Proton transfer in Guanine-Cytosine base pairs in B-DNA.

Diego Soler-Polo,[†] Jesús I. Mendieta-Moreno,[†] Daniel G. Trabada,[†] Jesús Mendieta,[‡] and José Ortega[*,†]

[†]*Departamento de Física Teórica de la Materia Condensada and Condensed Matter Physics Center (IFIMAC), Facultad de Ciencias, Universidad Autónoma de Madrid, E-28049 Madrid, Spain*

[‡]*Departamento de Biotecnología, Universidad Francisco de Vitoria, E-28223 Pozuelo de Alarcón (Madrid), Spain*

E-mail: jose.ortega@uam.es

Abstract

A double proton transfer reaction in a Guanine-Cytosine (GC) base pair has been proposed as a possible mechanism for rare tautomer (G*C*) formation and thus a source of spontaneous mutations. We analyze this system with *free energy* calculations based on extensive Quantum Mechanics / Molecular Mechanics simulations to properly consider the influence of the DNA biomolecular environment. We find that, although the G*C* rare tautomer is metastable in the gas phase, it is completely unstable in the conditions found in cells. Thus, our calculations show that a double proton reaction cannot be the source of spontaneous point mutations. We have also analyzed the intra-base H transfer reactions in Guanine. Our results show that the DNA environment

gives rise to a large free energy difference between the rare and canonical tautomers. These results show the key role of the DNA biological environment for the stability of the genetic code.

Proton-transfer reactions are ubiquitous throughout biochemical reactions and have therefore arisen much interest both in experimental and theoretical studies. In particular, the significance of proton-transfer reactions in DNA was already highlighted by Watson and Crick in 1953,[1] as they noted that "spontaneous mutations may be due to a base occasionally occurring in one of its less likely tautomeric forms." The reason for this is that tautomerization of a nucleotide alters the adenine-thymine/guanine-cytosine (A-T/G-C) correspondence. The rare tautomers of the DNA bases differ from their canonical forms by the position of the H atoms, see Figure 1. Thus, for instance, the rare tautomer G* does not pair up with cytosine, C, but instead with thymine, T.[2] Therefore, if at the time of DNA replication the polymerase runs over the rare tautomeric G* form of guanine, a substitution mutation will appear in the DNA sequence. It is then clear why it would be of interest to know the probability of these tautomerizations to take place by proton transfer, the mean life of the rare tautomeric forms and the reaction mechanisms of these processes. After many years of investigations on this topic[2–20] the relationship between proton transfer reactions and spontaneous point mutations is still under debate. The main reason behind this is the difficulty to properly assess the influence of the complex biomolecular environment on these reactions.

The rare tautomers of the different DNA bases can be obtained from their canonical forms via a double proton transfer (DPT) along the DNA hydrogen bonds,[2] see Figure 1. In the gas phase (*i.e.* considering only the interaction between the two DNA bases), different studies for the potential energy or free energy landscapes indicate that there is no local minimum corresponding to a A*T* rare tautomer[3,7,10] (or there is only a very shallow minimum[12,14]). Thus, DPT do not contribute to the presence of A* or T* rare tautomers. In the case of GC, however, the rare tautomer G*C* shown in Figure 1 is metastable in vacuo, with values for the direct and reverse energy barriers between the canonical and rare tautomers that could

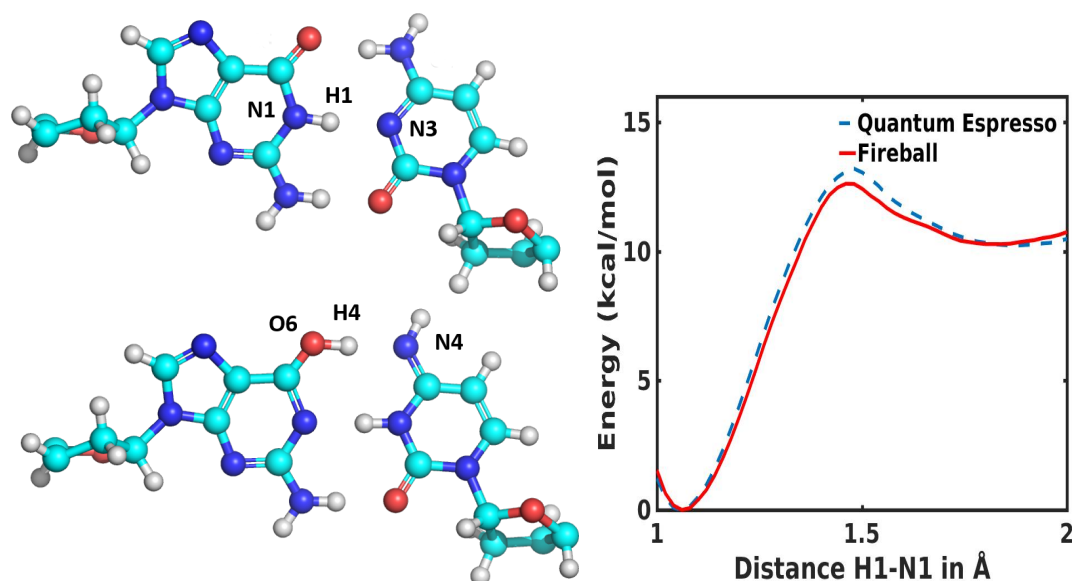yield the formation of spontaneous point mutations.[3,7,14]

Figure 1: Proton transfer and rare tautomers. Left: the canonical (GC, top) and rare (G*C*, bottom) tautomers; in the GC $\leftrightarrow$ G*C* DPT reaction the H atoms H1 and H4 jump along the N1-H1-N3 and O6-H4-N4 hydrogen bonds. Right: potential energy (in kcal/mol) for the GC $\leftrightarrow$ G*C* reaction as a function of the reaction coordinate $d_1 = |\vec{r}_{H1} - \vec{r}_{N1}|$ (in Å), calculated using the QUANTUM ESPRESSO and FIREBALL DFT codes, see text.

The influence of the solvent, DNA backbone and base stacking on these reactions has been analyzed in a number of works.[4,7,10,16,17,19–21] In particular, the effect of the aqueous environment has been investigated using implicit models for the solvent or including a few water molecules around the DNA bases (microhydration) explicitly in the calculation.[7,10,19,20] These studies indicate that the water molecules around the DNA bases have an important effect changing the relative stability of rare and canonical tautomers; also, the water molecules can participate in proton transfer reactions, interchanging protons with the DNA bases in a Grotthus-like fashion.[10,17] It is also interesting to mention that in the original idea proposed by Löwdin for DPT in DNA,[2] the transitions between canonical and rare tautomers are due to quantum tunneling of the protons along the hydrogen bonds. There are some recent studies[22–24] on quantum nuclear effects using path integral molecular dynamics (PIMD).[25] A recent study[26] using an open quantum systems approach indicates that quantum tunneling does not play any role in DPT reactions in DNA, so that the jump of the protons should

be entirely due to thermal activation. The combination of having a large system and a high temperature makes the quantum nuclear effects negligible.

A different situation arises when one considers proton-transfer in excited states or in ionized base-pairs. Several studies show how the barrier in this case drops dramatically and the inter-base tautomerization can become spontaneous.[27–31] These results suggest that proton-transfer could be a deactivation pathway for excited base-pairs,[27,32] which would subsequently go back to the canonical tautomer after decaying to the ground state. Thus, the tautomerization by proton-transfer could be considered, more than a source of spontaneous mutations, a defensive mechanism against radiation damage.[33,34]

Intra-base proton transfer can also induce the formation of rare tautomers. In this case, the energy difference between the canonical and rare tautomers is small in the gas phase, but there is a high energy barrier for the transition between the different isomers, *e.g.* G $\leftrightarrow$ G*; this barrier is reduced when one or two water molecules are included in the calculation.[5,6,35]

DNA is a very large macromolecule, with a complex atomic structure. At the conditions found in the cells, it generally presents the double helix B-DNA form that is stabilized by the interactions with the aqueous solvent; also, at these conditions DNA is a dynamic system that displays significant atomic motions. Thus, the DNA environment, *i.e.* solvent, DNA backbone, double helix interactions and atomic dynamics at room temperature (RT), must have an important impact on the properties of the DNA bases,[34] and in particular in the H transfer reactions taking place on them. In order to properly assess the influence of this environment the free energy of the system must be considered.[34,36–38]

In this work we analyze how a realistic biomolecular environment, including the double helix DNA backbone in B-form and water molecules and counterions of the solvent, modifies the *free energy* of the tautomerization reactions for the GC base pair at RT. In our approach we combine free energy techniques and extensive Quantum Mechanics / Molecular Mechanics (QM/MM) molecular dynamics (MD) simulations. Since the calculation of the free energy at RT requires an appropriate sampling of a large number of different atomic

configurations, it is particularly important that the QM/MM method presents both a good accuracy and a good computational efficiency. We use FIREBALL / AMBER,[39] a QM/MM technique that combines the AMBER[40,41] MM force fields and simulation programs with the Density Functional Theory (DFT) method FIREBALL.[42–44] The computational efficiency of this approach is related to the use of an optimized basis set of Numerical Atomic-like Orbitals (NAOs)[45,46] together with a practical tabulation-interpolation scheme.[42,45] The basis set used in the present work was optimized comparing with benchmark results for the dissociation curves for different molecular complexes relevant to biomolecular systems (the S66x8[47] and IHBx8[48] data sets). We use the BLYP exchange-correlation functional[49,50] and the D3-BJ correction[51,52] to include dispersion effects. See Methods for further details. Figure 1 (right) shows the potential energy profile for GC in the gas-phase as a function of the distance N1-H1, in comparison with the accurate plane-waves result, calculated using the code QUANTUM ESPRESSO (QE);[53,54] in both calculations we use the BLYP functional with D3-BJ dispersion corrections. The good agreement between both profiles shows that we can use FIREBALL / AMBER in our calculations, with a critical speedup in the MD simulations while retaining a good accuracy, thus allowing for the extensive sampling required for the free energy calculations (see below for further comparisons).
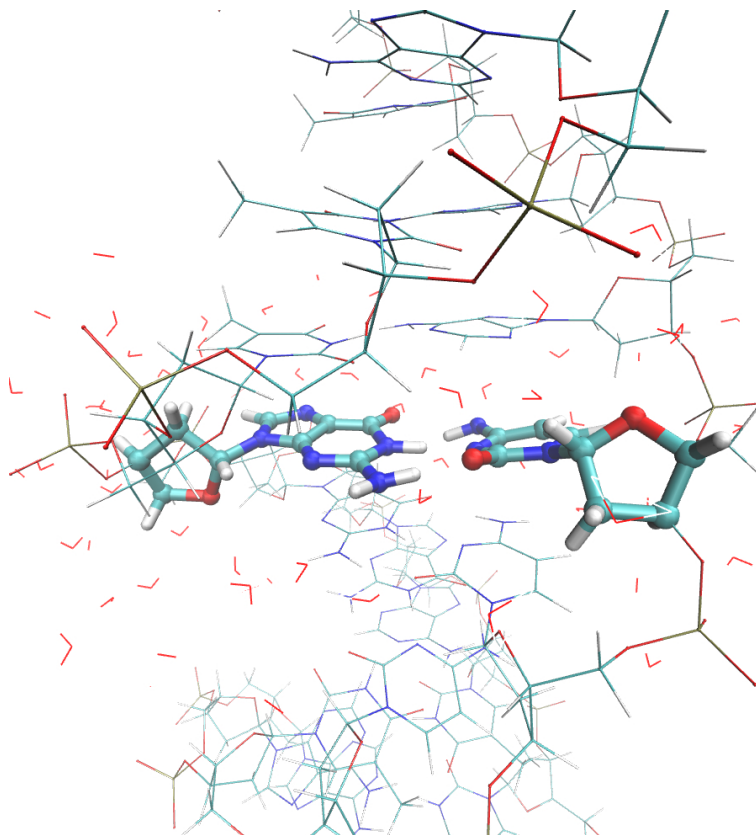
Figure 2: Snapshot of the system. The DNA double helix is simulated using two DNA strands that contain 12 DNA base pairs. The QM region, which corresponds to the complementary guanine and cytosine nucleosides (51 atoms in total), is highlighted. The rest of the atoms in the DNA strands, water molecules and counterions (Na^+) are included in the MM region. For clarity, in this figure the solvent is represented by only a few water molecules around the QM region.

Figure 2 shows the system used in our QM/MM free energy calculations. The DNA double helix is simulated by means of two DNA strands that contain 12 base pairs (sequence: TTAGGGTTAGGG). The QM region consists of the guanine and cytosine nucleobases and corresponding deoxyriboses, 51 atoms in total, in the middle of the DNA double helix; in some calculations we have also included three water molecules in the QM region, see below. The rest of the DNA, water molecules and counterions (Na^+) are included in the MM region, that contains $\sim 1.1 \times 10^4$ atoms. The overall charge of the system is neutral. We employed the ff14SB force field[55] incorporated in the AMBER package, TIP3P water molecules[56] and the Langevin thermostat[57] for the simulation of the Canonical Ensemble. The initial

positions of the water molecules and Na^+ counterions are set by the *tleap* software included in AMBER,[41] and the system is then thermalized at T = 300 K, first with a long MM simulation (20 ns) followed by a 100 ps QM/MM simulation from which we extract different conformations as seeds for our production runs, see below.

We have used three different free energy methods; two of them, the Weighted Histogram Analysis Method (WHAM)[58] and Umbrella Integration (UI),[59] are based on the Umbrella Sampling (US) technique, while the third one is based on Steered Molecular Dynamics (SMD) and the Jarzynski equality[60,61] (JE), see below. We have used as reaction coordinates (see Figure 1) $d_1 = |\vec{r}_{H1} - \vec{r}_{N1}|$, $d_2 = |\vec{r}_{H4} - \vec{r}_{N4}|$, $q_1 = |\vec{r}_{H1} - \vec{r}_{N1}| - |\vec{r}_{H1} - \vec{r}_{N3}|$, $q_2 = |\vec{r}_{H4} - \vec{r}_{N4}| - |\vec{r}_{H4} - \vec{r}_{O6}|$, and $d_3 = |\vec{r}_{H1} - \vec{r}_{O6}|$. Other reaction coordinates such as $d_1 + d_2$, $d_1' = |\vec{r}_{H1} - \vec{r}_{N3}|$ or $d_2' = |\vec{r}_{H4} - \vec{r}_{O6}|$ were also tested. Given a reaction coordinate, which is a function of the coordinates of the atoms in the system, $\hat{q} \equiv \hat{q}(\{\vec{r}_\alpha\})$, the free energy profile for different values, q, of the reaction coordinate is

$$F(q) = -k_B T \ln P(q), \tag{1}$$

where $P(q)$ is the probability density,

$$P(q) = \left\langle \delta(\hat{q} - q) \right\rangle \equiv \frac{1}{Z} \int d^N r \, \exp\left(-\beta U\right) \delta(\hat{q} - q); \tag{2}$$

in this equation the integral extends over the coordinates of all the atoms in the system, $\{\vec{r}_\alpha\}$, Z is the configurational partition function, U is the potential energy, $U \equiv U(\{\vec{r}_\alpha\})$, and $\beta = 1/(k_B T)$. Similarly, a free energy map $F(q_1, q_2)$ can be defined as a function of two reaction coordinates, etc.

In long-enough simulations ergodicity holds and $P(q)$ could be obtained directly by *e.g.* constructing the histograms associated to the different values that $\hat{q}$ takes during the dynamics. In practice, however, the time scales accessible in MD or Monte Carlo simulations are almost always far below the times for which ergodicity can be observed. In the US

technique[62] the idea is to enhance sampling at different values of the reaction coordinate by adding artificial "bias" potentials V_j. Different simulations are performed for a set of "windows", $\{V_j\}$, and all together the different US simulations provide appropriate sampling over all the relevant values of the reaction coordinate. This information, however, is biased by the V_j potentials, and an unbiasing procedure is required (see Methods).

A different approach to calculate free energy differences consists of performing a set of SMD trajectories, computing the work for each trajectory and then using the non-equilibrium Jarzynski equality (JE):[60,61]

$$\exp\left(-\beta\Delta F\right) = \langle\exp\left(-\beta W\right)\rangle, \tag{3}$$

where W is the work associated to a trajectory and the $\langle\,\rangle$ denote the average over trajectories starting from different configurations taken from a canonical ensemble for the the initial state.
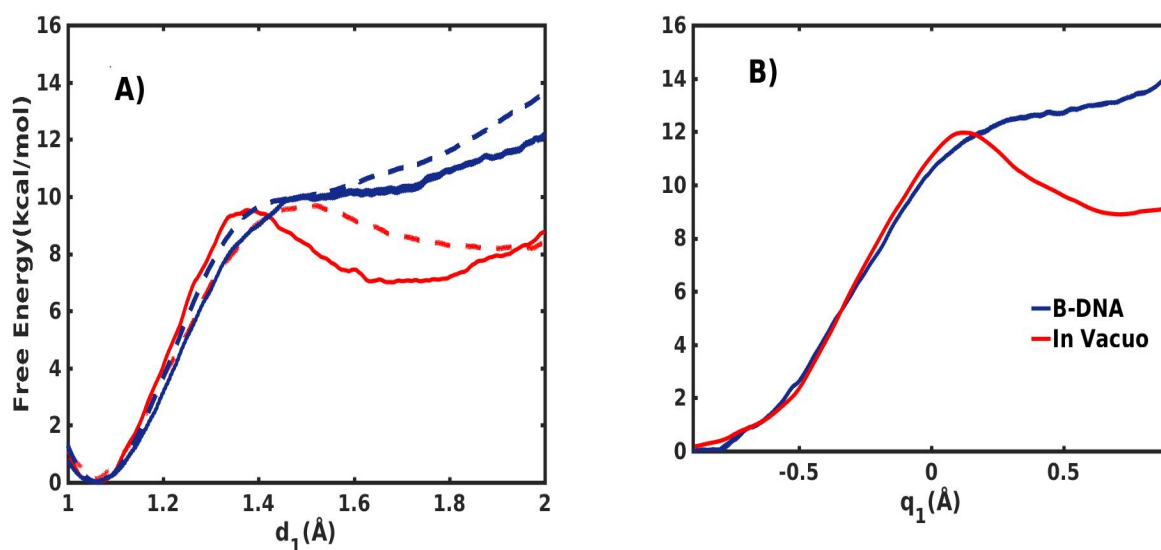

Figure 3: Free energy profiles in the gas-phase (red lines) and in the DNA environment (blue lines), as a function of the reaction coordinates d_1 (a) or q_1 (b). In (a) the dashed lines correspond to the case in which the 3 water molecules closer to the O6-H4-N4 hydrogen bond are included in the QM region, see text.

Figure 3 shows the free energy profiles for the GC $\leftrightarrow$ G*C* reaction in the gas phase, F_G, and in the DNA environment, F_E, as a function of the d_1 or q_1 reaction coordinates. In

both figures the canonical tautomer is on the left and the rare tautomer on the right. The gas phase free energy profile $F_G(d_1)$ is shifted to smaller values as compared to the potential energy profile shown in Figure 1: the energy barrier is reduced from 13.1 kcal/mol to 9.9 kcal/mol and the energy difference between G*C* and CG changes from 10.4 kcal/mol to 7.5 kcal/mol; also, the position of the top of the barrier and the G*C* equilibrium distance are shifted to slightly smaller d_1 values. Both $F_G(d_1)$ and $F_G(q_1)$ yield very similar results, with slightly lower free energy values for the d_1 case due to the extra freedom that this reaction coordinate provides. For comparison, we mention that $F_G(q_1)$ has also been analyzed by Xiao et al.[14] by means of DFT calculations with the BLYP exchange-correlation functional, with very similar results. In particular, we obtain a slightly smaller free energy barrier (12.2 vs. 13.6 kcal/mol) and energy difference between G*C* and GC (8.9 vs. 9.4 kcal/mol), probably due to the inclusion of dispersion corrections[51,52] in our calculations.

Figure 3 shows that the rare tautomer G*C* is metastable in the gas phase, with a free energy barrier of $\sim$ 2.4–3.3 kcal/mol for the reverse reaction. The DNA environment has a profound impact on the stability of the G*C* tautomer, changing qualitatively the free energy profile to the point that the rare tautomer is no longer a metastable configuration. This important result is confirmed by the free energy map shown in Fig. 4: $F_E(q_1, q_2)$ presents only one minimum, corresponding to the canonical tautomer, with a sharp uphill slope when the system goes away from that minimum. Neither the rare tautomer, nor the single proton transfer configurations (upper-left and lower-right regions), are accessible structures. In particular, the G*C* tautomer (upper-right corner in the map) is completely unstable. We have further corroborated this result performing dozens of MD simulations starting from configurations corresponding to the G*C* rare tautomer (conveniently thermalized using bias potentials). These simulations show how the system reverts to the canonical tautomer in an ultra-fast motion, in full agreement with the free energy landscape shown in Figure 4. Thus, a DPT reaction cannot be a source of spontaneous point mutations due to the appearance of G* or C* rare tautomers. We have also analyzed the evolution of the atomic

charges as a function of the reaction coordinates, both in SMD and US simulations, finding that the charges on the H1 and H4 atoms remain practically constant across the map, with a value of $\sim +0.15e$ (sum of the proton charge, $+e$ and an electron charge of $\sim -0.85e$), see Supplementary Material. Therefore, it would be more appropriate to speak of hydrogen atom transfer in this case.[16]
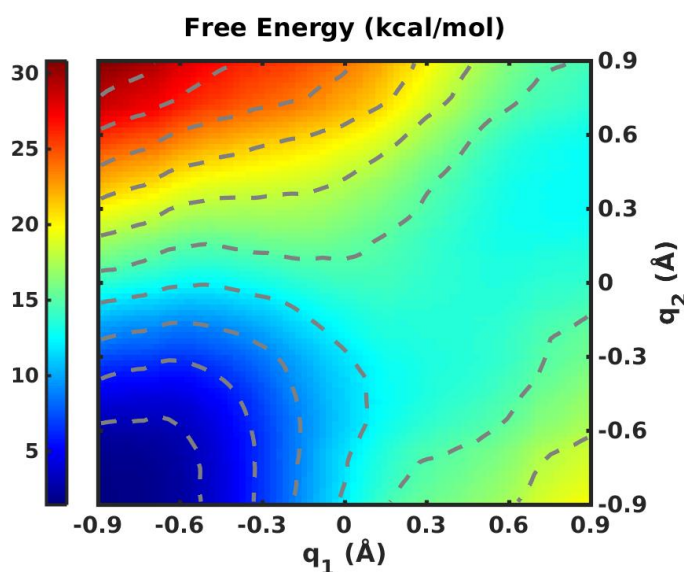

Figure 4: Free Energy (kcal/mol) map in the DNA environment as a function of the reaction coordinates q_1 and q_2 (in Å). There is only one local minimum, the canonical tautomer GC.

We have analyzed in more detail the effect of the few water molecules surrounding the hydrogen bonds that participate in the DPT. In the DNA environment there are no water molecules nearby the central hydrogen bond N1-H1-N3, but there are always 2-3 water molecules close to the lateral hydrogen bond O6-H4-N4. We have calculated the free energy profile $F(d_1)$ including in the QM region 3 water molecules around the O6-H4-N4 atoms both for the gas phase and DNA environment cases. The initial positions of these water molecules, alongside the rest of the system, are extracted from previously thermalized QM/MM simulations (in which these 3 molecules belonged to the MM region). These configurations are further thermalized including now these molecules in the QM region and different configurations are then extracted to be employed as seeds for the productions runs. As shown in Figure 3(a), the inclusion of these water molecules in the gas phase reduces the stability of

the rare tautomer and the barrier for the reverse reaction, although the rare tautomer is still

meta-stable, in agreement with previous results.[17–20,35] In the case of the DNA environment,

including these water molecules explicitly in the QM region yields almost identical free energy profile as obtained when these molecules are included in the MM region, with a slightly

higher value of the free energy for rare tautomer configurations. These results indicate that

the water molecules around the lateral hydrogen bond are an important factor reducing the

stability of the rare tautomer, but also highlight the importance of the rest of the DNA

environment, which must be fully included to predict the complete instability of the rare
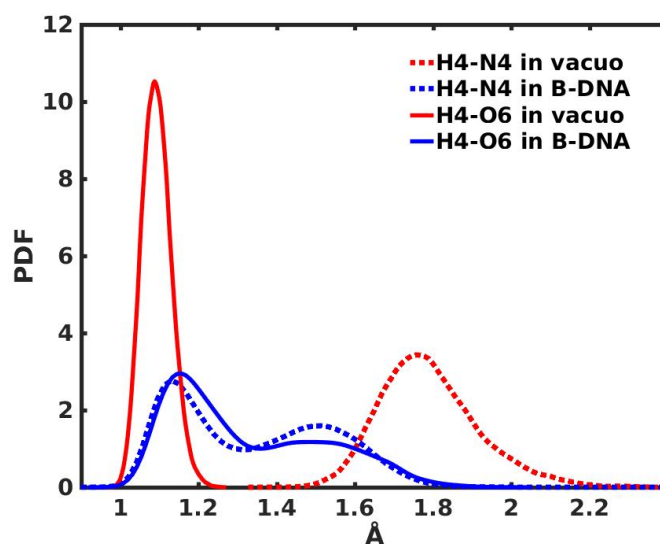
tautomer.

Figure 5: Probability distribution functions for the distances H4-N4 and H4-O6 (see Fig. 1) for the rare tautomer at RT; the distance H1-N1 (d_1) is restrained to a value ~ 1.9 Å using a bias potential. Red lines: gas phase; blue lines: DNA environment (with 3 water molecules around the O6-H4-N4 hydrogen bond included in the QM region). Distances in Å.

Figure 5 shows the probability distribution functions (PDFs) for the distances H4-O6

and H4-N4 in the rare tautomer configuration, *i.e.* adding a flat-bottom bias potential that

restrains the distance H1-N1 to be ~ 1.7–2.1 Å. To construct the PDFs we perform long

US simulations of 50 ps and use the Kernel Density Estimator[63] over the time series of the

distances. In the gas phase the rare tautomer is fully formed, with a H4-O6 distance ~ 1.1 Å

and a H4-N4 distance oscillating around ~ 1.8 Å. In the DNA environment, however, there is a drastic change in the PDFs. The H4-N4 and H4-O6 PDFs are now both spread between ~ 1.0-1.8 Å, and each of them is a superposition of two broad peaks centered at ~ 1.15 Å and ~ 1.5 Å. This situation is an indication of a low-barrier hydrogen bond (LBHB) in which the H atom can easily move between the O and N atoms; these short and strong hydrogen bonds typically enhance the stability of certain transition states in enzymatic reactions and help the catalytic activity of the protein.[64–66] Thus, when the covalent bond H1-N3 is formed in the biomolecular environment, instead of forming the corresponding O6-H4 covalent bond (see Fig. 1), the GC system prefers to form a O6-H4-N4 LBHB. This is due, in part, to the interaction with the surrounding water molecules, but also to the influence of the rest of the environment (*e.g.* see Fig. 3(a)). We remark that if the restraining bias potential is removed, the GC system reverts in an ultra-fast process to the canonical tautomer; thus, the LBHB discussed here is completely unstable (existing only in the virtual scenario of H1 being bonded to N3). On the other hand, if we restrain the distance H4-N4 to be ~ 1.7–2.1 Å (instead of the distance H1-N1), the rare tautomer is fully formed for both the gas phase and the DNA environment cases.
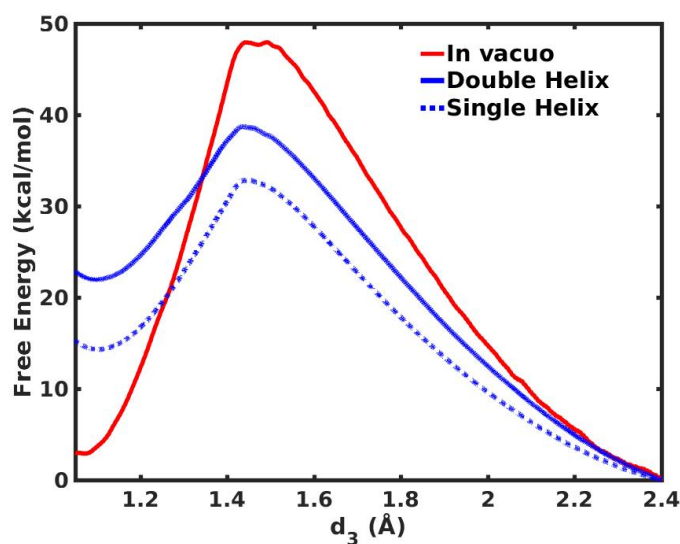

Figure 6: Free energy profiles (in kcal/mol) for the guanine intra-base proton transfer as a function of the distance $d_3 = |\vec{r}_{H1} - \vec{r}_{O6}|$ (in Å): gas-phase (red); DNA-environment, double-helix (blue, solid line); DNA-environment, single-helix (blue, dashed line).

We turn now our attention to intra-base proton transfer, which in principle could also yield rare tautomers. We have studied the transfer of H1 from N1 to O6 in guanine (see Figure 1), both for the single helix and double helix cases. It is relevant to study also the single helix case because, while DNA replication, that is the structure in which DNA is found. In vacuo, the energy difference between the two tautomers, G and G*, is quite small: 3.2 kcal/mol in our calculations, in excellent agreement with the value of 3.0 kcal/mol obtained with the plane-waves code Quantum Espresso. A very similar difference is obtained in the free energy profile, Figure 6, which also shows that there is a very large G $\leftrightarrow$ G* free energy barrier, $\sim$ 48 kcal/mol. The most important effect of the DNA environment is the large shift in the free energy difference between the canonical and rare tautomers, which is increased to 14.9 kcal/mol (single helix) or 22.4 kcal/mol (double helix). At the same time, the G $\leftrightarrow$ G* free energy barrier is reduced to 32.3 kcal/mol (single helix) or 38.8 kcal/mol (double helix), but it is still quite high.

In summary, we have analyzed H transfer reactions in guanine-cytosine base-pairs embedded in a realistic DNA biomolecular environment by means of free energy calculations based on extensive QM/MM MD simulations. We find that, although G*C* is metastable in the gas phase, this rare tautomer is not an accessible state in the DNA environment, *i.e.* the inter-base GC $\rightarrow$ G*C* DPT reaction does not take place. We have also analyzed the intra-base H transfer in Guanine for DNA in single helix or double helix configurations. Our calculations indicate that the DNA environment changes dramatically the relative stability between the canonical and rare tautomers. The QM/MM free energy calculations presented here for both inter-base and intra-base H transfer reactions demonstrate the important role of the DNA environment enhancing the stability of the genetic code against spontaneous mutations.

METHODS

Free energy techniques. The free energy methods WHAM[58] and UI[59] are based on the

Umbrella Sampling (US) technique[62] to properly sample the corresponding phase space. In this technique, bias potentials V_j are used to enhance sampling at different values of the reaction coordinate and an unbiasing procedure is required. Notice that if a system with potential energy U, $U \equiv U(\{\vec{r}_\alpha\})$, is modified by adding a potential $V_j(\hat{q})$, then the probability density $\tilde{P}_j(q)$ obtained from the biased simulation is

$$\tilde{P}_j(q) = \frac{1}{Z_j} \int d^N r \ e^{-\beta(U+V_j)} \ \delta(\hat{q} - q) = \frac{Z}{Z_j} \ e^{-\beta V_j(q)} \ P_j(q), \tag{4}$$

where Z_j is the configurational partition function for the system with potential energy $U + V_j$ and $P_j(q)$ is the *unbiased* probability density $P(q)$ obtained from a simulation with bias potential V_j.

In practice, $P_j(q)$ will only provide a good estimation for $P(q)$ for values within the j-window. In WHAM the different $P_j(q)$ are combined with certain weights to have a window-independent value for $P(q)$

$$P(q) = \sum_j c_j(q) \ P_j(q) \tag{5}$$

(with $\sum_j c_j(q) = 1$); these weights are calculated minimizing the statistical error for the calculation of $P(q)$.[58] As equation (4) shows, $P_j(q)$ depends on the unknown ratios Z_j/Z which in turn depend on $P(q)$; this problem is solved iteratively until self-consistency is achieved.[67] Finally, $F(q)$ is obtained from equation (1).

In Umbrella Integration[59] the derivatives dF_j/dq are calculated from the probability densities $P_j(q)$ so that the terms Z_j/Z vanish:

$$\frac{dF_j}{dq}(q) = -k_B T \frac{d \ln \tilde{P}_j}{dq}(q) - \frac{dV_j}{dq}(q), \tag{6}$$

see equation (4), and $\tilde{P}_j(q)$ are approximated by gaussian functions. Then one combines the

derivatives dF_j/dq to calculate the gradient of the Free Energy,

$$\frac{dF}{dq}(q) = \sum_j c_j(q) \, \frac{dF_j}{dq}(q) \tag{7}$$

which is numerically integrated to obtain $F(q)$.

In the calculation of the free energy profiles with SMD-JE we use time-dependent bias potentials $V(t) = k\big(\hat{q} - \mu(t)\big)^2$ to generate the SMD trajectories, with $\mu(t) = vt + q_0$. The work for each trajectory is given by

$$W = -2vk \int_0^t \big(\hat{q}\{\vec{r}_\alpha(t')\} - vt' - q_0\big)dt'. \tag{8}$$

To avoid directionality effects, we have performed SMD trajectories starting from the canonical tautomer and also starting from the rare tautomer. The initial configurations for each SMD trajectory are independent samples taken from restrained equilibrated dynamics.

In the calculation of the WHAM and UI free energy profiles we have used harmonic bias potentials $V_j = k\big(\hat{q} - \mu_j\big)^2$, with $k = 2000$ kcal/mol . Å^{-2}, and 100 windows, μ_j, along the reaction coordinate. For each window we run MD simulations at RT to generate 5×10^3 configurations out of which we use the last 3×10^3 to calculate the free energy profiles. In all the US and SMD simulations we use a time step $\Delta t = 0.5$ fs. The initial configurations for each window are obtained from SMD trajectories departing from equilibrated structures. The free energy map (Figure 4) was calculated by using a total of 30×30 US simulations of 1 ps (2000 snapshots), which gives a total of 1.8×10^6 structures. We use Alan Grossfield's code to solve the WHAM self-consistent equations.[67]

When using the Jarzinsky equality, we have typically employed six trajectories from the canonical tautomer to the rare tautomer and other six with the opposite directionality, from the rare tautomer to the canonical one. Each trajectory consists of 2000 steps, with timesteps of $\Delta t = 0.5$ fs, as above. Slower SMD trajectories of 5000 and 10000 have been also performed in order to check that the results are converged.

The free energy profiles $F_G(d_1)$ and $F_E(d_1)$ shown in Figure 3(a) have been calculated using the WHAM technique; identical results are obtained with UI. The calculation of these free energy profiles using the SMD-JE technique yields also almost identical results. Notice that WHAM and UI are based on US, while SMD-JE is a completely different approach (based on SMD simulations and the Jarzynski equality). The fact that we obtain practically the same free energy profiles using these different techniques is an excellent confirmation regarding the accuracy of these techniques. In this case, the SMD-JE technique is more efficient than WHAM or UI. Thus, the free energy profiles for the case with 3 water molecules included in the QM region were calculated using SMD-JE. Also, the free energy profiles for the intra-base proton transfer were calculated as well by means of the SMD-JE approach. As mentioned above, the free energy map (Figure 4) was calculated with WHAM.

QM/MM. The free energy results are obtained from extensive QM/MM MD simulations at RT. For this purpose we have used FIREBALL / AMBER.[39] FIREBALL[42–44] is a local-orbital Density Functional Theory (DFT) Molecular Dynamics (MD) technique based on a self-consistent extension of the Harris-Foulkes approach.[68] We also mention that FIREBALL is a fully real-space technique (no need for super-cells), which is a suitable property for non-periodic complex systems and QM/MM implementations. The numerical atomic-like orbitals (NAOs) are short-range and are optimized[46] in order to have a reduced number of orbitals in the basis set. These NAOs are generated from atomic calculations with given cut-off radii, $R_c(\mu)$, and atomic charges, q_μ, in the different orbitals; R_c is the confinement distance at which the orbitals vanish. In particular, we have used two s-orbitals ($ss*$) for H, sp^3-orbitals for C, and sp^3d^5-orbitals for N and O; the optimized basis set yields Mean Absolute Deviations (MAD) of 0.70 kcal/mol and 1.28 kcal/mol as compared to the S66×8 and IHBx8 data sets,[47,48] respectively.

QE. Regarding the test calculations performed using the QE code (*e.g.* Fig. 1(right)), we have used a cut-off of 80 Ry, and norm-conserving Martins-Troullier pseudopotentials.[69]

Supporting Information

Map of the average charge on the H1 and H4 hydrogen atoms as a function of the reaction coordinates q_1 and q_2.

This information is available free of charge via the Internet at http://pubs.acs.org
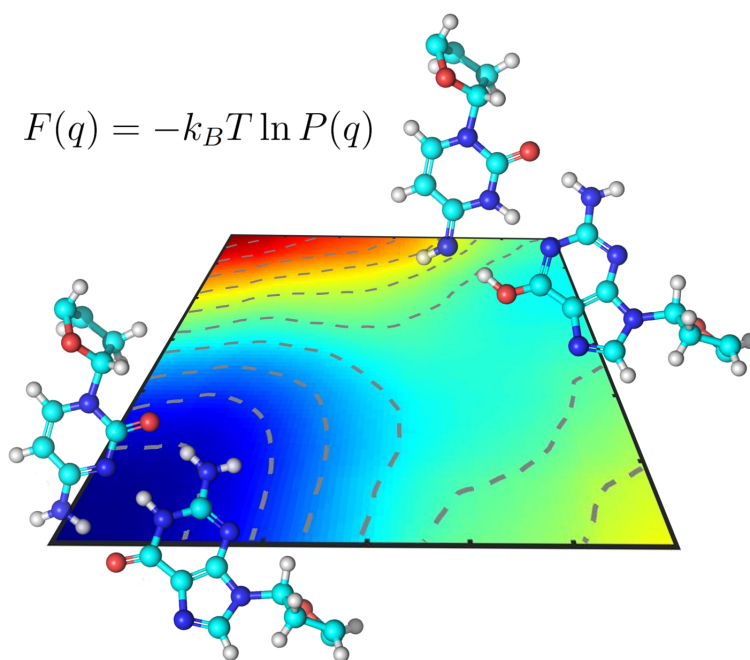
Acknowledgements

Figure 7: For Table of Contents Only

References

(1) Watson, J. D.; Crick, F. H. Genetical Implications of the Structure of DNA. *Nature* **1953**, *171*, 964–967.

(2) Löwdin, P.-O. Proton Tunneling in DNA and its biological implications. *Reviews of Modern Physics* **1963**, *35*, 1034–1042.

(3) Florian, J.; Leszczynski, J. Spontaneous DNA mutations induced by proton transfer in the guanine.cytosine base pairs: An energetic perspective. *Journal of the American Chemical Society* **1996**, *118*, 3010–3017.

(4) Zoete, V.; Meuwly, M. Double proton transfer in the isolated and DMA-embedded guanine-cytosine base pair. *Journal of Chemical Physics* **2004**, *121*, 4377–4388.

(5) Gorb, L.; Leszczynski, J. Intramolecular Proton Transfer in Mono- and Dihydrated Tautomers of Guanine: An ab Initio Post HartreeFock Study. *Journal of the American Chemical Society* **1998**, *120*, 5024–5032.

(6) Gorb, L.; Podolyan, Y.; Leszczynski, J. A theoretical investigation of tautomeric equilibria and proton transfer in isolated and monohydrated cytosine and isocytosine molecules. *Journal of Molecular Structure: THEOCHEM* **1999**, *487*, 47–55.

(7) Gorb, L.; Podolyan, Y.; Dziekonski, P.; Sokalski, W. A.; Leszczynski, J. Double-Proton Transfer in AdenineThymine and GuanineCytosine Base Pairs. A Post-HartreeFock ab Initio Study. *Journal of the American Chemical Society* **2004**, *126*, 10119–10129.

(8) Villani, G. Theoretical investigation of hydrogen transfer mechanism in the adenine-thymine base pair. *Chemical Physics* **2005**, *316*, 1–8.

(9) Villani, G. Theoretical investigation of hydrogen transfer mechanism in the guanine-cytosine base pair. *Chemical Physics* **2006**, *324*, 438–446.

(10) Ceron-Carrasco, J. P.; Requena, A.; Zuniga, J.; Michaux, C.; Perpete, E. A.; Jacquemin, D. Intermolecular Proton Transfer in Microhydrated GuanineCytosine Base Pairs: a New Mechanism for Spontaneous Mutation in DNA. *The Journal of Physical Chemistry A* **2009**, *113*, 10549–10556.

(11) Villani, G. Theoretical investigation of hydrogen atom transfer in the cytosine-guanine base pair and its coupling with electronic rearrangement. Concerted vs stepwise mechanism. *Journal of Physical Chemistry B* **2010**, *114*, 9653–9662.

(12) Villani, G. Theoretical investigation of hydrogen atom transfer in the adeninethymine base pair and its coupling with the electronic rearrangement. Concerted vs. stepwise mechanism. *Physical Chemistry Chemical Physics* **2010**, *12*, 2664.

(13) Villani, G. Theoretical investigation of the hydrogen atom transfer in the hydrated A-T base pair. *Chemical Physics* **2012**, *394*, 9–16.

(14) Xiao, S.; Wang, L.; Liu, Y.; Lin, X.; Liang, H.; Xiao, S.; Wang, L.; Liu, Y.; Lin, X.; Liang, H. Theoretical investigation of the proton transfer mechanism in guanine- cytosine and adenine-thymine base pairs. *Journal of Chemical Physics* **2012**, *137*, 1–8.

(15) Villani, G. Theoretical investigation of hydrogen atom transfer in the hydrated CG base pair. *Molecular Physics* **2013**, *111*, 201–214.

(16) Villani, G. Coupling between hydrogen atoms transfer and stacking interaction in adenine-thymine/guanine-cytosine complexes: A theoretical study. *Journal of Physical Chemistry B* **2014**, *118*, 5439–5452.

(17) Jacquemin, D.; Zuniga, J.; Requena, A.; Ceron-Carrasco, J. P. Assessing the Importance of Proton Transfer Reactions in {DNA}. *Accounts of Chemical Research* **2014**, *47*, 2467–2474.

(18) Tolosa, S.; Sansón, J. A.; Hidalgo, A. Theoretical thermodynamic study of the adeninethymine tautomeric equilibrium: Electronic structure calculations and steered molecular dynamic simulations. *International Journal of Quantum Chemistry* **2017**, *117*, 1–9.

(19) Romero, E. E.; Hernandez, F. E. Solvent effect on the intermolecular proton transfer of the Watson and Crick guanine-cytosine and adenine-thymine base pairs: A polarizable continuum model study. *Physical Chemistry Chemical Physics* **2018**, *20*, 1198–1209.

(20) Tolosa, S.; Sansón, J. A.; Hidalgo, A. Mechanisms for guaninecytosine tautomeric equilibrium in solution via steered molecular dynamic simulations. *Journal of Molecular Liquids* **2018**, *251*, 308–316.

(21) Cerón-Carrasco, J. P.; Jacquemin, D. DNA spontaneous mutation and its role in the evolution of GC-content: assessing the impact of the genetic sequence. *Physical Chemistry Chemical Physics* **2015**, *17*, 7754–7760.

(22) Pérez, A.; Tuckerman, M. E.; Hjalmarson, H. P.; Von Lilienfeld, O. A. Enol tautomers of Watson-Crick base pair models are metastable because of nuclear quantum effects. *Journal of the American Chemical Society* **2010**, *132*, 11510–11515.

(23) Daido, M.; Kawashima, Y.; Tachikawa, M. Nuclear quantum effect and temperature dependency on the hydrogen-bonded structure of base pairs. *Journal of Computational Chemistry* **2013**, *34*, 2403–2411.

(24) Markland, T. E.; Ceriotti, M. Nuclear quantum effects enter the mainstream. *Nature Reviews Chemistry* **2018**, *2*, 0109.

(25) Marx, D.; Parrinello, M. Ab initio path integral molecular dynamics: Basic ideas. *Journal of Chemical Physics* **1996**, *104*, 4077–4082.

(26) Godbeer, a. D.; Al-Khalili, J. S.; Stevenson, P. D. Modelling proton tunnelling in the adeninethymine base pair. *Physical Chemistry Chemical Physics* **2015**, *17*, 13034–13044.

(27) Groenhof, G.; Schäfer, L. V.; Boggio-Pasqua, M.; Goette, M.; Grubmüller, H.; Robb, M. A. Ultrafast deactivation of an excited cytosine-guanine base pair in DNA. *Journal of the American Chemical Society* **2007**, *129*, 6812–6819.

(28) Chen, H.-Y.; Kao, C.-L.; Hsu, S. C. N. Proton Transfer in GuanineCytosine Radical Anion Embedded in B-Form DNA. *Journal of the American Chemical Society* **2009**, *131*, 15930–15938.

(29) Bucher, D. B.; Schlueter, A.; Carell, T.; Zinth, W. Watson-Crick Base Pairing Controls Excited - State Decay in Natural DNA. *Angewandte Chemie - International Edition* **2014**, *53*, 11366–11369.

(30) Zhang, Y.; De La Harpe, K.; Beckstead, A. A.; Improta, R.; Kohler, B. UV-Induced Proton Transfer between DNA Strands. *Journal of the American Chemical Society* **2015**, *137*, 7059–7062.

(31) Kinz-Thompson, C.; Conwell, E. Proton Transfer in AdenineThymine Radical Cation Embedded in B-Form DNA. *The Journal of Physical Chemistry Letters* **2010**, *1*, 1403–1407.

(32) Zhang, Y.; De La Harpe, K.; Beckstead, A. A.; Martínez-Fernández, L.; Improta, R.; Kohler, B. Excited-State Dynamics of DNA Duplexes with Different H-Bonding Motifs. *Journal of Physical Chemistry Letters* **2016**, *7*, 950–954.

(33) Sauri, V.; Gobbo, J. P.; Serrano-Pérez, J. J.; Lundberg, M.; Coto, P. B.; Serrano-Andrés, L.; Borin, A. C.; Lindh, R.; Merchán, M.; Roca-Sanjuán, D. Proton/hydrogen transfer mechanisms in the guanine-cytosine base pair: Photostability and tautomerism. *Journal of Chemical Theory and Computation* **2013**, *9*, 481–496.

(34) Mendieta-Moreno, J. I.; Trabada, D. G.; Mendieta, J.; Lewis, J. P.; Gómez-Puertas, P.; Ortega, J. Quantum Mechanics/Molecular Mechanics Free Energy Maps and Nonadiabatic Simulations for a Photochemical Reaction in DNA: Cyclobutane Thymine Dimer. *The Journal of Physical Chemistry Letters* **2016**, *7*, 4391–4397.

(35) Podolyan, Y.; Gorb, L.; Leszczynski, J. Ab Initio Study of the Prototropic Tautomerism of Cytosine and Guanine and Their Contribution to Spontaneous Point Mutations. *International Journal of Molecular Sciences* **2003**, *4*, 410–421.

(36) Das, S.; Nam, K.; Major, D. T. Rapid Convergence of Energy and Free Energy Profiles with Quantum Mechanical Size in Quantum Mechanical-Molecular Mechanical Simulations of Proton Transfer in DNA. *Journal of Chemical Theory and Computation* **2018**, *14*, 1695–1705.

(37) Stevens, D. R.; Hammes-Schiffer, S. Exploring the Role of the Third Active Site Metal Ion in DNA Polymerase η with QM/MM Free Energy Simulations. *Journal of the American Chemical Society* **2018**, *140*, 8965–8969.

(38) Paul, S.; Paul, T. K.; Taraphder, S. Reaction Coordinate, Free Energy, and Rate of Intramolecular Proton Transfer in Human Carbonic Anhydrase II. *Journal of Physical Chemistry B* **2018**, *122*, 2851–2866.

(39) Mendieta-Moreno, J. I.; Walker, R. C.; Lewis, J. P.; Gómez-Puertas, P.; Mendieta, J.; Ortega, J. fireball/ amber : An Efficient Local-Orbital DFT QM/MM Method for Biomolecular Systems. *Journal of Chemical Theory and Computation* **2014**, *10*, 2185–2193.

(40) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynam-

ics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications* **1995**, *91*, 1–41.

(41) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *Journal of Computational Chemistry* **2005**, *26*, 1668–1688.

(42) Lewis, J. P.; Jelínek, P.; Ortega, J.; Demkov, A. A.; Trabada, D. G.; Haycock, B.; Wang, H.; Adams, G.; Tomfohr, J. K.; Abad, E.; Wang, H.; Drabold, D. A. Advances and applications in the FIREBALL ab initio tight-binding molecular-dynamics formalism. *Physica Status Solidi (B) Basic Research* **2011**, *248*, 1989–2007.

(43) Jelínek, P.; Wang, H.; Lewis, J. P.; Sankey, O. F.; Ortega, J. Multicenter approach to the exchange-correlation interactions in ab initio tight-binding methods. *Physical Review B - Condensed Matter and Materials Physics* **2005**, *71*, 1–9.

(44) Lewis, J. P.; Glaesemann, K. R.; Voth, G. A.; Fritsch, J.; Demkov, A. A.; Ortega, J.; Sankey, O. F. Further developments in the local-orbital density-functional-theory tight-binding method. *Physical Review B* **2001**, *64*, 195103.

(45) Sankey, O. F.; Niklewski, D. J. Ab initio multicenter tight-binding model for molecular-dynamics simulations and other applications in covalent systems. *Physical Review B* **1989**, *40*, 3979–3995.

(46) Basanta, M. A.; Dappe, Y. J.; Jelínek, P.; Ortega, J. Optimized atomic-like orbitals for first-principles tight-binding molecular dynamics. *Computational Materials Science* **2007**, *39*, 759–766.

(47) Rezáč, J.; Riley, K. E.; Hobza, P. S66: A well-balanced database of benchmark interaction energies relevant to biomolecular structures. *Journal of Chemical Theory and Computation* **2011**, *7*, 2427–2438.

(48) Rezáč, J.; Hobza, P. Advanced Corrections of Hydrogen Bonding and Dispersion for Semiempirical Quantum Mechanical Methods. *Journal of Chemical Theory and Computation* **2012**, *8*, 141–151.

(49) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A* **1988**, *38*, 3098–3100.

(50) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B* **1988**, *37*, 785–789.

(51) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *Journal of Chemical Physics* **2010**, *132*.

(52) Smith, D. G.; Burns, L. A.; Patkowski, K.; Sherrill, C. D. Revised Damping Parameters for the D3 Dispersion Correction to Density Functional Theory. *Journal of Physical Chemistry Letters* **2016**, *7*, 2197–2203.

(53) Giannozzi, P.; Baroni, S.; Bonini, N.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Chiarotti, G. L.; Cococcioni, M.; Dabo, I.; Dal Corso, A.; De Gironcoli, S.; Fabris, S.; Fratesi, G.; Gebauer, R.; Gerstmann, U.; Gougoussis, C.; Kokalj, A.; Lazzeri, M.; Martin-Samos, L.; Marzari, N.; Mauri, F.; Mazzarello, R.; Paolini, S.; Pasquarello, A.; Paulatto, L.; Sbraccia, C.; Scandolo, S.; Sclauzero, G.; Seitsonen, A. P.; Smogunov, A.; Umari, P.; Wentzcovitch, R. M. QUANTUM ESPRESSO: A modular and open-source software project for quantum simulations of materials. *Journal of Physics Condensed Matter* **2009**, *21*.

(54) Giannozzi, P.; Andreussi, O.; Brumme, T.; Bunau, O.; Buongiorno Nardelli, M.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Cococcioni, M.; Colonna, N.; Carnimeo, I.; Dal Corso, A.; de Gironcoli, S.; Delugas, P.; DiStasio, R. A.; Ferretti, A.; Floris, A.; Fratesi, G.; Fugallo, G.; Gebauer, R.; Gerstmann, U.; Giustino, F.; Gorni, T.;

Jia, J.; Kawamura, M.; Ko, H.-Y.; Kokalj, A.; Küçükbenli, E.; Lazzeri, M.; Marsili, M.; Marzari, N.; Mauri, F.; Nguyen, N. L.; Nguyen, H.-V.; Otero-de-la Roza, A.; Paulatto, L.; Poncé, S.; Rocca, D.; Sabatini, R.; Santra, B.; Schlipf, M.; Seitsonen, A. P.; Smogunov, A.; Timrov, I.; Thonhauser, T.; Umari, P.; Vast, N.; Wu, X.; Baroni, S. Advanced capabilities for materials modelling with Quantum ESPRESSO. *Journal of Physics: Condensed Matter* **2017**, *29*, 465901.

(55) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* **2015**, *11*, 3696–3713.

(56) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983**, *79*, 926–935.

(57) Bussi, G.; Parrinello, M. Stochastic thermostats: comparison of local and global schemes. *Computer Physics Communications* **2008**, *179*, 26–29.

(58) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE weighted histogram analysis method for freeenergy calculations on biomolecules. I. The method. *Journal of Computational Chemistry* **1992**, *13*, 1011–1021.

(59) Kästner, J.; Thiel, W. Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: "umbrella integration". *Journal of Chemical Physics* **2005**, *123*.

(60) Jarzynski, C. Nonequilibrium Equality for Free Energy Differences. *Physical Review Letters* **1997**, *78*, 2690–2693.

(61) Jarzynski, C. Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Physical Review E* **1997**, *56*, 5018–5035.

(62) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* **1977**, *23*, 187–199.

(63) Murray, R. Remarks on Some Nonparametric Estimates of a Density Function. *The Annals of Mathematical Statistics* **1956**, *Volume 27*, 832–837.

(64) Schiott, B.; Iversen, B. B.; Madsen, G. K. H.; Larsen, F. K.; Bruice, T. C. On the electronic nature of low-barrier hydrogen bonds in enzymatic reactions. *Proceedings of the National Academy of Sciences* **1998**, *95*, 12799–12802.

(65) Ishikita, H.; Saito, K. Proton transfer reactions and hydrogen-bond networks in protein environments. *Journal of the Royal Society Interface* **2014**, *11*.

(66) Cleland, W. W.; Frey, P. A.; Gerlt, J. A. The Low Barrier Hydrogen Bond in Enzymatic Catalysis. *Journal of Biological Chemistry* **1998**, *273*, 25529–25532.

(67) Grossfield, A. WHAM: the weighted histogram analysis method, version 2.0.9.1, Http://membrane.urmc.rochester.edu/wordpress/?page_id=126.

(68) Demkov, A. A.; Ortega, J.; Sankey, O. F.; Grumbach, M. P. Electronic structure approach for complex silicas. *Physical Review B* **1995**, *52*, 1618–1630.

(69) Troullier, N.; Martins, J. L. Efficient pseudopotentials for plane-wave calculations. *Physical Review B* **1991**, *43*, 1993–2006.