Modelling One-Electron Oxidation Potentials and Hole Delocalization in Double-Stranded DNA by Multilayer and Dynamic Approaches

Jesús Lucia-Tamudo,[†] Sergio Díaz-Tendero,^{*,†,‡,¶} and Juan J. Nogueira^{*,†,‡}

 †Department of Chemistry, Universidad Autónoma de Madrid, 28049, Madrid, Spain
 ‡Institute for Advanced Research in Chemistry (IAdChem), Universidad Autónoma de Madrid, 28049 Madrid, Spain

¶Condensed Matter Physics Center (IFIMAC), Universidad Autónoma de Madrid, 28049 Madrid, Spain

E-mail: sergio.diaztendero@uam.es; juan.nogueira@uam.es

Abstract

The amount of innovative applications for DNA nowadays is growing quickly. Its use as a nanowire or electrochemical biosensor leads to the need for a deep understanding of the charge transfer process along the strand, as well as its redox properties. These features are computationally simulated and analyzed in detail throughout this work by combining molecular dynamics, multilayer schemes and the Marcus theory. The one-electron oxidation potential and the hole delocalization have been analyzed for six DNA double strands that cover all possible binary combinations of nucleotides. The results have revealed that the one-electron oxidation potential decreases with respect to the single-stranded DNA, giving evidence that the greater rigidity of a double helix induces an increase in the capacity of storing the positive charge generated upon oxidation. In addition, the hole is mainly stored in nucleobases with large reducer character, i.e., purines, especially when those are arranged in a stacked configuration in the same strand. From the computational point of view, the sampling needed to describe biological systems implies a significant computational cost. Here, we show that a small number of representative conformations generated by a clustering analysis provides accurate results when compared with those obtained from the sampling, reducing considerably the computational cost.

1 Introduction

The applications of DNA have been evolved from storing genetic information in living organisms to having innovative applications such as DNA computation, DNA-templated synthesis, molecular detection, and the use of DNA as a nanowire.¹ DNA computation uses DNA as a molecular computer by leveraging the hybridization ability of DNA strands.^{2,3} DNA-templated synthesis utilizes DNA as a template to synthesize materials with specific properties.⁴ Molecular detection allows for the detection of specific molecules through DNA hybridization.^{5–15} In addition, using DNA as a nanowire enables the construction of electronic devices based on DNA strands.^{16,17}

Regarding molecular detection and DNA nanowires, charge transfer between the DNA strand and analyte/electrode is a crucial process. Nucleobases play a primary role in charge transfer, and obtaining precise values of redox properties like one-electron oxidation potential is essential since they usually undergo oxidation rather than reduction.^{18–31} The reducer ability of nucleobases in water follows a specific order: $G > A > T \sim C > U^{32}$ (Purines: G = guanine and A = adenine; Pyrimidines: T = thymine, C = cytosine and U = uracil). In fact, we have shown a clear correlation between the number of atoms in a nucleobase that participate in delocalizing the positive charge and the relative reducer character of nucleobases in a previous work.³¹ After nucleobase oxidation, hole transport along the DNA strand occurs, and nucleobases contribute to translating the positive charge. Two proposed mechanisms, tunnelling and hopping, explain this charge transport.^{33–36} Tunnelling involves

hole delocalization along multiple nucleobases, while hopping is a multistep process where the charge is localized in one nucleobase and moves through consecutive jumps. From previous works, it has been suggested that hopping is the predominant mechanism for single-stranded DNA (ss-DNA) in the electronic ground state, mainly due to solvent effects that stabilize charge delocalization.^{32,37–39} Additionally, also in ss-DNA, the charge tends to be held in a single nucleobase or, in some cases, in more than one, depending on the solvent and nucleobase type. Some of our previous works have focused on studying the redox properties and hole distribution in ss-DNA molecules. However, these aspects have not been discussed for double helices yet.

The most important disadvantage of modelling biological systems with a large number of degrees of freedom is the high computational cost that is required. Since there are a large amount of geometries to be considered from the sampling, the calculations involve a significant amount of time. In order to overcome this limitation, there exists some techniques that reduce the number of geometries to analyze, *e.g.*, a clustering analysis. This methodology classifies the total number of conformations in a specific number of groups in terms of the similarity between geometries, although they can also be classified based on the similarity in a specific property. This means that different snapshots from the dynamic simulations with similar geometry form a collective of frames, called a cluster, which are considered to have similar properties. This clustering classification is often performed based on the calculation of the root-mean-square deviation (RMSD) between all configurations. Then, the centroid structure of each cluster can be determined and, as an approximation, one can consider that the properties of the centroid structure will accurately represent the properties of the whole ensemble of geometries composing the cluster. As a result, the number of conformations to be taken into account in the calculations of the desired property is considerably reduced.

In this work we have determined the one-electron oxidation potential and the hole delocalization along the strand of all the possible binary combinations of model double-stranded DNA (ds-DNA) (see Figure 1). In addition, we discuss the different factors that influende the magnitude of these two properties. Finally, a clustering study has been conducted in order to determined whether the redox properties computed for a reduced number of representative conformations are in good agreement with the properties computed for a large ensemble of geometries.



Figure 1: Graphical view of the general form of the systems under study in this chapter. Each nucleobase is associated to a colour. The full sequence for each system is displayed in the right panel. Cyan refers to the nucleobases that form the CAM-B3LYP layer and to the sugar and phosphate described in the xTB layer.

2 Methods and Computational Details

The computation of the one-electron oxidation potential and the delocalization properties of the strands were conducted using a similar procedure than the one employed in previous studies on ss-DNA.^{32,39} A conformational sampling was carried out using classical molecular dynamics (MD) followed by quantum-mechanics/molecular-mechanics (QM/MM) MD simulations. Subsequently, the properties were computed from a selected ensemble of geometries, obtained from the sampling, through electronic-structure calculations. These calculations were performed using a QM1/QM2/Continuum approach in combination with the Marcus theory. In the following, the computational details of this protocol are explained.

The nab application provided by the AmberTools 22 package $^{40-42}$ was used to build the initial geometries of the ds-DNA strands. Each double helix was composed of 24 nucleotides arranged in a specific pattern, as shown in Figure 1. Thus, each strand was composed by 12 nucleotides. In order to analyze all possible binary combinations of nucleotides in one strand, six ds-DNA models were built where the complementary strand fully matched with the one with direction $5' \leftarrow 3'$ (see Figure 1). The ds-DNA molecules were solvated in a truncated octahedron box with a buffer of 16 Å, and the tleap program implemented in AmberTools 22 was used for this purpose. The ff90bsc0 force field 43,44 in combination with the bsc1 dihedral correction 45 was employed to describe the DNA molecules, while the TIP3P force field 46 modelled the interaction description of the water molecules. To counteract the negative charge of each strands, 22 sodium cations were added using the parameters described by Joung and Cheatham. 47

The exploration of the configurational space was conducted through classical MD simulations $^{48-50}$ using the CUDA version of the pmemd program, which is implemented in the AMBER 20 package. $^{40-42}$ The simulations began with a 10000-step minimization, where the steepest-descent algorithm was used for the first 5000 steps, 51 followed by the Newton-Raphson algorithm for the subsequent 5000 steps. 52 To regulate the temperature, a constant volume (*NVT*) progressive heating up to 300 K was performed for 500 ps, applying

a Langevin thermostat with a collision frequency of 2 ps⁻¹. Afterwards, an additional 500 ps simulation was conducted at a constant temperature of 300 K (using a canonical (*NVT*) ensemble). In the next phase, a 1 ns simulation was run in the isothermal-isobaric (*NPT*) ensemble to balance the system volume and achieve the desired density. Finally, a production simulation of 200 ns was performed in the (*NPT*) ensemble, and 200 equidistant snapshots were fetched. Throughout all simulations within the (*NPT*) ensemble, the Berendsen barostat with isotropic position scaling and a pressure relaxation time of 2 ps were employed to maintain a constant pressure of 1 bar. During the entire dynamic protocol, the particle-mesh Ewald method with a grid spacing of 1.0 Å was used to compute electrostatic interactions, while a cutoff of 10 Å was applied for non-bonded interactions. The SHAKE algorithm $^{53-55}$ was utilized to constrain the hydrogen-containing bonds, and a time step of 2 fs was used for the heating, equilibration, and production stages.

From the classical MD simulations, 200 geometries were selected for each strand as initial conditions to run QM/MM MD simulations in order to refine the structure of the relevant region of the system. These simulations were carried out for both the neutral and the cationic strands to apply in the next step the Marcus model. These QM/MM MD simulations were evolved for 100 steps in the (*NPT*) ensemble using the ORCA/AMBER interface.⁵⁶ The QM region comprises eight nucleobases, four adjacent nucleobases in one strand and their complementary ones in the opposite, and was described using the xTB model⁵⁷ and the 6-311G(d)^{58,59} basis set. Finally, the last geometry obtained from each QM/MM MD simulation was employed to compute the vertical ionization energy (VIE) and the vertical attachment energy (VAE) for each system. Using the Marcus theory,^{60–65} the one-electron oxidation potential $\Delta \mathcal{E}_{red}$ was computed using Eq. 1:

$$\Delta G_{red} = \frac{1}{2} \Big(\langle \Delta U_{\mathcal{N} \to \mathcal{N}^+}(\mathbf{r}; \mathbf{R}) \rangle_{\mathcal{N}} - \langle \Delta U_{\mathcal{N}^+ \to \mathcal{N}}(\mathbf{r}; \mathbf{R}) \rangle_{\mathcal{N}^+} \Big) - G(e_{(gas)}^-) =$$

$$= \frac{1}{2} \Big(\langle VIE \rangle_{\mathcal{N}} - \langle VAE \rangle_{\mathcal{N}^+} \Big) - G(e_{(gas)}^-)$$
(1)

to calculate the Gibbs free energy from its average VIE and VAE. Notice that $G(e_{(gas)}^{-})$ is

the free energy of the electron in the gas phase. In addition, this free energy can be related to the redox potential through Eq. 2.

$$\Delta \mathcal{E}_{red} = \frac{\Delta G_{red}}{n_e F} - \mathcal{E}_{red,SHE}^0 \tag{2}$$

where $\mathcal{E}^{0}_{red,SHE}$ refers to the reduction potential of the standard hydrogen electrode, the reference electrode selected for this work.^{32,39} In fact, the term $G(e^{-}_{(gas)})$ in Eq. 2 is included because it is also taken into account in the value of $\mathcal{E}^{0}_{red,SHE} = 4.28$ V.

These calculations were performed employing a hybrid QM1/QM2/Continuum approach, where the QM1/QM2 interaction was described by an electrostatic embedding. Specifically, the VIEs and VAEs were determined for the QM1 region, consisting of the eight nucleobases previously mentioned, using the CAM-B3LYP/6-311G(d) level of theory. The nucleotides with nucleobases that are not involved in the QM region were excluded from the final calculation, while the phosphates and sugars of the QM nucleobases were in the second layer QM2 described bu the DFTB approach with the GFN2-xTB scheme.⁵⁷ The effects of the solvent were accounted for using the ALPB continuum solvation model,⁶⁶ which is compatible with DFTB. All computations were carried out using the ORCA 5.0.3 package.⁵⁶

To analyze the localization of the hole, the molecular charge difference of each nucleobase in the QM1 region upon ionization of the neutral species in each geometry was calculated. The Löwdin charges⁶⁷ were employed for the charge calculations, and the analysis was conducted using custom scripts. The intermolecular delocalization number, denoted as n, was defined as the number of nucleobases among which the positive charge is distributed after ionization. To determine n, the eight nucleobases considered in the QM1 region were first ordered based on increasing hole charge, and then an empirical equation (Eq. 3) was applied. The technical explanation and the details of this empirical equation can be found in reference 32.

$$n = M - \sum_{i=1}^{M-1} \left[1 - \left(\frac{\Delta q_i}{\sum_{j=i}^M \Delta q_j} \right) (M - i + 1) \right]$$
(3)

In a similar spirit, Pipek and Mezey proposed an alternative method to quantify the delocalization of a positive charge within a system.⁶⁸ They derived an index using the gross atomic Mulliken population of the set of orbitals in each atom. In order to compare our empirical Eq. 3, we adapted their formula by incorporating the partial charge of each nucleobase in the considered QM region (see Eq. 4).

$$n' = \sum_{i=1}^{m} \frac{1}{\left(\frac{\Delta q_i}{\sum_{j=1}^{m} \Delta q_j}\right)^2}$$
(4)

Finally, a clustering analysis was carried out to investigate whether reducing the number of geometries could yield similar results to studying the complete set of geometries considered throughout the trajectories. This analysis was performed using the cpptraj tool implemented in the AMBER 20 package.^{40–42} Thus, a convergence analysis was conducted to determine the number of clusters required to obtain a converged value of VIE, VAE, $\Delta \mathcal{E}_{red}$, and n (in both neutral and cationic trajectories).

3 Results

3.1 One-electron oxidation potential

We begin the discussion by examining the redox power of the systems addressed in this work. To this end, the one-electron oxidation potentials of the double helices have been determined using above described computational protocol. In a previous work, such potential was studied in homogeneous ss-polyX systems, where the following reducing capacity order was concluded: ss-polyG > ss-polyA > ss-polyT > ss-polyC.³² This relative order was related to the extent of the π -system of each nucleobase. In this way, those strands derived from purine nucleobases (G and A), with a larger π -system, exhibited greater reducing power,

i.e., a lower one-electron oxidation potential than the strands formed by pyrimidine bases. Additionally, we also studied heterogeneous ss-polyXY systems with binary combinations of nucleotides.³⁹ The simulations showed that the resulting one-electron oxidation potential was, roughly speaking, a linear combination of the potentials of the homogeneous strands weighted by the occurrence of each nucleobase in the heterogeneous strand.

We will refer to the ds-DNA molecules investigated in this work as ds-poly(XY-X'Y'), where X and Y represent the binary combination of nucleobases appearing in the $5' \rightarrow 3'$ direction strand. The complementary strand $3 \rightarrow 5'$ is thereby determined since the systems have been modeled without mismatches. Thus, X' and Y' are the complementary nucleobases of X and Y, respectively. In the case of ds-poly(XY-X'Y') systems, the reducing power is presented in Figure 2 in terms of the one-electron oxidation potential. Generally, there is a noticeable decrease in the oxidation potentials compared to the ss-polyXY systems studied previously.^{32,39} Moreover, the trend observed in the case of heterogeneous ss-DNA, where the redox potential was a linear combination of the potentials of the pure strands weighted by the abundance of each nucleobase in the heterogeneous strand, is not observed. Therefore, the behavior exhibited by ss-DNA systems cannot be extended to ds-DNA systems. This indicates that the intermolecular interactions between strands lead to a greater stabilization of the resulting positive charge, significantly increasing its capacity to undergo oxidation. This larger stabilization of the generated positive hole can be attributed to the larger rigidity of ds-DNA with respect to ss-DNA, which would allow delocalization in consecutive stacked nucleobases ($\pi - \pi$ stacking). On the contrary, it may be due to the possibility of delocalize the charge in two nucleobases, paired through hydrogen bonds (G-C; A-T) and establishing a large π region to hold it. These assumptions are further investigated in the next section, where the charge delocalization along the DNA strand is analyzed.

There are certain patterns common to all the single strands. A careful analysis of Figure 2 reveals a different in the potential between strands with and without guanine bases. The one-electron oxidation potential is significantly higher in DNA molecules that lack guanine



Figure 2: One-electron oxidation potential predicted for ss-polyX (solid lines) and dspoly(XY-X'Y') (bars) in aqueous phase. Black values inside the bars are the one-electron oxidation potential of the corresponding ds-DNA models, while coloured values next to the horizontal lines are those for homogeneous ss-DNA. Standard deviations for each strand is represented by vertical lines. Notice that $\Delta \mathcal{E}_{red}$ for ss-polyX are taken from reference 32.

nucleobases. This is consistent with the hierarchy of reducing power among nucleobases, where guanine tops the list. Therefore, it can be stated that guanine plays a predominant role in increasing the reducing power of a ds-poly(XY-X'Y') system, in the same way it does in ss-DNA. If we compare the two systems in which guanine is absent (ds-poly(AA-TT) and ds-poly(AT-TA)), we can observe that the redox potential is practically identical. Therefore, there does not appear to be a relationship between the sequence of the strands and the potential, but rather between the overall composition (abundance of each nitrogenous base) and the potential. On the other hand, the situation is clearly different when guanine is present. Firstly, ds-poly(GG-CC) and ds-poly(GC-CG) have the same percentage of guanines, but in the first one the guanines occupies adjacent positions in the same strand. This distribution seems to favor the oxidation process of the strand since the oxidation potential of ds-poly(GG-CC) is smaller than that of ds-poly(GC-CG), where the guanines are arranged diagonally in opposite strands (0.73 vs 0.91 V). The same trend is observed for dspoly(GA-CT) and ds-poly(GT-CA) if one focuses on the position of the purine nucleobases. In ds-poly(GA-CT), the purines guanine and adenine are adjacent to each other in the same strand and, consequently, it has a greater reducing power (lower oxidation potential) than ds-poly(GT-CA) (0.80 vs 0.98 V). Therefore, it seems that the position of the nucleobases in the strands does have a relevant effect when guanine is present, while it is not important in other cases.

3.2 Charge delocalization along the DNA strand

In order to understand the differences in one-electron oxidation potentials based on the arrangement of nucleobases along a strand, we studied the distribution of the hole among the nucleobases considered in the QM region. As explained earlier, oxidation results in the generation of a positive charge in the DNA double helix. This charge can be delocalized along the strand among several nucleobases or localized in only one of them. Precisely, the two most accepted mechanisms for charge transport in DNA molecules revolve around this idea. On one hand, tunnelling advocates for transport based on the delocalization of the positive charge among several nucleobases simultaneously, evolving over time from one side of the strand to the other. On the other hand, the hopping mechanism states that transport occurs through jumps of the localized hole from one nucleobase to another. Therefore, studying the distribution of the hole among the nucleobases of the considered systems could help elucidate the dominant mechanism in DNA charge transport.

Figure 3 shows the charge distribution analysis along each double helix investigated here. Specifically. Figure 3a displays the two different delocalization numbers explained above, n' and n. It can be seen that, although the value of n' extracted from each system (see Eq. 4) is lower than that of n (see Eq. 3), the relative delocalization order among strands is maintained independently on the way of computing the delocalization number. In general, the variation of the intermolecular delocalization numbers is less important for double strands



Figure 3: a) Intermolecular delocalization number n (red) or n' (blue) for each ds-DNA system considered. b) Graphical representation of the percentage of positive hole stored in each nucleobase of a specific system.

than for single strands. All the n values for ds-DNA fall within the range of 1.6-2.0, while in the case of ss-DNA it has been previously computed a slightly wider interval range going from 1.56 to 2.40.³⁹ The explanation for this behaviour is related to the composition of the double helix. In ds-DNA, the number of purines (large delocalization) always equals the number of pyrimidines (small delocalization) and, therefore, the obtained n values are not as large as those reached in ss-DNA, where only purines can be present.

Figure 3a,b shows that the positive charge is mainly located on the most reducing nucleobase, i.e., guanine, or on adenine when guanine is not part of the strand. In addition, when two purines are stacked in the same strand the charge delocalization is larger than when they are located in a diagonal arrangement in different strands. This can be observed when compared n for ds-poly(AA-TT) vs ds-poly(AT-TA) (1.93 vs 1.83), ds-poly(GG-CC) vs ds-poly(GC-CG) (1.92 vs 1.65), and ds-poly(GA-CT) vs ds-poly(GT-CA) (1.70 vs 1.57). This can be easily explained by the more efficient orbital overlap between adjacent stacked nucleobases than between diagonal interacting ones, allowing for a larger delocalization of the positive hole.

A combined analysis of Figures 2 and 3 reveals that there is no a direct correlation between the one-electron oxidation potential and the intermolecular delocalization number, as could be expected. Instead, as we previously showed,³² the reducing power of the strands is dominated by the intramolecular hole delocalization within a single nucleobase rather than the intermolecular one. In other words, the magnitude of the oxidation potential is dominated by the specific nature of the nucleobases composing the strand. For example, the double strands with the lowest oxidation potentials are those that contains guanine, while the guanine-free strands, ds-poly(AA-TT) and ds-poly(AT-TA), present larger potentials (see Figure 2). The intermolecular delocalization is more related with the reduction of the oxidation potential when going from the isolated nucleobase to the strand.³² Therefore, a correlation between the oxidation potential and the intermolecular delocalization number n is found only when strands having the same nucleobase composition are compared. Specifically, the lower oxidation potentials (larger reducing power) is related with slightly larger hole intermolecular delocalization numbers. For example, ds-poly(GA-CT) ($\Delta \mathcal{E}_{red} = 0.7$ V and n = 1.7) vs ds-poly(GT-CA) ($\Delta \mathcal{E}_{red} = 0.9$ V and n = 1.6) or ds-poly(GG-CC) ($\Delta \mathcal{E}_{red} = 0.7$ V and n = 1.9) vs ds-poly(GC-CG) ($\Delta \mathcal{E}_{red} = 0.9$ V and n = 1.7).

In all the ds-DNA strands investigated, it is observed that the positive charge storage occurs mainly in purines, and delocalization does not extend to more than two nucleobases. These features are typical of the mechanism known as hopping, in which transport occurs through jumps between nucleobases. However, since the hole charge is not completely located on one nucleobase, the tunneling mechanism cannot be completely ruled out and it is very likely that both mechanisms might operate simultaneously.

3.3 Clustering Analysis

The results discussed in previous sections require the computation of the redox properties on a large number of geometries for each system, leading to a significant high computational cost. To investigate possible reduction of such cost, we performed a clustering analysis. This technique classifies the ensemble of configurations in terms of the RMSD between their geometries. Conformations separated in the configuration space by small (large) RMSD will belong to the same (different) cluster. Finally, the structure corresponding to the centroid of each cluster is computed and associated with the closest configuration of the system, which becomes the representative conformation of the cluster. The properties of the cluster, therefore, are associated to the properties of the representative geometry. Thus, the total set of geometries from the dynamics for each system was grouped into a certain number of clusters, varying from 2 to 10, and the properties were calculated only for the representative structures of these clusters (see Figure 4a-e). Specifically, for the trajectories of the neutral (cationic) DNA strands, the VIE (VAE) and the delocalization number n_{VIE} (n_{VAE}) were computed. Then, the average value of each property \mathcal{A} is determined using Eq. 5, where the values of the property in each centroid $\mathcal{A}(i)$ of each cluster i = 1, ..., l are weighted by the fraction of geometries included in each cluster (P(i)):

$$\langle \mathcal{A} \rangle = \sum_{i=1}^{l} P(i)\mathcal{A}(i)$$
 (5)



Figure 4: Convergence of a) VIE, b) VAE, c) $\Delta \mathcal{E}_{red}$, d) n_{VIE} , and e) n_{VAE} with respect to the number of clusters. f) Bar plot of the MRUEs for each system in terms of the VIE, VAE, delocalization number for the neutral and cationic species (n_{VIE} and n_{VAE}) and one-electron oxidation potential. The MRUEs are calculated with respect to the values obtained from the dynamics using 200 snapshots.

Figure 4 shows the variation of each computed property as a function of the number of representative structures considered in the calculation. In the case of ds-poly(AA-TT) and ds-poly(GA-CT), convergence is reached after considering 7 geometries in the calculation, while for the remaining systems (in both neutral and cationic trajectories), 6 representative geometries are sufficient. Once convergence was achieved, the average values of VIE, VAE,

and the intermolecular delocalization number of the neutral (n_{VIE}) and cationic (n_{VAE}) systems were determined. The resulting values of the VIE and the VAE were used to calculate $\Delta \mathcal{E}_{red}$. Figure 4(f) shows the mean relative unsigned error (MRUE) of these properties with respect to those obtained from the complete trajectory, employing 200 snapshots, as explained above. It can be observed that these errors are considerably low, not exceeding 6%. This reflects that the value of all these properties can be estimated quite accurately using only a small number of representative geometries from the dynamics, specifically with 6-7 conformations. In conclusion, the redox properties of DNA strands can be studied with relatively low computational effort, allowing to apply high-level electronic-structure methods to obtain more accurate properties.

4 Conclusions

An extensive computational study has been carried out on the redox properties of DNA double-helix model systems, as well as on how a vacancy is distributed along this structure. In general terms, a significant increase in the reducer character of the nucleobases has been observed when they are part of a DNA double strand, compared to when they are arranged in a single strand. Intermolecular interactions between bases from different strands are capable of stabilizing to a greater extent the hole generated in the oxidation process, thereby facilitating the occurrence of this phenomenon. Moreover, a clear relationship between the reducing power of the strands and their nucleobase composition and arrangement has been observed. Thus, helices with a higher percentage of purines are more reducing than those with a higher percentage of pyrimidines, especially when the purines are stacked in the same strands rather than in a diagonal disposition in different strands.

The intermolecular delocalization number has proven to be relatively small for ds-DNA (always smaller than 2). In fact, intermolecular delocalization is smaller than in the case of single strands, due to the presence of both purine and pyrimidine nucleobases. In general,

the positive charge is localized exclusively on purine-derived nucleobases, which is consistent with the presence of a more extended π -system in comparison with pyrimidines. The charge, therefore, is delocalized mainly between only two adjacent purine nucleobases (stacked in the same strand or diagonally arranged), making the hopping mechanism predominant, as outlined in previous works for ss-DNA. Furthermore, when the adjacent purine nucleobases are different, there is a clear predominance of guanine in hosting most of the positive charge. Moreover, for strands with the same nucleobase composition, larger intermolecular delocalization of the hole is related to smaller one-electron oxidation potentials, i.e., to a stronger reducing power.

Finally, it has been demonstrated that the number of geometries to be considered in this type of calculations can be reduced using a clustering approach. With a very small number of conformations (6 or 7), the redox properties are satisfactorily similar to those obtained with a larger number of frames (200) selected from the trajectories. Thus, higher-level electronic-structure methods can be applied to these representative geometries in order to obtain more accurate results.

5 Data Availability

The nab and tleap toolkits from the AmberTools 22 package⁴⁰⁻⁴² were used to generate the topology and the coordinate files for the MD simulations. The PMEMD.CUDA module of the AMBER 20^{69,70} (https://ambermd.org/) software were used to perform the classical MD simulations. Afterwards, the SANDER module of the AMBER 20 software⁴⁰⁻⁴² was used in combination with the ORCA 5.0.3 package⁵⁶ (https://orcaforum.kofo.mpg.de/app.php/portal) to conduct the QM/MM MD simulations. With the aim of automatically generating the QM/MM input files, the MoBioTools toolkit (https://github.com/mobiochem/MoBioTools) was used.⁷¹ QM1/QM2/ALPB computations for the redox properties and hole delocalization were performed with the ORCA 5.0.3 software.⁵⁶ Homemade scripts were generated

with Python 3 (https://www.python.org/) in order to analyze the hole delocalization of the strands. The cpptraj application from the AmberTools 22 package was used to perform the clustering analyses. Finally, trajectories were visualized with the Visual Molecular Dynamics (VMD, https://www.ks.uiuc.edu/Research/vmd/).⁷²

Acknowledgement

We acknowledge the generous allocation of computer time at the Centro de Computación Científica at the Universidad Autónoma de Madrid (CCC-UAM) and at the Red Española de Supercomputación (RES). This work was partially supported by the MICINN – Spanish Ministry of Science and Innovation – Projects PID2022-138470NB-I00 and PID2020-117806GA-I00 funded by MCIN/AEI/10.13039/501100011033, and the 'María de Maeztu' (CEX2018-000805-M) Program for Centers of Excellence in R&D. J.J.N. acknowledge the Comunidad de Madrid for funding through the Attraction of Talent Program (Grant reference 2022-5A/BMD-24244). J.L.T. acknowledges the FPU19/02292 grant from the Spanish Ministry of University. Additionally, J.L.T kindly appreciates the design of the Table of Contents figure, created by Henar Mateo-delaFuente.

References

- Condon, A. Designed DNA Molecules: Principles and Applications of Molecular Nanotechnology. Nat. Rev. Genet. 2006, 7, 565–575.
- (2) Braich, R.; Chelyapov, N.; Johnson, C.; Rothemund, P.; Adleman, L. Solution of a 20-variable 3-SAT Problem on a DNA computer. <u>Science (New York, N.Y.)</u> 2002, <u>296</u>, 499–502.
- (3) Xiong, X.; Zhu, T.; Zhu, Y.; Cao, M.; Xiao, J.; Li, L.; Wang, F.; Fan, C.; Pei, H.

Molecular convolutional neural networks with DNA regulatory circuits. <u>Nat. Mach.</u> Intell. **2022**, <u>4</u>, 1–11.

- (4) O'Reilly, R. K.; Turberfield, A. J.; Wilks, T. R. The Evolution of DNA-Templated Synthesis as a Tool for Materials Discovery. Acc. Chem. Res. 2017, 50, 2496–2509.
- (5) Zhai, J.; Cui, H.; Yang, R. DNA based biosensors. Biotechnol. Adv. **1997**, 15, 43–58.
- (6) Saidur, M.; Aziz, A. A.; Basirun, W. Recent advances in DNA-based electrochemical biosensors for heavy metal ion detection: A review. <u>Biosens. Bioelectron.</u> 2017, <u>90</u>, 125–139.
- (7) Minunni, M.; Tombelli, S.; Mascini, M.; Bilia, A.; Bergonzi, M. C.; Vincieri, F. An optical DNA-based biosensor for the analysis of bioactive constituents with application in drug and herbal drug screening. Talanta 2005, 65, 578–585.
- (8) Bu, N.-N.; Tang, C.-X.; He, X.-W.; Yin, X.-B. Tetrahedron-structured DNA and functional oligonucleotide for construction of an electrochemical DNA-based biosensor. ChemComm. 2011, 47, 7689–7691.
- (9) Liu, A.; Wang, K.; Weng, S.; Lei, Y.; Lin, L.; Chen, W.; Lin, X.; Chen, Y. Development of electrochemical DNA biosensors. Trends Anal. Chem. 2012, 37, 101–111.
- (10) Izadi, Z.; Sheikh-Zeinoddin, M.; Ensafi, A. A.; Soleimanian-Zad, S. Fabrication of an electrochemical DNA-based biosensor for Bacillus cereus detection in milk and infant formula. Biosens. Bioelectron. **2016**, 80, 582–589.
- (11) Drummond, T. G.; Hill, M. G.; Barton, J. K. Electrochemical DNA sensors. <u>Nat.</u> <u>Biotechnol.</u> 2003, <u>21</u>, 1192–1199.
- (12) Paleček, E.; Fojta, M.; Jelen, F. New approaches in the development of DNA sensors: hybridization and electrochemical detection of DNA and RNA at two different surfaces.

<u>Bioelectrochemistry</u> **2002**, <u>56</u>, 85–90, Extended Abstracts of the XVIth International Symposium on Bioelectrochemistry and Bioenergetics Part 2.

- (13) Zhang, S.; Wang, K.; Li, K.-B.; Shi, W.; Jia, W.-P.; Chen, X.; Sun, T.; Han, D.-M. A DNA-stabilized silver nanoclusters/graphene oxide-based platform for the sensitive detection of DNA through hybridization chain reaction. <u>Biosens. Bioelectron.</u> 2017, <u>91</u>, 374–379.
- (14) Dai, N.; Kool, E. T. Fluorescent DNA-based enzyme sensors. <u>Chem. Soc. Rev.</u> 2011, <u>40</u>, 5756–5770.
- (15) Zhou, C.; Zou, H.; Sun, C.; Ren, D.; Chen, J.; Li, Y. Signal amplification strategies for DNA-based surface plasmon resonance biosensors. <u>Biosens. Bioelectron.</u> 2018, <u>117</u>, 678–689.
- (16) Berlin, Y. A.; Burin, A. L.; Ratner, M. A. DNA as a molecular wire. <u>Superlattices</u> Microstruct. **2000**, 28, 241–252.
- (17) Wohlgamuth, C. H.; McWilliams, M. A.; Slinker, J. D. DNA as a Molecular Wire: Distance and Sequence Dependence. Anal. Chem. 2013, 85, 8634–8640.
- (18) Kissinger, P. T. Biosensors—a perspective. Biosens. Bioelectron. 2005, 20, 2512–2516.
- (19) Mehrotra, P. Biosensors and their applications A review. J. Oral. Biol. Craniofac. <u>Res.</u> 2016, 6, 153–159.
- (20) D'Annibale, V.; Nardi, A. N.; Amadei, A.; D'Abramo, M. Theoretical Characterization of the Reduction Potentials of Nucleic Acids in Solution. <u>J. Chem. Theory Comput.</u> **2021**, <u>17</u>, 1301–1307.
- (21) Psciuk, B. T.; Lord, R. L.; Munk, B. H.; Schlegel, H. B. Theoretical Determination of One-Electron Oxidation Potentials for Nucleic Acid Bases. <u>J. Chem. Theory Comput.</u> 2012, 8, 5107–5123.

- (22) Faraggi, M.; Broitman, F.; Trent, J. B.; Klapper, M. H. One-Electron Oxidation Reactions of Some Purine and Pyrimidine Bases in Aqueous Solutions. Electrochemical and Pulse Radiolysis Studies. J. Phys. Chem. **1996**, 100, 14751–14761.
- (23) Jovanovic, S. V.; Simic, M. G. One-electron redox potentials of purines and pyrimidines.J. Phys. Chem. **1986**, 90, 974–978.
- (24) Crespo-Hernández, C. E.; Close, D. M.; Gorb, L.; Leszczynski, J. Determination of Redox Potentials for the Watson-Crick Base Pairs, DNA Nucleosides, and Relevant Nucleoside Analogues. J. Phys. Chem. B 2007, 111, 5386–5395.
- (25) Seidel, C. A. M.; Schulz, A.; Sauer, M. H. M. Nucleobase-Specific Quenching of Fluorescent Dyes. 1. Nucleobase One-Electron Redox Potentials and Their Correlation with Static and Dynamic Quenching Efficiencies. J. Phys. Chem. **1996**, 100, 5541–5553.
- (26) Steenken, S.; Jovanovic, S. V. How Easily Oxidizable Is DNA? One-Electron Reduction Potentials of Adenosine and Guanosine Radicals in Aqueous Solution. <u>J. Am. Chem.</u> <u>Soc.</u> 1997, <u>119</u>, 617–618.
- (27) Steenken, S.; Jovanovic, S. V.; Bietti, M.; Bernhard, K. The Trap Depth (in DNA) of 8-Oxo-7,8-dihydro-2'deoxyguanosine as Derived from Electron-Transfer Equilibria in Aqueous Solution. J. Am. Chem. Soc. **2000**, 122, 2373–2374.
- (28) Wang, J.; Yang, S.; Zhang, Y. One-electron oxidation and redox potential of nucleobases and deoxyribonucleosides computed by QM/MM simulations. <u>Chem. Phys. Lett.</u> 2020, 739, 136948.
- (29) Zhang, Y.; Xie, P.; Yang, S.; Han, K. Ionization and Electron Attachment for Nucleobases in Water. J. Phys. Chem. B 2019, 123, 1237–1247.
- (30) Paukku, Y.; Hill, G. Theoretical Determination of One-Electron Redox Potentials for DNA Bases, Base Pairs, and Stacks. J. Phys. Chem. A 2011, 115, 4804–4810.

- (31) Lucia-Tamudo, J.; Cárdenas, G.; Anguita-Ortiz, N.; Díaz-Tendero, S.; Nogueira, J. J. Computation of Oxidation Potentials of Solvated Nucleobases by Static and Dynamic Multilayer Approaches. J. Chem. Inf. Model. **2022**, 62, 3365–3380.
- (32) Lucia-Tamudo, J.; Díaz-Tendero, S.; Nogueira, J. J. Intramolecular and intermolecular hole delocalization rules the reducer character of isolated nucleobases and homogeneous single-stranded DNA. Phys. Chem. Chem. Phys. **2023**, 25, 14578–14589.
- (33) Boon, E. M.; Barton, J. K. Charge transport in DNA. <u>Curr. Opin. Struct. Biol.</u> 2002, 12, 320–329.
- (34) Delaney, S.; Barton, J. K. Long-Range DNA Charge Transport. <u>J. Org. Chem.</u> 2003, 68, 6475–6483.
- (35) Fujitsuka, M.; Majima, T. Hole and excess electron transfer dynamics in DNA. <u>Phys.</u> Chem. Chem. Phys. **2012**, 14, 11234–11244.
- (36) Giese, B.; Wessely, S.; Spormann, M.; Lindemann, U.; Meggers, E.; Michel-Beyerle, M. E. On the Mechanism of Long-Range Electron Transfer through DNA. Angew. Chem. Int. Ed. 1999, 38, 996–998.
- (37) Voityuk, A. A. Charge transfer in DNA: Hole charge is confined to a single base pair due to solvation effects. J. Chem. Phys. 2005, 122, 204904.
- (38) Rooman, M.; Wintjens, R. Sequence and conformation effects on ionization potential and charge distribution of homo-nucleobase stacks using M06-2X hybrid density functional theory calculations. J. Biomol. Struct. Dyn. 2014, 32, 532–545.
- (39) Lucia-Tamudo, J.; Alcamí, M.; Díaz-Tendero, S.; Nogueira, J. J. One-Electron Oxidation Potentials and Hole Delocalization in Heterogeneous Single-Stranded DNA. ChemRxiv 2023, preprint, DOI: 10.26434/chemrxiv-2023-csvcp.

- (40) Case, D.; Aktulga, H.; Belfon, K.; Ben-Shalom, I.; Brozell, S.; Cerutti, D.; III, T. C.; Cisneros, G.; Cruzeiro, V.; Darden, T.; Duke, R.; Giambasu, G.; Gilson, M.; Gohlke, H.; Goetz, A.; Harris, R.; Izadi, S.; Izmailov, S.; Jin, C.; Kasavajhala, K.; Kaymak, M.; King, E.; Kovalenko, A.; Kurtzman, T.; Lee, T.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Machado, M.; Man, V.; Manathunga, M.; Merz, K.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; O'Hearn, K.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Rahnamoun, A.; Roe, D.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C.; Skrynnikov, N.; Smith, J.; Swails, J.; Walker, R.; Wang, J.; Wei, H.; Wolf, R.; Wu, X.; Xue, Y.; York, D.; Zhao, S.; Kollman, P. Amber 2021. 2021.
- (41) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An overview of the Amber biomolecular simulation package. Wiley Interdiscip. Rev. Comput. Mol. Sci. 2013, 3, 198–210.
- (42) Case, D. A.; Cheatham III, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz Jr., K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. J. Comput. Chem. **2005**, 26, 1668–1688.
- (43) Wang, J.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? <u>J. Comput. Chem.</u> 2000, <u>21</u>, 1049–1074.
- (44) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of alpha/gamma Conformers. Biophys. J. 2007, 92, 3817–3829.
- (45) Ivani, I.; Dans, P. D.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goñi, R.; Balaceanu, A.; Portella, G.; Battistini, F.; Gelpí, J. L.; González, C.; Vendruscolo, M.; Laughton, C. A.; Harris, S. A.; Case, D. A.; Orozco, M. Parmbsc1: A refined force field for DNA simulations. <u>Nat. Methods</u> **2016**, <u>13</u>, 55–58.
- (46) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L.

Comparison of simple potential functions for simulating liquid water. <u>J. Chem. Phys.</u> **1983**, 79, 926–935.

- (47) Joung, I. S.; Cheatham, T. E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. <u>J. Phys. Chem. B</u> 2008, 112, 9020–9041.
- (48) Poltev, V. In <u>Handbook of Computational Chemistry</u>; Leszczynski, J., Ed.; Springer Netherlands: Dordrecht, 2016; pp 1–48.
- (49) Adcock, S. A.; McCammon, J. A. Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins. Chem. Rev. 2006, 106, 1589–1615.
- (50) Braun, E.; Gilmer, J.; Mayes, H. B.; Mobley, D. L.; Monroe, J. I.; Prasad, S.; Zuckerman, D. M. Best Practices for Foundations in Molecular Simulations [Article v1.0]. Living J. Comp. Mol. Sci. **2018**, 1, 5957–5957.
- (51) Meza, J. C. Steepest descent. Wiley Interdiscip. Rev. Comput. Stat. 2010, 2, 719–722.
- (52) Galántai, A. The theory of Newton's method. <u>J. Comput. Appl. Math.</u> 2000, <u>124</u>, 25–44.
- (53) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. <u>J.</u> Comput. Phys. **1977**, 23, 327–341.
- (54) Hammonds, K. D.; Heyes, D. M. Shadow Hamiltonian in classical NVE molecular dynamics simulations: A path to long time stability. J. Chem. Phys. 2020, 152, 024114.
- (55) Yoneya, M.; Berendsen, H. J. C.; Hirasawa, K. A Non-Iterative Matrix Method for Constraint Molecular Dynamics Simulations. <u>Mol. Simul.</u> 1994, <u>13</u>, 395–405.
- (56) Neese, F.; Wennmohs, F.; Becker, U.; Riplinger, C. The ORCA quantum chemistry program package. J. Chem. Phys. 2020, 152, 224108.

- (57) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. <u>J. Chem. Theory</u> Comput. **2019**, 15, 1652–1671.
- (58) Petersson, G. A.; Bennett, A.; Tensfeldt, T. G.; Al-Laham, M. A.; Shirley, W. A.; Mantzaris, J. A complete basis set model chemistry. I. The total energies of closed-shell atoms and hydrides of the first-row elements. J. Chem. Phys. **1988**, 89, 2193–2218.
- (59) Petersson, G. A.; Al-Laham, M. A. A complete basis set model chemistry. II. Openshell systems and the total energies of the first-row atoms. <u>J. Chem. Phys.</u> 1991, <u>94</u>, 6081–6090.
- (60) Marcus, R. A. On the Theory of Oxidation-Reduction Reactions Involving Electron Transfer. I. J. Chem. Phys. 1956, 24, 966–978.
- (61) Marcus, R. A. On the Theory of Oxidation-Reduction Reactions Involving Electron Transfer. III. Applications to Data on the Rates of Organic Redox Reactions. <u>J. Chem.</u> <u>Phys.</u> 1957, <u>26</u>, 872–877.
- (62) Marcus, R. A. On the Theory of Oxidation-Reduction Reactions Involving Electron Transfer. V. Comparison and Properties of Electrochemical and Chemical Rate Constants. J. Phys. Chem. **1963**, 67, 853–857.
- (63) Marcus, R. A. On the theory of electron-transfer reactions. VI. Unified treatment for homogeneous and electrode reactions. J. Chem. Phys. **1965**, <u>43</u>, 679–701.
- (64) Marcus, R. A. Electrostatic Free Energy and Other Properties of States Having Nonequilibrium Polarization. I. J. Chem. Phys. 1956, 24, 979–989.
- (65) Marcus, R. A. On the Theory of Oxidation-Reduction Reactions Involving Electron

Transfer. II. Applications to Data on the Rates of Isotopic Exchange Reactions. <u>J.</u> Chem. Phys. **1957**, 26, 867–871.

- (66) Ehlert, S.; Stahn, M.; Spicher, S.; Grimme, S. Robust and Efficient Implicit Solvation Model for Fast Semiempirical Methods. <u>J. Chem. Theory Comput.</u> 2021, <u>17</u>, 4250– 4261.
- (67) Löwdin, P.-O.; Shull, H. Natural Orbitals in the Quantum Theory of Two-Electron Systems. Phys. Rev. 1956, 101, 1730–1739.
- (68) Pipek, J.; Mezey, P. G. A fast intrinsic localization procedure applicable for ab initio and semiempirical linear combination of atomic orbital wave functions. <u>J. Chem. Phys.</u> 1989, 90, 4916–4926.
- (69) Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. J. Chem. Theory Comput. **2012**, 8, 1542–1555.
- (70) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. J. Chem. Theory Comput. **2013**, 9, 3878–3888.
- (71) Cárdenas, G.; Lucia-Tamudo, J.; Mateo-delaFuente, H.; Palmisano, V. F.; Anguita-Ortiz, N.; Ruano, L.; Pérez-Barcia, A.; Díaz-Tendero, S.; Mandado, M.; Nogueira, J. J. MoBioTools: A toolkit to setup quantum mechanics/molecular mechanics calculations. J. Comput. Chem. **2023**, 44, 516–533.
- (72) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. <u>J. Mol.</u> Graph. **1996**, 14, 33–38.