

# Using Constrained Density Functional Theory to Track Proton Transfers and to Sample Their Associated Free Energy Surface

Chenghan Li and Gregory A. Voth\*



Cite This: *J. Chem. Theory Comput.* 2021, 17, 5759–5765



Read Online

ACCESS |



Metrics & More

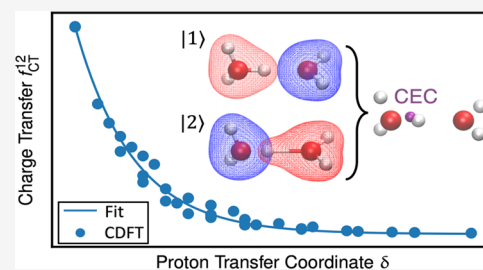


Article Recommendations



Supporting Information

**ABSTRACT:** Ab initio molecular dynamics (AIMD) and quantum mechanics/molecular mechanics (QM/MM) methods are powerful tools for studying proton solvation, transfer, and transport processes in various environments. However, due to the high computational cost of such methods, achieving sufficient sampling of rare events involving excess proton motion—especially when Grotthuss proton shuttling is involved—usually requires enhanced free energy sampling methods to obtain informative results. Moreover, an appropriate collective variable (CV) that describes the effective position of the net positive charge defect associated with an excess proton is essential both for tracking the trajectory of the defect and for the free energy sampling of the processes associated with the resulting proton transfer and transport. In this work, such a CV is derived from first principles using constrained density functional theory (CDFT). This CV is applicable to a broad array of proton transport and transfer processes as studied via AIMD and QM/MM simulations.



## INTRODUCTION

The accurate and efficient delineation of proton transport (PT) and its associated mechanism continues to be fundamentally important in chemistry, biology, and materials science.<sup>1–3</sup> Excess proton transport in aqueous and biomolecular environments involves rearranging covalent and hydrogen bonds, which is known as the Grotthuss hopping or shuttle mechanism.<sup>4,5</sup> Due to this chemically reactive nature of the process, the ab initio molecular dynamics (AIMD) method,<sup>6,7</sup> which treats the electronic degrees of freedom explicitly and “on the fly” along with the dynamics of the nuclei, provides one popular approach for modeling PT at an atomistic level. Among the various possible electronic structure methods, density functional theory (DFT) represents a powerful approach for implementing AIMD, as DFT has a reasonable balance between accuracy and computational efficiency.

However, even when using the generalized gradient approximation (GGA) level of DFT,<sup>8</sup> the high computational cost of AIMD typically limits the sampling of the MD to within the subnanosecond timescale for systems containing several hundreds of electrons. This limitation can prevent an adequate sampling of rare events, such as PT involving weak acids that are commonly found as protonatable amino acids in protein channels and transporters. In such cases, the high reaction barrier for proton dissociation from weak acids results in timescales that usually exceed nanoseconds; thus, enhanced free energy sampling methods to bias the PT process are necessary to obtain statistically and physically meaningful results.

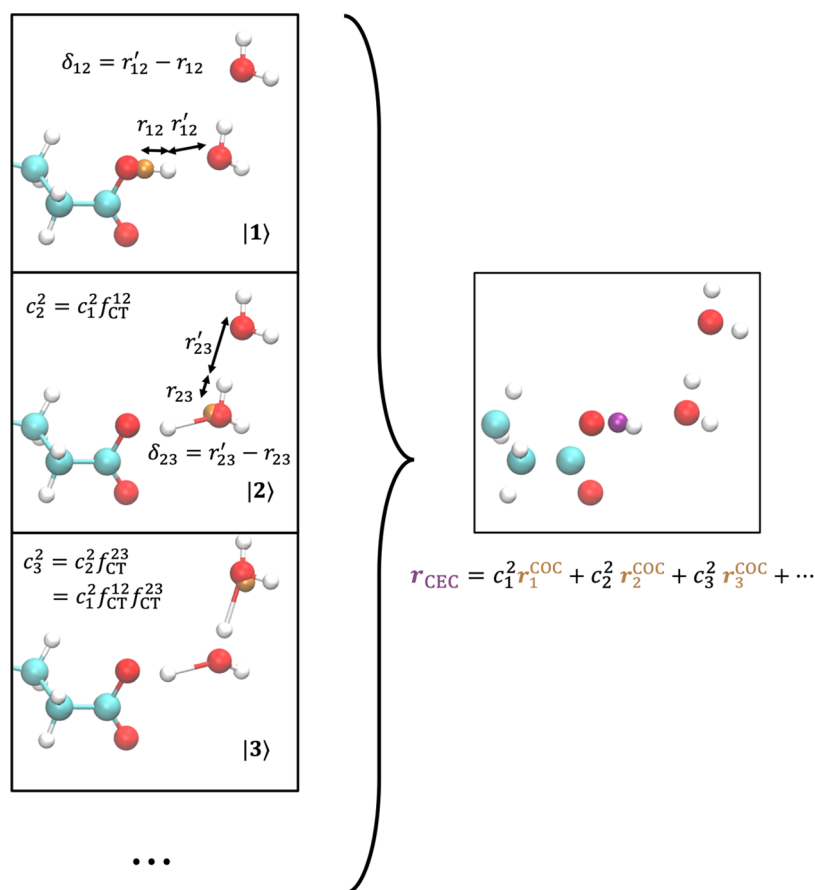
Common enhanced sampling methods, such as umbrella sampling,<sup>9</sup> metadynamics,<sup>10</sup> and adaptive biasing force,<sup>11</sup> add

bias to one or more collective variables (CVs) to accelerate the sampling along these degrees of freedom. In the context of PT, identifying an appropriate CV that represents the position of the net positive charge defect associated with the excess proton is important. Due to frequent bond breaking and forming events in Grotthuss proton shuttling, the identity of the charge carrier species (hydronium-like or protonated weak acids) is dynamically changing, and the excess protonic charge defect tends to be distributed among several solvation shells instead of localizing on a central hydronium structure or on a weak acid. As such, a CV cannot be associated with any specific “proton” in the system but is more appropriately assigned in some way to be the charge defect associated with the excess proton, often referred to as the “center of excess charge” (CEC). For AIMD simulations, there are several CEC definitions that have been proposed, namely, mCEC,<sup>12</sup> the proton indicator,<sup>13</sup> and the more recent rCEC.<sup>14</sup> However, a more rigorous definition based on ab initio theory is preferred. In this work, we present a variant of the CEC definition derived from a diabatic electronic structure method, the constrained DFT (CDFT),<sup>15</sup> and apply it to two case studies: an excess proton in water and glutamic acid in water. We conclude by unraveling the collective motions encoded in the CEC via computing its IR spectrum as well as illustrating its

Received: June 20, 2021

Published: September 1, 2021





**Figure 1.** Illustration of CEC calculation for Glu in water. For the sake of clarity, only three diabatic states are shown, even though 20–30 states (on average) will be resolved in the condensed phase by searching up to three solvation shells of the hydrated excess proton. The COC in each state is rendered by an orange sphere. The resulting CEC as a linear combination of COCs is rendered in purple in the right panel. The  $f_{\text{CT}}^{12}$  and  $f_{\text{CT}}^{23}$  are computed using eq 12 as a function of  $\delta_{12}$  and  $\delta_{23}$ , respectively.

ability to accelerate the sampling of proton transfer when combined with metadynamics.

### THEORY OF CONSTRAINED DFT

The CDFT framework was proposed for solving the electronic structure of a system subject to the following constraint on electron density

$$\int w(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} = N \quad (1)$$

where  $w(\mathbf{r})$  is the weighting function that defines the constraint and  $N$  is the constraint target value. The constrained lowest-energy state can be obtained from an optimization problem via the standard method of Lagrange multipliers

$$E(N) = \min_{\rho} \max_{\lambda} \left( E[\rho(\mathbf{r})] + \lambda \left( \int w(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} - N \right) \right) \quad (2)$$

Herein,  $E[\rho(\mathbf{r})]$  is the density functional, which in this work is BLYP<sup>16,17</sup> and  $\omega$ B97X,<sup>18</sup> and  $\lambda$  is the Lagrange multiplier. The electron density determined from eq 2 thus deviates from the constraint-free adiabatic ground-state density, making it a so-called diabatic state. As molecules in the condensed phase sample more compact geometries on average, the promolecule formalism approach<sup>15,19</sup> was employed and the system was partitioned into two molecular fragments, A and B. The constraint target value in eq 2 was then calculated from the

total promolecule density by summing up the ground-state electronic density of the two fragments as if they were independent

$$N \equiv \int w(\mathbf{r})(\rho_A(\mathbf{r}) + \rho_B(\mathbf{r}))d\mathbf{r} \quad (3)$$

Here, the Becke population<sup>20</sup> scheme was used to define the weighting function as

$$w(\mathbf{r}) = \sum_I w_I^{\text{Becke}}(\mathbf{r}) \quad (4)$$

where the summation index  $I$  refers to the atoms of protonated species in each diabatic state, i.e., a hydronium or a neutral glutamate. The expected behavior of the promolecule constraint is that the resulting diabatic electronic density will resemble as much as possible the superposition of two pure fragments, such as, e.g., a pure water and a pure hydronium in the case of the Zundel cation  $\text{H}_5\text{O}_2^+$ .

The coupling between two diabatic states is calculated from the integral using the Kohn–Sham determinant  $|\Phi_i\rangle$ <sup>15,21</sup>

$$H_{12} = \langle \Phi_1 | \hat{H} | \Phi_2 \rangle \quad (5)$$

A  $2 \times 2$  Hamiltonian can be constructed using the diabatic energies from eq 2 as diagonal terms

$$\mathbf{H} = \begin{pmatrix} E_1 & H_{12} \\ H_{21} & E_2 \end{pmatrix} \quad (6)$$

Similarly, the overlap matrix is defined as

$$\mathbf{S} = \begin{pmatrix} 1 & \langle \Phi_1 | \Phi_2 \rangle \\ \langle \Phi_2 | \Phi_1 \rangle & 1 \end{pmatrix} \quad (7)$$

Then, the so-called CDFT configurational interaction (CI)<sup>15,19</sup> can be performed by solving the generalized eigenvalue problem

$$\mathbf{H}\mathbf{c} = \mathbf{E}\mathbf{S}\mathbf{c} \quad (8)$$

The resulting eigenfunctions,  $\mathbf{c} = \{c_1, c_2\}$ , determine the degree to which each of the two diabatic states contributes to the CI ground state of the system; we will use the  $\mathbf{c}$  vector to define our CEC, as described in the next section.

## THEORY OF CENTER OF EXCESS CHARGE

Assuming that each diabatic state defines a bonding topology, e.g., in Figure 1, state  $|1\rangle$  defines a neutral glutamate and two neutral water molecules, while state  $|2\rangle$  defines a hydronium, a charged Glu, and a neutral water molecule as two of the possible topologies. Given the bonding topology of every diabatic state, the “diabatic” CEC within state  $|i\rangle$  is simply the center of charge (COC) of the species that carries protonic charges, i.e., the hydronium or protonated weak acid in  $|i\rangle$  in Figure 1, such that

$$\mathbf{r}_i^{\text{COC}} = \sum_I q_I^i \mathbf{r}_I \quad (9)$$

Here, we assume that the diabatic excess charges are associated with atomic positions  $\mathbf{r}_I$  and are modeled by fixed charges  $q_I^i$  (the charge of atom  $I$  in state  $|i\rangle$ ). The fixed charge values are taken in this example from the CHARMM 36 force field<sup>22</sup> and a prior paper<sup>23</sup> and are also summarized in Table S1. In diabatic states with well-defined bonding topologies, the fixed force field charges are considered to be a reasonable description of the system, while the excess charge delocalization and the polarization due to the excess proton are characterized by the  $\mathbf{c}$  vector. Since the coefficient vector  $\mathbf{c}$  obtained from eq 8 represents the population of each diabatic state in the final CI ground state, the “adiabatic” CEC is naturally defined as the weighted average of each diabatic CEC (i.e., COC)

$$\mathbf{r}_{\text{CEC}} = \sum_i c_i^2 \mathbf{r}_i^{\text{COC}} \quad (10)$$

In this sense, the  $c_i$ 's measure the extent of excess charge transfer. Accordingly, we define the charge-transfer factor in what follows to represent the excess charge distribution between state  $|i\rangle$  and state  $|j\rangle$

$$f_{\text{CT}}^{ij} = \frac{c_j^2}{c_i^2} \quad (11)$$

Due to the extended searching space introduced by the multiplier  $\lambda$ , the CDFT calculation (eq 2) is typically more expensive in comparison to the adiabatic electronic structure method used in AIMD. Therefore, we adopted an approximation of the ground-state vector  $\mathbf{c}$  to compute the CEC on the fly in the AIMD simulations. It was found that an exponential function of proton transfer coordinate  $\delta$  can provide a good fit for the charge-transfer factor

$$f_{\text{CT}}^{ij} \approx \exp(-k(\delta_{ij} + \delta_0)) \quad (12)$$

The  $\delta_{ij}$  is defined here as the difference between two O–H distances

$$\delta_{ij} = r'_{ij} - r_{ij} \quad (13)$$

where  $r_{ij}$  denotes the distance between the shared proton and the proton donor oxygen in state  $|i\rangle$  and  $r'_{ij}$  denotes the distance between the proton and the proton acceptor in state  $|j\rangle$ . The parameters  $k$  and  $\delta_0$  were calibrated to match the exact  $f_{\text{CT}}^{ij}$  from CDFT-CI calculations between protonated species and water in the gas phase using BLYP or  $\omega$ B97X functional, in this case by a least-squares fitting. The training configurations used are described in the Simulation Details section and provided as Supporting Information. The list of fitting parameters for the CEC is provided in Table 1. Further details pertinent to the parametrization procedure can be found in the Simulation Details section.

**Table 1. Fitted Parameters of the CEC for Water and Glutamic Acid**

	CDFT functional	$k$ ( $\text{\AA}^{-1}$ )	$\delta_0$ ( $\text{\AA}$ )
$\text{H}_3\text{O}^+ - \text{H}_2\text{O}$	BLYP	4.234	0
	$\omega$ B97X	4.898	0
$\text{Glu} - \text{H}_2\text{O}$	BLYP	2.946	0.5361

To generalize the CEC to the condensed phase environment, we assume that the solvating waters of the hydronium or protonated acid propagate the excess charge to further solvation shells following the same exponential rule (eq 12) as illustrated in Figure 1. The diabatic states for the CEC calculation were selected by searching three solvation shells of hydrogen bond acceptors using a 2.5  $\text{\AA}$  criterion for the O–H distance, as described in more detail in ref 24. The further solvation shells were found to be not needed because the computed  $c_i$  values already diminished for the fourth shell. After resolving all of the charge-transfer factors between each proton donor–acceptor pair, the approximated  $c_i^2$  was then computed from  $f_{\text{CT}}^{ij}$  by applying the normalization condition ( $\sum_i c_i^2 = 1$ )

$$c_1^2 = \frac{1}{1/c_1^2} = \frac{1}{\sum_i c_i^2/c_1^2} = \frac{1}{1 + \sum_{i \neq 1} f_{\text{CT}}^{1i}} \quad (14a)$$

$$c_i^2 = c_1^2 f_{\text{CT}}^{1i} \quad (14b)$$

The  $c_1^2$  was computed directly from the charge-transfer factors (third equal sign in eq 14a), while other  $c_i$ 's were available via eq 14b. In summary, eqs 9 and 10 define the CEC, while eqs 12, 14a, and 14b yield an approximation to the exact CDFT-CI  $\mathbf{c}$ , thereby facilitating CEC calculations at a reasonable computational cost.

Another way of viewing the CEC is the dipole moment of excess charges. Following eqs 9 and 10, we have

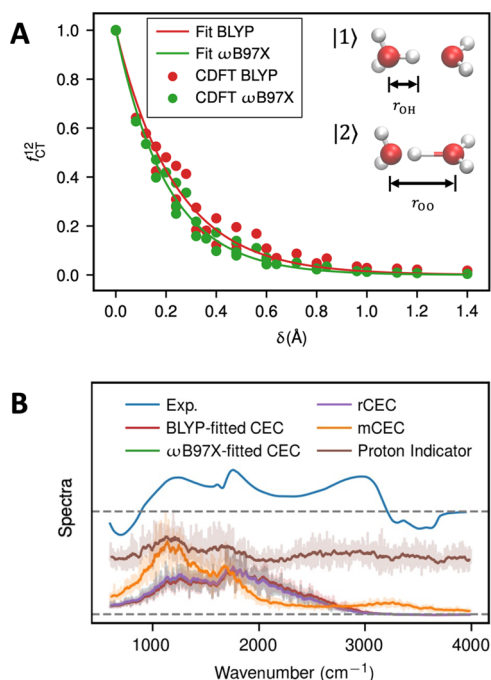
$$\mathbf{r}_{\text{CEC}} = \sum_i c_i^2 \sum_I q_I^i \mathbf{r}_I = \sum_I \left( \sum_i c_i^2 q_I^i \right) \mathbf{r}_I = \sum_I q_I^{\text{ex}} \mathbf{r}_I = \boldsymbol{\mu}^{\text{ex}} \quad (15)$$

where we define the excess charge of atom  $I$  as its weighted average charge  $\sum_i c_i^2 q_I^i$ . Hence, the excess charge contribution to the IR spectrum can be calculated directly from the CEC velocity correlation function

$$\begin{aligned}
 A(\omega) &\propto \int \langle \dot{\mu}^{\text{ex}}(0) \dot{\mu}^{\text{ex}}(t) \rangle e^{-i\omega t} dt \\
 &= \int \langle \dot{r}_{\text{CEC}}(0) \dot{r}_{\text{CEC}}(t) \rangle e^{-i\omega t} dt
 \end{aligned}
 \quad (16)$$

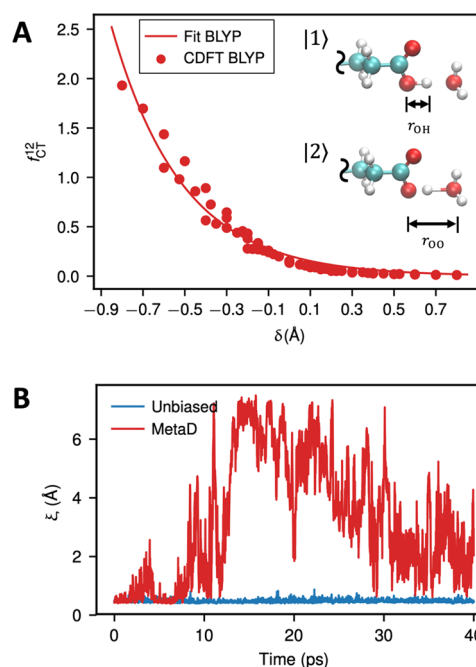
## SIMULATION DETAILS

The CDFT calculations were conducted for hydronium–water (Figure 2A) and glutamate–water (Figure 3A); the electronic



**Figure 2.** (A) Charge-transfer factor between hydronium and water calculated by CDFT with BLYP and  $\omega$ B97X functionals and fitted curves. (B) Calculated AIMD excess charge spectrum of a hydrated excess proton in water using various CEC definitions. The experimental IR spectrum is the acid solution spectrum subtracted by the pure water spectrum, taken from ref 40. The computed CEC spectra are shown in transparent, while the running averages performed using a  $33 \text{ cm}^{-1}$  window are shown in solid colors. The proton indicator spectrum intensity was scaled by 0.5 for better presentation. Note that the  $\omega$ B97X-fitted CEC (green) overlaps with the rCEC curve (purple), which very nearly overlaps with the BLYP-fitted CEC (red).

structure settings were identical to those used for AIMD simulations, as shown below. Both the  $\omega$ B97X and the BLYP functionals were used for hydronium–water CDFT calculations, while the latter was adopted for glutamate–water. For hydronium–water, the total promolecule density of states |1> and |2> was calculated by adding the ground-state density of the hydronium defined in that state plus the water density, i.e., the hydronium and water were chosen to be the fragments A and B, respectively, in eq 3. For Glu–water, the promolecule density of state |1> is the sum of neutral Glu density plus the neutral water density, and for state |2>, it is the sum of deprotonated Glu plus the hydronium. Established literature values<sup>25</sup> for the atomic radii, including  $0.75 \text{ \AA}$  for carbon,  $0.32 \text{ \AA}$  for hydrogen,  $0.63 \text{ \AA}$  for oxygen, and  $0.71 \text{ \AA}$  for nitrogen, were used for calculating the Becke population. The fitting procedure involving the hydronium–water molecular pair was based on a training set comprised of a series of fixed values for



**Figure 3.** (A) Charge-transfer factor between glutamate and water calculated by CDFT with the BLYP functional and its fitted curve. (B) Time series of the proton disassociation CV  $\xi$  in an unbiased AIMD (blue) and a metadynamics run (red) of glutamate solution.

the oxygen–oxygen distance,  $r_{\text{OO}}$ , including 2.2, 2.4, 2.6, 2.8, 3.0, and 3.2  $\text{\AA}$ . For each value of  $r_{\text{OO}}$  in this set, 6 shared proton positions were sampled evenly from  $r_{\text{OH}} = 0.9 \text{ \AA}$  to  $r_{\text{OH}} = r_{\text{OO}}/2$ , resulting in  $6 \times 6 = 36$  data points. The training set for glutamate–water consists of 7 oxygen–oxygen distances evenly distributed from 2.2 to 2.8  $\text{\AA}$  and 9 oxygen–hydrogen distance values ranging from 1.0  $\text{\AA}$  to  $r_{\text{OO}} - 1.0 \text{ \AA}$ , resulting in  $7 \times 9 = 63$  data points. The CDFT calculations were performed by the CDFT implementation<sup>26</sup> in CP2K<sup>27,28</sup> combined with Libxc.<sup>29</sup> In CDFT calculations, the setup for the DFT part follows the same as for the AIMD described below. The convergence criterion for the CDFT was chosen to be  $10^{-3}$  au.

The AIMD simulation of the excess proton in water was performed for 128 water molecules and 1 excess proton in a  $15.64 \text{ \AA} \times 15.64 \text{ \AA} \times 15.64 \text{ \AA}$  box. The electronic structure was described by the BLYP-D3 density functional<sup>30</sup> with Goedecker–Teter–Hutter pseudopotentials.<sup>31</sup> The Gaussian and plane waves (GPW) method<sup>32</sup> was used, the Kohn–Sham orbitals were expanded using the Gaussian basis set TZV2P, and the electronic density was expanded in plane waves with a cutoff of 400 Ry. The orbital transformation (OT) method<sup>33,34</sup> with the direct inversion in the iterative subspace (DIIS) minimizer was used as the self-consistent field (SCF) method with a convergence criterion of  $10^{-6}$  au. An experiment directed simulation (EDS) correction<sup>35</sup> was also employed as a minimal add-on bias to correct the overly strong hydrogen bonding in most DFT functionals. It was found that excess proton and water diffusion better match experimental values after including the EDS correction. This EDS method for excess protons in water, which follows earlier work for pure water,<sup>36</sup> is briefly summarized in the Supporting Information, and more details will be published in the future.

The system was first equilibrated in the constant NVT ensemble at 298 K for 30 ps, and then it was switched to the constant NVE ensemble for 200 ps for collecting non-

thermostatted dynamical data. A timestep of 0.5 fs was used to integrate the system MD. All of the AIMD simulations were carried out with the CP2K program package, coupled with a modified version of PLUMED2<sup>37</sup> for the EDS correction. Both the BLYP-fitted and the  $\omega$ B97X-fitted CEC parameters were employed in the analysis of AIMD of proton in water.

The Glu–water system consisted of 1 neutral glutamate with 110 water molecules in a  $16 \text{ \AA} \times 16 \text{ \AA} \times 16 \text{ \AA}$  box. We set up the electronic structure calculation similar to that used for the proton–water system, except that no EDS correction was used. The unbiased and well-tempered metadynamics<sup>38</sup> simulations were conducted with the constant NVT ensemble at 300 K in CP2K with PLUMED2 for computing and biasing the CEC. To be consistent with the underlying density functional used to perform AIMD simulations, the BLYP-fitted hydronium–water and glutamate–water CEC parameters were used. The collective variable used in metadynamics was the minimum distance between the CEC and the two Glu carboxyl oxygens:

$$\xi = \text{softmin}(|r_{\text{CEC}} - r_{\text{OE1}}|, |r_{\text{CEC}} - r_{\text{OE2}}|) \quad (17)$$

where softmin is a smooth version of minimum function

$$\text{softmin}(r_1, r_2) = -\frac{1}{\kappa} \ln(e^{-\kappa(r_1-r_2)/2} + e^{-\kappa(r_2-r_1)/2}) + \frac{r_1 + r_2}{2} \quad (18)$$

where  $\kappa = 40 \text{ \AA}^{-1}$ . The Gaussians in metadynamics were deposited every 50 fs with an initial height of 0.2 kcal/mol and a width of 0.1  $\text{\AA}$ . The bias factor of  $\gamma = 12$  was used to account for a roughly 9 kcal/mol proton dissociation barrier.

## RESULTS

The CDFT-calculated charge-transfer factors  $f_{\text{CT}}^{12}$  and the fitted curves using eq 12 are shown in Figure 2A. The  $\omega$ B97X-based value of the charge-transfer factor was found to decay more quickly compared to that calculated using the BLYP functional. This finding is not surprising because the range-separated hybrid functional  $\omega$ B97X produces less charge delocalization in comparison to the GGA BLYP functional. However, for both cases, the exponential function (eq 12) provides a good fit. Given that  $c_1^2 = c_2^2$  when the proton is equally shared between two water molecules, the  $\delta_0$  parameter was set to be zero. Interestingly, the fitted parameter  $k$  from  $\omega$ B97X was found to share a similar value with the one used in the rCEC parameters<sup>14</sup> ( $4.898 \text{ \AA}^{-1}$  vs  $4.984 \text{ \AA}^{-1}$ ), which is based on the multistate empirical valence bond method (MS-EVB),<sup>23,24,39</sup> which better justifies the use of the rCEC variable for AIMD simulations.

As shown in an earlier study,<sup>14</sup> the IR spectrum of the excess proton charge provides a systematic measure of the CEC dynamics by revealing the encoded collective motions in that CV. Due to the similarity in their parameters, we found that the CDFT-CECs in general align with the rCEC spectrum (Figure 2B), again justifying the use of the MS-EVB-derived CEC in AIMD simulations.<sup>14</sup> It should be noted that although the two functionals exhibited different charge-transfer behaviors (Figure 2A), the BLYP-fitted CEC and the  $\omega$ B97X-fitted CEC produced very similar spectra, as shown in Figure 2B; this correlation implies that the CEC parametrized by one functional may be applied to AIMD simulations using other functionals. Compared to the experimental IR difference spectrum, the CEC spectra reproduce the acid continuum  $600\text{--}3200 \text{ cm}^{-1}$ , which is the signature feature of the acid solutions that arise from the hydrated excess proton. In

particular, the proton transfer mode (PTM) at around  $1200 \text{ cm}^{-1}$  and the flanking water bending at around  $1750 \text{ cm}^{-1}$  are effectively reproduced in the excess proton CEC spectra, indicating that these modes are well reflected in the encoded collective motions of CEC. It should be noted that the CEC spectrum decays in the range of  $2300\text{--}3200 \text{ cm}^{-1}$ , as opposed to the peak present in the experimental spectrum. This difference is due to the decaying excess proton charge in outer solvation shells, thereby reducing the intensity of this region, which is associated with the red-shifted O–H stretching in the second and third solvation shells of the excess proton, as detailed in a prior paper.<sup>14</sup> Importantly, the CEC spectrum decays to zero at the same position as the experimental difference spectrum at around  $3200 \text{ cm}^{-1}$ , revealing that the CEC excludes any bulk-like water O–H stretching in its motions. In summary, the CEC mostly represents the inner core motions of the protonated water complex, including the PTM and the flanking water bending, smoothly scales down its weight in the outer solvation shells, and is completely shut off for the bulk-like waters, which is the ideal behavior of a CV to focus only on PT and its related collective motions different from bulk-water fluctuations.

By contrast, the mCEC exhibits a nondiminishing intensity in the pure water region ( $3200\text{--}4000 \text{ cm}^{-1}$ ), revealing its sensitivity to the bulk-water motions. This is an unfavorable behavior as illustrated previously<sup>14</sup> in that the use of mCEC in enhanced sampling will bias the bulk-like water motions and may result in artificial water autoionizations, distant from the excess proton reaction center. The proton indicator, on the other hand, has nonzero absorption over the full frequency range, as a result of its discontinuity.<sup>14</sup> This nondifferentiability prevents its application in MD-based enhanced free energy sampling simulations due to an energy conservation problem when biasing such a CV, which can lead to a violation of detailed balance.<sup>41</sup>

We then studied glutamate in water to examine the CEC as a collective variable to be used in enhanced free energy sampling. Accordingly, we first parametrized the CEC from CDFT calculations of charge transfer between glutamate and water. The  $\delta_0$  parameter in eq 12 is needed for this case to account for the asymmetry in the PT between Glu and water. Figure 3A shows that the exponential function indeed outlines the charge-transfer behavior for Glu–water, suggesting that this functional form can be applied to PT involving other weak acids. We note that other forms of switching functions could also be employed if eq 12 is deficient in providing an accurate fit to the CDFT behavior for a particular system.

The proton disassociation barrier for Glu is large,<sup>42</sup> given the experimental  $\text{pK}_a$  of around 4.2. Thus, dissociation tends to be a rare event compared to the achievable timescale of AIMD simulations. In Figure 3B, we show the sampling efficiency gained from a metadynamics run using the CV  $\xi$ , representing the distance between the CEC and the closest carboxyl oxygen (see Simulation Details for the definition). During the 40 ps AIMD run, the unbiased simulation was found to sample only the free energy well corresponding to a protonated Glu, while the metadynamics drives the CEC to easily sample the dissociation of the proton from Glu and almost completes a “round trip”.

## CONCLUSIONS

In this work, we developed an appropriate collective variable (CV) to define the charge defect location and transfer

properties for proton transfer and transport processes, resulting in a more rigorous ab initio definition of the center of excess charge. We also showed that the charge-transfer behavior of CDFT can be approximated by an exponential function. We further examined the encoded collective motions in this newly defined CEC via calculating its IR spectrum. The full acid continuum was reproduced, suggesting the ability of this new CEC CV for capturing the excess proton motions without any contamination from other irrelevant degrees of freedom.

We also simulated a glutamate–water solution as an example, illustrating the use of the new CEC in enhanced free energy sampling for amino acid ionization in water. An AIMD metadynamics run driving the CEC was found to explore a much larger CV space than an unbiased AIMD run, providing efficient sampling of the proton disassociation of Glu.

We note that the present CDFT formalism is suitable for describing any charge-transfer reaction in which the bonding topology changes between the diabatic states; thus, it is not limited to the proton transfer processes studied herein. Therefore, the method can be generalized to identify appropriate CVs for other charge-transfer reactions, e.g., ATP or GTP hydrolysis.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.1c00609>.

Description of the experiment directed simulation (EDS) correction for hydrated excess proton AIMD simulations; table of atomic charges used to compute the center of charge (COC) (PDF)

Coordinate file used for fitting CEC parameters for water–water (XYZ)

Coordinate file used for fitting CEC parameters for glutamate–water (XYZ)

## ■ AUTHOR INFORMATION

### Corresponding Author

**Gregory A. Voth** – Department of Chemistry, Chicago Center for Theoretical Chemistry, James Franck Institute, and Institute for Biophysical Dynamics, University of Chicago, Chicago, Illinois 60637, United States; [orcid.org/0000-0002-3267-6748](https://orcid.org/0000-0002-3267-6748); Email: [gavoth@uchicago.edu](mailto:gavoth@uchicago.edu)

### Author

**Chenghan Li** – Department of Chemistry, Chicago Center for Theoretical Chemistry, James Franck Institute, and Institute for Biophysical Dynamics, University of Chicago, Chicago, Illinois 60637, United States

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jctc.1c00609>

### Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This research was supported in part by the National Institute of General Medical Sciences (NIGMS) of the U.S. National Institutes of Health (NIH) through Grant R01 GM053148 and in part by the U.S. Department of Energy (DOE), Office of Basic Energy Sciences, Separation Science Program of the Division of Chemical Sciences, Geosciences, and Biosciences under Award Number DE-SC0018648. The computational resources for this research were provided by the University of Chicago Research Computing Center (RCC).

## ■ REFERENCES

- (1) Decoursey, T. E. Voltage-gated proton channels and other proton transfer pathways. *Physiol. Rev.* **2003**, *83*, 475–579.
- (2) Wraight, C. A. Chance and design - Proton transfer in water, channels and bioenergetic proteins. *Biochim. Biophys. Acta* **2006**, *1757*, 886–912.
- (3) Swanson, J. M. J.; Maupin, C. M.; Chen, H. N.; Petersen, M. K.; Xu, J. C.; Wu, Y. J.; Voth, G. A. Proton solvation and transport in aqueous and biomolecular systems: Insights from computer simulations. *J. Phys. Chem. B* **2007**, *111*, 4300–4314.
- (4) de Grotthuss, C. J. T. Sur la décomposition de l'eau et des corps qu'elle tient en dissolution à l'aide de l'électricité galvanique. *Ann. Chim.* **1806**, *58*, 54–73.
- (5) Knight, C.; Voth, G. A. The curious case of the hydrated proton. *Acc. Chem. Res.* **2012**, *45*, 101–109.
- (6) Marx, D. Proton transfer 200 years after von Grotthuss: Insights from ab initio simulations. *ChemPhysChem* **2006**, *7*, 1848–1870.
- (7) Voth, G. A. Computer simulation of proton solvation and transport in aqueous and biomolecular systems. *Acc. Chem. Res.* **2006**, *39*, 143–150.
- (8) Gillan, M. J.; Alfe, D.; Michaelides, A. Perspective: How good is DFT for water? *J. Chem. Phys.* **2016**, *144*, No. 130901.
- (9) Torrie, G. M.; Valleau, J. P. Non-Physical Sampling Distributions in Monte-Carlo Free-Energy Estimation - Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187–199.
- (10) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.
- (11) Comer, J.; Gumbart, J. C.; Henin, J.; Lelievre, T.; Pohorille, A.; Chipot, C. The Adaptive Biasing Force Method: Everything You Always Wanted To Know but Were Afraid To Ask. *J. Phys. Chem. B* **2015**, *119*, 1129–1151.
- (12) König, P. H.; Ghosh, N.; Hoffmann, M.; Elstner, M.; Tajkhorshid, E.; Frauenheim, T.; Cui, Q. Toward theoretical analysis of long-range proton transfer kinetics in biomolecular pumps. *J. Phys. Chem. A* **2006**, *110*, 548–563.
- (13) Pezeshki, S.; Lin, H. Adaptive-Partitioning QM/MM for Molecular Dynamics Simulations: 4. Proton Hopping in Bulk Water. *J. Chem. Theory Comput.* **2015**, *11*, 2398–2411.
- (14) Li, C. H.; Swanson, J. M. J. Understanding and Tracking the Excess Proton in Ab Initio Simulations; Insights from IR Spectra. *J. Phys. Chem. B* **2020**, *124*, 5696–5708.
- (15) Kaduk, B.; Kowalczyk, T.; Van Voorhis, T. Constrained Density Functional Theory. *Chem. Rev.* **2012**, *112*, 321–370.
- (16) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic-Behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- (17) Lee, C. T.; Yang, W. T.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron-Density. *Phys. Rev. B* **1988**, *37*, 785–789.
- (18) Chai, J. D.; Head-Gordon, M. Systematic optimization of long-range corrected hybrid density functionals. *J. Chem. Phys.* **2008**, *128*, No. 084106.
- (19) Wu, Q.; Kaduk, B.; Van Voorhis, T. Constrained density functional theory based configuration interaction improves the prediction of reaction barrier heights. *J. Chem. Phys.* **2009**, *130*, No. 034109.

- (20) Becke, A. D. A Multicenter Numerical-Integration Scheme for Polyatomic-Molecules. *J. Chem. Phys.* **1988**, *88*, 2547–2553.
- (21) Wu, Q.; Van Voorhis, T. Extracting electron transfer coupling elements from constrained density functional theory. *J. Chem. Phys.* **2006**, *125*, No. 164105.
- (22) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone phi, psi and Side-Chain chi(1) and chi(2) Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273.
- (23) Biswas, R.; Tse, Y. L. S.; Tokmakoff, A.; Voth, G. A. Role of Presolvation and Anharmonicity in Aqueous Phase Hydrated Proton Solvation and Transport. *J. Phys. Chem. B* **2016**, *120*, 1793–1804.
- (24) Wu, Y.; Chen, H.; Wang, F.; Paesani, F.; Voth, G. A. An improved multistate empirical valence bond model for aqueous proton solvation and transport. *J. Phys. Chem. B* **2008**, *112*, 467–482.
- (25) Pyykkö, P.; Atsumi, M. Molecular Single-Bond Covalent Radii for Elements 1-118. *Chem. – Eur. J.* **2009**, *15*, 186–197.
- (26) Holmberg, N.; Laasonen, K. Efficient constrained density functional theory implementation for simulation of condensed phase electron transfer reactions. *J. Chem. Theory Comput.* **2017**, *13*, 587–601.
- (27) VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. QUICKSTEP: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.* **2005**, *167*, 103–128.
- (28) Kühne, T. D.; Iannuzzi, M.; Del Ben, M.; Rybkin, V. V.; Seewald, P.; Stein, F.; Laino, T.; Khaliullin, R. Z.; Schutt, O.; Schiffmann, F.; Golze, D.; Wilhelm, J.; Chulkov, S.; Bani-Hashemian, M. H.; Weber, V.; Borstnik, U.; Taillefumier, M.; Jakobovits, A. S.; Lazzaro, A.; Pabst, H.; Müller, T.; Schade, R.; Guidon, M.; Andermatt, S.; Holmberg, N.; Schenter, G. K.; Hehn, A.; Bussy, A.; Belleflamme, F.; Tabacchi, G.; Gloss, A.; Lass, M.; Bethune, I.; Mundy, C. J.; Plessl, C.; Watkins, M.; VandeVondele, J.; Krack, M.; Hutter, J. CP2K: An electronic structure and molecular dynamics software package - Quickstep: Efficient and accurate electronic structure calculations. *J. Chem. Phys.* **2020**, *152*, No. 194103.
- (29) Lehtola, S.; Steigemann, C.; Oliveira, M. J. T.; Marques, M. A. L. Recent developments in LIBXC - A comprehensive library of functionals for density functional theory. *SoftwareX* **2018**, *7*, 1–5.
- (30) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, No. 154104.
- (31) Hartwigsen, C.; Goedecker, S.; Hutter, J. Relativistic separable dual-space Gaussian pseudopotentials from H to Rn. *Phys. Rev. B* **1998**, *58*, 3641–3662.
- (32) Lippert, G.; Hutter, J.; Parrinello, M. A hybrid Gaussian and plane wave density functional scheme. *Mol. Phys.* **1997**, *92*, 477–487.
- (33) VandeVondele, J.; Hutter, J. An efficient orbital transformation method for electronic structure calculations. *J. Chem. Phys.* **2003**, *118*, 4365–4369.
- (34) Weber, V.; VandeVondele, J.; Hutter, J.; Niklasson, A. M. Direct energy functional minimization under orthogonality constraints. *J. Chem. Phys.* **2008**, *128*, No. 084113.
- (35) White, A. D.; Voth, G. A. Efficient and Minimal Method to Bias Molecular Simulations with Experimental Data. *J. Chem. Theory Comput.* **2014**, *10*, 3023–3030.
- (36) Calio, P. B.; Hocky, G. M.; Voth, G. A. Minimal Experimental Bias on the Hydrogen Bond Greatly Improves Ab Initio Molecular Dynamics Simulations of Water. *J. Chem. Theory Comput.* **2020**, *16*, 5675–5684.
- (37) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604–613.
- (38) Barducci, A.; Bussi, G.; Parrinello, M. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **2008**, *100*, No. 020603.
- (39) Schmitt, U. W.; Voth, G. A. The computer simulation of proton transport in water. *J. Chem. Phys.* **1999**, *111*, 9361–9381.
- (40) Biswas, R.; Carpenter, W.; Fournier, J. A.; Voth, G. A.; Tokmakoff, A. IR spectral assignments for the hydrated excess proton in liquid water. *J. Chem. Phys.* **2017**, *146*, No. 154507.
- (41) Ferrarotti, M. J.; Bottaro, S.; Perez-Villa, A.; Bussi, G. Accurate Multiple Time Step in Biased Molecular Simulations. *J. Chem. Theory Comput.* **2015**, *11*, 139–146.
- (42) Nelson, J. G.; Peng, Y.; Silverstein, D. W.; Swanson, J. M. J. Multiscale Reactive Molecular Dynamics for Absolute pK(a) Predictions and Amino Acid Deprotonation. *J. Chem. Theory Comput.* **2014**, *10*, 2729–2737.