

# Solvent Fluctuations Drive the Hole Transfer in DNA: a Mixed Quantum-Classical Study

Tomáš Kubař, Ulrich Kleinekathöfer and Marcus Elstner

## Supporting Information

### A. Selected data presented as graphics – in tabular format

| # of A's | survival probability | rate constant (ns <sup>-1</sup> ) |
|----------|----------------------|-----------------------------------|
| 1        | 0.029 ± 0.079        | 361 ± 184                         |
| 2        | 0.109 ± 0.195        | 237 ± 154                         |
| 3        | 0.179 ± 0.350        | 154 ± 99                          |
| 4        | 0.231 ± 0.263        | 119 ± 85                          |
| 5        | 0.250 ± 0.252        | 102 ± 68                          |
| 7        | 0.316 ± 0.252        | 75 ± 45                           |
| 10       | 0.410 ± 0.213        | 52 ± 29                           |
| 14       | 0.602 ± 0.212        | 29 ± 19                           |

TABLE I: Survival probability after 20 ps and rate constant of hole transfer in GT<sub>n</sub>GCG. *Data for Figure 4.*

| # of A's | fully MD-based | constant $T_{ij}$ | constant $\varepsilon_i$ |
|----------|----------------|-------------------|--------------------------|
| 1        | 361 ± 184      | 460 ± 191         | 4.1 ± 4.0                |
| 2        | 237 ± 154      | 309 ± 187         | 0.18 ± 0.25              |
| 3        | 154 ± 99       | 203 ± 116         | 0.027 ± 0.083            |
| 4        | 119 ± 85       | 159 ± 85          |                          |
| 5        | 102 ± 68       | 117 ± 62          |                          |
| 7        | 75 ± 45        | 98 ± 53           |                          |
| 10       | 52 ± 29        | 76 ± 36           |                          |
| 14       | 29 ± 19        | 46 ± 26           |                          |

TABLE II: Rate constant of hole transfer (ns<sup>-1</sup>) in GT<sub>n</sub>GCG with the restricted models. *Data for Figure 5.*

| # of A's | full QM/MM sim. | 'in vacuo' | statistical sim. |
|----------|-----------------|------------|------------------|
| 1        | 361 ± 184       | 99 ± 78    | 248 ± 77         |
| 2        | 237 ± 154       | 85 ± 55    | 162 ± 46         |
| 3        | 154 ± 99        | 76 ± 49    | 128 ± 33         |
| 4        | 119 ± 85        | 69 ± 51    | 90 ± 25          |
| 5        | 102 ± 68        | 55 ± 38    | 70 ± 25          |
| 6        | 75 ± 45         | 43 ± 36    | 45 ± 15          |
| 10       | 52 ± 29         | 31 ± 22    | 26 ± 10          |
| 14       | 29 ± 19         | 19 ± 14    | 12 ± 5           |

TABLE III: Rate constant of hole transfer (ns<sup>-1</sup>) in GT<sub>n</sub>GCG – the effect of environment and the influence of correlation of CT parameters. *Data for Figures 8 and 10, respectively.*

## B. Simulation details

### Starting structure

- Two GC base pairs were added to both 5'- and 3'-end of each studied DNA sequence (for instance, GTGGG yielded GGGTGGGGG). This was to guarantee the stability of the DNA species, which might have been brought in danger otherwise (so-called end effects). These additional end-base pairs were removed from consideration after the MD simulation was done.
- The starting structure of each DNA species studied was a B-DNA double helix created by using the MAKE-NA server (<http://structure.usc.edu/make-na/server.html>).
- A rectangular periodic box was then set up with the minimum distance of the box walls from the molecule of 10 Å. The box was filled with TIP3P water so that the density of  $1 \text{ g}\times\text{cm}^{-3}$  was reached; the corresponding number of water molecules was between 2566 and 4778, depending on the DNA species.
- An appropriate number (16 to 42) of water molecules was substituted by sodium counterions to achieve electroneutrality of the entire system.

### Equilibration

- Steepest-descents minimization – 100 steps.
- MD at constant volume – 2 fs time step, total 10,000 steps. initial velocities generated at 10 K, then coupling to thermal bath by means of Berendsen thermostat (reference temperatures – 10 K for DNA and 300 K for solvent, relaxation time 0.2 ps). bonds involving hydrogen atoms constrained to the equilibrium lengths using SHAKE.
- MD at constant pressure – 2 fs time step, total 500,000 steps. coupling to thermal bath by means of Nosé-Hoover thermostat (ref. temp. 300 K, time constant 0.5 ps), pressure maintained by means of Parrinello-Rahman barostat (ref. pressure 1 bar, time constant 0.5 ps). bonds involving hydrogen atoms constrained to the equilibrium lengths using LINCS.

### Production

- The same setup as in the last equilibration step.

### C. Discussion of the trapping constant $\tau$

The rate with which the hole disappears from the system obviously depends on the trapping rate constant  $\tau$ , or rather its order of magnitude. A rather large value of  $\tau = 1000 \text{ ns}^{-1}$  was used in our simulations, to obtain detectable annihilation of the hole even with the longest adenine bridges, on this short time scale (20 ps) accessible computationally. Giese (Top. Cur. Chem. **2004**, 236, 27) reports an estimate  $\tau \approx 0.1 \text{ ns}^{-1}$  at pH 5 and mentions that the reaction is faster at pH 7, due to increased nucleophilic character of water. The influence of  $\tau = 1 \dots 1000 \text{ ns}^{-1}$  on annihilation rate is illustrated in Fig. 1. However, we are interested merely in the behavior of adenine bridges as an obstruction to CT, or in their ‘permeability’ for the hole, for the time being. Indeed, the application of slower (and possibly more physical) trapping rates will merely slow down the flow of hole out of the adenine bridge, but it will probably not affect the mentioned permeability whatsoever.

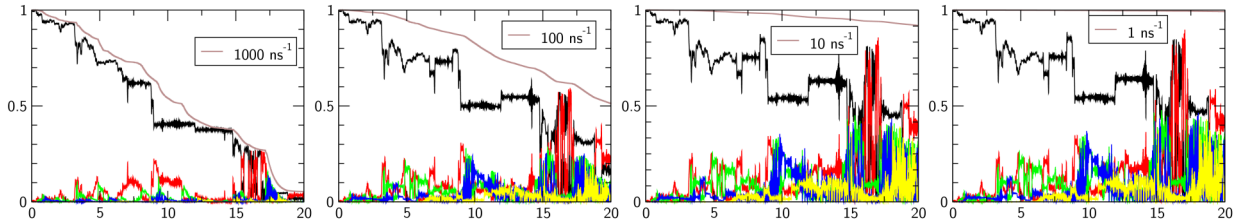


FIG. 1: The effect of trapping constant  $\tau$  on the survival of hole, in a 20ps simulation. Plots labeled by the value of  $\tau$ . (Brown – survival of hole; other colors – occupation of individual nucleobases.)

### D. Discussion of convergence of charge-transfer rates

| number of A's | $4 \times 25$ |     |     |     | $2 \times 50$ |     | $1 \times 100$ |
|---------------|---------------|-----|-----|-----|---------------|-----|----------------|
| 1             | 383           | 449 | 319 | 302 | 416           | 311 | 361            |
| 2             | 246           | 219 | 244 | 240 | 232           | 242 | 237            |
| 3             | 211           | 140 | 151 | 167 | 176           | 134 | 154            |
| 4             | 95            | 139 | 107 | 136 | 117           | 121 | 119            |
| 5             | 99            | 86  | 118 | 110 | 92            | 114 | 102            |
| 7             | 69            | 66  | 77  | 85  | 67            | 81  | 75             |
| 10            | 58            | 52  | 44  | 55  | 55            | 49  | 52             |
| 14            | 32            | 26  | 29  | 28  | 29            | 29  | 29             |

TABLE IV: Rate of hole transfer ( $\text{ns}^{-1}$ ) averaged over varied number of individual 20ps simulations. Used were 4 groups à 25 simulations and 2 groups à 50 simulations, for all bridge lengths considered, in comparison with the average from all 100 simulations ( $1 \times 100$ ).

### E. Survival of charge in GTGGG if either site energies or electronic couplings are taken constant

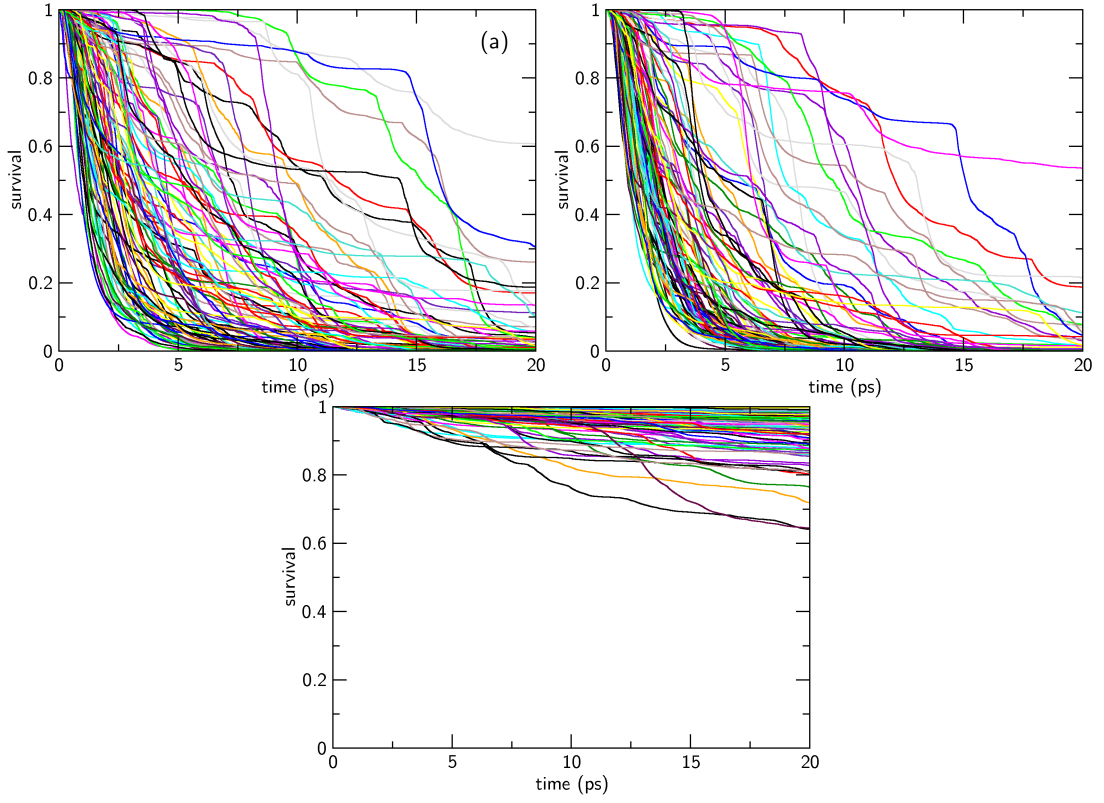


FIG. 2: Survival of hole in GTGGG. Data from dynamical simulation (left), on-site energies from simulation while electronic couplings constant (right), and electronic couplings from simulation while on-site energies constant (bottom).

### F. Survival of charge in GTGGG without the QM/MM treatment ('in vacuo')

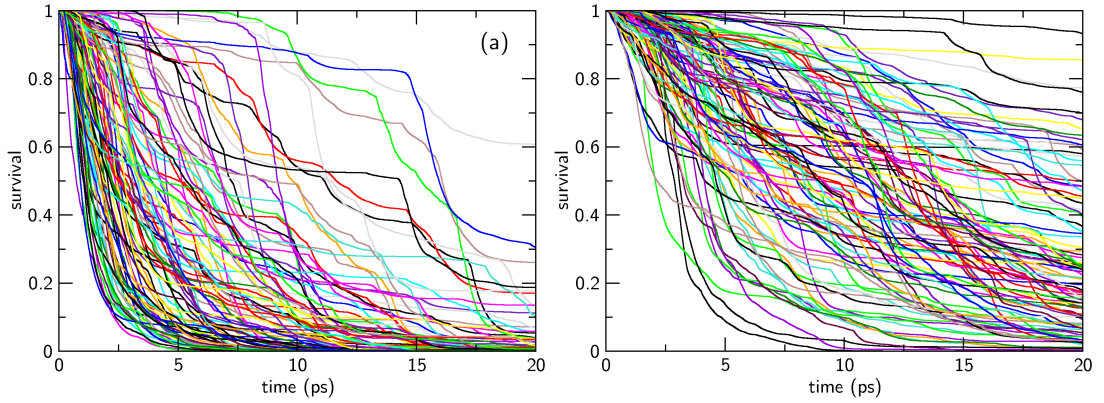


FIG. 3: Survival of hole in GTGGG. CT parameters calculated with the QM/MM treatment of environment (left) and without that ('in vacuo', right).





## H. Distributions of CT parameters along MD trajectories

The distribution of site energies was normal, cf. the table on this page for mean values and standard deviations. The electronic couplings were distributed normally around zero because positive and negative values of the same magnitude are obtained with equal probabilities;\* correct distributions are then obtained as positive branches of normal distribution centered in zero with the standard deviations presented in the table.

| GTGGG               |                 |          |             | GT <sub>5</sub> GGG  |                 |          |             | GT <sub>7</sub> GGG  |                 |          |             |
|---------------------|-----------------|----------|-------------|----------------------|-----------------|----------|-------------|----------------------|-----------------|----------|-------------|
| eV                  | $\varepsilon_i$ |          | $T_{i,i+1}$ | eV                   | $\varepsilon_i$ |          | $T_{i,i+1}$ | eV                   | $\varepsilon_i$ |          | $T_{i,i+1}$ |
| base                | mean            | st. dev. | st. dev.    | base                 | mean            | st. dev. | st. dev.    | base                 | mean            | st. dev. | st. dev.    |
| G                   | 5.649           | 0.414    | 0.0153      | G                    | 5.842           | 0.463    | 0.0242      | G                    | 5.962           | 0.465    | 0.0219      |
| A                   | 5.997           | 0.365    | 0.0200      | A                    | 6.492           | 0.381    | 0.0728      | A                    | 6.541           | 0.430    | 0.0659      |
| G                   | 5.861           | 0.375    | 0.0465      | A                    | 6.705           | 0.375    | 0.0777      | A                    | 6.764           | 0.414    | 0.0735      |
| G                   | 5.847           | 0.389    | 0.0358      | A                    | 6.770           | 0.384    | 0.0789      | A                    | 6.879           | 0.405    | 0.0723      |
| G                   | 5.652           | 0.422    |             | A                    | 6.729           | 0.397    | 0.0805      | A                    | 6.941           | 0.389    | 0.0749      |
| GT <sub>2</sub> GGG |                 |          |             | A                    | 6.577           | 0.403    | 0.0226      | A                    | 6.988           | 0.381    | 0.0763      |
| eV                  | $\varepsilon_i$ |          | $T_{i,i+1}$ | G                    | 6.226           | 0.446    | 0.0567      | A                    | 6.960           | 0.389    | 0.0834      |
| base                | mean            | st. dev. | st. dev.    | G                    | 6.163           | 0.468    | 0.0403      | A                    | 6.813           | 0.400    | 0.0240      |
| G                   | 5.649           | 0.427    | 0.0199      | G                    | 5.924           | 0.493    |             | G                    | 6.435           | 0.438    | 0.0477      |
| A                   | 6.177           | 0.400    | 0.0777      | GT <sub>14</sub> GGG |                 |          |             | G                    | 6.370           | 0.465    | 0.0411      |
| A                   | 6.247           | 0.365    | 0.0225      | eV                   | $\varepsilon_i$ |          | $T_{i,i+1}$ | G                    | 6.109           | 0.498    |             |
| G                   | 5.970           | 0.411    | 0.0444      | base                 | mean            | st. dev. | st. dev.    | GT <sub>10</sub> GGG |                 |          |             |
| G                   | 5.989           | 0.414    | 0.0385      | G                    | 6.264           | 0.503    | 0.0200      | eV                   | $\varepsilon_i$ |          | $T_{i,i+1}$ |
| G                   | 5.798           | 0.430    |             | A                    | 6.871           | 0.454    | 0.0692      | base                 | mean            | st. dev. | st. dev.    |
| GT <sub>3</sub> GGG |                 |          |             | A                    | 7.105           | 0.444    | 0.0690      | G                    | 6.201           | 0.525    | 0.0241      |
| eV                  | $\varepsilon_i$ |          | $T_{i,i+1}$ | A                    | 7.224           | 0.438    | 0.0667      | A                    | 6.748           | 0.444    | 0.0669      |
| base                | mean            | st. dev. | st. dev.    | A                    | 7.309           | 0.424    | 0.0695      | A                    | 6.974           | 0.422    | 0.0718      |
| G                   | 5.739           | 0.430    | 0.0230      | A                    | 7.360           | 0.395    | 0.0697      | A                    | 7.077           | 0.414    | 0.0713      |
| A                   | 6.356           | 0.384    | 0.0864      | A                    | 7.407           | 0.370    | 0.0732      | A                    | 7.132           | 0.408    | 0.0678      |
| A                   | 6.503           | 0.381    | 0.0885      | A                    | 7.431           | 0.351    | 0.0671      | A                    | 7.164           | 0.386    | 0.0697      |
| A                   | 6.427           | 0.370    | 0.0231      | A                    | 7.442           | 0.346    | 0.0689      | A                    | 7.186           | 0.370    | 0.0726      |
| G                   | 6.100           | 0.414    | 0.0476      | A                    | 7.445           | 0.359    | 0.0696      | A                    | 7.197           | 0.362    | 0.0772      |
| G                   | 6.081           | 0.416    | 0.0358      | A                    | 7.442           | 0.378    | 0.0757      | A                    | 7.202           | 0.356    | 0.0810      |
| G                   | 5.864           | 0.430    |             | A                    | 7.434           | 0.397    | 0.0773      | A                    | 7.143           | 0.370    | 0.0796      |
| GT <sub>4</sub> GGG |                 |          |             | A                    | 7.417           | 0.405    | 0.0879      | A                    | 6.963           | 0.395    | 0.0252      |
| eV                  | $\varepsilon_i$ |          | $T_{i,i+1}$ | A                    | 7.344           | 0.411    | 0.0815      | G                    | 6.555           | 0.446    | 0.0509      |
| base                | mean            | st. dev. | st. dev.    | A                    | 7.162           | 0.408    | 0.0275      | G                    | 6.476           | 0.482    | 0.0394      |
| G                   | 5.869           | 0.430    | 0.0210      | G                    | 6.756           | 0.479    | 0.0461      | G                    | 6.196           | 0.525    |             |
| A                   | 6.356           | 0.422    | 0.0702      | G                    | 6.642           | 0.482    | 0.0366      |                      |                 |          |             |
| A                   | 6.560           | 0.414    | 0.0791      | G                    | 6.326           | 0.517    |             |                      |                 |          |             |
| A                   | 6.596           | 0.395    | 0.0838      |                      |                 |          |             |                      |                 |          |             |
| A                   | 6.495           | 0.375    | 0.0240      |                      |                 |          |             |                      |                 |          |             |
| G                   | 6.111           | 0.444    | 0.0491      |                      |                 |          |             |                      |                 |          |             |
| G                   | 6.079           | 0.441    | 0.0384      |                      |                 |          |             |                      |                 |          |             |
| G                   | 5.847           | 0.460    |             |                      |                 |          |             |                      |                 |          |             |

\* The ‘A|A’ base-pair steps are a slight exclusion to this observation, with a distribution centered around a non-zero value, giving rise to a more complicated distribution:  $P(x) \propto \exp(a(x - x_0)^2) + \exp(a(-x - x_0)^2)$ . For sake of simplicity and because the deviation is quite small, we considered normal distributions centered around zero in these cases as well.