

## How abasic sites impact hole transfer dynamics in GC-rich DNA sequences

Received 00th January 20xx,  
Accepted 00th January 20xx

Marina Corbella<sup>a</sup>, Alexander A. Voityuk<sup>\*b,c</sup>, and Carles Curutchet<sup>\*a</sup>

DOI: 10.1039/x0xx00000x

www.rsc.org/

Changes in DNA charge transfer properties upon the creation of apurinic and apyrimidinic sites has been used to monitor DNA repair processes, given that such lesions generally reduce charge transfer yields. However, because these lesions translate into distinct intra and extrahelical conformations depending on the nature of the unpaired base and its DNA context, it is unclear the actual impact of such diverse conformations on charge transfer. Here we combine classical molecular dynamics, quantum/molecular mechanics (QM/MM) calculations, and Kinetic Monte Carlo simulations to investigate the impact of abasic sites on the structure and hole transfer (HT) properties of DNA. We consider both apurinic and apyrimidinic sites in polyG and polyGC sequences and find that most situations lead to intrahelical conformations where HT rates are significantly slowed down due to the energetic disorder induced by the abasic void. In contrast, the presence of an unpaired C flanked by C bases leads to an extrahelical conformation where stacking among G sites is reduced, leading to an attenuation of electronic couplings and a destabilization of hole states. Interestingly, this leads to an asymmetric HT behavior, given that the 5' to 3' transfer along the G strand is slowed down by one order of magnitude while the opposite 3' to 5' transfer remains similar to that estimated for the reference polyG sequence. Our simulations thus suggest that electrochemical monitoring of DNA repair process following changes in charge transfer properties can miss repair events linked to abasic sites adopting extrahelical conformations.

### 1. Introduction

The loss of a purine or pyrimidine base in DNA, resulting in the creation of an abasic site, is one of the most frequent DNA lesions, with a ratio of formation reaching thousands of times per day in a human cell.<sup>1</sup> Abasic sites are formed by spontaneous or enzymatic hydrolysis of the N-glycosidic bond,<sup>2–4</sup> by chemical modifications of nucleic bases, or by physical agents like UV or  $\gamma$  radiation.<sup>3</sup> Events leading to abasic sites often involve guanines, and the structure and stability of G-rich quadruplex-forming DNA sequences, for example, has been shown to be strongly affected by such defects.<sup>5–8</sup> In general, sequences containing abasic sites can adopt unusual conformations, making them fragile and susceptible to breaks.<sup>9,10</sup> Luckily, cells are equipped with an elaborate repair machinery to neutralize the proliferation of cytotoxic and promutagenic defects. Among such repair mechanisms, base excision repair (BER) starts with the formation of an abasic site due to the excision of a damaged or mispaired nucleobase by the action of a DNA glycosylase.<sup>11,12</sup> The process then continues with the recognition of the abasic site by an

apurinic/apyrimidinic (AP) endonuclease<sup>13,14</sup> and finishes with the replacement of the missing nucleotide. Other repair mechanisms, which rely on the direct dealkylation by an alkyltransferase or oxidative dealkylation by AlkB enzymes, have recently been studied in detail using multiscale simulations.<sup>15–18</sup> The study of the different steps involved in the repair process is crucial and may have vast consequences in cancer research.

On the other hand, there is a growing interest in the synthetic incorporation of abasic sites due to its application to the development of biosensors. Abasic site biosensors have been used to detect analytes using fluorescent and electrochemical probes by selectively binding to the unpaired nucleobase,<sup>19,20</sup> while electrochemical monitoring of DNA repair processes is also possible following the change on the charge transfer (CT) properties associated with the creation of abasic sites.<sup>21</sup> Indeed, the CT properties of DNA have been found to be strongly sensitive to the presence of abasic sites, which in general lead to a decrease in the CT yield.<sup>22,23</sup> How abasic sites modulate CT kinetics on DNA duplexes, however, has been only briefly explored.<sup>24</sup> Many electrochemical studies, for example, have used abasic sites as a brake for the CT process. The Slinker group proved the DNA repair function of two DNA glycosylases by measuring the current through a DNA monolayer after and before the damaged base excision, showing a substantial decrease in redox signal after the formation of the abasic site.<sup>21</sup> This group also demonstrated an attenuation of the CT yield in the presence of an abasic site in

<sup>a</sup> Departament de Farmàcia i Tecnologia Farmacèutica i Físicoquímica and Institut de Química Teòrica i Computacional (IQTC-UB), Universitat de Barcelona, 08028 Barcelona, Spain. E-mail: carles.curutchet@ub.edu; Tel: +34-934-024-557

<sup>b</sup> Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain

<sup>c</sup> Institució de Química Computacional i Catàlisi and Departament de Química, Universitat de Girona, Campus Montivili, 17071 Girona, Spain. E-mail: alexander.voityuk@icrea.cat; Tel: +34-972-418-926

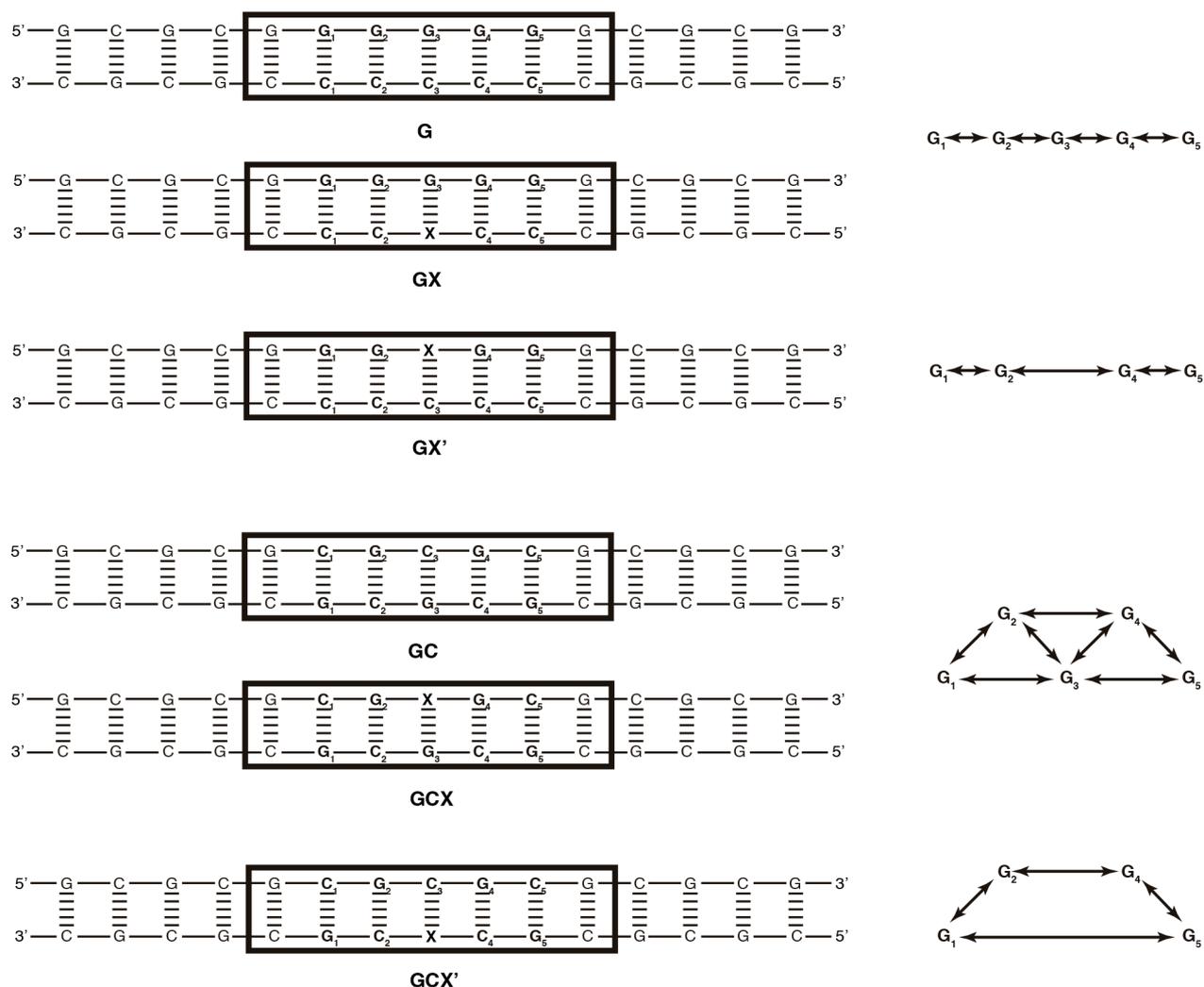


Fig. 1 Sequences of the duplex DNA systems considered (left) and corresponding Kinetic Monte Carlo schemes (right) used to simulate hole transfer dynamics. Nucleobases enclosed in the boxes were included in the quantum-mechanical region in QM/MM calculations.

a 17-mer AT-rich duplex.<sup>24</sup> Given the well-known dependence of CT properties on DNA sequence, the absence of a nucleobase is expected to directly modulate CT rates and pathways. However, the unusual conformations induced by the defect can lead to concomitant changes in the CT parameters (site energies and couplings) of proximate nucleobases, thus indirectly modulating CT kinetics. Indeed, a variety of studies indicate that the unpaired nucleobase can adopt intra and extrahelical conformations, depending on the nature of the unpaired base and its context in the DNA sequence.<sup>25–37</sup>

In this study, we combine classical molecular dynamics (MD) simulations with semiempirical INDO/S quantum/molecular mechanics (QM/MM) calculations to investigate how an abasic site influences the structure and hole transfer (HT) kinetics of six model GC-rich DNA duplexes. In particular, we investigate the impact of a guanine (G) or cytosine (C) deletion in 5'-GXG-3' and 5'-CXC-3' contexts and compare our results with the analogous unmodified DNA 15-mers. We aim at understanding

how these defects distort the double helix structure, the site energies and electronic couplings of the proximate nucleobases, and the overall HT process, depending on the DNA sequence. Our results indicate that in situations where the unpaired nucleobase adopts intrahelical conformations, the overall HT process is slowed down by ~1-2 orders of magnitude. However, when an unpaired C adopts an extrahelical conformation the transfer becomes similar to the reference DNA in the 5' to 3' transfer direction along the G strand, but it is slowed down by 1 order of magnitude in the opposite direction, thus introducing a preferential route for oxidative damage.

## 2. Methods

### 2.1 MD simulations

MD simulations at room temperature (300 K) were performed for the six 15-mer DNA duplexes shown in Fig. 1, terminated at

both 3' and 5' ends by hydroxyl groups. These systems represent polyG and polyGC central sequences where either a G or a C base has been deleted leading to an abasic site. The sequences were capped with GCGC moieties to keep base pairing in the ending regions and thus prevent strand slippage, which has been found to be enhanced by the presence of abasic sites.<sup>38</sup> We adopted the prototypical abasic site, tetrahydrofuran (X), commonly used as a model for enzymatically generated abasic sites, which exist as mixtures of  $\alpha$  and  $\beta$  hemiacetals, due to its chemical stability.<sup>28,37</sup> MD simulations were performed starting from standard B-DNA fiber geometries created using the NAB module of the Amber12 software,<sup>39</sup> eventually modified in order to delete a C or G nucleobase and generate the abasic site. The systems were neutralized adding Na<sup>+</sup> ions<sup>40</sup> and solvated in a truncated octahedron box of TIP3P water (buffer region 14 Å). Two sets of simulations were run either adopting the parmbsc0<sup>41</sup> and the recently developed parmbsc1<sup>42</sup> force fields for DNA, the latter introducing some improvements, for example, related to excessive terminal fraying. The tetrahydrofuran residue was modelled using parmbsc0 and parmbsc1 parameters for the phosphate and sugar moieties, the latter capped with a hydrogen atom, which charge was adjusted to give a global charge of -1. The solvent and counterions were first energy minimized during 1000 steps, followed by 2500 steps of further minimization of the full system. Then, the systems were gradually thermalized up to 300 K during 200 ps with weak constraints in the nucleic acids (10 kcal/mol·Å<sup>2</sup>), followed by 200 ns production runs at 1 bar. For the GX' system, production runs were extended up to 350 ns. All simulations were done using the Amber 12 suite of programs<sup>39</sup> using an integration time step of 2 fs, the SHAKE algorithm to restrain bonds involving hydrogen, periodic boundary conditions, the Particle Mesh Ewald approach, and a nonbonded cutoff of 10 Å. For each system, we extracted a total of 1000 structures at equal intervals from the last 100 ns of MD trajectories to be used as input in QM/MM calculations. In addition, we performed a cluster analysis based on the root-mean square deviations (RMSD) along the last 100 ns of trajectory. From this analysis, a representative structure for each system was extracted from the most populated cluster.

## 2.2 QM/MM calculations

QM/MM calculations were performed using the semiempirical INDO/S method on the structures extracted from the MD trajectories. Then, we applied Koopman's approximation to estimate HT energies and electronic couplings.<sup>43</sup> Previous benchmark studies have shown that application of this scheme at the INDO/S level of theory<sup>44,45</sup> provides surprisingly good estimates of both HT energies and couplings in DNA stacks compared to high-level MS-CASPT2 data,<sup>46</sup> while keeping a moderate computational cost that allows accounting for structural dynamics, which significantly affect CT properties in

DNA.<sup>43</sup> In particular, the semiempirical INDO/S method has been shown to perform well in calculations of HT parameters, while the MNDO scheme and related approaches like AM1 or PM3 considerably underestimate couplings values.<sup>44</sup> We note here that the MS-CASPT2 benchmarks were obtained using a level shift parameter of 0.2, whereas no IPEA shift was used. In all calculations the QM region included the 7 base pairs in the central region of the duplex (excluding the sugar and backbone atoms), as indicated in Fig. 1, where the C1' atom of the sugar unit was replaced by a hydrogen atom with a 1.08 Å bond length. Electrostatic embedding effects exerted by the environment (DNA, solvent, and counterions) were considered adopting the charges as defined in the force field used for the MD simulations. Thus, we do not account for explicit polarization effects in the MM region. MM polarization has been shown to introduce a considerable impact on reorganization energies,<sup>47</sup> but in this study we do not compute this quantity, as explained in the next section. Instead, we expect minor changes on HT couplings, which are mediated by orbital overlap,<sup>48</sup> and on HT energies, as here the energies of the radical cation states are not explicitly calculated. In fact, they are obtained from an MD simulation of the neutral QM system including 7 base pairs, so explicit MM polarization is expected to introduce small variations on the relative energies of the sites.

Hole states associated to the G sites in each duplex were represented by six or seven highest-occupied molecular orbitals (HOMOs) of the neutral species (HOMO → HOMO-5/HOMO-6), while CT couplings were estimated using the Fragment Charge Difference method (FCD):<sup>49</sup>

$$V_{12} = \frac{(E_2 - E_1)|\Delta q_{12}|}{\sqrt{(\Delta q_1 - \Delta q_2)^2 + 4\Delta q_{12}^2}} \quad (1)$$

For hole transfer (HT), adiabatic splittings  $E_2 - E_1$  were estimated using the one-electron energies of the HOMOs of neutral systems. Assuming a two-state-model, free energy differences for the HT reaction were then estimated as  $|\Delta G^0| = \sqrt{(E_2 - E_1)^2 - 4V_{12}^2}$ . In Eq. (1),  $\Delta q_1$  and  $\Delta q_2$  are the difference of hole charges on the donor and acceptor sites in the adiabatic states of interest and  $\Delta q_{12}$  is the corresponding off-diagonal term. The calculation of these quantities and the underlying approximations are described in previous studies.<sup>49,50</sup>

## 2.3 Hole transfer kinetics

HT rates for the forward and reverse reactions shown in Fig. 1 were estimated using Marcus theory:

$$k_{HT} = \frac{2\pi}{\hbar} |V|^2 \frac{1}{\sqrt{4\pi k_B T \lambda}} e^{-(\lambda + \Delta G^0)^2 / 4\lambda k_B T} \quad (2)$$

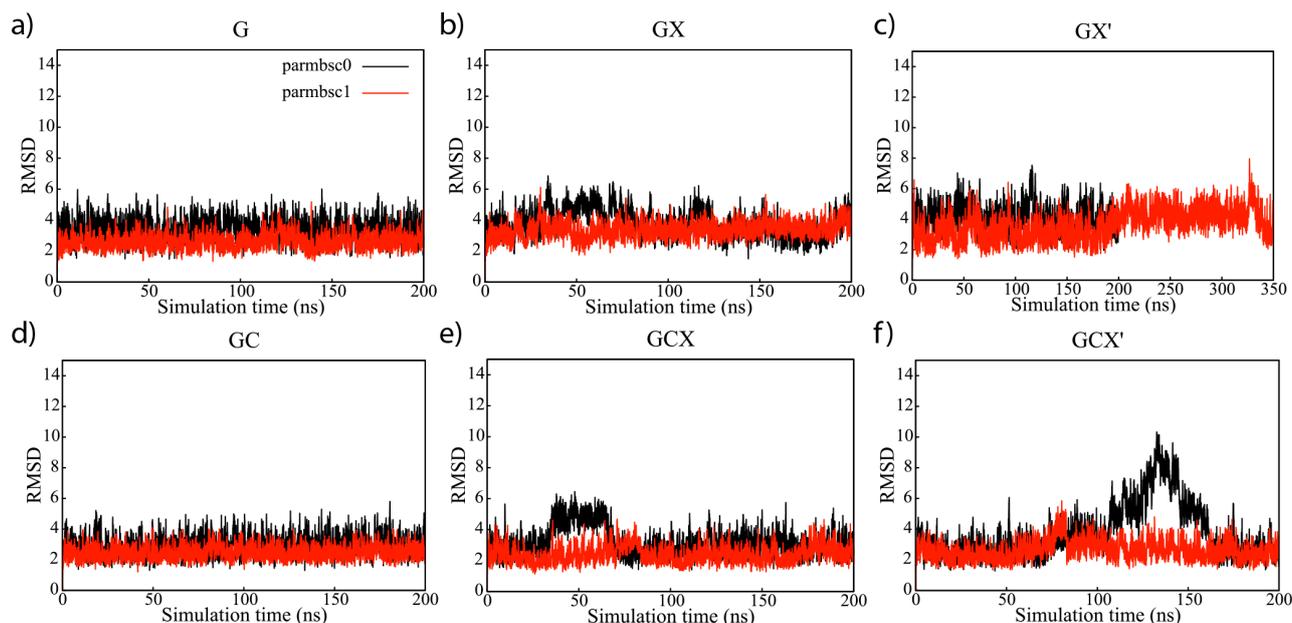


Fig. 2 Positional root-mean square deviations of DNA duplexes along MD trajectories: a) G, b) GX, c) GX', d) GC, e) GCX and f) GCX'.

The driving force of the HT reaction ( $\Delta G^0$ ) was estimated from the energy difference in donor/acceptor ionization potentials, and the reorganization energy ( $\lambda$ ), which includes internal and solvent terms, was assumed to be 0.7 eV in all cases.<sup>51</sup> The latter value is a compromise between an experimental estimate  $0.6 \pm 0.1$  eV derived for hole transfer in oligonucleotides<sup>52</sup> and values  $\sim 0.8$ -1.5 eV obtained in several computational studies.<sup>53-55</sup> Whereas a different choice of  $\lambda$  would change the absolute HT timecales we predict, the relative HT rates between standard DNA and the analogous sequences including an abasic site are more robust and essentially unaffected by the value of  $\lambda$ . All rates were computed using the energies and squared couplings averaged over the MD trajectories.

The effective rates corresponding to the overall HT process  $G_1 \rightarrow G_5$  and  $G_5 \rightarrow G_1$  were simulated using a Kinetic Monte Carlo (KMC) algorithm.<sup>56</sup> HT rates were used to generate a cumulative probability distribution function (CPD) for each possible step. This function is associated with a transition  $i \rightarrow n$  according to Eq. (3), where  $\sum_m k_{im}$  is the sum of transfer rates from the same origin  $i$ .

$$CPD_{i(n)} = \frac{\sum_{p=1}^n k_{ip}}{\sum_m k_{im}} \quad (3)$$

The Monte Carlo hopping algorithm consists of the following steps: 1) Compute the CPD function for each allowed transfer. 2) Start at a given site  $i$  ( $G_1$  or  $G_5$ ), choose a random number  $0 < r_1 \leq 1$  and hop to site  $j$ , where  $CPD_{ij} \geq r_1 > 0$ . 3) Compute the waiting time  $t_w$  at site  $i$ , generating a new random number  $0 < r_2 \leq 1$  as  $t_w = -(\ln(r_2)/\sum_m k_{im})$ . 4) Choose a new random number and repeat steps 2 and 3 recording the total accumulated waiting time. 5) Stop if the final site has been

reached or if the total time exceeds a fixed value  $t_{\max}$ . For each system, we averaged HT times over  $10^6$  realizations of the KMC algorithm.

### 3. Results and discussion

#### 3.1 DNA structure

The impact of abasic sites on the structure and flexibility of DNA has been investigated both using theoretical models and NMR spectroscopy.<sup>25-37</sup> The bulk of these studies indicate that the overall B-DNA structure of the duplex is maintained in the presence of an abasic site. However, the lesion increases structural flexibility, and the local conformation of the unpaired base depends on its nature and its sequence context. Thus, whereas purine unpaired nucleobases adopt intrahelical conformations that preserve  $\pi$  stacking, pyrimidines stack poorly and tend to be extrahelical, although intrahelical conformations have also been identified when the pyrimidine is flanked by two guanines. Thus, it is important to assess the reliability of the structural models obtained here from MD simulations before investigating the impact of abasic sites on hole transfer dynamics.

In Fig. 2 we show the positional RMSD of the DNAs computed with respect to the initial B-DNA structure along the MD simulations performed using either the parmbc1 or the parmbc0 force fields. In all cases, neglecting terminal base pairs only lead to minor variations in RMSD values, indicating that terminal fraying was small. If we focus on the standard DNA sequences without the lesion (G and GC systems), simulations performed using the refined parmbc1 potential leads to RMSD values  $\sim 1$  Å lower than those obtained using the parmbc0, thus keeping the structures closer to the initial B-DNA. The presence of an abasic site (GX, GX', GCX and GCX'

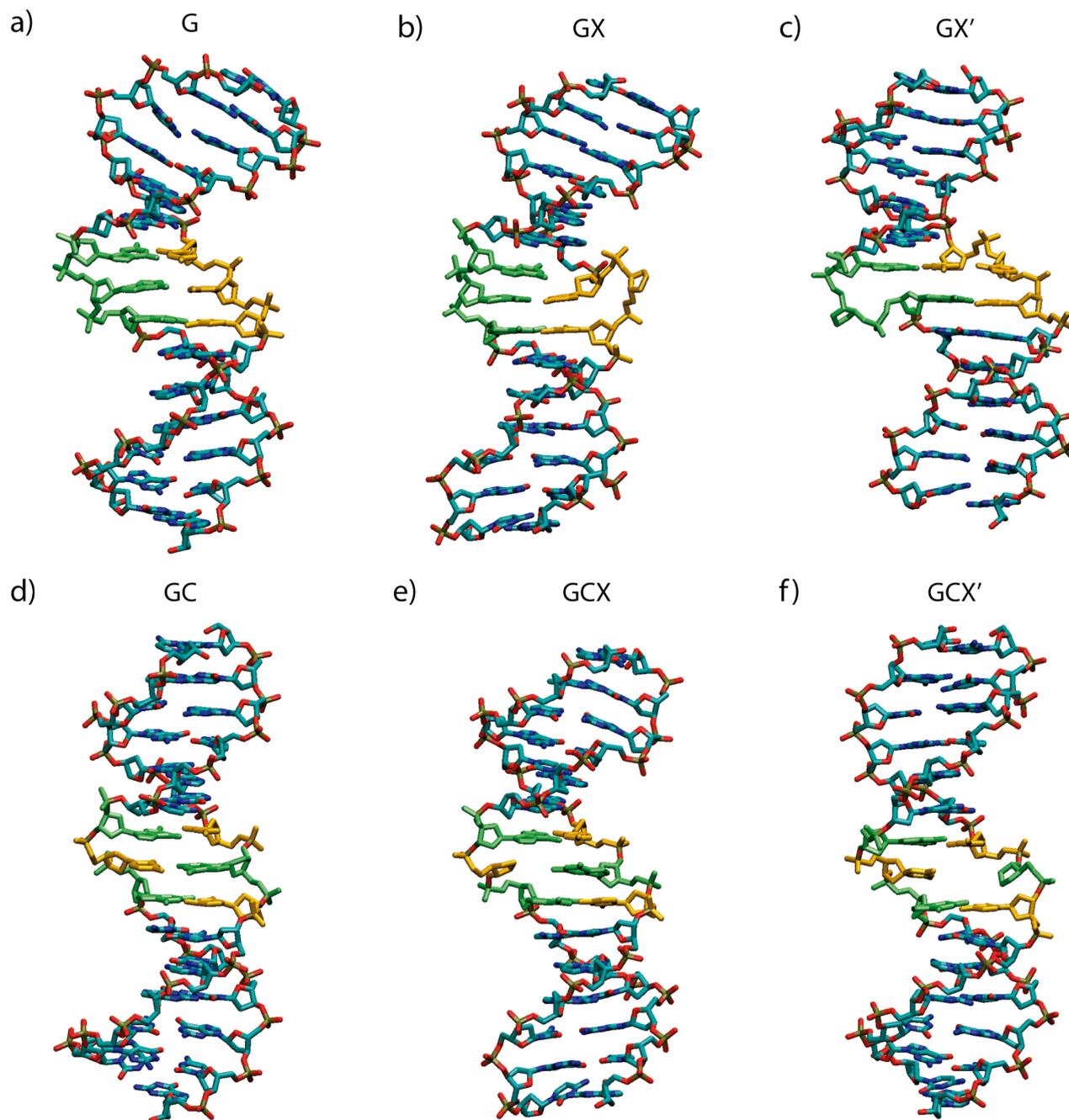


Fig. 3 Representative structures of DNA duplexes extracted from MD simulations based on the parmbsc1 force field: a) G, b) GX, c) GX', d) GC, e) GCX and f) GCX'.

systems) also leads to differences in the structures described by both force fields, again with parmbsc1 displaying more stable structures characterized by lower RMSD values.

For the GCX system, both force fields lead to similar intrahelical position for the unpaired guanine site, as shown in Fig. 3, with the sugar of the abasic site flipping in and out of the duplex, in accord with NMR data and theoretical simulations performed for an unpaired G in the same CGC<sup>25,27</sup> or similar CGT contexts.<sup>36</sup> When the unpaired guanine is flanked by other G bases (GX system), both force fields again predict intrahelical conformations, in accord with previous studies of unpaired guanines in GGG<sup>27,37</sup> or GGA sequences.<sup>28</sup>

In this case, however, the simulation performed using parmbsc1 shows a strong disruption of the Watson-Crick hydrogen bonds between the G<sub>2</sub> and C<sub>2</sub> base pairs adjacent to X, with the unpaired G<sub>3</sub> and C<sub>2</sub> forming non-Watson-Crick hydrogen bonds. This pattern is also briefly exchanged with similar interactions between G<sub>3</sub> and C<sub>4</sub> (the base 5' to X). The observation of this interaction between the unpaired base and a nucleobase adjacent to X agrees with previous predictions for this GGG sequence<sup>27</sup> and was also observed for an unpaired cytosine in a GCA context.<sup>28</sup> In a recent study, NMR and constrained MD simulations also pointed to such non-Watson-Crick pattern for the same GGG sequence when a 2'-

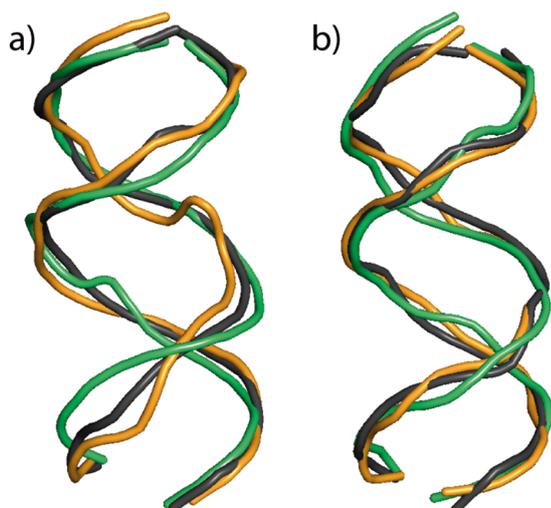


Fig. 4 Backbone overlay of representative structures of DNA duplexes extracted from MD simulations based on the parmbsc1 force field. a) G (black), GX (orange) and GX' (green), b) GC (black), GCX (orange) and GCX' (green).

deoxyribose abasic site was considered, whereas, for its chemically stable analogue tetrahydrofuran, as used here, the standard Watson-Crick pattern was kept.<sup>37</sup> We note however that the sequence of the DNA in that study differs from the one considered here and contained also an additional 8-oxoguanine lesion.

On the other hand, when the unpaired site is a cytosine in a GCG context (GCX' system) the parmbsc1 potential describes a stable intrahelical conformation, in agreement with previous predictions,<sup>27</sup> whereas parmbsc0 leads to both intrahelical and extrahelical conformations characterized by large distortions of the overall stacking pattern. Intrahelical conformations have

previously been observed for an unpaired C in a similar GCA context<sup>28</sup> or for an unpaired T flanked by guanines.<sup>31</sup>

Finally, when the unpaired cytosine is located in a CCC sequence (GX' system), both force fields initially predict a stable intrahelical conformation. However, with the parmbsc1 potential at ~180 ns, the orphan base C<sub>3</sub> flips out of the helix, which collapses allowing the flanking bases to stack. More in detail, the carbonyl group of the extrahelical C<sub>3</sub> establishes a hydrogen bond with the amino group of C<sub>2</sub>, which keeps its Watson-Crick hydrogen bonds with G<sub>2</sub> thus leading to a triplex-like G<sub>2</sub>:C<sub>2</sub>:C<sub>3</sub> structure, in which the G<sub>4</sub>:C<sub>4</sub> base pair is mainly stacked with the C<sub>2</sub>:C<sub>3</sub> pair. Extension of the simulations for a total of 350 ns confirmed the stability of this conformation, in agreement with previous NMR data<sup>25</sup> and theoretical predictions<sup>27</sup> pointing out to such extrahelical configuration.

Overall, our analysis indicates that the parmbsc1 force field leads to more stable structures that nicely agree with previous theoretical and experimental data on the intra and extrahelical position of the unpaired nucleobase.<sup>25–37</sup> Moreover, in Fig. 4 we show superimposition of the backbone of the DNA duplexes studied, indicating that the abasic site in the GX, GX', GCX and GCX' duplexes does not alter the global B-like conformation, in accord again with previous studies. Thus the description of the structural impact of abasic sites seems to be quite robust when the refined parmbsc1 force field is adopted, and in the following, we use the trajectories sampled with this potential to investigate the impact of these lesions on hole transfer dynamics.

### 3.2 Charge transfer kinetics

HT in DNA is known to be rather sensitive to structural deformations of the double helix, especially those modifying the degree of stacking between nucleobases. We have thus computed the energies of the hole states as well as the

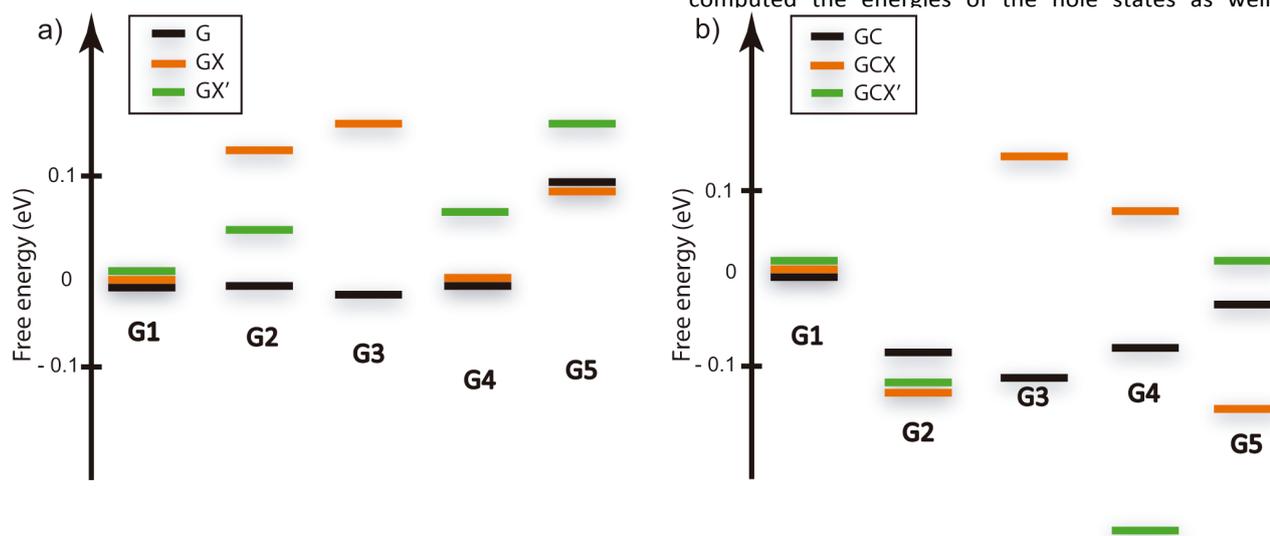


Fig. 5 Representation of the free energies computed using QM/MM calculations for the radical cation states localized on guanine sites. a) G, GX and GX' systems, b) GC, GCX and GCX' systems.

corresponding HT couplings for the guanines located in the central regions of the duplexes, as shown in Fig. 1, in order to investigate how this lesion affects HT dynamics. In Fig. 5 we show schematically the relative energy of different guanine sites in each system, whereas in Table 1 we report the energy differences among sites and the corresponding electronic coupling parameters and HT times. Based on the HT times for individual transfer steps, we then used KMC simulations to simulate the overall HT times reported in Table 2. We simulated both forward and backward processes, starting from  $G_1$  and recording the time needed to reach  $G_5$  or the inverse transfer process.

Table 1. Electronic couplings, free energy differences and forward and backward HT times computed for the DNA sequences.

	$V$ (meV)	$\Delta G$ (meV)	$\tau_F$ (ns)	$\tau_B$ (ns)
GC				
$G_1 - G_2$	24	-80	$1.9 \times 10^{-2}$	0.42
$G_1 - G_3$	10	-100	$6.9 \times 10^{-2}$	3.39
$G_2 - G_3$	9	-18	0.40	0.82
$G_2 - G_4$	8	7	0.83	0.64
$G_3 - G_4$	24	21	0.12	$5.4 \times 10^{-2}$
$G_3 - G_5$	11	69	1.60	0.11
$G_4 - G_5$	10	54	1.33	0.17
GCX				
$G_1 - G_2$	22	-119	$1.1 \times 10^{-2}$	1.13
$G_1 - G_3$	6	123	15.59	0.13
$G_2 - G_3$	21	248	29.59	$1.9 \times 10^{-3}$
$G_2 - G_4$	2	188	1085.78	0.71
$G_3 - G_4$	32	-67	$1.23 \times 10^{-2}$	0.17
$G_3 - G_5$	9	-261	$8.47 \times 10^{-3}$	222.07
$G_4 - G_5$	8	-185	$3.31 \times 10^{-2}$	44.39
GCX'				
$G_1 - G_2$	23	-106	$1.2 \times 10^{-2}$	0.75
$G_2 - G_4$	9	-153	$4.4 \times 10^{-2}$	16.54
$G_4 - G_5$	10	275	281	$6.3 \times 10^{-3}$
$G_1 - G_5$	0.3	13	848	505.82
G				
$G_1 - G_2$	91	-7	$4.8 \times 10^{-3}$	$6.3 \times 10^{-3}$
$G_2 - G_3$	91	-10	$4.6 \times 10^{-3}$	$6.6 \times 10^{-3}$
$G_3 - G_4$	102	10	$5.3 \times 10^{-3}$	$3.6 \times 10^{-3}$
$G_4 - G_5$	86	94	$4.3 \times 10^{-2}$	$1.1 \times 10^{-3}$
GX				
$G_1 - G_2$	92	135	$9.5 \times 10^{-2}$	$5.1 \times 10^{-4}$
$G_2 - G_3$	73	23	$1.3 \times 10^{-2}$	$5.5 \times 10^{-3}$
$G_3 - G_4$	117	-157	$2.2 \times 10^{-4}$	$9.8 \times 10^{-2}$
$G_4 - G_5$	106	87	$2.4 \times 10^{-2}$	$8.2 \times 10^{-4}$
GX'				
$G_1 - G_2$	94	43	$1.2 \times 10^{-2}$	$2.3 \times 10^{-3}$
$G_2 - G_4$	40	27	$4.9 \times 10^{-2}$	$1.7 \times 10^{-2}$
$G_4 - G_5$	61	89	$7.7 \times 10^{-2}$	$2.5 \times 10^{-3}$

In the absence of any lesion, the HT transfer process in the GC system follows several pathways involving both intra and interstrand hops, given the relatively similar energies and electronic couplings involved, although interstrand hops are slightly favored. In this case, the overall forward ( $G_1 \rightarrow G_5$ ) and backward ( $G_5 \rightarrow G_1$ ) times are  $\sim 2$  ns. In the G system the transfer only involves intrastrand hops, characterized by much

larger electronic coupling values  $\sim 100$  meV, thus leading to a faster transfer with times 0.2 and 0.04 ns for forward and backward processes.

Table 2. Overall hole transfer times computed using Kinetic Monte Carlo simulations for  $G_1 \rightarrow G_5$  and  $G_5 \rightarrow G_1$  transfers.

$\tau$ (ns)	G	GX'	GX	GC	GCX'	GCX
$G_1 \rightarrow G_5$	0.21	1.84	2.73	2.20	282.93	104.70
$G_5 \rightarrow G_1$	0.038	0.023	2.71	1.74	302.95	149.26

In the polyGC sequence, the presence of the abasic site (GCX and GCX' systems) does not lead to major structural changes in the double helix, as shown in Fig. 3, with the unpaired base adopting an intrahelical conformation. In the GCX' sequence, indeed, the electronic couplings relevant for the HT process ( $V_{12}$ ,  $V_{24}$ , and  $V_{45}$ ) remain very close to the GC reference system, and the main impact of the lesion is a significant stabilization  $\sim 150$  meV of  $G_4$  with respect of  $G_2$ . This stabilization slows down hops from  $G_2$  to the neighboring sites, and together with the lack of the intermediate  $G_3$  leads to a slower overall time of  $\sim 300$  ns compared to the GC duplex.

In the GCX system, in contrast, the unpaired  $G_3$  is partially displaced toward the abasic site void (see Fig. 3), leading to a better stacking with  $G_2$  and  $G_4$  which increases  $V_{23}$  and  $V_{34}$  interstrand couplings from 9 to 21 meV and from 24 to 32 meV, respectively, as found previously for a similar sequence.<sup>57</sup> On the other hand,  $V_{24}$  decreases from 8 to 2 meV, because in this case there is no  $C_3$  mediating the coupling between  $G_2$  and  $G_4$ . In this case, however, the abasic site induces an important destabilization of the hole states on  $G_3$  and  $G_4$ . Thus, although interstrand couplings are enhanced compared to the reference GC system, the HT process is again considerably slower with total times  $\sim 100$ -150 ns due to the larger time needed to access those sites.

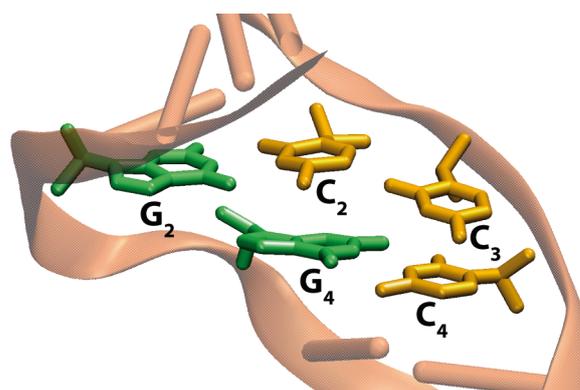


Fig. 6 Structure of the abasic site region for the GX' system.

In the G system, the abasic site leads to qualitatively different structural deformations in GX and GX'. As discussed previously, in GX the Watson-Crick hydrogen bonds between the  $G_2$  and  $C_2$  base pairs are disrupted, with the unpaired  $G_3$  forming non-Watson-Crick hydrogen bonds with  $C_2$  3' to the abasic site, as shown in Fig. 3. This conformational change does not alter much the couplings among guanines compared to the G system, which remains close to  $\sim 100$  meV, but leads to an

important energetic destabilization of  $G_2$  and  $G_3$  hole states and a corresponding increase in the times needed to reach these sites. Thus, the overall HT times, both for the forward and backward pathways, are increased to  $\sim 2$  ns compared to the 0.04–0.2 ns obtained for the same sequence without the lesion.

In the  $GX'$  system, on the other hand, the lesion induces an important conformational change, with unpaired  $C_3$  adopting an extrahelical conformation and the sugar of the abasic site being extruded from the helix. As discussed in the previous section, this leads to the triplex like  $G_2:C_2:C_3$  structure shown in Fig. 6, in which the  $G_4:C_4$  base pair is mainly stacked with the  $C_2:C_3$  pair, but also to some degree with the  $G_2:C_2$  pair. Because the stacking between  $G_1$  and  $G_2$  is not significantly modified, the  $V_{12}$  intrastrand coupling remains similar to the value estimated for G and GX, around 90 meV. The  $V_{24}$  and  $V_{45}$  couplings, however, are approximately halved compared to the standard  $\sim 100$  meV coupling value for intrastrand G-G contacts because the stacking between  $G_2/G_4$  and  $G_4/G_5$  is reduced for these pairs. This reduced stacking explains the partial destabilization of  $G_2$ ,  $G_4$  and  $G_5$  shown in Fig. 4, given the well-known ability of GG motifs to stabilize hole states.<sup>58</sup> KMC simulations indicate that destabilization of these sites, specially of  $G_5$ , induces a significant asymmetry in overall HT times for the forward  $G_1 \rightarrow G_5$  and the backward  $G_5 \rightarrow G_1$  process, for which we estimate time constants of 1.8 ns and 23 ps, respectively, compared to 210 ps and 38 ps estimated for the G system without lesion. Overall, thus, the partial disruption of the stacking interactions along the G strand in  $GX'$  leads to fast forward HT times similar to the reference G system, whereas the backward process is significantly slowed down by one order of magnitude, inducing a directional asymmetry in HT dynamics.

## Conclusions

Apurinic and apyrimidinic sites constitute one of the most frequent DNA lesions. Indeed, the repair of other lesions by the BER machinery leads also to the formation of an abasic site after enzymatic hydrolysis of the N-glycosidic bond. Electrochemical monitoring of DNA repair processes is thus often performed following the modulation of charge transfer properties associated with the creation of abasic sites. Despite the fact that several studies indicate a general attenuation of charge transfer yield in the presence of abasic sites, the fact that such lesions lead to diverse alterations in the DNA structure depending on the nature of the unpaired nucleobase and the DNA context suggests that the consequences on charge transfer dynamics are also diverse. In this study, we have investigated the impact of unpaired cytosine and guanine sites on the structure of polyG and polyGC sequences in atomic detail using MD simulations. Individual HT rates are then derived from Marcus theory based on MD-averaged energies and couplings estimated using QM/MM calculations, and the impact of abasic sites on overall HT dynamics are modeled using a Kinetic Monte Carlo scheme.

We find that MD simulations based on the parmbsc0 force field lead to large structural fluctuations and unusual conformations for the DNA sequences considered. In contrast, those performed using the recently refined parmbsc1 potential point to small alterations on the global B-DNA structure, with abasic sites adopting both intrahelical ( $GX$ ,  $GXC$  and  $GXC'$ ) and extrahelical ( $GX'$ ) conformations, in agreement with previous studies based on NMR spectroscopy and theoretical models. When the DNA adopts intrahelical conformations, the electronic interactions among guanines remain similar to the reference sequence without lesion. However, the abasic void induces a significant disorder in the energies of the hole states, which translate into HT longer times by  $\sim 1$ – $2$  orders of magnitude. In the  $GX'$  sequence adopting an extrahelical conformation, in contrast, the unpaired cytosine and the abasic sugar are extruded from the DNA and the double helix collapses. This leads to a reduced stacking among sites in the G strand, leading to a destabilization of the hole states over  $G_2$ ,  $G_4$  and  $G_5$  and a significant attenuation of electronic coupling values between these sites. Interestingly, this translates into an asymmetric behavior in HT kinetics, given that the  $5'$  to  $3'$  transfer along the guanine strand ( $G_1 \rightarrow G_5$ ) is slowed down by one order of magnitude, whereas HT in the opposite  $3'$  to  $5'$  direction remains similar to that estimated for the reference sequence.

Overall, our results thus show that abasic sites giving rise to intrahelical conformations generally slow down HT dynamics due to energetic disorder induced by the presence of the abasic void, whereas extrahelical conformations lead to less trivial implications on HT dynamics, including asymmetries in the directionality of HT. Thus, our results suggest that electrochemical monitoring of DNA repair process following changes in charge transfer properties can miss repair events linked to abasic sites adopting extrahelical conformations.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

We are grateful to the Consorci de Serveis Universitaris de Catalunya for providing access to computational resources (Molecular Recognition project). Financial support from the Agència de Gestió d'Ajuts Universitaris i de Recerca from Generalitat de Catalunya (GENCAT; 2014SGR1189, 2014SGR931, and 2017SGR1746) and from the Spanish Agencia Estatal de Investigación and the European Regional Development Fund (AEI/FEDER/UE; grants CTQ2012-36195, CTQ2015-69363-P, and CTQ2017-89924-P) are acknowledged. C.C. is a Serra Húnter Fellow (Generalitat de Catalunya), and M.C. acknowledges a fellowship from the Spanish Ministerio de Economía y Competitividad (MINECO; grant BES-2013-064088).

## References

- 1 S. Boiteux and M. Guillet, *DNA Repair (Amst.)*, 2004, **3**, 1–12.
- 2 T. Lindahl and B. Nyberg, *Biochemistry*, 1972, **11**, 3610–3618.
- 3 T. Lindahl, *Nature*, 1993, **26**, 709–715.
- 4 B. Demple and L. Harrison, *Annu. Rev. Biochem.*, 1994, **63**, 915–948.
- 5 V. Esposito, L. Martino, G. Citarella, A. Virgilio, L. Mayol, C. Giancola and A. Galeone, *Nucleic Acids Res.*, 2010, **38**, 2069–2080.
- 6 P. Školáková, K. Bednářová, M. Vorlíčková and J. Sagi, *Biochem. Biophys. Res. Commun.*, 2010, **399**, 203–208.
- 7 A. Virgilio, L. Petraccone, V. Esposito, G. Citarella, C. Giancola and A. Galeone, *Biochim. Biophys. Acta - Gen. Subj.*, 2012, **1820**, 2037–2043.
- 8 J. Beckett, J. Burns, C. Broxson and S. Tornaletti, *Biochemistry*, 2012, **51**, 5257–5268.
- 9 J. Malina and V. Brabec, *Chem. - A Eur. J.*, 2014, **20**, 7566–7570.
- 10 M. M. Greenberg, *Acc. Chem. Res.*, 2014, **47**, 646–655.
- 11 H. E. Krokan, R. Standal and G. Slupphaug, *Biochem. J.*, 1997, **325**, 1–16.
- 12 O. D. Schärer and J. Jiricny, *BioEssays*, 2001, **23**, 270–281.
- 13 G. Barzilay and I. D. Hickson, *BioEssays*, 1995, **17**, 713–719.
- 14 H. Gattuso, E. Durand, E. Bignon, C. Morell, A. G. Georgakilas, E. Dumont, C. Chipot, F. Dehez and A. Monari, *J. Phys. Chem. Lett.*, 2016, **7**, 3760–3765.
- 15 D. Fang, R. L. Lord and G. A. Cisneros, *J. Phys. Chem. B*, 2013, **117**, 6410–6420.
- 16 D. Fang and G. A. Cisneros, *J. Chem. Theory Comput.*, 2014, **10**, 5136–5148.
- 17 B. Wang, D. Usharani, C. Li and S. Shaik, *J. Am. Chem. Soc.*, 2014, **136**, 13895–13901.
- 18 M. G. Quesne, R. Latifi, L. E. Gonzalez-Ovalle, D. Kumar and S. P. de Visser, *Chem. - A Eur. J.*, 2014, **20**, 435–446.
- 19 J. K. Ahn, K. S. Park, B. Y. Won and H. G. Park, *Biosens. Bioelectron.*, 2015, **67**, 590–594.
- 20 Y. Pang, Z. Xu, Y. Sato, S. Nishizawa and N. Teramae, *ChemBioChem*, 2012, **13**, 436–442.
- 21 M. a. McWilliams, F. H. Anka, K. J. Balkus and J. D. Slinker, *Biosens. Bioelectron.*, 2014, **54**, 541–546.
- 22 A. K. Boal, E. Yavin, O. A. Lukianova, V. L. O'Shea, S. S. David and J. K. Barton, *Biochemistry*, 2005, **44**, 8397–8407.
- 23 M. C. Buzzeo and J. K. Barton, *Bioconjug. Chem.*, 2008, **19**, 2110–2112.
- 24 M. a. McWilliams, R. Bhui, D. W. Taylor and J. D. Slinker, *J. Am. Chem. Soc.*, 2015, **137**, 11150–11155.
- 25 P. Cuniasse, G. V. Fazakerley, W. Guschlbauer, B. E. Kaplan and L. C. Sowers, *J. Mol. Biol.*, 1990, **213**, 303–314.
- 26 C. A. Gelfand, G. E. Plum, A. P. Grollman, F. Johnson and K. J. Breslauer, *Biochemistry*, 1998, **37**, 7321–7327.
- 27 L. Ayadi, C. Coulombeau and R. Lavery, *Biophys. J.*, 1999, **77**, 3218–3226.
- 28 D. Barsky, N. Foloppe, S. Ahmadi, D. M. Wilson III and A. D. MacKerell, *Nucleic Acids Res.*, 2000, **28**, 2613–2626.
- 29 M. P. Singh, G. C. Hill, D. Peoc'h, B. Rayner, J.-L. Imbach and J. W. Lown, *Biochemistry*, 1994, **33**, 10271–10285.
- 30 I. Goljer, S. Kumar and P. H. Bolton, *J. Biol. Chem.*, 1995, **270**, 22980–22987.
- 31 Y. Coppel, N. Berthet, C. Coulombeau, C. Coulombeau, J. Garcia and J. Lhomme, *Biochemistry*, 1997, **36**, 4817–4830.
- 32 K. Y. Wang, S. A. Parker, I. Goljer and P. H. Bolton, *Biochemistry*, 1997, **36**, 11629–11639.
- 33 R. D. Beger and P. H. Bolton, *J. Biol. Chem.*, 1998, **273**, 15565–15573.
- 34 L. Ayadi, D. Forget, A. Martelli, J.-F. Constant, M. Demeunynck and C. Coulombeau, *Theor. Chem. Acc.*, 2000, **104**, 284–289.
- 35 S. T. Hoehn, *Nucleic Acids Res.*, 2001, **29**, 3413–3423.
- 36 J. Chen, F.-Y. Dupradeau, D. A. Case, C. J. Turner and J. Stubbe, *Nucleic Acids Res.*, 2008, **36**, 253–262.
- 37 J. Zálešák, M. Lourdin, L. Krejčí, J.-F. Constant and M. Jourdan, *J. Mol. Biol.*, 2014, **426**, 1524–1538.
- 38 R. Y. Au, K. S. Ng, L. M. Chi and S. L. Lam, *J. Phys. Chem. B*, 2012, **116**, 14781–14787.
- 39 D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Götz, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, R.M. Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman (2012), *AMBER 12*, University of California, San Francisco.
- 40 D. E. Smith and L. X. Dang, *J. Chem. Phys.*, 1994, **100**, 3757–3766.
- 41 A. Pérez, I. Marchán, D. Svozil, J. Sponer, T. E. Cheatham, C. A. Lughton and M. Orozco, *Biophys. J.*, 2007, **92**, 3817–3829.
- 42 I. Ivani, P. D. Dans, A. Noy, A. Pérez, I. Faustino, A. Hospital, J. Walther, P. Andrio, R. Goñi, A. Balaceanu, G. Portella, F. Battistini, J. L. Gelpí, C. González, M. Vendruscolo, C. A. Lughton, S. A. Harris, D. A. Case and M. Orozco, *Nat. Methods*, 2016, **13**, 55–8.
- 43 A. A. Voityuk, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2013, **3**, 515–527.
- 44 A. A. Voityuk, *Chem. Phys. Lett.*, 2006, **427**, 177–180.
- 45 A. A. Voityuk, *J. Chem. Phys.*, 2008, **128**, 115101.
- 46 L. Blancafort and A. A. Voityuk, *J. Phys. Chem. A*, 2006, **110**, 6426–6432.
- 47 E. Vladimirov, A. Ivanova and N. Rösch, *J. Phys. Chem. B*, 2009, **113**, 4425–4434.
- 48 C. Curutchet and A. A. Voityuk, *J. Phys. Chem. C*, 2012, **116**, 22179–22185.
- 49 A. A. Voityuk and N. Rösch, *J. Chem. Phys.*, 2002, **117**, 5607–5616.
- 50 N. Rösch and A. A. Voityuk, in *Top. Curr. Chem.*, 2004, vol. 237, pp. 37–72.
- 51 K. Siri Wong and A. A. Voityuk, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2012, **2**, 780–794.
- 52 W. B. Davis, S. Hess, I. Naydenova, R. Haselsberger, A.

- Ogrodnik, M. D. Newton and M.-E. Michel-Beyerle, *J. Am. Chem. Soc.*, 2002, **124**, 2422–2423.
- 53 D. N. LeBard, M. Lilichenko, D. V. Matyushov, Y. A. Berlin and M. A. Ratner, *J. Phys. Chem. B*, 2003, **107**, 14509–14520.
- 54 K. Siriwong, A. A. Voityuk, M. D. Newton and N. Rösch, *J. Phys. Chem. B*, 2003, **107**, 2595–2601.
- 55 T. Kubař and M. Elstner, *J. Phys. Chem. B*, 2009, **113**, 5653–5656.
- 56 J. Singh, E. R. Bittner, D. Beljonne and G. D. Scholes, *J. Chem. Phys.*, 2009, **131**, 194905.
- 57 M. Corbella, A. a Voityuk and C. Curutchet, *J. Phys. Chem. Lett.*, 2015, **6**, 3749–3753.
- 58 H.-A. Wagenknecht, S. R. Rajsiki, M. Pascaly, E. D. A. Stemp and J. K. Barton, *J. Am. Chem. Soc.*, 2001, **123**, 4400–4407.