

# Nonadiabatic QM/MM Simulations of Fast Charge Transfer in *Escherichia coli* DNA Photolyase

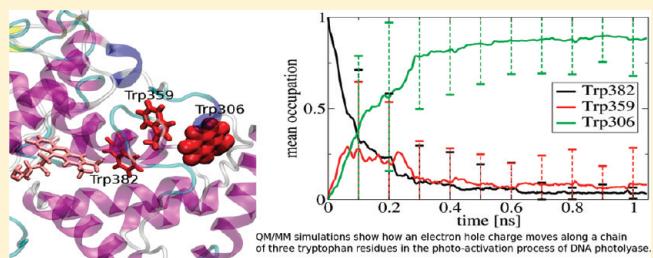
Paul Benjamin Woiczikowski, Thomas Steinbrecher, Tomáš Kubář, and Marcus Elstner\*

Department for Theoretical Chemical Biology, Institute for Physical Chemistry, Karlsruhe Institute of Technology, Kaiserstrasse 12, D-76131 Karlsruhe, Germany

Supporting Information

Web-Enhanced

**ABSTRACT:** In this report, we study the photoactivation process in *Escherichia coli* DNA photolyase, involving long-range electron transport along a conserved chain of Trp residues between the protein surface and the flavin adenine dinucleotide (FAD) cofactor. Fully coupled nonadiabatic (Ehrenfest) quantum mechanics/molecular mechanics (QM/MM) simulations allow us to follow the time evolution of charge distributions over the natural time scale of multiple charge transfer events and conduct rigorous statistical analysis. Charge transfer rates in excellent agreement with experimental data are obtained without the need for any system-specific parametrization. The simulations are shown to provide a more detailed picture of electron transfer than a classical analysis of Marcus parameters. The protein and solvent both strongly influence the localization and transport properties of a positive charge, but the directionality of the process is mainly caused by solvent polarization. The time scales of charge movement, delocalization, protein relaxation and solvent reorganization overlap and lead to nonequilibrium reaction conditions. All these contributions are explicitly considered and fully resolved in the model used and provide an intricate picture of multistep biochemical electron transfer in a flexible, heterogeneous environment.

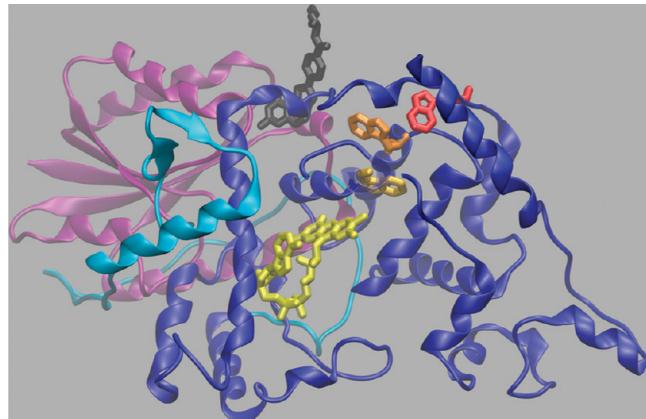


QM/MM simulations show how an electron hole charge moves along a chain of three tryptophan residues in the photo-activation process of DNA photolyase.

## INTRODUCTION

Photolyases (PLs), or photoreactivating DNA repair enzymes, are members of the PL/cryptochrome superfamily of blue-light driven flavoproteins. They utilize near-UV light to repair DNA bases damaged by radiation. Their commonly accepted reaction mechanism involves recognition and binding of a damaged DNA strand containing dimerized bases, “flipping out” of the base dimer to bring it into contact with the flavin adenine dinucleotide (FAD) redox cofactor, absorption of light, and a cyclic electron transfer (ET) reaction resulting in the recovery of the single bases.<sup>1–3</sup> PLs contain a nonessential second chromophor, e.g., methenyltetrahydrofolate (MTHF) in the case of *Escherichia coli* PL, which functions as a light-harvesting antenna and increases the efficiency of repair under limiting light conditions. Depending on the type of UV-damage they repair, PLs are further subdivided into CPD (cyclobutane pyrimidine dimer) and 6–4 (pyrimidine-[6–4]-pyrimidone) types.

The function of PL was elucidated more than 50 years ago,<sup>6</sup> in one of the first examples of *in vivo* enzymology<sup>7</sup> and PLs, and the homologous cryptochromes have become the subject of intense study over the past decade.<sup>8–12</sup> The members of the enzyme family are widely, if unevenly, distributed in organisms. PLs are found in many microbes and some nonmammalian animals, while cryptochromes are found in animals (including humans), plants and bacteria.<sup>13</sup> Much experimental effort has been put into studying the mechanism of PLs,<sup>14</sup> which is of bisubstrate ordered sequential type with the unique twist of using a near-UV photon



**Figure 1.** Cartoon representation of *E. coli* DNA PL from the crystal structure by Park et al.<sup>4</sup> The FAD cofactor (yellow), MTHF antenna molecule (black), and the three conserved Trp residues 382, 359, and 306 (gold, orange, red, respectively) participating in the photoactivation process are displayed. The two-domain structure of the enzyme, with an N-terminal  $\alpha/\beta$ -domain (purple) and an FAD binding  $\alpha$ -helical domain (blue) connected by a long interdomain loop (cyan) is indicated. Figure made using VMD<sup>5</sup> and Povray-3.6.

**Received:** May 20, 2011

**Revised:** July 7, 2011

**Published:** July 27, 2011

as the required second substrate. In the course of these studies, a second mechanism of action, called photoactivation, was found in which an electron is transferred along a chain of evolutionary conserved tryptophan residues to the partially oxidized flavin cofactor. The physiological significance of this second mechanism is unclear for PLs and has been suggested for cryptochromes.<sup>15,16</sup>

We have selected the photoactivation process in PL as the electron-transfer (ET) reaction to study here. PL photoactivation is a complex, multistep charge transfer (CT) process coupled to the very heterogeneous environment of a solvated protein; it is therefore on the limits of processes that can be characterized by classical CT theory, but well within the scope of modern molecular simulation methods. As both the relevant times and transfer distances are on the nanoscale,<sup>17–19</sup> coupled quantum-mechanical and molecular dynamics (MD) calculations can be employed to gain detailed insight into the relevant physical parameters and degrees of freedom. Photoactivation, in contrast to photoreactivation (the DNA repair process) which requires a DNA–enzyme complex model, has the additional benefit of occurring within a stable macromolecule of known structure, thereby reducing the sampling requirements.

Several previous theoretical studies, both classical and quantum mechanical, dealt with PL (reviewed in ref 20; more recent works include refs 21–25). Modeling studies from the Stuchebrukhov group dealt with the photoreactivation process for PL (both from *E. coli* and *Aspergillus nidulans*).<sup>26–29</sup> Of particular interest to the study here, Cheung et al. have studied possible tunneling pathways in PL at atomic resolution using an extended Hückel methodology.<sup>30</sup> Our choice of quantum mechanics (QM) system (see below) is based on data presented therein. Combined classical/QM simulations were reported before,<sup>31,32</sup> but here we present the first account of a direct, fully coupled QM/molecular mechanics (MM) simulation of the complete CT process of photoactivation, in order to address several remaining unsolved questions.

The foremost question regards the exact reaction mechanism of photoactivation. Hopping transport is assumed to be more efficient than direct superexchange between FAD and Trp306, but little evidence of transient oxidation at Trp359 and Trp382 exists.<sup>10</sup> Furthermore, the study of ET between sites with vastly different environments provides an interesting contrast to the widely studied CT in DNA, where charge carrying nucleobases have nearly identical surroundings. Finally, the question of how long-distance ET can occur so fast (<100 ps) in this protein ties the system into the ongoing wider discussion of ET in complex environments and the role of protein relaxation versus nonequilibrium ET.

CT processes have been among the most intensely studied phenomena in biophysics for a long time.<sup>33–37</sup> The recent focus on ET in nanostructures and the emerging field of molecular electronics have renewed scientific interest in the underlying fundamental reaction steps and the models used to describe them. Molecular simulations, with their ability to explore free energy landscapes of arbitrary shape, have proven themselves to be valuable tools in this area, and many contributions in the past years have used them to study CT in unusual and intractable environments.

The theoretical foundation of CT remains Marcus' seminal theory<sup>38–41</sup> and its extensions.<sup>42–46</sup> Within the assumption of weak electronic coupling between the donor and the acceptor, the rate of CT can be obtained in the high-temperature nonadiabatic

limit as

$$k_{\text{DA}} = \frac{T_{ij}^2}{\hbar} \sqrt{\frac{\pi}{\lambda k_{\text{B}} T}} \exp\left(-\frac{(\Delta G + \lambda)^2}{4\lambda k_{\text{B}} T}\right) \quad (1)$$

$T_{ij}$  is the electronic coupling between the donor and the acceptor state,  $\Delta G$  is the free energy difference between the initial and the final state, and  $\lambda$  is the reorganization energy, i.e., the energy required to change the molecular structure of the initial state to that corresponding to the relaxed product state without CT taking place.  $\lambda$  is usually decomposed into two parts  $\lambda = \lambda_s + \lambda_i$ : the outer-space reorganization energy  $\lambda_s$  represents the rearrangement of solvent and protein upon the redistribution of charge, and the inner-space reorganization energy  $\lambda_i$  describes the change of the molecular structure of the donor and the acceptor themselves.

$\lambda$  and  $\Delta G$  determine the activation free energy, and therefore a significant amount of work has concentrated on the question of how to compute these values using MD simulations based on classical force fields.<sup>47–53</sup> In most simulations, linear response theory has been applied, which allows the computation of  $\lambda$  from the average value of the donor–acceptor energy gap (for recent reviews, see, e.g., refs 54 and 55). Alternatively, it can be obtained using free energy perturbation (FEP) approaches.<sup>56,57</sup> While standard force fields use a fixed charge distribution, it has been shown repeatedly that electronic polarizability has a distinct impact on the computed quantities.<sup>49,55,58–60</sup> Particularly,  $\lambda$  is affected, while  $\Delta G$  seems to be quite insensitive to electronic polarization.<sup>55</sup> Therefore, estimates of  $\lambda$  based on fixed charge models have been proposed to be scaled by the high-frequency dielectric constant,<sup>53</sup> which can be rationalized by considering the Pekar factor of the Marcus continuum formula.<sup>55</sup> The studies so far have shown that  $\lambda$  is overestimated by 30 to 50% when computed with a fixed charge model. The scaling of values obtained from a fixed-charge model is often preferred over performing MD simulations using a fully polarizable force field for reasons of computational efficiency.

MM polarization has two effects, it reduces both the energy gap and the fluctuations.<sup>61</sup> For certain donors and acceptors, the proper account of charge distribution seems to be important besides MM polarization.<sup>62</sup> Therefore, there is an increasing amount of QM<sup>54</sup> and QM/MM calculations on this subject.<sup>63–66</sup> Note, however, that most ab initio or density functional theory (DFT) approaches are limited in system size quite severely. As Warshel and co-workers pointed out, computational methods have to balance the accuracy given by high-level QM methods with the need for extensive sampling.<sup>64</sup> Therefore, more efficient QM methods are highly desirable.

This is particularly the case when the electronic couplings  $T_{ij}$  have to be evaluated for extended systems, involving large donor–bridge–acceptor complexes. The calculation of  $T_{ij}$  for small systems is already quite an involved task<sup>67,68</sup> and the calculation for extended systems can only be accomplished using more approximate quantum methods such as the tunneling currents approach developed by the Stuchebrukhov group.<sup>69–72</sup> A large amount of work has focused on the role of structural fluctuations, analyzing  $T_{ij}$  over trajectories from classical MD simulations. While fluctuations are averaged out in the thermodynamical calculations for  $\lambda$  and  $\Delta G$  (parameters describing the accessibility of the transition state), they have been shown to play a vital role in understanding CT in complex systems such as DNA and proteins.<sup>73</sup> From this perspective, the relative time scales of the

fluctuations and the CT become crucial and characterize different mechanisms (such as gated CT). For recent reviews of this extensive field, see, e.g., refs 74 and 75.

The time scale of CT in a polar environment is mainly determined by the solvent reorganization energy and the driving force. These quantities can be computed using FEP-type calculations that assume the relaxation of the solvent to be faster than the CT reaction. Interestingly, such a case has been described recently for the extremely fast initial ET step in the bacterial reaction center,<sup>76</sup> illustrating the limits of conventional ET theory.<sup>77</sup> Such a scenario has been argued already by Warshel and co-workers,<sup>50</sup> questioning the assumption of equilibrium in the conventional calculation of  $\lambda_s$ . Instead, the semiclassical trajectory approach, developed by Warshel and co-workers over the past decades,<sup>47,48,78,79</sup> seems to be more appropriate. Such studies allow the propagation of electronic and nuclear degrees of freedom simultaneously and do not rely on any assumption of equilibration or relevant time scales.

In recent years, we have developed a methodology that allows us to compute CT parameters along classical MD simulations by treating the relevant electronic degrees of freedom (hole or electron) using the time-dependent Schrödinger (Kohn–Sham (KS)) equations. This leads to a direct simulation of the propagation of the electronic wave function, coupled to a realistic description of the protein and solvent dynamics via a combined QM/MM approach.<sup>80</sup> Note that similar direct approaches have been successfully applied to study CT in DNA recently.<sup>81–83</sup> In this work, we combine classical MD simulations on the nanosecond time scale with a very efficient, approximate DFT QM method to study electron rearrangement on its natural time scale along a chain of tryptophan residues embedded in an atomistic, fully flexible solvated protein environment.

This report is organized as follows: We first present results of extended classical MD simulation of PL in different charge states and show that conformational changes with significant effect on the CT efficiency occur on the nanosecond time scale. The effect of the dynamic distribution of solvent around the different stepping stone amino acids is also considered. Then, building on the large conformational ensemble generated by the simulations, we turn to computing the CT parameters quantum-mechanically and analyzing transfer coherence parameters and coupling and reorganization energy distributions. Then, the main achievement of this study, electron–ion dynamics (EID) charge distribution trajectories from Ehrenfest-QM/MM-dynamics simulations will be presented. Previously unavailable data on the kinetics and CT sequence can be derived from these simulations, as well as significantly improved estimates on CT rates. A final comparison of CT kinetics and thermodynamics parameters, obtained with different simulation methods, is provided in the Discussion, along with a generalization of our results for the wider field of modeling ET in nanostructures.

## METHODOLOGY

**Protein Structure and Simulation Set-Up.** The protein structure of *E. coli* DNA PL used in this study is based on the X-ray crystal structure by Park et al.,<sup>4</sup> which was used in previous studies as well.<sup>20,31</sup> The redox cofactor FAD is decomposed into riboflavin and adenosine-5'-diphosphate (ADP) parts. For the latter, force field parameters from ref 84 are used, whereas the riboflavin and the light harvesting cofactor 5,10-methylentetrahydrofolate (MTHF) are parametrized

**Table 1. Simulation Labels<sup>a</sup>**

simulation	q(FAD)	q(1)	q(2)	q(3)
A	0	0	0	0
B	-1	0	0	0
C	-1	+1	0	0
D	-1	0	+1	0
E	-1	0	0	+1

<sup>a</sup> Tryptophan residues Trp382, Trp359, and Trp306 are denoted as 1, 2, and 3, respectively. Total fragment charges  $q$  for the Trp residues and the riboflavin part of FAD are listed for the various simulations in units of  $e$ .

according to the gaff forcefield.<sup>85</sup> Atomic charges for both cofactors (including both neutral and negatively charged forms of FAD) are determined from RESP calculations.<sup>86</sup> Additionally, charges for a Trp side-chain radical cation species need to be obtained. To maintain consistency with existing amino acid partial charge parametrization, these were determined by adding atomic charge differences to the respective atoms of a neutral Trp residue. The charge differences are obtained from RESP calculations for both the neutral and positive charged forms of the model molecule 3-methylindole (skatole). Atom labels and corresponding atomic charges of neutral and charged species for the Trp residue can be found in Figure S1 and Table S1 in the Supporting Information. Note that bond lengths, angles and dihedrals of the cationic Trp species are not altered compared to the neutral moiety. However, the relaxation of the Trp geometry due to the charge state, i.e., the internal reorganization energy  $\lambda_i$ , is determined separately by quantum chemical calculations and treated as a constant parameter for the calculation of Marcus rates as well as for the nonadiabatic EID simulations.

The protein is solvated in a rectangular box with 33,753 water molecules using the TIP3P model.<sup>87</sup> Periodic boundary conditions (PBCs) are applied. All simulations are carried out with the GROMACS software package.<sup>88</sup> Amber parm99<sup>89</sup> force field parameters with the extension of improved protein backbone parameters ff99SB<sup>90</sup> are used for the protein parts. The LINCS algorithm<sup>91</sup> is employed to keep bonds involving hydrogen at fixed length. After a standard heating-minimization protocol followed by a 500 ps equilibration phase, which is discarded afterward, 30 ns MD simulations are performed with a time step of 2 fs. Snapshots of the molecular structures were saved every 1 ps for which the CT parameters were calculated with the SCC-DFTB-FO approach described below.

Five classical MD simulations of PL in solution have been carried out with the following setup: (A) the cofactor FAD [For the sake of simplicity, the Riboflavin part will be referred to as FAD in the following] and the 3 Trp residues carry no charge, (B) FAD is negatively charged, and (C–E) FAD carries a charge of -1 and either one of the Trp residues 382, 359, or 306 a charge of +1. In what follows, Trp382, Trp359, and Trp306 are denoted as site 1, 2, and 3, respectively. Table 1 summarizes the notation that is used throughout the paper.

**QM Region Selection.** An important point in the setup of QM/MM simulations is the appropriate choice of the QM region. We have, guided by previous data and chemical intuition, selected the three evolutionary conserved Trp residues in Table 1. The involvement of these three residues is strongly supported by spectroscopic and point mutation experimental data (summarized, e.g., in ref 10) as well as theory.<sup>30</sup> There is no

compelling experimental evidence for the involvement of other residues in the CT, and a hopping process along Trp382, 359, and 306 is the accepted mechanism for photoactivation.<sup>14</sup> Nevertheless, we have reexamined the structure of PL (based on PDB code 1DNP) with the intent to identify possible alternate pathways. Among all aromatic residues that have at least one atom within 7 Å of the FAD cofactors and the three Trp side chains, several need to be considered: His294, Phe297, Phe300, Phe350, and Tyr365 lie along the ET path, but are unlikely to play an important role due to their spatial orientation and higher redox potentials. Tyr464 could be suggested as an alternate end point in the ET chain, but it was shown that electron localization on it would be unfavorable.<sup>92</sup> Trp384 lies close to FAD and could be considered as an alternate initial electron donor to FAD and starting point for a different ET chain. Furthermore, possible ET between Trp382 and Trp316 cannot be ruled out by visual inspection. For the last two residues, Trp316 and 384, no evidence of their involvement in the CT process has been found so far, and it was shown in a previous theoretical study that they would not influence transport along the Trp chain.<sup>30</sup> Therefore, our present system provides a minimal but sufficient choice of QM system to study the photoactivation of PL.

**Free Energy Differences via Thermodynamic Integration.** In this study, free energy differences are computed via the thermodynamic integration (TI) method.<sup>93</sup> Two chemical states A and B, described by the two potential functions  $V_A$  and  $V_B$ , are coupled via an additional coordinate  $\Lambda$ . [We have chosen  $\Lambda$  instead of the more common  $\lambda$  to avoid confusion with the reorganization energy.] In the context of ET processes, states A and B are defined as the charge being completely localized on one amino acid side-chain at the beginning as well as the end of the CT process. Therefore, the functions  $V_A$  and  $V_B$  differ only in the atomic partial charges of the CT-active side-chains, and  $V(\Lambda)$  is constructed by linear mixing. The free energy difference between states A and B is obtained by numerically integrating:

$$\begin{aligned} \Delta G &= G_B - G_A \\ &= \int_0^1 d\Lambda \left\langle \frac{\partial V(\Lambda)}{\partial \Lambda} \right\rangle_{\Lambda} \approx \sum_i w_i \left\langle \frac{\partial V(\Lambda)}{\partial \Lambda} \right\rangle_{\Lambda_i} \end{aligned} \quad (2)$$

The angular brackets in eq 2 denote Boltzmann-weighted averages, obtained from MD simulations using the GROMACS software package.<sup>88</sup>

The free energy change is usually decomposed into an inner and an outer contribution. The internal part contains changes of the geometry as well as of the electronic structure of the CT-active species. The outer part is composed of global conformational changes in the protein as well as changes of the solvent structure, i.e., solvent polarization. The internal part of  $\Delta G$  cannot be described by MM force fields correctly, but this contribution to  $\Delta G$  is vanishingly small in the case of CT between identical species (Trp side-chains) and can be neglected.

**Electronic Structure. The Coarse Grained DFT Approach.** The methodology for a direct simulation of CT in complex systems has been described in detail recently.<sup>80</sup> Here, we will only briefly mention the most important concepts. We start from a

charge-neutral closed-shell system with  $N$  electrons, for which the DFT ground-state density  $\rho_0$  and the KS orbitals  $\Psi_i^0$  have been determined. The total DFT energy can be written in terms of the KS matrix elements:<sup>94</sup>

$$E^N[\rho_0] = \sum_i^{N/2} 2\langle \Psi_i^0 | H[\rho_0] | \Psi_i^0 \rangle + E_{DC}[\rho_0] \quad (3)$$

with

$$\begin{aligned} E_{DC}[\rho_0] = & -\frac{1}{2} \int \int' \frac{1}{|r-r'|} \rho_0(r) \rho_0(r') d^3r d^3r' \\ & - \int v_{XC}[\rho_0] \rho_0 d^3r + E_{XC}[\rho_0] + V_{NN} \end{aligned} \quad (4)$$

$E_{DC}[\rho_0]$  contains the DFT “double-counting” and nucleus–nucleus repulsion terms  $V_{NN}$ . The state of the system containing a hole is described as containing  $N - 1$  electrons, and its energy can be approximated by an expansion with respect to the density  $\rho = \rho_0 + \delta\rho$  around the  $N$ -electron density  $\rho_0$  up to the second-order ( $n_{HOMO} = 1$ ,  $n_i = 2$  otherwise):

$$\begin{aligned} E^{N-1}[\rho] \approx & \sum_i^{N/2} n_i \langle \Psi_i | H[\rho_0] | \Psi_i \rangle + E_{DC}[\rho_0] \\ & + \frac{1}{2} \int \int' \left( \frac{1}{|r-r'|} + f_{XC}[\rho_0] \right) \delta\rho(r) \delta\rho(r') d^3r d^3r' \end{aligned} \quad (5)$$

with  $f_{XC}$  being the second derivative of the DFT exchange-correlation energy ( $f_{XC} = \delta^2 E_{XC} / \delta \rho^2$ ).

Using the frozen-core approximation:

$$\Psi_i \approx \Psi_i^0 \quad (6)$$

we can approximate the total energy as (using the highest occupied molecular orbital (HOMO)  $\Psi_h^0$ )

$$E^{N-1}[\rho] \approx E^N[\rho_0] - \langle \Psi_h^0 | H[\rho_0] | \Psi_h^0 \rangle + E^{2nd} \quad (7)$$

The three contributions to the total energy are then further approximated as follows:

- The total energy of the neutral reference system is computed as the molecular-mechanics total energy

$$E^N[\rho_0] = E^{MM} \quad (8)$$

- To compute the matrix elements involving the hole wave function,  $\Psi_h^0$ , we use an expansion in terms of fragment orbitals  $\phi_m$ , which are obtained as DFT (KS) orbitals of the molecular fragments. These fragments are the side-chains of the involved tryptophanes.<sup>95</sup>

$$\Psi_h = \sum_{m=1}^M a_m^i \phi_m \quad (9)$$

In the simplest case, only one orbital per fragment is included for hole transfer, namely the HOMO  $\phi_m$  of the fragment  $m$ . [For more complicated cases, more orbitals per site can be easily included; see, e.g., ref 96.] The wave function of the hole is then

expanded as<sup>95</sup>

$$\begin{aligned}\langle \Psi_h^0 | H[\rho_0] | \Psi_h^0 \rangle &= \sum_{mn} a_m^h a_n^{h*} \langle \phi_n | H[\rho_0] | \phi_m \rangle \\ &= \sum_{mn} a_m^h a_n^{h*} H_{mn}^0\end{aligned}\quad (10)$$

The diagonal matrix elements  $\varepsilon_m^0 = T_{mm}^0 = H_{mm}^0$  represent effective ionization potentials, and the off-diagonal elements are the so-called CT or hopping matrix elements  $T_{ij}^0 = H_{mn}^0$ . The index "0" refers to the charge-neutral reference density  $\rho_0$ .

- In a similar fashion, the differential density  $\delta\rho = \rho - \rho_0$  in eq 6 can be decomposed into contributions located on the individual molecular fragments:

$$\delta\rho = \sum_m \delta\rho_m \quad (11)$$

and the last term in eq 6,  $E^{2nd}$ , can be written as

$$E^{2nd} = \frac{1}{2} \sum_{nm} \int \int' \left( \frac{1}{|r-r'|} + f_{XC}[\rho_0] \right) \delta\rho_m(r) \delta\rho_n(r') d^3r d^3r' \quad (12)$$

Applying a monopole approximation for the integral, the expression can be further simplified to

$$E^{2nd} = \frac{1}{2} \sum_m U_m \Delta Q_m^2 + \frac{1}{2} \sum_{m \neq n} \frac{\Delta Q_m \Delta Q_n}{R_{mn}} \quad (13)$$

where  $\Delta Q_m$  denotes the fraction of the hole charge localized on site  $m$ . The first term involves the Hubbard parameter  $U$  (chemical hardness), which determines the electron–electron repulsion on site  $m$ . The interaction between neighboring sites is described as simple Coulomb repulsion of the hole portions on fragments  $m$  and  $n$ . Note that this notation is used only for simplicity, since it is easily possible to project the fragment charges  $Q_m$  to atomic charges  $q_\alpha$  and compute the interaction energy much more accurately. This second-order term contains much of the well-known DFT self-interaction error, and we apply a simple self-interaction correction (SIC) by scaling  $E^{2nd}$  by a factor of  $\Gamma = 1/6$  (see ref 80 for more detail).  $U$  is determined using DFT calculations, and we find a value of  $U = 5.33 \text{ eV}/e^2$  for the tryptophan side chain.

Now, the total energy reads

$$\begin{aligned}E^{N-1}[\rho] &= E^{MM} - \sum_m a_m^* a_m \varepsilon_m^0 - \sum_{m \neq n} a_m^* a_n T_{ij} \\ &\quad + \Gamma \left( \frac{1}{2} \sum_m U_m \Delta Q_m^2 + \frac{1}{2} \sum_{m \neq n} \frac{\Delta Q_m \Delta Q_n}{R_{mn}} \right)\end{aligned}\quad (14)$$

**QM/MM Coupling.** In the case of hole transfer in PL, only the three tryptophan side chains are treated quantum mechanically (QM region), while the remainder of the system is described with classical molecular mechanics (MM region). Certain parts of the interaction between the QM and MM regions occur in all three energy contributions introduced above. In the coarse-grained methodology described above, a part of the QM region is already treated with  $E^{MM}$ . Therefore, to describe the interaction of this part with the MM region, the MM energy expression is simply extended to the entire

(QM+MM) system. The CT parameters  $\varepsilon_m^0$  and  $T_{ij}^0$  are computed in the presence of the solvent and the remainder of the protein. Therefore, the electrostatic potential (ESP) at site  $m$ :

$$ESP_m = \sum_{MMcharges} q_\alpha^{MM} / R_{m\alpha} \quad (15)$$

is included in the computation of the CT parameters.<sup>95,97</sup> Now the self-consistent diagonal elements  $\varepsilon_m^0$  [note that only the CT parameters  $\varepsilon_m^0$  are referred to as site energies] are obtained as

$$\varepsilon_m = -\varepsilon_m^0 + U_m \Delta Q_m + \sum_{n \neq m} \frac{\Delta Q_n}{R_{mn}} \quad (16)$$

Three major factors determine the site energies  $\varepsilon_m^0$ , and hence the energetics of CT:

- 1 *Static differences:* Differences in ionization potentials of the charge-carrying sites strongly influence CT. This does not apply here since the three charge carrying side-chains in PL are identical, but is important if, e.g., different amino acid residues, such as Tyr and Trp, are involved in CT (this effect is also crucial for the sequence-dependent CT in DNA).
- 2 *QM/MM coupling, i.e., fluctuations of  $\varepsilon_m^0$ :* The electrostatic interaction with the environment, i.e., the electrostatic potential  $ESP_m$  affects the site energy substantially. This can include a static shift, which would drive the CT and also a dynamic component<sup>97</sup> in which the dynamics of the MM environment leads to considerable fluctuations of the site energies (e.g., about 0.4 eV in DNA).
- 3 *Effect of the hole charge:* The positive hole charge polarizes the environment. This leads to a rearrangement of the surroundings and a large change in the ESP. In turn, this change leads to a stabilization of the hole charge due to a significant decrease of the respective site energy.

Note that eq 16 accounts for the complete interaction with the environment, but it does not represent the total energy since the  $E^{MM}$  term is missing. When comparing energy differences, e.g., from a hole between sites  $\varepsilon_m$  and  $\varepsilon_n$ , this does not reflect the full total energy difference. Nevertheless, this energy difference gives a first insight into the energetics of the system and is a critical factor that drives the dynamics of the system. As we have shown in detail before, the dynamics is determined by the ESP, since fluctuations induced by solvent and protein will determine the energies of the CT-active sites. Further, the presence of the hole charge at site  $m$  polarizes the (protein and water) environment. This leads to a dramatic decrease of the ESP at site  $m$ . For instance, we have shown the site energy to be lowered by 1 to 2 eV for hole transfer in DNA. This polarization of environment has been conceptualized as a "polaron",<sup>98</sup> as it is likewise a dynamic entity following the hole charge.

**Internal Reorganization Energy.** The inner-sphere reorganization is not included in the equations above, since the dynamics of atoms is propagated using classical MD simulations with force field parameters derived for neutral nucleobases (except for the atomic charges, which are updated in every step to include the projected hole charge). In this study, the effect of internal reorganization for each site  $m$  is approximated as a (small) correction  $\lambda_i^m$  to the diagonal elements  $\varepsilon_m$  [note that in Marcus

theory,  $\lambda_i$  is the overall internal reorganization energy for both donor and acceptor]:

$$\varepsilon'_m = \varepsilon_m - \lambda_i^m \Delta Q_m \quad (17)$$

Compared to the SIC scaling parameter  $\Gamma$ , the inclusion of  $\lambda_i^m$  is a minor correction, and it may be considered as an additional localization force.

**Calculation of Coarse-Grained CT Parameters.** In previous publications, we have shown in great detail that CT parameters computed for DNA bases at the SCC-DFTB level agree excellently with those obtained with higher-level DFT, HF, and CASPT2 calculations.<sup>95</sup> Recently, Voityuk and co-workers have shown that DFT provides reasonable CT parameters for hole transfer in DNA.<sup>96</sup> Nevertheless, extensive benchmark calculations have to be performed for every new application. For the PL system, benchmark calculations of energies and MOs for the highest occupied levels of a single skatole molecule (used as a model for the tryptophan side chain) were performed with DFTB, DFT, and HF. As a result, an excellent performance of DFTB was confirmed for this application. The data can be found in Table SII in the Supporting Information. In previous work, we have shown that the QM/MM coupling at the DFTB level of theory reproduces the DFT values very well.<sup>95</sup> Therefore, the zeroth-order part is well treated using SCC-DFTB for the calculation of the CT parameters.

**Non-Adiabatic Simulations.** Applying the Lagrangian formalism, coupled equations of motion for the electronic and ionic degrees of freedom were derived, as described in our previous work.<sup>80</sup> This leads to classical equations of motion for the nuclei (atoms):

$$M_\alpha \ddot{R}_\alpha = \frac{\partial E^{\text{MM}}(q_\alpha^{\text{ion}}, q_A^0)}{\partial R_\alpha} \quad (18)$$

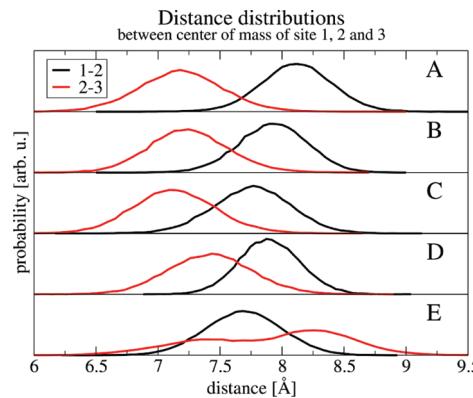
with the QM charges  $q_\alpha^{\text{ion}}$  of the  $(N - 1)$ -electron system and the MM charges  $q_A^0$  representing the charge-neutral system; and effective time-dependent Kohn–Sham (TD-KS) equations (DFT analogue of the time-dependent Schrödinger equation) for the electronic degrees of freedom:

$$\dot{a}_m = i \sum_n a_n T_{ij} \quad (19)$$

with

$$T_{ij} = \begin{cases} \varepsilon_m, & \text{if } n = m \\ -T_{ij}^0, & \text{if } n \neq m \end{cases} \quad (20)$$

Only the diagonal elements  $\varepsilon_m$  are affected by the second-order term (eq 16), while the off-diagonal elements are the hopping matrix elements  $T_{ij}^0$ . Since the two sets of equations are coupled, the electronic degrees of freedom are propagated using the Runge–Kutta integration, with a variable time step, between two time steps of the propagation of the nuclei. An MD time step of 1 fs was chosen. Effectively, the propagation of the electronic degrees of freedom leads to a new charge distribution, which is projected onto the classical particles (atomic partial charges) in the following MD time step. The propagation of the classical system then proceeds using classical Newtonian mechanics.<sup>80</sup> Effectively, the electronic dynamics leads to an update of MM charges before every MD time step. In this way, the dynamics of the hole charge is driven by the TD-KS equations, while the dynamics of the molecular system is determined by MM by using



**Figure 2.** Structural analysis of simulations A–E: Distribution of distances between sites 1 and 2 as well as between sites 2 and 3.

the updated atomic charges obtained from the integration of the TD-KS equation.

## RESULTS

**Structural Analysis.** An important structural parameter controlling the CT process is the distance between the tryptophan side chains, i.e., sites 1, 2, and 3, since the rate of tunneling CT depends exponentially on distance. As can be seen in Figure 2, the relevant distances for both steps  $1 \rightarrow 2$  and  $2 \rightarrow 3$  are distributed normally.

Generally, the distances are significantly beyond van der Waals contact distances, exhibiting maxima between 7.2 Å and 8.2 Å. This is more than twice the stacking distance between base pairs in double-stranded DNA, over which fast CT is known to proceed. In comparison, large distances between the Trp residues should result in smaller electronic couplings, rendering a slower CT process in this protein.

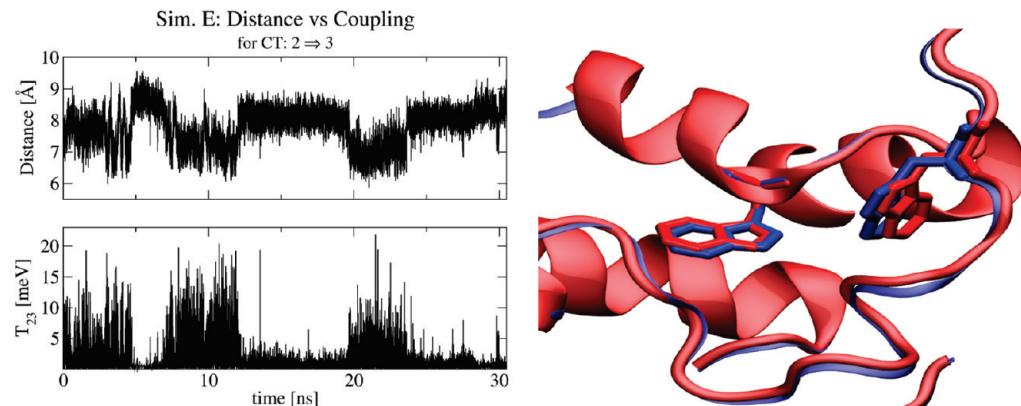
The average distance  $2 \rightarrow 3$  is 0.5–1.0 Å smaller than that for  $1 \rightarrow 2$  in simulations A, B, C and D. The distance  $2 \rightarrow 3$  in simulation E differs from the previous cases in that it exhibits a bimodal distribution with maxima around 7.4 and 8.2 Å. Apparently, rearrangements between two (or more) conformations occur during the MD simulation once site 3 is charged.

A direct CT from site 1 to site 3 appears unlikely, due to the large distance  $1 \rightarrow 3$  of about 11.5 Å.

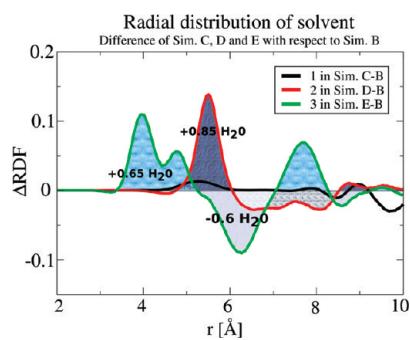
Additionally, the root-mean-square deviation (rmsd) with respect to the X-ray structure PDB ID 1DNP<sup>4</sup> and the rms fluctuation for the complete protein sequence during MD simulations were measured and show a fairly stable macromolecular structure (Figure S2 in the Supporting Information).

**Structural Changes with a Hole on Site 3.** As mentioned above, the distance between sites 2 and 3 in simulation E is distributed bimodally. The time series of the distance as well as of the corresponding electronic coupling is shown in Figure 3. A strong correlation of these properties is apparent: Small electronic coupling is observed in time intervals where the distance is large, denoted as E2 in the following, while a larger coupling of up to 20 meV occurs in intervals with a shorter distance (E1).

There exist two metastable local conformations, and there are frequent transitions between them in the course of simulation E. The two conformations obtained as average structures from the E1 and E2 parts of the trajectory are visualized in Figure 3. Their comparably small rmsd value of 0.44 Å illustrates the rather



**Figure 3.** Two conformations of the protein are observed in simulation E. Left: Time series of the distance between sites 2 and 3 as well as the corresponding electronic coupling  $T_{23}$ . Right: The averaged structure of the conformations E1 and E2 obtained as averages of the split trajectories.



**Figure 4.** Solvent distribution around the CT sites. Plots show the change in radial distribution functions when a site becomes charged, compared to simulation B. While little change is seen for site 1, H<sub>2</sub>O molecules are able to move significantly closer to sites 2 and 3 once one of them carries the hole charge, indicating the formation of a small solvent polaron in either case.

subtle character of the difference between E1 and E2. Still, the distance between sites 2 and 3 is 1 Å smaller in E1 than in E2, and the difference is sufficient to affect the CT process substantially. Further analysis did not allow us to identify any specific protein conformational change as the cause of the different structures of the CT region. These data can be found in Figure S3 in the Supporting Information

**Water Distribution.** The importance of aqueous solvent interacting with localized charges in CT active systems has been noted previously, notably in the context of CT in double-stranded DNA.<sup>80</sup> A hole charge can cause strong orientation polarization of the surrounding solvent, and the resulting induced ESP strongly influences the energetics of CT. Therefore the effect of polar solvent must be studied in biomolecular CT, and we will analyze this solvent contribution in the following in detail. The distribution of solvent around the three CT-active tryptophan side-chains is different due to their different locations inside the protein.

Site 1 is located 6–7 Å away from the FAD cofactor in the central region of the protein, and there is no water until 8 Å away from it. Therefore the hole located on site 1 would be expected to be only slightly affected by the solvent. Sites 2 and 3 are substantially closer to the surface of the protein and thus to water, which is reflected by the distribution of water setting on at about 5.5 and 4 Å, respectively. On average, six water molecules

can be found within 6 Å distance of site 3 once it is charged. By contrast, only one water molecule is located in the neighborhood of site 2 once it is charged (compare Figure S4 in the Supporting Information).

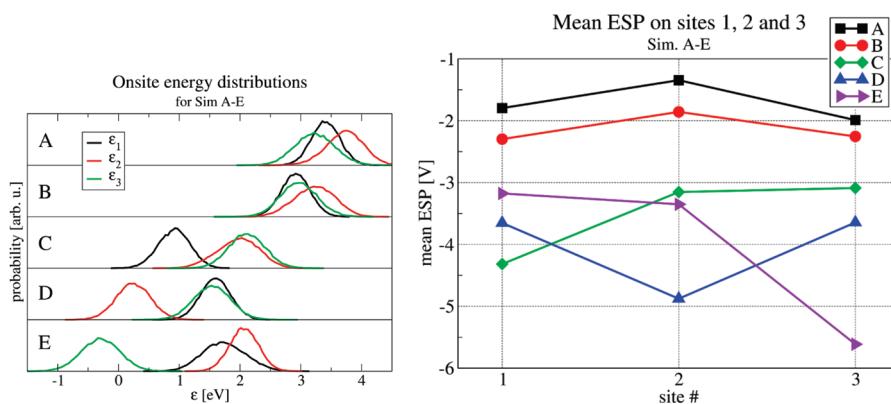
A plot of the change in water distribution after introducing charges (Figure 4) illustrates the formation of a solvent polaron around the positively charged sites 2 and 3. For instance, the difference in water distribution for the hole located on the site 3 is negative between 7 and 5 Å, while it turns positive between 5 and 3 Å; this can be understood as a rearrangement of water molecules moving closer to the hole charge or as a movement of site 3 toward the solvent. As expected, there is no such effect for site 1 in simulation C.

**Statistics of CT Parameters.** The quantities of interest in a study of CT are the site energy  $\varepsilon_i$  (eq 16), which is crucial for the energetics of CT, and the electronic coupling  $T_{ij}$ , which determines the probability of a transfer. In contrast to the site energy, it is well-known from previous studies on CT that  $T_{ij}$  depends sensitively on the structure of the CT system itself, while the effect of solvent is of minor importance.<sup>97,100</sup>

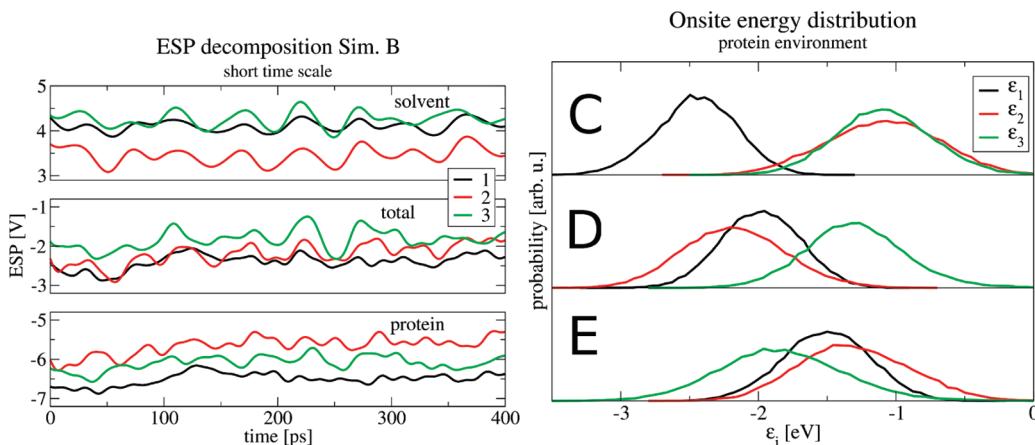
The main factors determining the hole transfer are (i) the fluctuations of the environment, which lead to fluctuations of the site energies  $\varepsilon_i$ ,<sup>97,100</sup> and (ii) the polarization of the environment due to the hole charge, which leads to a stabilization of the hole charge due to the induced ESP.

We evaluated the site energies  $\varepsilon_i$  and the electronic couplings  $T_{ij}$  for simulations A through E. We will first analyze the distributions of  $\varepsilon_i$  and the ESP, which are given in Figure 5 (mean values and standard deviations are presented in Table SIII in the Supporting Information).

**Protein and Water Fluctuations.** Factor (i) can be readily inferred from the simulation A, where all sites including FAD are chosen to be charge neutral. As shown in Figure 5 (left), the site energies are distributed normally with standard deviations of 0.25 to 0.5 eV. The fluctuations of site 1 (0.3 eV) are smaller than those of site 3 (0.5 eV). This is due to the fact that site 3 is more exposed to the solvent, and therefore the site energy is driven by the solvent fluctuations, while the fluctuations of site 1 are dominated by the protein fluctuations, which are slightly smaller but still significant. Note further that charging FAD (simulation B) and then also the Trp side-chains (in simulation C–E) does not change the magnitude of the fluctuations, but only leads to a shift in the energy.



**Figure 5.** Distribution of site energies  $\epsilon_m^0$  (left) and mean ESP (right) for simulations A–E. Note the direct connection between  $\epsilon_m^0$  and  $ESP_m$ .



**Figure 6.** Total ESP versus time in simulation B, as well as the ESP components due to the protein and solvent environment (left). Distribution of site energies  $\epsilon_m^0$  for the protein only in simulations C, D, and E, neglecting the effect of the water solvent (right).

To address the origin of the fluctuations of site energy, which is directly connected to the ESP (eq 15), a 10 ps MD simulation with the setup of simulation B was performed. The ESP was calculated in 1 fs intervals and decomposed into contributions by the solvent and by the protein environment; the results are presented in Figure 6. The ESP on site 3 is largely controlled by the solvent component, and this is also illustrated by the correlation coefficient of almost 0.8 given in Table 2. Contrary to that, the electronic structure of sites 1 and 2 is affected more markedly by the adjacent parts of the protein. Moreover, quite regular fluctuations of the ESP due to the solvent with a period of 40–50 fs are revealed.

The corresponding power spectra of site energy as well as ESP time series can be found in the Supporting Information (Figure S5). In summary, two distinct bands are observed in the power spectrum of site energy: a band around  $800\text{ cm}^{-1}$ , which can be assigned to the fluctuations of the environment, whereas the other band around  $1800\text{ cm}^{-1}$  corresponds to the internal stretching modes of the tryptophan side chains. The decomposition of ESP reveals bands around  $750\text{ cm}^{-1}$  in the spectrum of solvent as well as around  $800\text{ cm}^{-1}$  in the spectrum of the protein environment. Both movements manifest themselves in a combined band in the power spectrum of the site energy at around  $800\text{ cm}^{-1}$ , corresponding to a oscillation period of about 40 fs.

**Table 2. Correlation of the Total ESP and Its Components: Protein and Solvent Environment Obtained from a 10 ps Trajectory of Simulation B**

$\rho(x,y)$	1	2	3
total, solvent	0.56	0.53	0.77
total, protein	0.70	0.69	0.33

A less distinct band at  $1700\text{ cm}^{-1}$  in the power spectrum of ESP originates from Amide I modes of the protein backbone.

*Polarization of the Environment by the Hole Charge.* As discussed above, the polarization of the environment due to the hole charge will lead to a drastic change in the ESP at site  $m$ , where the hole is localized. As a result, the site energy  $\epsilon_m^0$  is drastically lowered by more than 1 eV. This leads to a (self-) trapping of the hole charge. Such a phenomenon is called a “polaron” in solid state physics and has been introduced to DNA hole transfer by Conwell and co-workers.<sup>98</sup> We have discussed this effect in great detail for hole transfer in DNA using the same methodology as described above,<sup>80</sup> and report a similar finding for a protein here (see Figure 5 (right)). The mean ESP values at sites 1–3 are shown for simulations A–E. They are quite similar for simulations A and B and only slightly shifted due to the negative charge on FAD in simulation B. However, simulations

**Table 3.** Mean ESPs on CT Sites 1–3 for Simulations A–E Decomposed into (i) Protein and (ii) Solvent Environments<sup>a</sup>

Sim.	ESP <sub>1</sub>		ESP <sub>2</sub>		ESP <sub>3</sub>	
	protein	solvent	protein	solvent	protein	solvent
A	−5.3	3.5	−4.8	3.4	−5.5	3.5
B	−6.6	4.3	−5.9	4.0	−6.2	3.9
C	−7.7	3.3	−6.3	3.1	−6.4	3.3
D	−7.2	3.5	−7.4	2.5	−6.6	3.0
E	−6.7	3.5	−6.4	3.1	−7.3	1.7

<sup>a</sup> All values in V.

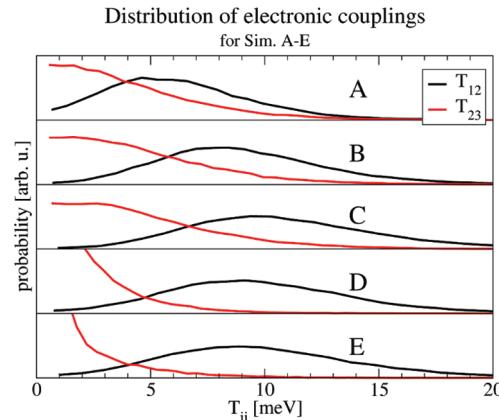
C–E show a very different behavior: At the site where the charge is located, the ESP decreases by 1 to 2 eV. Since  $\text{ESP}_m$  and  $\varepsilon_m^0$  are directly related, the energy shift of the occupied site in simulations C–E is also seen in the  $\varepsilon_m^0$  distributions in Figure 5 (left). It is important to note that this polarization effect leads to a self-stabilization or self-localization of the hole charge (see also the detailed discussion in ref 80).

In Figure 5 (left), the site energies are ca. 0.5 eV smaller in simulation B than in simulation A, which is due to the negative charge of the FAD cofactor. The decrease is larger for sites 1 and 2 (0.5 eV) than for site 3 (0.3 eV) because the latter is located further away from the FAD moiety.

The introduction of a hole charge on site 1 (simulation C) leads to a strong decrease of  $\varepsilon_1$  by 2 eV (compared to simulation B). Also, the adjacent sites 2 and 3 experience a drop of 1.3 and 0.9 eV, respectively. This is a consequence of the orientation polarization of the environment (solvent and protein), undergoing conformational reorganization due to the hole charge and thereby inducing a strong electric field back on the site 1 as well as the adjacent sites.

Not surprisingly, in each case, the strongest drop of site energy is observed for the site on which the hole charge is located. This drop in the  $\varepsilon_m^0$  results from a corresponding drop in the  $\text{ESP}_m$ , as shown in Figure 5 (right). The site energies for the sites carrying the hole charge in simulations C, D, and E decrease from site 1 to site 2 by 0.6 eV and further from site 2 to site 3 by another 0.5 eV. This indicates a downhill-type CT reaction, in which the hole charge is most strongly stabilized on site 3. From the average site energies, we can estimate a driving force of  $\Delta E_{1 \rightarrow 2} = -0.7$  eV CT step and of  $\Delta E_{2 \rightarrow 3} = -0.5$  eV for the second CT step (Supporting Information, Table SIII) by taking the differences of the lowest  $\varepsilon_m^0$  for simulations C–E. Note, however, that the differences of  $\varepsilon_m^0$  neglect a part of the energy, as discussed above, as well as entropic effects. Therefore, these values can be used only as a first estimate, and a more rigorous calculation of  $\Delta G_{m \rightarrow n}$  will be presented in the next section. In a similar fashion, the  $\varepsilon_m^0$  differences within one simulation can be used to estimate the sum of  $\Delta G_{m \rightarrow n} + \lambda$  ( $\lambda$ : reorganization energy), as discussed in the Supporting Information.

**Role of Water and Protein.** The negative charge located on the FAD (included as part of the “protein” here) induces negative ESP, stabilizing the (positive) hole charge on the tryptophan side chains. By contrast, the solvent induces a positive ESP in simulation B, therefore destabilizing the hole by 0.8, 0.6, and 0.4 eV for sites 1, 2, and 3, respectively. However, the effect of the protein part (FAD) prevails. It should be noted that ESPs can only be given with respect to an arbitrary zero, therefore any interpretation should focus on ESP differences only.



**Figure 7.** Distribution of electronic couplings  $T_{ij}$  for simulations A–E.

Introducing the hole charge in simulations C–E leads to a decrease of ESP induced by protein and solvent (Table 3). The hole stabilization of the protein part is more pronounced for sites 1 and 2. By contrast, the solvent part of ESP on site 1 is hardly affected by the location of the hole, since the values in simulations C–E are rather similar. For site 3, the solvent potential drops considerably by 1.3–1.6 V in simulation E compared to simulations C and D. Apparently, a hole charge on site 1 is predominantly stabilized by the protein, whereas the solvent plays the key role on site 3. Both solvent and protein environments play an important role for site 2.

The fluctuations of ESP in both solvent and protein environments are of comparable magnitude of 0.3 to almost 0.5 eV, with the exception of the ESP fluctuations on site 1, which are reduced considerably compared to sites 2 and 3. This indicates a less flexible environment around site 1.

The simulations show an important functional role of water, both from introducing fluctuations and stabilization. It introduces fluctuations of all three site energies, thereby enhancing the hole dynamics. Solvent predominantly stabilizes site 3 and, to a lesser degree, site 2, contributing to making the transfer exothermic.

Figure 6 shows the distribution of site energies  $\varepsilon_m^0$  when the effect of solvent is neglected, i.e., the water molecules are not included in the calculation of the ESP entering eq 16. For simulation C, we still see strong stabilization of site 1, where the charge is located, while for simulations D and E this stabilization is absent. This is due to the fact that the stabilization of charge on sites 2 and 3 is mostly due to water polarization. As a result, sites 2 and 3 are higher in energy than site 1, and the CT would be endothermic without the effects of water. This leads us to the conclusion that the hole transfer in PL is driven by the surrounding water.

**Electronic Couplings  $T_{ij}$ .** Figure 7 displays the distributions of electronic couplings  $T_{12}$  and  $T_{23}$  for simulations A–E. While  $T_{12}$  shows a Gaussian-like distribution around a maximum at 5–10 meV, the coupling  $T_{23}$  is much smaller on average and seems to be distributed around zero. This distribution of the couplings suggests that the first step of the CT can be described well using average couplings, while the transfer from site 2 to 3 may be dominated by the coupling fluctuations, giving rare conformations with high  $T_{23}$  values an outsize influence on CT.

The effect of fluctuations of electronic couplings in biomolecules (especially proteins) was investigated in great detail by Beratan, Skourtis, and co-workers, who studied a large number of

CT systems in proteins and DNA.<sup>75,101</sup> As described in ref 101, the coherence parameter C

$$C = \frac{\langle T_{ij} \rangle^2}{\langle T_{ij}^2 \rangle} = \frac{1}{1 + \frac{\sigma^2}{\langle T_{ij} \rangle^2}} \quad (21)$$

introduces a good measure for the importance of fluctuations. On the basis of eq 21,  $C \approx 1$  indicates that CT is mainly controlled by the average coupling, while  $C \ll 1$  indicates that the transfer is dominated by coupling fluctuations, i.e., strong dynamical disorder. The coherence parameter for electronic couplings  $T_{12}$  and  $T_{23}$  of both CT steps 1→2 and 2→3 is given in Table 4. While C for both couplings is larger than typical values in DNA,<sup>102,103</sup> the transfer 1→2 appears to be considerably more “coherent” than 2→3, and the coherence of the second step drops markedly in simulation E. The importance of fluctuations of  $T_{ij}$  will be addressed again below.

**CT Rates Based on Marcus Theory.** The mean squared electronic coupling  $\langle T_{ij}^2 \rangle$  is used for the donor–acceptor coupling in eq 1. [As indicated in Table 4, the coherence parameter is smaller than 1, i.e., the transfer is mainly dominated by coupling fluctuations, thus  $\langle T_{ij}^2 \rangle$  is used in eq 1 rather than  $\langle T_{ij} \rangle^2$ .]

In this study,  $\lambda_s$  for the charge shift reaction  $A^+ + B \rightarrow A + B^+$  is calculated from two ensembles (i.e., hole charge either on site 1 or 2) via a re-evaluation of the total system energy with the set of atomic charges corresponding to the other charge state. Then, the reorganization energies for the forward and backward CT ( $\lambda_{sf}$  and  $\lambda_{sb}$ , respectively) follow as

$$\begin{aligned} \lambda_{sf} &= \langle E_{A^+B} \rangle_{AB^+} - \langle E_{A^+B} \rangle_{A^+B} \\ \lambda_{sb} &= \langle E_{AB^+} \rangle_{A^+B} - \langle E_{AB^+} \rangle_{AB^+} \end{aligned} \quad (22)$$

with the subscripts of E denoting the potential functions used and the subscripts of the angular brackets indicating the conformational ensemble used for Boltzmann averaging. Recently, this protocol has been applied successfully to determine  $\lambda_s$  for the hole transfer in DNA.<sup>56</sup>  $\Delta G$  for all single CT steps is obtained via TI simulations. An internal reorganization energy  $\lambda_i$  of 0.36 eV was calculated at the B3LYP 6-311 G(d,p) level. Both values for  $\lambda$  and  $\Delta G$  are presented in Table 5.

**Table 4. Coherence Parameter for Both Electronic Couplings  $T_{12}$  and  $T_{23}$ , for Simulations A–E**

	A	B	C	D	E
1→2	0.77	0.84	0.85	0.84	0.82
2→3	0.62	0.60	0.62	0.50	0.35

**Table 5. Reaction Rates Calculated with eq 1 According to Marcus Theory of ET<sup>a</sup>**

CT step	$\lambda_{sf}$	$\lambda_{sb}$	$\Delta G$	no scaling		scaled $\lambda_s^b$		scaled $\lambda_s^c$	
				$k_f$	$k_b$	$k_f$	$k_b$	$k_f$	$k_b$
1→2	0.9	1.6	-0.40	$5.0 \times 10^9$	$1.2 \times 10^0$	$5.4 \times 10^{10}$	$9.4 \times 10^1$	$2.9 \times 10^{11}$	$2.2 \times 10^3$
2→3	1.5	1.9	-0.15	$2.4 \times 10^4$	$6.3 \times 10^{-1}$	$1.7 \times 10^6$	$1.6 \times 10^2$	$4.2 \times 10^7$	$9.9 \times 10^3$
1→3	1.4	2.3	-0.62	$1.2 \times 10^3$	$9.0 \times 10^{-13}$	$3.3 \times 10^4$	$4.3 \times 10^{-10}$	$2.9 \times 10^5$	$3.5 \times 10^{-8}$

<sup>a</sup>  $T_{ij}$  is taken from Table SIII in the Supporting Information. Note that the internal reorganization energy  $\lambda_i$  is included in the rate evaluation as  $\lambda_{f/b} = \lambda_{s_{f/b}} + \lambda_i$ .  $\lambda$  and  $\Delta G$  are in eV, and  $k$  is in  $s^{-1}$ . For comparison, experimental rates of  $k_{1\rightarrow 2} = 1.1 \times 10^{11} s^{-1}$  and  $k_{2\rightarrow 3} = 3.3 \times 10^{10} s^{-1}$  are available.<sup>10</sup>

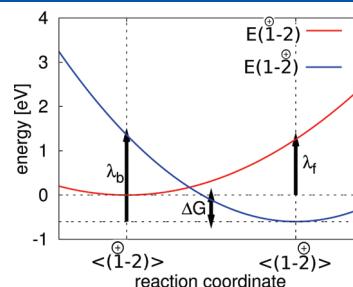
<sup>b</sup>  $\lambda_s$  scaled with 1/1.4. <sup>c</sup>  $\lambda_s$  scaled with 1/2.

The scheme depicted in Figure 8 illustrates the energy profile of the first CT step 1↔2. Note that this is already a step beyond the framework of Marcus' original theory, as the curvatures of the PES corresponding to the initial and the final states are different due to the different  $\lambda$  for the forward and backward reactions.

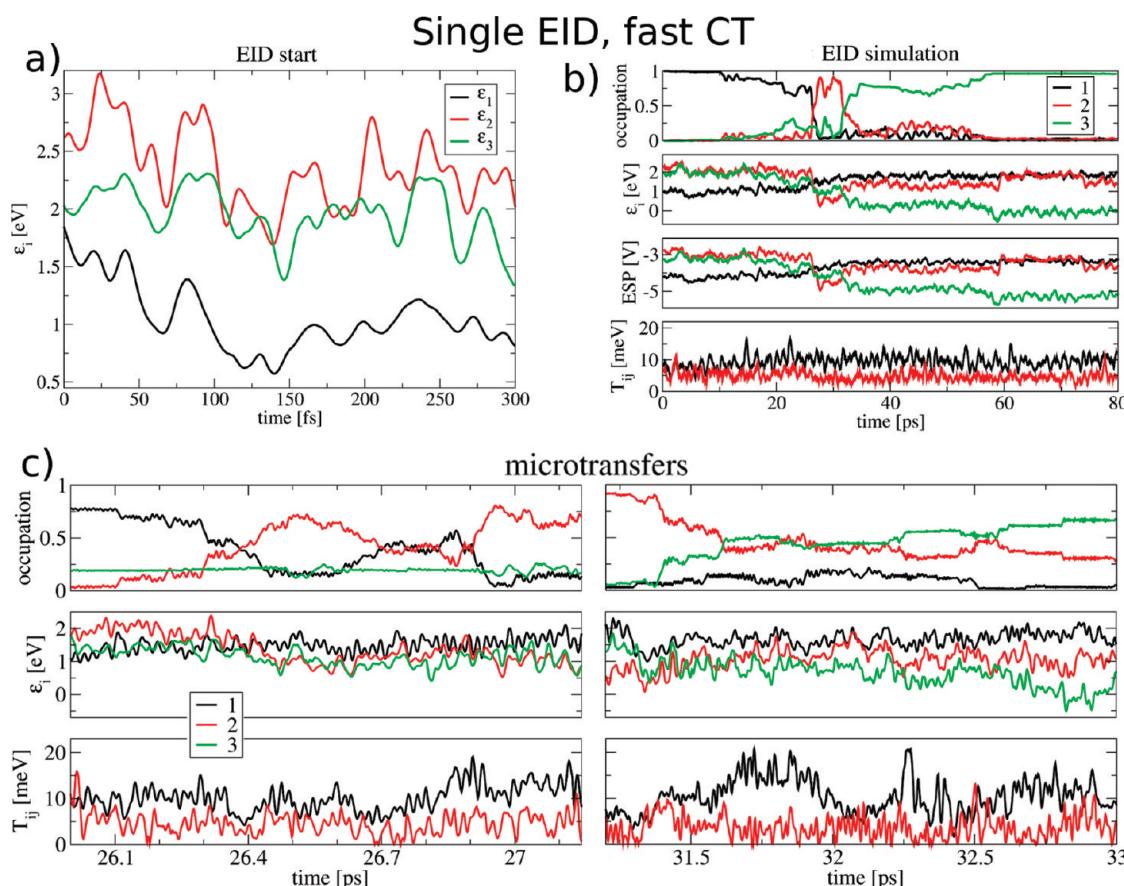
Evidently, the backward CT 1↔2 is disfavored because of the large reaction free energy of 0.4 eV and the reorganization energy amounting to almost 2 eV. These values enter the exponential term in eq 1, hence the rate will be very low. Combining the parameters  $\lambda$  and  $\Delta G$  with the donor–acceptor coupling evaluated from the  $T_{ij}$  values (Table SIII in the Supporting Information), the corresponding Marcus rates were calculated with eq 1 (Table 5).

The results should only be discussed in a qualitative manner because of the exponential dependence of the rate of CT on the energies used, giving modest changes of the parameters outside effects. As indicated in Table 5,  $\lambda_{sb}$  is significantly larger than  $\lambda_{sf}$  resulting in extremely slow back CT rates. The rate for 1→2 is 5 orders of magnitude larger than that for 2→3 since  $\lambda + \Delta G$  is about 1 eV smaller, plus the coupling  $T_{12}$  is five times larger than  $T_{23}$ .

It is well-known that reorganization energies computed from nonpolarizable force field simulations overestimate  $\lambda_s$  significantly.<sup>60,104–106</sup> Therefore, the resulting values of  $\lambda_s$  should be scaled by the inverse optical dielectric constant of the medium,  $\epsilon_\infty$ , which is not immediately accessible for the heterogeneous environment of a solvated protein. Various scaling factors in the range of 0.5–0.7 were suggested in the past,<sup>49,53,58,59,107</sup> and we present results for two different scaling factors in Table 5. The large effect that a change of  $\lambda$  has on  $k_r$  is readily apparent, indicating again that the results should be taken qualitatively only. We find that the scaling brings the rate for 1→2 into the picosecond range, while  $k_{2\rightarrow 3}$  remains much smaller.



**Figure 8.** Energetics of forward ( $k_f$ ) and backward ( $k_b$ ) hole transfer between sites 1 and 2 based on reorganization energy  $\lambda_{f/b} = \lambda_{s_{f/b}} + \lambda_i$  and  $\Delta G$  from Table 5. The equilibrium energy of the initial state (1) is set to zero.



**Figure 9.** Nonadiabatic Ehrenfest dynamics (EID) in PL, example of a fast CT: (a) Onsite energies for the first 300 fs, (b) CT parameters, occupations, and ESPs for all three sites along an 80 ps trajectory (running averages over 50 fs are used to smoothen the graphs), and (c) a more detailed cut-out for the two CT events at 27 and 32 ps (bottom).

An explanation for the surprisingly small rates for  $2 \rightarrow 3$  CT can be given: In the Marcus picture, the whole CT process  $1 \rightarrow 2 \rightarrow 3$  is decomposed into two separate, independent reactions, for which it is assumed that each initial state (i.e., hole on site 1 or 2, respectively) is fully equilibrated, that is, the surrounding was given enough time to stabilize the charge due to reorganization. However, this does not seem to be the case here. While the solvent reorganization is rather fast, namely within several hundreds of femtoseconds, additional relaxations occur on a substantially longer time scale of up to hundreds of picoseconds involving reorganization in the protein environment. This is clearly a factor for the second step  $2 \rightarrow 3$ .

In a previous study, Marcus rates of  $0.11 \text{ ns}^{-1}$  and  $0.45 \text{ ns}^{-1}$  were computed from MD simulations.<sup>31</sup> While these rates are comparably close to experimental data (off by about 1 order of magnitude), some fortuitous error cancellation could have played a role since the rates were obtained without considering two opposing effects,  $\lambda$ -scaling and  $\lambda_i$ . It is also notable that the significantly extended simulation length here resulted in somewhat different  $\Delta G$  estimates. Both the previous study as well as this report conclude that the assumption of a single  $\lambda$ -value for forward- and back-transfer is untenable and that two different  $\lambda$  values must be used; see, e.g., Table 5 or Table 2 in ref 31. In this regard, the rates presented here already represent an improved model beyond traditional Marcus theory.

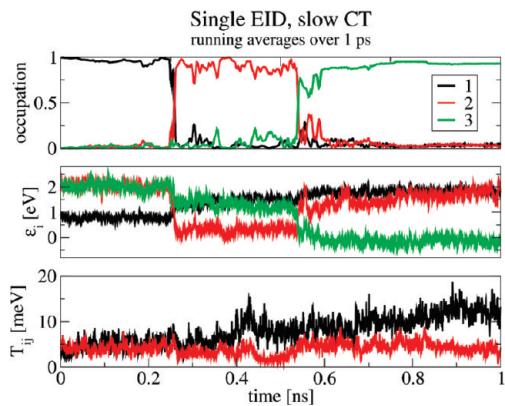
**Direct Dynamics.** The results in the previous section indicate that the CT in PL requires a theoretical description beyond

Marcus equilibrium theory. Therefore, we turn to conducting nonadiabatic coupled EID simulations to study charge redistribution directly. The setup assumes the photoactivation process to proceed according to the following mechanism:

CT is initiated by electronic excitation of the FAD cofactor (possibly via excitation transfer from MTHF) followed by a fast hole transfer to the adjacent tryptophan residue site 1 (Trp382). At this point, further dynamics occur solely in the electronic ground state. The cofactor remains negatively charged while the hole charge is free to move between site 1, the adjacent site 2 (Trp359), and site 3 (Trp306) located on the exterior of the protein. We will focus on this three-site hopping process, i.e., the direct CT  $1 \rightleftharpoons 2 \rightleftharpoons 3$ , starting with the hole charge completely located on site 1 and treating all three tryptophan residues with the QM/MM methodology described above. Further reactions like deprotonation of the Trp radical cation are not considered. No prior protein relaxation is considered, since the initial charge separation is assumed to be fast.

To aid visualization of the CT process, we provide an mp4-movie of a typical hole transfer trajectory available as a Web Enhanced Object accompanying this article. The amount of positive charge on each of the tryptophan residues is indicated by the size of the red van der Waals representation. Note that, in accord with the coarse-grained model, no assignment of excess charge to individual atoms is shown.

Simulations were started from randomly chosen initial structures from simulation A. In total, we conducted 42 simulations



**Figure 10.** Nonadiabatic Ehrenfest dynamics (EID) in PL, example of a slow CT: Hole probability, site energies, and electronic couplings between sites 1, 2, and 3 during 1 ns of a “slow” CT trajectory. To get a clearer picture, running averages over 1 ps are used to smoothen the graphs.

with a length of 1.05 ns each. The method does not presuppose sequential or direct transfer from site 1 to site 3, so the simulations are unbiased with regard to the detailed CT mechanism.

**Analysis of individual CT Events.** We will first discuss two examples of typical EID simulations, beginning with one in which CT occurs rapidly (Figure 9). The hole is stabilized on site 1 during the first 100–200 fs, the energy of site 1 drops by more than 1 eV, establishing an energy gap between site 1 and sites 2 and 3 (Figure 9a). The gap can be monitored as an effective reaction coordinate,<sup>48,108,109</sup> indicating the reorganization of protein and solvent environment due to the relocation of the positive charge, an effect discussed in detail above.

For CT from site 1 to site 2 to occur, the energy gap has to become small,  $\varepsilon_1 \approx \varepsilon_2$ , while the electronic couplings  $T_{12}$  have to assume sufficiently large values. This situation occurs for the first time after about 10 ps, when site 2 becomes gradually occupied until about 25 ps of simulation time (Figure 9b). Fractions of the hole charge on site 2 are also able to pass on directly to site 3. By accumulating more and more positive charge on site 2 and site 3,  $\varepsilon_2$  decreases. The partial hole charge on site 2 changes the polarization of the environment, leading to a partial reorganization, which in turn lowers the energy of site 2. The energy of site 1 increases simultaneously, and a crossing of  $\varepsilon_1$  and  $\varepsilon_2$  occurs at about 27 ps. Then,  $\varepsilon_2$  becomes the lowest energy state and the hole becomes almost completely localized on site 2 with minor fluctuations between sites 2 and 3. With the nearly complete charge on site 2, a substantial relaxation of site 2 might be suspected, as seen for site 1 after charge injection. However, this is not observed. Instead, a significant amount of charge is quickly transferred to site 3 as well. Therefore, the gap between sites 2 and 3 never becomes as large as that between 1 and 2 at the beginning of the simulation. This explains the quick 2→3 transfer and the residence time on site 2 of roughly 5 ps, which is significantly shorter than that on site 1.

The period from 25 to 35 ps is particularly interesting, showing very similar site energies (Figure 9c). There is no constant delocalization of charge between the sites, but rather an oscillation with short intervals of delocalization between 2 and 3, which occur for a fraction of a picosecond, similarly to the situation in DNA.<sup>80</sup> Until a simulation time of 60 ps, about 20% of the hole charge remains on sites 1 and 2, oscillating back and forth until the whole charge relaxes to site 3. The further drop in the energy

$\varepsilon_3$  after 60 ps establishes a gap of about 2 eV to the other sites and prevents any back-transfer. As shown below, this last step is coupled to a further relaxation of the protein. Figure 5 shows that the time dependence of site energies and ESP is nearly identical. They only differ by fast and small amplitude oscillations resulting from the vibrations of the sites, which are negligible compared to the oscillations introduced by the ESP (for further details see ref 97). Therefore, a drop in the site energy is always closely correlated to a drop in the ESP, i.e., it indicates a structural reorganization in the environment.

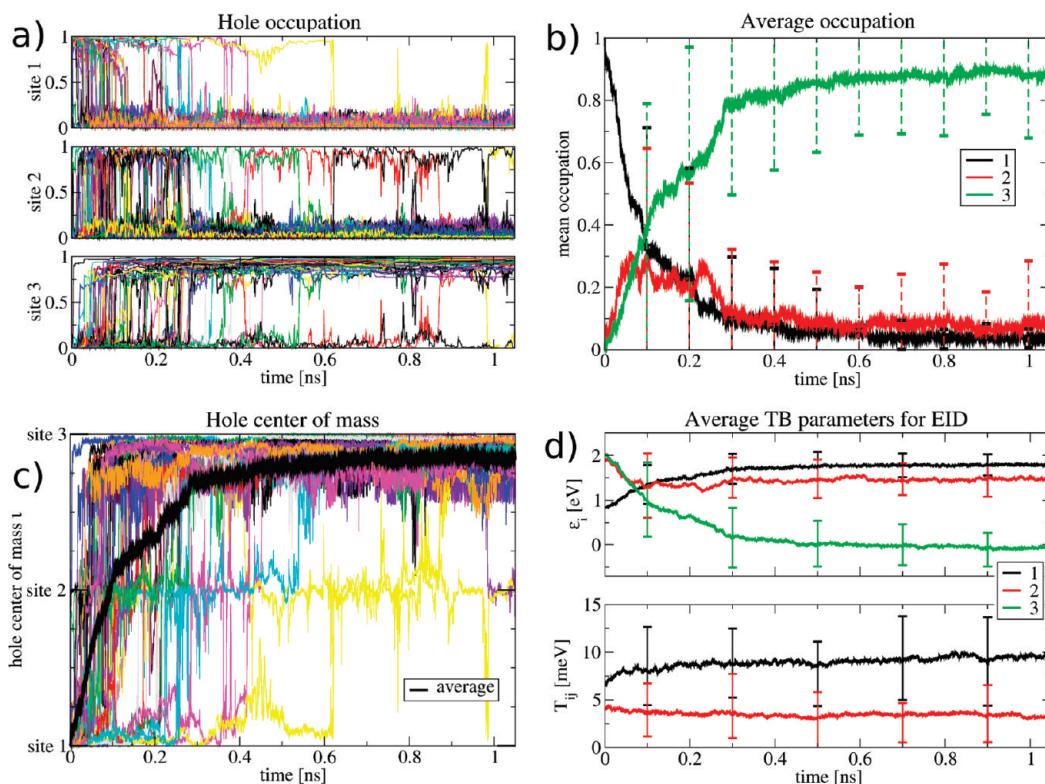
A “slow” CT trajectory shows quite different characteristics (Figure 10). Here, the hole moves from site 1 to 2 after about 0.25 ns, but does not pass on further to site 3. Instead, a shift in the site energies due to relaxation of the environment occurs. The site energies become clearly separated by a gap of about 1 eV (as in the discussion of the statistics of the site energies; see above) until a second CT event occurs at about 0.55 ns where the hole moves to site 3. In this case, the hole has a similar residence time on site 2 as it had on site 1.

Therefore, different CT mechanisms seem possible, depending on whether the hole stays at site 2 or moves directly on to site 3. This depends on the fact that at the same time the gap is vanishing and the couplings are nonzero. See ref 83 for a comparison to the situation in DNA CT (also, see Supporting Information for a summary comparison of the two CT systems). When the hole does not immediately move on to site 3, the environment relaxes and stabilizes the hole at site 2. This situation is implicitly assumed in Marcus theory, where each charge state is fully equilibrated to estimate the reorganization energy. In contrast, when the hole does move directly from site 1 to 3 with only a short residence time at site 2, the fact that  $\varepsilon_2$  is not lowered points to an incomplete reorganization, i.e., Marcus theory should not be applied in the way it has been above. Therefore, the localization/delocalization of charge plays an important role for the CT mechanism.

**Statistics.** Figure 11a shows the time series of occupation numbers for the three sites for all 42 EID simulations. In most of the simulations, CT 1→2 occurs within 120 ps, although in a few cases this first CT step takes up to 600 ps. As discussed above, a small fraction of charge remains on sites 1 and 2 until a final relaxation step of the protein. Even though the majority of simulations reveal the final CT 2→3 to be completed within 300 ps, several effective back transfers 3→2 are observed.

Figure 11b shows the averaged occupation number, indicating that site 2 is only transiently populated on average, up to an occupation of about 0.2, within the first 300 ps. The population of site 3 rises simultaneously with that of site 2 within the first 100 ps. During the first 100 ps, more than 60% of the charge is transferred to sites 2 and 3, and more than 30% is directly passed on to site 3. Although the charge is not substantially delocalized between the sites, the system dynamics allow the site energies to be sufficiently close for short periods, so that charge oscillations can occur. The residence times during the CT on average are not sufficient for the system to equilibrate in every charge state. In particular, site 2 seems to be occupied transiently (on average) so that its site energy is not significantly lowered due to environmental relaxation. This allows for a fast follow-up transfer to site 3. Note the large standard deviation bars for the curves in Figure 11b, indicating the variability of individual CT reactions, especially in the interval from 100 to 300 ps.

Figure 11c shows the time course of the hole center of mass,  $\iota = \sum_{i=1}^3 i \Delta Q_i$ , for the 42 simulations as well as the average  $\langle \iota \rangle$ .



**Figure 11.** EID statistics over 42 trajectories, each 1.05 ns long: (a) Site occupations for the three Trp side-chains involved in the CT for each single EID. Note, the different trajectories are realized by starting from different conformations obtained from simulation A. (b) Averaged occupations with standard deviations over all 42 single EID trajectories. (c) Movement of the hole center of mass  $t = \sum_{i=1}^3 i \Delta Q_i$  for each single EID as well as the average over all EIDs. (d) mean TB parameters, i.e., averages with standard deviations over all 42 EIDs (shown are running averages over 5 ps).

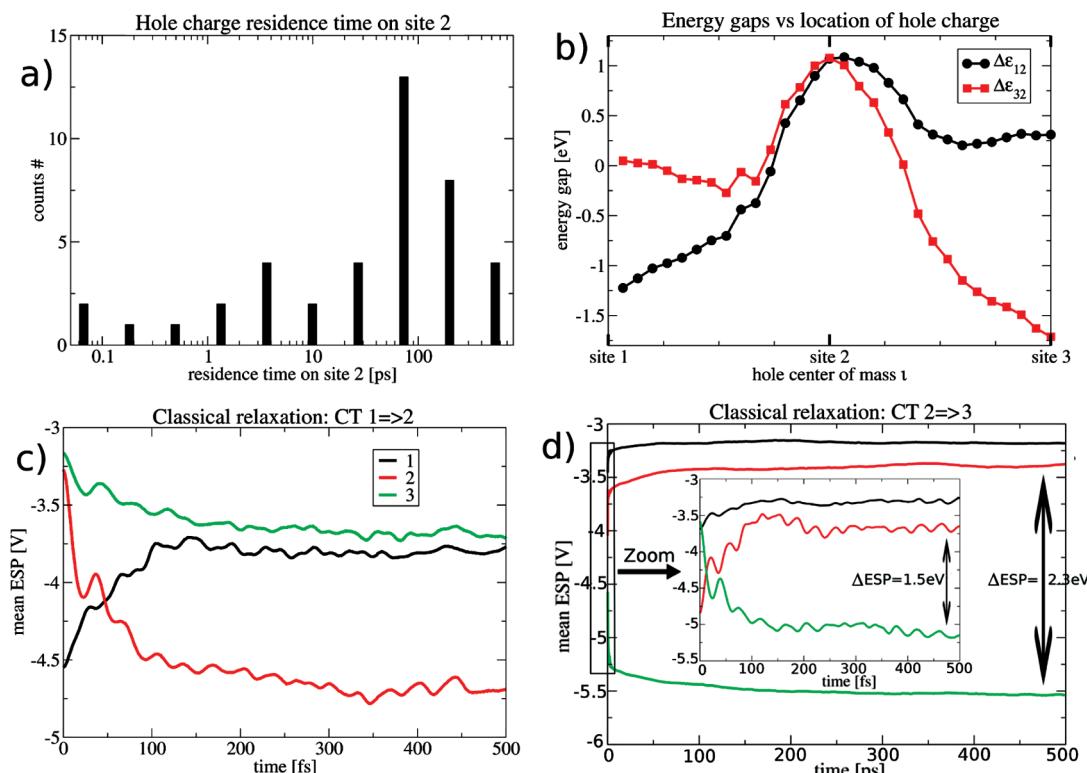
Before and after the CT steps  $1 \rightarrow 2$  and  $2 \rightarrow 3$ , the hole charge is typically localized on the respective sites. However, the hole charge is delocalized during the transfer process itself, which usually occurs within 1 to 2 ps. Therefore, the hole seems to be localized on one site most of the time, but delocalization, which is a result of the oscillation between 2 sites, occurs during the CT events. Note that the average did not reach a value of 3 even after 1 ns, indicating the slow progress of CT in some trajectories.

The time series of the average site energies and the couplings are shown in Figure 11d. It turns out that the site energy for the central site is never the lowest one on average, although there is a large overlap for the distribution of site energies until about 200 ps. The average site energy gap between the central site 2 and the site 3 amounts to 1.5 eV already after about 450 ps. Furthermore, the average couplings for both CT steps are rather comparable at the beginning of the simulation, i.e., they match the average values for simulation A (see Table SIII in the Supporting Information). However,  $T_{12}$  increases from 6 to 10 meV, while  $T_{23}$  decreases as the simulation proceeds. Both the large site energy gap between sites 2 and 3 as well as the rather small electronic coupling  $T_{23}$  lower the chances for successful back-transfers  $2 \leftarrow 3$ .

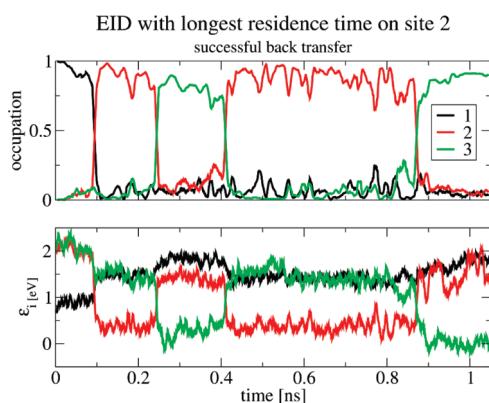
**Relaxation.** The analysis so far has shown that some trajectories allow for a fast transfer from site 1 to site 3, with site 2 being only temporarily occupied. For a more detailed analysis, we evaluated the residence times for site 2 (Figure 12a). The hole was defined as residing on site 2 if the occupation there exceeded 0.8.

A few very fast transfers can be seen, but the bulk of CT events show a residence time on site 2 on the order of 10–100 ps. This can be compared with the time scale of protein relaxation. Using classical MD simulations, we monitored the dynamics of the ESP at sites 1–3. Figure 12c shows the time course of the ESP after moving the hole charge from site 1 to site 2 (using the charge topology from simulation D for equilibrated conformations from simulation C). It takes about 200–300 fs for the ESP to converge closely to their new values. Therefore, only during the first few hundreds of fs after the transfer from  $1 \rightarrow 2$ , the site energies of sites 2 and 3 are sufficiently close to allow for a further transfer from site 2 to site 3. On the contrary, the relaxation for the step  $2 \rightarrow 3$  seems to be much slower, as the ESP converges only after several hundreds of picoseconds. This explains why back-transfer events for the last step are readily possible, as found, e.g., for the simulation with the longest residence time on site 2 (Figure 13). The classical relaxation of ESP is analyzed further by decomposing the ESP into protein and solvent components for both CTs  $1 \rightarrow 2$  and  $2 \rightarrow 3$ . Both solvent and protein response for the first CT  $1 \rightarrow 2$  are rather fast compared to those for CT  $2 \rightarrow 3$ . For the latter step, mainly the ESP relaxation stemming from the protein takes place on a nanosecond time scale. In general, the solvent response occurs on a considerably shorter time scale (several 100 fs) than the protein response (up to nanoseconds, maybe even more). The data can be seen in Figure S6 in the Supporting Information.

The explanation for the fast transfer can be found in the average energy gaps, evaluated in dependence of the hole center



**Figure 12.** Analysis of CT in PL: (a) Residence times at the central site 2, (b) energy gaps between the central site and sites 1 and 3, respectively, depending on the hole center of mass  $t$ , and (c,d) ESP relaxations from classical MD simulations for CTs 1 $\rightarrow$ 2 and 2 $\rightarrow$ 3, respectively.



**Figure 13.** Site occupation and energy levels for the simulation with the longest charge residence on site 2 (running averages over 5 ps). One successful back-transfer 2 $\rightarrow$ 3 is observed, after which the hole charge remains on the central site for almost half a nanosecond. Finally at 0.9 ns, the hole charge hops back again to site 3, where it undergoes a stronger relaxation (compared to the first CT 2 $\rightarrow$ 3 at 0.25 ns), thus effectively widening the energy gap  $\Delta\epsilon_{23}$ .

of mass  $t$ , as shown in Figure 12b. If site 2 is only partially occupied, the gap between  $\epsilon_2$  and  $\epsilon_3$  decreases, i.e., allows for a faster transfer to the next site. This indicates that the transfer cannot be modeled as consisting of two separate hopping steps, where the protein environment is allowed to relax fully after each transfer, as assumed above when computing the CT parameters for the Marcus model.

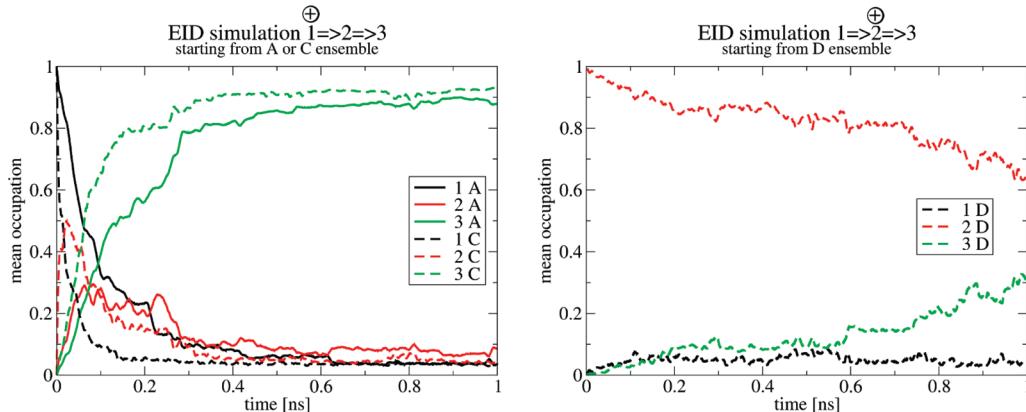
Two additional sets of simulations 1 $\rightarrow$ 2 $\rightarrow$ 3 emphasize this point further: In the first set, the simulations were started with

the hole charge placed on site 1, and the starting conformations were taken from the classical simulation C, where the environment has been equilibrated with a hole charge at site 1. The second set of simulations was started with the charge on site 2 and conformations from the classical simulation D, where the environment has been equilibrated with a hole charge at site 2. Both simulations begin with the charges in full equilibrium to their surroundings. Interestingly, the respective transfer rates are found to be slightly faster for the first simulation but significantly slower for the second (Figure 14). Apparently, a charge on site 1 is capable of quickly transferring along the Trp chain regardless of the starting conformation, while full equilibration of a charge on site 2 prevents fast CT. Therefore, we can conclude that CT on the order of 10–30 ps, as found experimentally, cannot involve full relaxation of the environment at intermediate steps. These findings are corroborated by additional simulations with a restricted QM system (see Supporting Information, Figures S7–S9).

**EID in Artificial Environments.** We have extended upon our EID direct dynamics simulation by applying a variety of more artificial models in order to better understand the details of the CT process. The simulations address the effect of restricted QM systems, constant couplings  $T_{ij}$  and only partial QM/MM coupling, and their results can be found in the Supporting Information.

## DISCUSSION

Our results render a novel picture of CT in DNA PL. QM/MM MD simulations with on-the-fly evaluation of CT parameters of unprecedented length allow us to identify subtle conformational changes in the protein that influence CT.



**Figure 14.** EIDs  $1 \rightarrow 2 \rightarrow 3$  starting from equilibrated systems. Left: CT initiated with hole charge on site 1 and starting conformation from simulations C and A, respectively (the latter for comparison, data taken from Figure 11b). Right: CT initiated with hole charge on site 2 and starting conformations taken from simulation D.

Deriving energetics parameters from a large conformational ensemble yields converged and statistically valid data. A description of the CT process both with Marcus theory using parameters from equilibrium simulation and with direct dynamics makes the comparison of both models possible. The good agreement of predicted rates from direct simulations with experimental results indicates that our choice of QM region, based on experimental data and theoretical predictions, is valid. In detail, the simulations presented here resolve the following points of interest:

**Protein Structure and Conformational Change.** The MD simulations reveal a protein structure that is overall quite stable over many nanoseconds, a feature not uncommon for ET proteins. The solvated protein was studied with different charge distribution models, corresponding to different substeps of the complete photoactivation process. Each of these charge distributions yields a stable protein structure and produces small but significant changes in the distance distribution between the Trp sites. In general, Trp residues are 7 to 8 Å apart, beyond van der Waals distance but close enough for ET to proceed. Especially interesting is the case of a charge localized on Trp306, which causes the protein to adopt a structure composed of at least two subconformations, which, while broadly similar, exhibit markedly different CT parameters.

Simulations of PL with different fixed charge distributions also shed light on the rearrangement of solvent around the charge-carrying amino acid side chains. Since the Trp residues 382, 359, and 306 are progressively closer to the protein surface and therefore to bulk water, it is not surprising to see a rising influence of solvent in that sequence, accompanied by a more pronounced formation of a solvent polaron around the positive hole charge. The mean ESP at each site, its breakdown into protein and solvent parts, and correlation coefficients produce a picture of the CT chain in which the first residue, Trp382, is located far away from solvent, next to the FAD cofactor, and well stabilized by the protein surrounding but only weakly affected by solvent. By contrast, the end state of the CT, Trp306, is mostly solvent exposed and stabilized, but the protein environment still plays an important role. This residue is the only one that reacts with a noticeable conformational change when charged, leading to an even larger solvent exposure. The central site, Trp 359, has properties in between those, and is strongly affected both by the protein environment and surrounding solvent. A complete CT event therefore

involves charge separation in the core of the protein, followed by a movement of a positive charge from the inner region where it is stabilized by the protein, toward the protein surface where energetic stabilization is provided by solvent. This CT process is strongly exothermic and leads to a stable final state of a Trp306 radical cation that can undergo further reaction, e.g., deprotonation by solvent. Our findings of energetically downhill transport are comparable to those obtained from electrostatics calculations.<sup>92</sup> Compared to these findings, our results show both substeps to be exothermic with a larger driving force ( $\Delta G$  overall  $-0.55$  eV), but a major difference between the studies is that full relaxation and dynamics of the protein were considered in our MD simulation.

**Structure, Fluctuation and Relaxation.** The relative importance of average structure versus dynamic disorder for CT in biomolecules has been discussed at length in the literature,<sup>65,73–75,101,110–112</sup> and PL remains an interesting addition to the array of test systems. The distribution of electronic couplings and corresponding coherence parameters obtained here (Figure 7, Table 4) show a CT that is fairly coherent, with  $C \approx 1$ . For step  $1 \rightarrow 2$ ,  $T_{ij}$  is normally distributed with a maximum of 5–10 meV, but the distribution maxima lie close to zero for  $2 \rightarrow 3$ . This makes the second CT step slightly less coherent and means that replacing mean-squared  $T_{ij}$  values with the square of the mean is at best permissible for the first ET substep. In simulations with couplings artificially held constant, we see that ET still occurs on a comparable time scale, but the kinetics of especially the second step change. Nevertheless, both couplings are non-zero in the majority of conformations, and the charge transport does not appear to be overly influenced by changes in  $T_{ij}$ . A gated-type mechanism in which only very rare conformational fluctuations produce high hopping probabilities can be ruled out. Instead, the CT is strongly influenced by the site energetics, emphasized by large  $\lambda$ -values and large negative  $\Delta G$ . This also makes the transport unidirectional, with very small rates for back-transfers, especially for the second substep.

A crucial role in understanding charge dynamics in PL is understanding the effect of environmental relaxation. We have shown in Figure 12 that typical residence times of the positive charge on site 2 are  $<100$  ps, but a full relaxation of the environment (mostly the protein, since solvent relaxes rapidly) occurs on a time scale longer than that. Coupled with the fact that charge delocalization prevents the complete charge from residing on site

2, this means that the full stabilization of the electron hole on site 2 as it occurs after multiple nanoseconds in simulation D (see Table 3) is unlikely to often occur in reality, even though there are simulations where the charge resides on site 2 for nanoseconds. This means that the two CT substeps cannot be treated independently; instead, a significant part of 2→3 transfers will occur from a nonequilibrium conformational ensemble that is closer to that of simulations B or C than D.

This also explains why very small Marcus rates are obtained for the second transfer substep. Marcus theory assumes a localized, fully equilibrated positive hole on site 2, which leads to dramatically smaller rates than for the first step. In reality, the surroundings do not have time to fully stabilize the hole since the relaxation time can be quite long, i.e., longer than the residence time of a hole on site 2. Therefore,  $\lambda + \Delta G$  is overestimated in the classical two-step Marcus picture. Moreover, slight delocalization reduces  $\lambda + \Delta G$  further, thus the energy barrier going from 2 to 3 decreases, and the actual rate can be considerably larger.

**Do We Need Fully Coupled QM/MM Simulations?** Early experiments showed that the ET between FAD and Trp306 occurs within 1  $\mu$ s or faster,<sup>113</sup> but recent experimental advances have produced more accurate rate estimates for the individual CT steps in PL. The first hopping step 1→2 occurs within 9 ps, and the 2→3 step follows within <30 ps.<sup>10,18,19,114</sup> This corresponds to rate constants of  $k_{1\rightarrow 2} = 110 \text{ ns}^{-1}$  and  $k_{2\rightarrow 3} = 33 \text{ ns}^{-1}$ . A comparison to our results shows that Marcus rates from Table 5 give a rate constant for the first step that is in excellent agreement with experiment if an appropriate scaling factor for  $\lambda$  is used. In contrast, the rate for the second step, even with scaling, is calculated to be lower by orders of magnitude. Marcus theory is applicable for the first, but apparently not for the second step of this hopping process, emphasizing again the importance of considering CT out of nonequilibrium conformational ensembles in the second step.

Apparently, while Marcus theory has undeniably been exceptionally successful in ET studies, there can be problems when it is applied to obtain rates from computationally determined parameters. CT in which nonequilibrium steps play a major role is incorrectly described. Depending on the scaling factor used, we obtain rates from 5 to 290 ns<sup>-1</sup> for  $k_{1\rightarrow 2}$ , spanning the range of the experimental value, but much smaller rates for  $k_{2\rightarrow 3}$ . Due to the exponential dependence of  $k$  on  $\lambda$  and  $\Delta G$  in eq 1, the rates are very sensitively dependent on the details of the energetical model, leading to large uncertainties.

ET rates obtained from fitting a kinetic model to the mean occupations from the EID simulations yield different rates, especially for the step 2→3. The most sophisticated model gives rates of  $k_{1\rightarrow 2} = 12.3 \text{ ns}^{-1}$  and  $k_{2\rightarrow 3} = 15.4 \text{ ns}^{-1}$ , plus a significant rate for 2→1 back-transfers (See the “Kinetic Models” Section in the Supporting Information for details of the fits and additional data). These rates are within an order of magnitude of experimental data, with no need for empirical scaling parameters. The EID dynamics with full coupling of electronic and nuclear degrees of freedom explicitly consider the relaxation properties and time scale of the environment and thereby describe both ET substeps correctly. We conclude that unlike classical Marcus theory, the EID model constitutes a superior protocol to simulate CT events in proteins, especially if the time scales of environmental relaxation and ET overlap.

## CONCLUSIONS

To summarize the above results and discussion, we present three main findings about CT in PL:

- Transfer along the Trp side chains is unidirectional along sites 1 → 2 → 3 with back transfer playing a minor role. The energetics are downhill and solvent driven. The solvent effectively soaks out the electron hole to the protein surface, and there is no need for preformed protein conformations to allow CT to occur. Nevertheless, the following reaction step is much slower than CT, and protein conformational change may play a role in stabilizing the charge once it is on site 3 long enough to allow deprotonation. The distinct protein structural change (Figure 3), which was observed on a much longer time scale than CT, is one possible mechanism for lowering the energy of site 3 so far that the charge is definitely trapped there.
- We find the formation of a polaron, i.e., strong stabilization of the electron hole on any site it is located on. Environmental fluctuations, both of solvent and protein, then strongly influence the site energies and make the dynamics fluctuation driven. In contrast, fluctuations of the electronic couplings  $T_{ij}$  are less important. For the transfer 1→2, the couplings can be replaced by their averages. For 2→3,  $T_{ij}$  fluctuations play a slightly larger role, but still do not influence CT very strongly. In the ongoing discussion about the relative importance of fluctuations of either the site energies or the electronic couplings in CT, this system clearly is an example controlled by energetic fluctuations.
- Energetic relaxation of the sites takes place on a time scale overlapping that of CT, similar to the case described in ref 76. Transfer out of an incompletely equilibrated conformational ensemble may be a reason why the transfer step between sites 2 and 3 is much faster than predicted by Marcus theory. Still, the majority of relaxation does occur faster than CT when judged by the energy gap between sites (see Figure 12). An additional phenomenon leading to faster transfers is that charge accumulation on sites 2 and 3 does not occur sequentially but concurrently (see occupation numbers in Figure 11). This indicates possible delocalization of the charge during transfer or the existence of CT trajectories in which the positive hole transfers directly on to site 3 without fully localizing on site 2. Neither incomplete relaxation nor delocalization are addressed in equilibrium CT theory, showing the need for direct dynamics simulations in which these effects are explicitly included.

## ASSOCIATED CONTENT

- S Supporting Information.** Includes details about atomic partial charges for the tryptophan residues, a comparison of molecular orbitals computed by DFTB, HF, and DFT, rmsd values, and additional structural analysis results for the protein MD simulations and solvent distribution. Furthermore, average ET parameters for the different simulations are available. Additional simulation results for ESP relaxation properties, EID dynamics in restricted QM systems of only two tryptophan residues and EID dynamics with fixed electronic couplings as well as with only partial QM/MM coupling are presented. We also give a comparison between CT in DNA and proteins based on previous simulation results and technical details about the kinetic models used to fit rate constants to average site occupation data.

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

**W Web Enhanced Feature.** An mpg-858 movie of a typical hole transfer trajectory.

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: marcus.elstner@kit.edu.

## ACKNOWLEDGMENT

The authors would like to thank Thorsten Koslowski for helpful discussion during the course of this project. Funding was provided by the DFG program “Quantum transport at the molecular scale”/EL 206/5-3. T.S. is indebted to the Baden-Württemberg Stiftung for the financial support of this research project by the Eliteprogramme for Postdocs.

## REFERENCES

- (1) Sancar, A. *Chem. Rev.* **2003**, *103*, 2203–2237.
- (2) Byrdin, M.; Sartor, V.; Eker, A.; Vos, M.; Auberta, C.; Brettel, K.; Mathisa, P. *Biochim. Biophys. Acta, Bioenergetics* **2004**, *1655*, 64–70.
- (3) Carell, T.; Burgdorf, L.; Kundu, L.; Cichon, M. *Curr. Opin. Chem. Biol.* **2001**, *5*, 491–498.
- (4) Park, H.; Kim, S.; Sancar, A.; Deisenhofer, J. *Science* **1995**, *268*, 1866–1872.
- (5) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33–38.
- (6) Kelner, A. *Proc. Natl. Acad. Sci. U.S.A.* **1949**, *35*, 73–79.
- (7) Sancar, A. *J. Biol. Chem.* **2008**, *283*, 32153–32157.
- (8) Moldt, J.; Pokorny, R.; Orth, C.; Linne, U.; Geisselbrecht, Y.; Marahiel, M.; Essen, L.; Batschauer, A. *J. Biol. Chem.* **2009**, *284*, 21670–21683.
- (9) Essen, L.; Klar, T. *Cell. Mol. Life Sci.* **2006**, *63*, 1266–1277.
- (10) Brettel, K.; Byrdin, M. *Curr. Opin. Struct. Biol.* **2010**, *20*, 693–701.
- (11) Kodali, G.; Siddiqui, S.; Stanley, R. *J. Am. Chem. Soc.* **2009**, *131*, 4795–4807.
- (12) Beukers, R.; Eker, A.; Lohman, P. *DNA Repair* **2008**, *7*, 530–543.
- (13) Lin, C.; Todo, T. *GenomeBiology* **2005**, *6*, 220.
- (14) Weber, S. *Biochim. Biophys. Acta, Bioenerg.* **2005**, *1707*, 1–23.
- (15) Giovani, B.; Byrdin, M.; Ahmad, M.; Brettel, K. *Nat. Struct. Biol.* **2003**, *10*, 489–490.
- (16) Biskup, T.; Schleicher, E.; Okafuji, A.; Link, G.; Hitomi, K.; Getzoff, E.; Weber, S. *Angew. Chem., Int. Ed.* **2009**, *48*, 404–407.
- (17) Byrdin, M.; Eker, A.; Vos, M.; Brettel, K. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 8676–8681.
- (18) Aubert, C.; Vos, M.; Mathis, P.; Eker, A.; Brettel, K. *Nature* **2000**, *405*, 586–590.
- (19) Lukacs, A.; Eker, A.; Byrdin, M.; Brettel, K.; Vos, M. *J. Am. Chem. Soc.* **2008**, *130*, 14394–14395.
- (20) Harrison, C.; O’Neil, L.; Wiest, O. *J. Phys. Chem. A* **2005**, *109*, 7001–7012.
- (21) Chatgilialoglu, C.; Guerra, M.; Kaloudis, P.; Houee-Levin, C.; Marignier, J.; Swaminathan, V.; Carell, T. *Chemistry* **2007**, *13*, 8979–8984.
- (22) Tachikawa, H.; Kawabata, H. *J. Phys. Chem. B* **2008**, *112*, 7315–7319.
- (23) Prytkova, T.; Beratan, D.; Skourtis, S. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 802–807.
- (24) Acocella, A.; Jones, G.; Zerbetto, F. *J. Phys. Chem. B* **2010**, *114*, 4101–4106.
- (25) Sadeghian, K.; Bocola, M.; Merz, T.; Schütz, M. *J. Am. Chem. Soc.* **2010**, *132*, 16285–16295.
- (26) Zheng, X.; Garcia, J.; Stuchebrukhov, A. *J. Phys. Chem. B* **2008**, *112*, 8724–8729.
- (27) Zheng, X.; Ly, N.; Stuchebrukhov, A. *Int. J. Quantum Chem.* **2007**, *107*, 3126–3131.
- (28) Medvedev, D.; Stuchebrukhov, A. *J. Theor. Biol.* **2001**, *210*, 237–248.
- (29) Antony, J.; Medvedev, D.; Stuchebrukhov, A. *J. Am. Chem. Soc.* **2000**, *122*, 1057–1065.
- (30) Cheung, M.; Daizadeh, I.; Stuchebrukhov, A.; Heelis, P. *Biophys. J.* **1999**, *76*, 1241–1249.
- (31) Krapf, S.; Koslowski, T.; Steinbrecher, T. *Phys. Chem. Chem. Phys.* **2010**, *12*, 9516–9525.
- (32) Masson, F.; Laino, T.; Röthlisberger, U.; Hutter, J. *Chem. Phys. Chem.* **2009**, *10*, 400–410.
- (33) Warshel, A. *Acc. Chem. Res.* **2002**, *35*, 385–395.
- (34) Hervas, M.; Navarro, J.; Rosa, M. D. L. *Acc. Chem. Res.* **2003**, *36*, 798–805.
- (35) Schultz, B.; Chan, S. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 23–65.
- (36) Gray, H.; Winkler, J. *Annu. Rev. Biochem.* **1996**, *65*, 537–561.
- (37) Marcus, R. *Rev. Mod. Phys.* **1993**, *65*, 599–610.
- (38) Marcus, R. *J. Chem. Phys.* **1956**, *24*, 966–978.
- (39) Marcus, R. *J. Chem. Phys.* **1956**, *24*, 979–989.
- (40) Marcus, R. *Annu. Rev. Phys. Chem.* **1964**, *15*, 155–196.
- (41) Marcus, R.; Sutin, N. *Biochim. Biophys. Acta* **1985**, *811*, 265–322.
- (42) Levich, V.; Dogonadze, R. *Dokl. Akad. Nauk SSSR* **1959**, *124*, 123–126.
- (43) Hush, N. *Trans. Faraday Soc.* **1961**, *57*, 557–580.
- (44) Hopfield, J. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 3640–3644.
- (45) Jortner, J. *J. Chem. Phys.* **1976**, *64*, 4860–4867.
- (46) Moser, C.; Keske, J.; Warncke, K.; Farid, R.; Dutton, P. *Nature* **1992**, *355*, 796–802.
- (47) Warshel, A. *J. Phys. Chem.* **1982**, *86*, 2218–2224.
- (48) Hwang, J.; Warshel, A. *J. Am. Chem. Soc.* **1987**, *109*, 715–720.
- (49) Alden, R.; Parson, W.; Chu, Z.; Warshel, A. *J. Am. Chem. Soc.* **1995**, *117*, 12284–12298.
- (50) Parson, W.; Chu, Z.; Warshel, A. *Biophys. J.* **1998**, *74*, 182–191.
- (51) Warshel, A.; Chu, Z.; Parson, W. *J. Photochem. Photobiol., A* **1994**, *82*, 123–128.
- (52) Nonella, M.; Schulten, K. *J. Phys. Chem.* **1991**, *95*, 2059–2067.
- (53) Schulten, K.; Tesch, M. *Chem. Phys.* **1991**, *158*, 421–446.
- (54) VandeVondele, J.; Sulpizi, M.; Sprak, M. *Chimia* **2007**, *61*, 155–158.
- (55) Blumberger, J. *Phys. Chem. Chem. Phys.* **2008**, *10*, 5651–5667.
- (56) Kubář, T.; Elstner, M. *J. Phys. Chem. B* **2009**, *113*, 5653–5656.
- (57) Koslowski, T.; Steinbrecher, T. *Z. Phys. Chem.* **2009**, *223*, 739–752.
- (58) King, G.; Warshel, A. *J. Chem. Phys.* **1990**, *93*, 8682–8692.
- (59) Ceccarelli, M.; Marchi, M. *J. Phys. Chem. B* **2003**, *107*, 5630–5641.
- (60) Blumberger, J.; Lamoureux, G. *Mol. Phys.* **2008**, *106*, 1597–1611.
- (61) Tipmanee, V.; Oberhofer, H.; Park, M.; Kim, K. S.; Blumberger, J. *J. Am. Chem. Soc.* **2010**, *132*, 17032–17040.
- (62) Blumberger, J.; Sprak, M. *Theor. Chem. Acc.* **2006**, *115*, 113–126.
- (63) Zeng, X.; Hu, H.; Hu, X.; Cohen, A. J.; Yang, W. *J. Chem. Phys.* **2008**, *128*, 124510.
- (64) Olsson, M.; Hong, G.; Warshel, A. *J. Am. Chem. Soc.* **2003**, *125*, 5025–5039.
- (65) Cascella, M.; Magistrato, A.; Tavernelli, I.; Carloni, P.; Rothlisberger, U. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 19641–19646.
- (66) Oberhofer, H.; Blumberger, J. *Angew. Chem., Int. Ed.* **2010**, *49*, 3631–3634.
- (67) Newton, M. *Chem. Rev.* **1991**, *91*, 767–792.
- (68) Priyadarshy, S.; Skourtis, S.; Risser, S.; Beratan, D. *J. Chem. Phys.* **1996**, *104*, 9473–9481.
- (69) Stuchebrukhov, A. *A. Theor. Chem. Acc.* **2003**, *110*, 291–306.
- (70) Stuchebrukhov, A. *Adv. Chem. Phys.* **2001**, *118*, 1–44.
- (71) Stuchebrukhov, A. *J. Chem. Phys.* **1996**, *105*, 10819–10829.
- (72) Stuchebrukhov, A. *J. Chem. Phys.* **1996**, *104*, 8424–8432.
- (73) Hatcher, E.; Balaeff, A.; Keinan, S.; Venkatramani, R.; Beratan, D. N. *J. Am. Chem. Soc.* **2008**, *130*, 11752–11761.
- (74) Beratan, D. N.; Skourtis, S. S.; Balabin, I. A.; Balaeff, A.; Keinan, S.; Venkatramani, R.; Xiao, D. *Acc. Chem. Res.* **2009**, *42*, 1669–1678.

- (75) Skourtis, S. S.; Waldeck, D. H.; Beratan, D. N. *Annu. Rev. Phys. Chem.* **2010**, *61*, 461–485.
- (76) Wang, H.; Lin, S.; Allen, J. P.; Williams, J. C.; Blankert, S.; Laser, C.; Woodbury, N. W. *Science* **2007**, *316*, 747–750.
- (77) Skourtis, S. S.; Beratan, D. N. *Science* **2007**, *316*, 703–704.
- (78) Warshel, A.; Hwang, J. *J. Chem. Phys.* **1986**, *84*, 4938–4957.
- (79) Warshel, A.; Parson, W. *Q. Rev. Biophys.* **2001**, *34*, 563–679.
- (80) Kubář, T.; Elstner, M. *J. Phys. Chem. B* **2010**, *114*, 11221–11240.
- (81) Grozema, F. C.; Tonzani, S.; Berlin, Y. A.; Schatz, G. C.; Siebbeles, L. D. A.; Ratner, M. A. *J. Am. Chem. Soc.* **2008**, *130*, 5157–5166.
- (82) Steinbrecher, T.; Koslowski, T.; Case, D. *J. Phys. Chem. B* **2008**, *112*, 16935–16944.
- (83) Kubář, T.; Kleinekathofer, U.; Elstner, M. *J. Phys. Chem. B* **2009**, *113*, 13107–13117.
- (84) Meagher, K.; Redman, L.; Carlson, H. *J. Comput. Chem.* **2003**, *24*, 1016–1025.
- (85) Wang, J.; Wolf, R.; Caldwell, J.; Kollman, P.; Case, D. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (86) Bayly, C.; Cieplak, P.; Cornell, W.; Kollman, P. *J. Phys. Chem.* **1993**, *97*, 10269–10280.
- (87) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
- (88) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701.
- (89) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049.
- (90) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct., Genet.* **2006**, *65*, 712–725.
- (91) Hess, B.; Bekker, H.; Berendsen, H.; Fraaije, J. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
- (92) Popovic, D.; Zmiric, A.; Zaric, S.; Knapp, E. *J. Am. Chem. Soc.* **2002**, *124*, 3775–3782.
- (93) Kirkwood, J. *J. Chem. Phys.* **1935**, *3*, 300–313.
- (94) Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*; Oxford University Press: New York/Oxford, 1989.
- (95) Kubář, T.; Woiczikowski, P. B.; Cuniberti, G.; Elstner, M. *J. Phys. Chem. B* **2008**, *112*, 7937–7947.
- (96) Woiczikowski, P. B.; Kubář, T.; Gutiérrez, R.; Cuniberti, G.; Elstner, M. *J. Chem. Phys.* **2010**, *133*, 035103.
- (97) Kubář, T.; Elstner, M. *J. Phys. Chem. B* **2008**, *112*, 8788–8798.
- (98) Conwell, E. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 8795–8799.
- (99) Félix, M.; Voityuk, A. A. *Int. J. Quantum Chem.* **2011**, *111*, 191–201.
- (100) Voityuk, A. A.; Siriwong, K.; Rösch, N. *Angew. Chem., Int. Ed.* **2004**, *43*, 624–627.
- (101) Balabin, I. A.; Beratan, D. N.; Skourtis, S. S. *Phys. Rev. Lett.* **2008**, *101*, 158102.
- (102) Woiczikowski, P. B.; Kubář, T.; Gutiérrez, R.; Caetano, R. A.; Cuniberti, G.; Elstner, M. *J. Chem. Phys.* **2009**, *130*, 215104.
- (103) Gutiérrez, R.; Caetano, R. A.; Woiczikowski, B. P.; Kubář, T.; Elstner, M.; Cuniberti, G. *Phys. Rev. Lett.* **2009**, *102*, 208102.
- (104) Sulpizi, M.; Raugei, S.; VandeVondele, J.; Carloni, P.; Sprik, M. *J. Phys. Chem. B* **2007**, *111*, 3969–3976.
- (105) Marchi, M.; Gehlen, J.; Chandler, D.; Newton, M. *J. Am. Chem. Soc.* **1993**, *115*, 4178–4190.
- (106) Kuharski, R.; Bader, J.; Chandler, D.; Sprik, M.; Klein, M.; Impey, R. *J. Chem. Phys.* **1988**, *89*, 3248–3257.
- (107) Ando, K. *J. Chem. Phys.* **2001**, *114*, 9470–9477.
- (108) Tachiya, M. *J. Phys. Chem.* **1993**, *97*, 5911–5916.
- (109) Tachiya, M. *J. Chem. Phys.* **2008**, *129*, 066102.
- (110) Balabin, I.; Onuchic, J. *Science* **2000**, *290*, 114–117.
- (111) Sterpone, F.; Ceccarelli, M.; Marchi, M. *J. Phys. Chem. B* **2003**, *107*, 11208–11215.
- (112) Blumberger, J.; Klein, M. *J. Am. Chem. Soc.* **2006**, *128*, 13854–13867.
- (113) Li, Y.; Heelis, P.; Sancar, A. *Biochemistry* **1991**, *30*, 6322–6329.
- (114) Byrdin, M.; Lukacs, A.; Thiagarajan, V.; Eker, A.; Brettel, K.; Vos, M. *J. Phys. Chem. A* **2010**, *114*, 3207–3214.