, 2615–2623.