Simulating nuclear and electronic quantum effects in enzymes

Lu Wang\textsuperscript{a}, Christine M. Isborn\textsuperscript{b}, Thomas E. Markland\textsuperscript{c}

\textsuperscript{a}Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, New Jersey 08854, USA
\textsuperscript{b}Chemistry and Chemical Biology, School of Natural Sciences, University of California, Merced, 5200 North Lake Road, Merced, California 95343, USA
\textsuperscript{c}Department of Chemistry, Stanford University, Stanford, California 94305, USA

Abstract

An accurate treatment of the structures and dynamics that lead to enhanced chemical reactivity in enzymes requires explicit treatment of both electronic and nuclear quantum effects. The former can be captured in \textit{ab initio} molecular dynamics (AIMD) simulations while the latter can be included by performing \textit{ab initio} path integral molecular dynamics (AI-PIMD) simulations. Both AIMD and AI-PIMD simulations have traditionally been computationally prohibitive for large enzymatic systems. Recent developments in streaming computer architectures and new algorithms to accelerate path integral simulations now make these simulations practical for biological systems, allowing elucidation of enzymatic reactions in unprecedented detail. In this chapter, we summarize these recent developments and discuss practical considerations for applying AIMD and AI-PIMD simulations to enzymes.

Keywords: density functional theory, \textit{ab initio} molecular dynamics, path integral molecular dynamics, QM/MM, nuclear quantum effects

1. Introduction

Tremendous effort has been devoted to using molecular simulation to unravel how enzymes catalyze chemical reactions with such remarkable efficiency and selectivity. A large amount of this work has utilized classical molecular mechanical (MM) methods combined with fixed charge and, more recently, polarizable empirical force fields \cite{92}. These classical methods have been successful in predicting and elucidating important properties ranging from ligand binding free energies to $pK_a$’s of active-site residues\cite{103, 81, 8}. However, empirical force fields are unable to describe the full electronic reorganization in enzyme active sites arising from bond making/breaking and charge transfer along short hydrogen bonds. In addition, most force fields do not include parametrization for transition metals or non-standard conditions.
ligands. For these systems, single-point electronic structure calculations and geometry optimizations (i.e. quenching to the local 0 K structure) have been highly useful in determining bonding topologies, identifying transition states and intermediates, as well as predicting reaction kinetics [37, 8, 101]. However, one must go beyond these snapshot-based methods to determine the full electron density redistribution as the enzyme’s structure fluctuates, which can lead to bond cleavage and proton movement.

*Ab initio* molecular dynamics (AIMD) simulations evolve the nuclei using forces generated from the instantaneous electronic structure obtained at each time step, which allows coupling between nuclear motion and electronic rearrangement. These simulations treat the electrons quantum mechanically and the nuclei classically. However, if light atoms are present, nuclear quantum effects (NQEs), such as tunneling and zero point energy (ZPE), can also play an important role. For example, kinetic isotope effects of over two orders of magnitude on the enzymes’ catalytic rates have been experimentally observed upon substituting hydrogen (H) by deuterium (D) [109, 63] and have been attributed to tunneling. Also, the ZPE in a typical oxygen–hydrogen (O–H) bond is ∼5 kcal/mol. As a result, including the ZPE can significantly alter the structure and dynamics of systems containing short hydrogen bonds, which are commonly observed in biomolecules where the protein fold can position groups much closer than typically seen in solution [25, 80]. To treat NQEs one can perform *ab initio* path integral molecular dynamics (AI-PIMD) simulations, which exactly include the effect of NQEs on static equilibrium properties of a given electronic surface [9, 78, 53]. The imaginary time path integral formalism on which PIMD simulations are founded also forms the basis for the approximate centroid molecular dynamics (CMD) [14, 57] and ring polymer molecular dynamics (RPMD) [28, 45] approaches to obtain quantum dynamics.

Performing AIMD simulations for enzyme active sites has traditionally been a formidable computational task because electronic structure calculations must be performed at each time step. Due to the need to perform many thousands of electronic structure calculations for each picosecond evolved, these simulations are typically performed using density functional theory (DFT) to generate the wave function, although cheaper semiempirical methods such as density functional tight binding have also been employed when longer time scales are required, albeit at a loss in accuracy [96, 38]. However, recent theoretical and algorithmic advances and novel computer processing architectures have greatly accelerated these simulations and allowed for trajectories of hundreds of picoseconds or even nanosecond time scales for systems on the order of 60–300 atoms. In particular, electronic structure calculations have been accelerated by the development of linear scaling [40, 10, 86, 11, 104] DFT methods and the advent of codes that use the massively parallel stream processing capabilities of graphical processing units (GPUs) [115, 116]. The former methods take advantage of the locality of the density matrix as well as the fact that both the Coulomb/Hartree and DFT exchange-correlation energies are functions of the local spin density, which allows for a linear growth in computational cost when the system size increases. An example of the latter, which will be the major focus of this article, is the TeraChem electronic structure program, which uses GPUs to accelerate the computation of the electronic wave function and has demonstrated speedups of over 100 fold compared to CPU-based codes [113, 116, 117, 53]. This speedup has enabled *ab initio* energy calculations on both the ground and excited states of systems
containing thousands of atoms, including polypeptides and proteins [115, 116, 117, 55, 68]. AI-PIMD simulations have typically required about 2 orders of magnitude more computational cost than the corresponding AIMD simulations. A large amount of this additional cost arises from the need to make many replicas of the system, each of which requires a separate electronic structure calculation. Hence a standard PIMD implementation requires 30–50 electronic structure calculations to evolve a single time step for a typical hydrogen containing system at room temperature [119, 9, 75]. In addition, the PIMD Hamiltonian contains high-frequency motions that limit the time step that can be employed and do not efficiently sample the full phase space if used directly. However, recent developments have significantly alleviated these issues, reducing the computational overhead for including NQEs in simulations [75, 74, 22, 19] and making it more efficient to extract isotope effects on free energy changes from these simulations [118, 21, 76].

Here, we briefly review some considerations when performing electronic structure calculations of large biological systems with a particular emphasis on the GPU-accelerated TeraChem code. We then outline how these electronic structure developments have recently been combined with the latest PIMD algorithms to allow AI-PIMD simulations, which include nuclear and electronic quantum effects, to be performed for enzyme active sites.

2. Electronic quantum effects in biological systems

Electronic structure codes that are written to take advantage of GPUs, such as TeraChem, have made it possible to perform DFT single-point energy calculations, geometry optimizations, and AIMD simulations on quantum mechanical (QM) regions of many hundreds to thousands of atoms. The ability to now perform such large calculations has provided a number of insights into biological processes. For example, large-scale geometry optimizations showed that changes in protein structure upon mutation are directly correlated with enzymatic methyl transfer efficiency [129]. In addition, AIMD simulations have been used to examine charge transfer and polarization in the BPTI protein [114], discover new pathways for glycine synthesis from primitive compounds [122], and determine amorphous indium phosphide nanostructures [131].

However, calculations of large systems have also highlighted deficiencies in DFT. In particular, semilocal DFT methods lead to large size-dependent errors and nonlocal exact exchange is necessary to fix these errors. As nonlocal exact exchange is extremely difficult to include in linear-scaling DFT methods, parallel stream processing computer hardware such as GPUs and software advances like those implemented in TeraChem are essential for large-scale exact exchange DFT calculations. While TeraChem is not linear scaling with the number of basis functions, \( N \), over a large range of practical system sizes it achieves scaling of around \( N^{1.5} \), which is considerably better than the formal DFT scaling of \( N^3 \) [72]. In this section we discuss the choice of basis set, DFT exchange correlation functional, and QM region in the hybrid quantum mechanics/molecular mechanics (QM/MM) calculations for biological systems.
2.1. Basis sets

The basis set describes the electronic wave functions and must balance the accuracy of a given property with computational cost in the quantum chemistry calculation. For example, calculating gas-phase bond dissociation energies with a high level of theory to sub-kcal/mol convergence may require approaching the complete basis set limit \[85, 49, 47\]. However, for very large QM calculations, using such a large basis set is often not necessary, and including diffuse functions may lead to difficulties converging the wave function. In many condensed phase applications one is concerned with properties that are less sensitive to basis set effects, e.g. transfer of a proton between similar chemical groups. For such geometry based properties, smaller (e.g. double zeta) basis sets can often be sufficient to obtain the desired accuracy \[123\]. Figure 1 shows the potential energy profile for moving a proton along the active-site hydrogen bond network in an enzyme KSI40N. Except for 6-31G, all the other basis sets produce potential energy curves in quantitative agreement with that predicted using the largest basis set aug-cc-pVDZ.

Figure 2 shows results from Kulik et al., who performed DFT structural optimizations of 58 proteins using TeraChem and showed that small basis sets such as MINI and STO-3G, while computationally efficient, lead to significant errors in predicted properties such as bond lengths and structural clashes as compared to the experimental crystal structures \[68\]. This problem can be alleviated by using a double-zeta basis set such as 3-21G, although this basis set is still relatively small and thus is likely subject to basis set superposition errors and may be too inflexible for modeling more subtle changes in polarization. For an accurate description of the electronic wave function and hydrogen bonding, the basis set should include polarization functions on all atoms. For systems with electrons far from the nucleus such as anions or excited states, diffuse functions should also be included.

2.2. Density functionals and the role of exact exchange

While DFT is formally exact in principle, owing to the Hohenberg-Kohn and Kohn-Sham theorems \[51, 65\], the functional form of the exchange-correlation energy is unknown and therefore it is necessary to make approximations for practical calculations \[88, 12\]. DFT methods in practice often approximate the electron exchange and correlation interactions by considering only semi-local properties of the electron density. For example, the generalized gradient approximation (GGA) functionals use the local spin density and its gradient at a given point in space \[89\], while meta-GGA functionals further include the kinetic energy of the density \[110\]. These semilocal functionals are often quite successful in regions of slowly varying electron density and can also benefit from cancellation of errors in the exchange and correlation functionals.

Dispersion is important for accurately modeling thermodynamic and substrate binding properties of enzymes. However, as dispersion is a long-range dynamical correlation effect, it is not accounted for in semilocal density functionals \[59, 42\]. A common technique to overcome this deficiency is to add an explicit, empirically parametrized attractive term to the DFT energy to represent interactions between atomic pairs, in which case '-D' is usually appended to the name of the functional \[127, 51, 41, 126\]. These DFT-D approaches require little additional computational cost and often work quite well. Nevertheless, the accuracy
of the DFT-D predictions depend on the size of the basis sets and are limited by the atom types included in the parametrization. An alternative, non-empirical approach is to compute the dispersion interaction between atoms based on the exchange-hole dipole model (XDM) [6, 58, 7]. This model also adds negligible computational cost to a DFT calculation, but its implementation in geometry optimizations and dynamical simulations is currently not practical because forces obtained from the XDM method have not been widely available.

The approximate exchange term in semilocal DFT methods produces errors in both the electron density and energy, which leads to the derivative discontinuity [87, 113, 66, 98], the many-electron self-interaction error [130, 81] and the delocalization error [82, 48, 132, 62]. These semilocal density functionals underestimate the energy gap between the highest-occupied and lowest-unoccupied molecular orbitals (HOMO and LUMO) [82, 27, 26, 106], which leads to severe problems as the size of the QM region increases. We have shown that for molecules in aqueous solution, semilocal DFT methods predict that the band gap approaches zero and eventually the ground state self-consistent field (SCF) calculation can no longer converge if more and more solvent molecules are included in the QM region [56]. Similar difficulties were found by Kulik et al. when performing semilocal DFT calculations on entire proteins [68]. Additionally, we have recently shown that approximate semilocal exchange functionals generate a size-dependent error in the ionization potential, with the size of the error increasing as the size of the system grows [125, 105].

Many of these inaccuracies can be improved by incorporating nonlocal exact exchange into the density functional via the generalized Kohn-Sham scheme [99]. Exact exchange is often included in approximate semilocal exchange methods through either global hybrid or range-separated hybrid (RSH) techniques. For global hybrids, such as the B3LYP functional [5], a fixed fraction of exact exchange (typically 20–50%) is included for all interelectronic distances. RSH methods vary the amount of exact exchange based on the interelectronic distance $r_{12} \equiv |r_1 - r_2|$, with the error function usually used to make the variation smooth for splitting the short- and long-range parts of the Coulomb operator

$$\frac{1}{r_{12}} = \frac{1}{r_{12}} - \text{erf}(\omega r_{12}) + \text{erf}(\omega r_{12}). \quad (1)$$

For a long-range corrected (LC) hybrid, full exact exchange is used at long range (the \text{erf}(\omega r_{12}) term), while semilocal DFT exchange dominates at short-range (the \(1 - \text{erf}(\omega r_{12})\) term) to provide good balance with semilocal DFT correlation. The ratio of local DFT exchange to nonlocal exact exchange is determined by a range separation parameter, $\omega$, given in atomic units of inverse bohrs, $a_0^{-1}$ [39, 53, 111, 128, 112, 4, 23]. The default value of $\omega$ is often between 0.2 and 0.5 $a_0^{-1}$, with smaller values leading to local DFT exchange having a longer range. A system-dependent ‘optimal’ $\omega$ can also be determined by tuning $\omega$ to enforce Koopmans’ theorem [97, 107, 8, 67, 2].

Long-range corrected hybrid functionals have the correct asymptotic limit for the exchange interaction, which often yields improved ionization potentials, band gaps, and excitation energies compared to experiment [94, 69, 106, 67, 95, 2], as well as improved convergence of the SCF ground state for systems containing large QM regions [68, 56]. However, although all long-range corrected hybrid functionals have the correct asymptotic behavior of
the exchange interaction, care must be taken to ensure that the range-separation parameter $\omega$ is large enough to fix the deficiencies of using approximate semilocal exchange at short range. We have recently shown that the choice of $\omega$ is key in determining the correct physical polarization in response to electron ionization. For example, for the ionization of a solvated ethene molecule, if $\omega$ is too small (below $\sim 0.4 \, a_0^{-1}$), the surrounding water molecules will unphysically contribute some of their own electron density to the ionization process, while for a larger $\omega$ they will correctly polarize in response to the electron being removed from the solute \cite{105}. This can be seen in the density differences for the neutral and cation systems shown in Figure 3. Thus the correct choice of functional, and particularly the treatment of exchange in the functional, is crucial for correctly describing processes that are key in enzymatic reactions such as polarization and charge-transfer. Hence, in addition to including dispersion through methods mentioned above, we recommend a long-range corrected hybrid functional such as LC-BLYP, LC-$\omega$PBE, or $\omega$B97, with an $\omega$ value in the range $\omega=0.4-0.6 \, a_0^{-1}$.

2.3. QM regions in QM/MM calculations

When simulating large biomolecules, it is desirable to accurately describe the region of interest, such as an enzyme’s active site, using a high-level QM method with a large basis set. Such quantum mechanical treatments are only feasible for a fraction of atoms in biomolecules due to the high computational cost of the QM approach. However, unlike an isolated system in vacuum, the electrostatic environment in the condensed phase strongly polarizes the QM system and therefore it is crucial to include the environmental effects for a correct description of the energy, geometry, and electron density distribution in the QM region. The hybrid QM/MM approach takes advantage of both the accuracy of the QM method and the speed of the MM method, allowing the MM charges to polarize the QM wave function \cite{124, 36}. The coupling between the QM and MM regions has been extensively investigated, see Refs. \cite{70, 11} for examples; and for recent reviews of the QM/MM method see Refs. \cite{96, 71, 100, 93}.

To combine large-scale GPU-accelerated quantum chemistry calculations with MM force fields, we have created an extensive TeraChem interface for QM/MM calculations \cite{54}. In this approach, TeraChem receives the point charges of the MM atoms and performs a QM calculation with the electrostatic embedding method to account for the polarization of the QM region due to the electrostatic MM environment. TeraChem then outputs the QM forces on each atom for the propagation of the QM/MM system. Communication between the QM and MM regions is either through files or the message passing interface (MPI). Although the TeraChem QM/MM interface was originally implemented with the AMBER molecular dynamics program \cite{15}, it is general enough to be used with any program that can accept quantum mechanical forces for molecular dynamics simulations. Indeed, the AI-PIMD discussed in the following section were performed using the TeraChem MPI interface to connect to an in-house PIMD code to perform the dynamics.

Ideally, the QM region in the QM/MM calculation will be large enough that the results are independent of QM region size. To test how large the QM region should be to compute converged properties, we performed excitepd state QM/MM calculations on photoactive
yellow protein (PYP) and computed its absorption spectrum using six different QM regions of increasing size, which range from the photoactive chromophore alone (22 QM atoms, 217 basis functions) to including the entire protein and counter ions (1935 QM atoms, 16827 basis functions) [54]. For all calculations the total system was the same, which included the entire protein, counter ions, and a 32 Å solvation sphere, and any atoms not included in the QM region were modeled with MM point charges. We found that although the electronic excitation was for the most part localized on the chromophore in all QM/MM calculations, a large QM region (723 QM atoms) was required to reproduce the energy of the largest QM system (the entire protein included in the calculation). Furthermore, the predicted excitation energy did not have a monotonic trend as the size of the QM region increased, but instead jumped around substantially (up by 0.4 eV when going from the 4th to 5th largest QM region) as additional protein residues were included in the QM region, see Figure 4. Overall, when sampling over many snapshots, this dependence of excitation energy on the size of the QM region led to significant differences in the final computed absorption spectrum [54].

The slow convergence of a local property, such as an excitation energy, with the size of the QM region suggests that the electrostatic interactions at the QM/MM boundary can cause large changes in the QM electron density. Therefore the QM region should be as large as possible to push this boundary away from the active site to obtain an accurate QM density. Others have arrived at similar conclusions about the large size required for the QM region, citing problems with charges at the QM/MM junction [108, 73, 52, 35, 34].

3. Incorporating nuclear quantum effects in AIMD simulations

The interplay between nuclear and electronic quantum effects can dramatically alter the structure and dynamics of hydrogen bonded systems. It has been shown that the distance between the donor and acceptor heavy atoms, R, plays a crucial role in determining the proton behavior in a hydrogen bond [90, 79]. In particular, when R is around or below 2.7 Å, the influence of NQEs becomes significant. For example, in liquid water, which has an average O–O distance of around 2.8 Å, the quantum nature of the hydrogen bonds leads to transient proton excursion events in which a proton is closer to the hydrogen bond acceptor than donor oxygen [18, 120].

Despite the importance of NQEs, most molecular simulations are performed treating the nuclei as classical particles. The path integral formalism of quantum mechanics provides an efficient way to incorporate nuclear quantum effects in molecular simulations [33]. By combining the path integral methods and on-the-fly electronic structure calculations, one can perform AI-PIMD simulations to examine the nuclear and electronic quantum fluctuations in chemical and biological systems. Below we provide the basic concepts, implementation, and examples for AI-PIMD simulations.

3.1. Path integral molecular dynamics

PIMD simulations allow the exact inclusion of NQEs on static equilibrium properties for a given electronic surface by exploiting the isomorphism between the quantum partition
function of a quantum mechanical system and the classical partition function of a “ring polymer”, which is a cyclic structure containing multiple copies of the classical system with adjacent copies (beads) connected by harmonic springs [9].

For a system of \( N \) classical particles of masses \( m_i \) the Hamiltonian is

\[
H = \sum_{i=1}^{N} \frac{p_i^2}{2m_i} + V(r_1, \ldots, r_N),
\]

(2)

where in AIMD the potential energy \( V(r_1, \ldots, r_N) \) is obtained from an electronic structure calculation. The path integral expression for the partition function for this Hamiltonian is

\[
Q_P = \frac{1}{(2\pi\hbar)^f} \int d^f p \int d^f r \ e^{-\beta_P H_P(p, r)},
\]

(3)

where \( f = 3NP \), \( P \) is the number of beads in the ring polymer, \( p \equiv \{p_i^{(j)}\}_{i=1,...,N} \), \( r \equiv \{r_i^{(j)}\}_{i=1,...,N} \) and \( \beta_P = \frac{\beta}{P} = \frac{1}{k_B T_P} \). The effective temperature, \( T_P = PT \), under which each bead evolves is thus \( P \) times higher than the physical temperature, allowing the copies of the system to sample areas of phase space that are not accessible to a classical particle. The PIMD Hamiltonian \( H_P(p, r) \) is,

\[
H_P(p, r) = \sum_{j=1}^{P} \left( \sum_{i=1}^{N} \frac{|p_i^{(j)}|^2}{2m_i} + \frac{1}{2} m_i \omega_P^2 |r_i^{(j)} - r_i^{(j-1)}|^2 \right) + V(r_1^{(j)}, \ldots, r_N^{(j)}),
\]

(4)

where \( \omega_P = 1/\beta_P \hbar \) and cyclic boundary conditions, \( j + P \equiv j \) are implied. \( H_P \) corresponds to the Hamiltonian of a classical ring polymer, in which adjacent beads are linked by harmonic springs with a force constant of \( m_i \omega_P^2 \). In PIMD simulations the system is evolved using Hamilton’s equations of motion from this Hamiltonian [9]. Note that PIMD simulations utilize the momentum term in Eq. 4 to sample phase space and hence the dynamics generated from this Hamiltonian are not exact quantum dynamics.

Static equilibrium properties of the quantum mechanical system can be exactly calculated as

\[
<A> = \frac{\int d^f p \int d^f r \ A e^{-\beta_P H_P}}{\int d^f p \int d^f r \ e^{-\beta_P H_P}}.
\]

(5)

3.2. Accelerating PIMD convergence using generalized Langevin equations

Eq. 3 is exact when \( P \to \infty \). In practice, a finite number of replicas \( P \) is used to converge the properties to the desired accuracy. A useful indication as to the number of replicas is \( P > 3\hbar \omega_{\text{max}} \) where \( \omega_{\text{max}} \) is the highest frequency present in the system. This criteria can be understood as stating that the role of \( P \) is to increase the effective thermal energy of each replica \( (P k_B T) \) to be larger than the energy level spacings in the system \( (\hbar \omega_{\text{max}} \text{ for a harmonic oscillator}) \), i.e. allowing each replica to approach the classical limit [119, 75]. For example, the number of replicas used for convergence is typically \( P \geq 32 \) for a hydrogen-containing system at 300 K. Since each replica of the system requires an \textit{ab initio}
calculation to evaluate its energy and forces, the computational cost of PIMD simulations is therefore at least $P$ times higher than the corresponding classical simulation.

Computational overhead due to the convergence requirement can be significantly reduced using a generalized Langevin equation (GLE) thermostat \cite{17, 20, 19}. By coupling the path integral system to a GLE thermostat (the PI+GLE approach), one can tune the correlated noise such that convergence of the potential energy, $\langle V \rangle$, to its quantum mechanical expectation value can be achieved with considerably fewer path integral beads \cite{17, 20}. However, the PI+GLE approach does not guarantee fast convergence of the quantum kinetic energy, $\langle T \rangle$, as the centroid virial estimator for the kinetic energy (Eq. 7) \cite{50, 13} involves correlations between the beads and the ring-polymer centroid. To allow rapid convergence of the kinetic energy, Ceriotti and coworkers have recently introduced the PIGLET method, which enforces the necessary correlations to accelerate the convergence of the centroid virial kinetic energy estimator \cite{19}. The PIGLET approach allows the convergence of thermodynamic properties and isotope effects that depend on the kinetic energy (see discussions in the next section) using $P = 6$ for hydrogen-containing systems at 300 K. This reduces the computational cost by over 5 fold compared to a typical PIMD simulation \cite{19, 18}. Parameters for the GLE thermostat can be conveniently generated using the GLE4MD input generator \cite{16}.

### 3.3. Extracting thermodynamic isotope effects from PIMD simulations

Isotope substitution techniques have been widely used in fields ranging from molecular biology to atmospheric chemistry \cite{64}. Computer simulations that accurately and efficiently predict isotope effects form an important complement to experiments and have enabled detailed analysis of the reaction mechanism and thermodynamic equilibrium in a wide variety of chemical and biological processes \cite{64}.

Since isotopes of the same element have almost identical electronic structures, the change in structural and thermodynamic properties due to isotope substitution arise entirely from the quantum mechanical nature of the nuclei. The central quantity for calculating these thermodynamic isotope effects is the free energy change upon isotope substitution, which can be computed exactly using PIMD simulations for a given electronic surface. For example, the free energy change upon exchanging D for H in a given system $k$, $\Delta A_k$, is directly related to the quantum kinetic energies of the isotopes by the thermodynamic integration \cite{21}

$$\Delta A_k = \int_{m_H}^{m_D} d\mu \frac{T_k(\mu)}{\mu}. \quad (6)$$

Here $m_H$ and $m_D$ are the masses of H and D, respectively. $T_k(\mu)$ is the quantum kinetic energy of a hydrogen isotope of mass $\mu$ and can be determined from AI-PIMD simulations using the centroid virial estimator \cite{50, 13},

$$\langle T_k(\mu) \rangle = \frac{3}{2\beta} + \frac{1}{2P} \sum_{j=1}^{P} \langle (r^{(j)} - \bar{r}) \cdot \frac{\partial V}{\partial r^{(j)}} \rangle, \quad (7)$$

where $V$ is the potential energy of the system and $\bar{r} = \sum_{j=1}^{P} r^{(j)}/P$ is the centroid position.
From Eq. 6, one can in principle perform multiple PIMD simulations with varying $\mu$ and obtain $\Delta A_k$ by carrying out the integration. This process can be greatly accelerated by using the free energy perturbation (FEP) method [21], which allows one to extract $\langle T_k(\mu) \rangle$ from a single simulation of the most abundant isotope of hydrogen. From the FEP approach [21],

$$\langle T_k(\mu) \rangle = \frac{\langle T_k(m_H) e^{-h(\mu/m_H;x)} \rangle_{m_H}}{\langle e^{-h(\mu/m_H;x)} \rangle_{m_H}} \tag{8}$$

with

$$h(\alpha; r) = \frac{(\alpha - 1) \beta m_H \omega^2}{2P} \sum_{j=1}^{P} (r^{(j)} - r^{(j+1)})^2. \tag{9}$$

In the following, we use $pK_a$ as an example and demonstrate how one can efficiently extract thermodynamic isotope effects from PIMD simulations. Let us consider the side-chain $pK_a$ of an amino acid X in an enzyme upon H/D substitution. In H$_2$O, this $pK_a$ corresponds to the chemical equilibrium where the neutral amino acid, $XH_{EnzH}$, dissociates into the anion $X^-_{EnzH}$ and the proton $H^+$ (shown in the scheme below). In D$_2$O, we denote the neutral and ionized species of X as $XD_{EnzD}$ and $X^-_{EnzD}$, respectively. Note that in D$_2$O all labile protons in the enzyme are exchanged to D, and thus $X^-_{EnzH}$ and $X^-_{EnzD}$ are not necessarily identical. These chemical reactions can be connected using the thermodynamic cycle:

$$\begin{align*}
XH_{EnzH} & \rightleftharpoons X^-_{EnzH} + \ H^+ & \Delta A_1 \\
XD_{EnzD} & \rightleftharpoons X^-_{EnzD} + \ D^+ & \Delta A_2
\end{align*}$$

Assuming the free energy changes of the aforementioned dissociation processes are $\Delta A_1$ and $\Delta A_2$, then the $pK_a$ isotope effect is

$$\Delta pK_a^{Enz} \equiv pK_a^{EnzD} - pK_a^{EnzH} = \frac{\Delta A_2 - \Delta A_1}{2.303k_BT} = \frac{\Delta A_{Enz} - \Delta A_{Enz-} - \Delta A_H}{2.303k_BT}. \tag{10}$$

$\Delta A_{Enz}$ and $\Delta A_{Enz-}$ are the free energy changes upon converting $XD_{EnzD}$ and $X^-_{EnzD}$ to $XH_{EnzH}$ and $X^-_{EnzH}$, respectively, and can be efficiently calculated from PIMD simulations using Eq. 8 along with the FEP method (Eq. 9).

To obtain the absolute value of $\Delta pK_a^{Enz}$ one also needs $\Delta A_H$, which is the free energy change upon converting D to H in aqueous solution. It can be computed from PIMD simulations of protons in liquid water using the same level of QM description as that for the enzyme. However, it is often computationally costly to properly sample the solvent configurations around the proton in order to determine an accurate $\Delta A_H$. Alternatively, a useful metric is to compare $\Delta pK_a^{Enz}$ to a reference state such as the $pK_a$ isotope effect of the amino acid X in aqueous solution, $\Delta pK_a^{Sol}$. This results in an excess isotope effect,
\[ \Delta \Delta pK_a \equiv \Delta pK_a^{\text{Enz}} - \Delta pK_a^{\text{Sol}}, \]
which represent the extra amount of NQEs arising from the unique enzyme environment as compared to aqueous solution.

Similar to the enzyme case, \( \Delta pK_a^{\text{Sol}} \) can be computed from the thermodynamic cycle

\[
\begin{align*}
XH_{\text{Sol}} & \leftrightarrow X_{\text{Sol}}^+ + H^+ & \Delta A_3 \\
\Delta A_{\text{Sol}} & \uparrow & \Delta A_{\text{H}} \\
XD_{\text{Sol}} & \leftrightarrow X_{\text{Sol}}^- + D^- & \Delta A_4 
\end{align*}
\]

\( XH_{\text{Sol}} \) and \( XD_{\text{Sol}} \) denote the neutral amino acid \( X \) in \( H_2O \) and \( D_2O \), respectively, while \( X_{\text{Sol}}^+ \) represent the ionized species. Therefore,

\[
\Delta pK_a^{\text{Sol}} = pK_a^{\text{Sol}D} - pK_a^{\text{Sol}H} = \frac{\Delta A_4 - \Delta A_3}{2.303 k_B T} = \frac{\Delta A_{\text{Sol}} - \Delta A_{\text{H}}}{2.303 k_B T}. \quad (11)
\]

Comparing \( \Delta pK_a^{\text{Enz}} \) and \( \Delta pK_a^{\text{Sol}} \) makes the \( \Delta A_{\text{H}} \) term cancel, and thus the excess isotope effect is

\[
\Delta \Delta pK_a = \Delta pK_a^{\text{Enz}} - \Delta pK_a^{\text{Sol}} = \frac{\Delta A_{\text{Enz}} - \Delta A_{\text{Enz}} - \Delta A_{\text{Sol}}}{2.303 k_B T}. \quad (12)
\]

In this way, \( \Delta \Delta pK_a \) can be extracted from PIMD simulations by computing the free energy changes upon isotope substitution in the enzyme and solution environment.

### 3.4. Electronic and nuclear quantum fluctuations in biological hydrogen bond networks

Combining the technical advances introduced earlier, we now demonstrate how one can employ AI-PIMD simulations to elucidate the interplay between electronic and nuclear quantum effects in enzymes. We have recently studied the behavior of an active-site hydrogen bond network in an enzyme mutant KSI\(^{D40N}\) by interfacing with TeraChem and taking advantage of the PIGLET algorithm \([19]\) and the FEP method \([21]\). As the majority of computational cost comes from \textit{ab initio} force evaluations on the \( P \) ring-polymer beads, we have parallelized these calculations across GPUs with one GPU dedicated for each bead. These methods allow us to achieve a simulation speed that is three orders of magnitude higher than existing AI-PIMD approaches. In addition, in TeraChem the speed of electronic structure evaluations scales almost linearly with the number of GPUs. Therefore, one can further speedup the simulations by assigning one node for each bead and using multiple GPUs on each node to carry out electronic structure calculations.

In KSI\(^{D40N}\), the side chain groups of the active-site residues form a triad structural motif, in which residue Tyr57 sits at the center and forms two hydrogen bonds with the hydroxyl groups of residues Tyr16 and Tyr32 (Figure 5(a)). In several high-resolution crystal structures \([41, 102]\), the hydrogen bonded O57–O16 and O57–O32 distances are found to be around 2.6 Å, noticeably shorter than those commonly observed in hydrogen bonded systems such as liquid water.

In order to choose the proper DFT functional and basis set for the QM region, we have calculated the potential energy profiles for proton transfer within the tyrosine triad. We
define the proton transfer coordinate as \( \nu_{16} = d_{O16,H16} - d_{O57,H16} \), where \( d_{Oi,Hj} \) is the distance between the oxygen O of residue i and hydrogen H of residue j (Figure 5(a)), and a quantity \( \Delta E_{\nu=0} \) which represents the energy required to move the proton from its equilibrium position to the perfectly shared position, \( \nu_{16} = 0 \). As shown in Figure 5 all density functionals tested generate results that qualitatively agree with each other, although GGA functionals BLYP and PBE underestimate \( \Delta E_{\nu=0} \) by about 1.5 kcal/mol. Hybrid functionals B3LYP and PBE0 produce almost identical potential energy profiles, with the CAM-B3LYP RSH method in excellent agreement with the global hybrids. Dispersion corrections (D3 in this case) lead to very little change. Our basis set studies in Figure 1 demonstrate that 6-31G* is the smallest basis set that is able to produce the correct energy profile for proton transfer. Based on these tests, we choose to treat the QM region at the B3LYP-D3 level [5, 43] with the 6-31G* basis set.

In our AI-PIMD simulations, the QM region includes the tyrosine triad (47 atoms and 174 electrons) and is treated quantum mechanically in both the electronic and nuclear degrees of freedom. We have shown that increasing the size of the QM region by incorporating nearby residues does not significantly change the potential energy profile for proton transfer [121]. As environmental fluctuations are indispensable for a correct description of the QM region, especially the O–O distances, we include the rest of the protein as well as solvent molecules and counter-ions in the MM region (over 52,000 atoms), which is described using the AMBER03 [30] force field and the TIP3P water model [60]. All bonds across the QM/MM boundary are capped with hydrogen link atoms in the QM region, which are constrained along the bisecting bonds and do not interact with the MM region. The total interactions in the QM/MM system include 1) forces between QM atoms, 2) forces between MM atoms and 3) electrostatic and van der Waals interactions between QM and MM atoms. Our simulation program obtains forces within the QM region and the electrostatic QM/MM interactions through an interface with TeraChem [117, 54], while the forces within the MM region and the Lennard-Jones QM/MM interactions are acquired via an MPI interface with the LAMMPS molecular dynamics package [91]. A snapshot of the QM/MM system is shown in Figure 5(a).

The impact of NQEs can be understood by comparing AIMD and AI-PIMD simulations of KSI\(^{D40N} \). Figure 5(b) shows that NQEs significantly decrease the free energy required to move the proton from Tyr16 to Tyr57 (\( \nu_{16} \geq 0 \)), thus allowing the proton to be quantum mechanically delocalized in the active-site hydrogen bond network. This is manifested by the wide spread of the protons’ ring polymer beads, which represents their position uncertainty (Figure 5(a)). This quantum delocalization leads to a 10,000 fold increase in the acidity dissociation constant of Tyr57 and a \( \Delta \Delta pK_a \) of 0.50 compared to tyrosine in solution, in excellent agreement with experiment [121].

Due to the short R in the active-site hydrogen bond network, the ZPE possessed by the O–H bonds and \( \Delta E_{\nu=0} \) are both about 5 kcal/mol (Figure 6), which facilitates proton movements along \( \nu_{16} \). Hence the interplay between nuclear and electronic quantum effects qualitatively and quantitatively changes the proton behavior from classical hydrogen bonding to quantum delocalization.
4. Outlook

With the development of new algorithms and their combination with novel computer streaming architectures, path integral methods and AIMD simulations can now be combined to elucidate the movements of electrons and light nuclei in unprecedented detail during biological processes. In addition, recent developments will accelerate methods such as CMD [14, 57] and RPMD [28, 45] to obtain approximate quantum dynamics. These methods are not amenable to acceleration using the PIGLET approach. However, very recently it has been shown how the ring-polymer contraction (RPC) approach [75, 32], which exploits the separation of the forces acting on each copy of the system (Eq. 1) into rapidly and slowly varying parts, can be extended to AI-PIMD and RPMD simulations [77, 61].

As was illustrated in Sec. 3.4, AI-PIMD simulations can already provide novel insights into the structure and functional roles of short hydrogen bonds (R ≤ 2.6 Å) in biological systems. Another area where AI-PIMD simulations may be able to offer important insights is the study of proton coupled electron transfer (PCET) reactions, which play a fundamental role in photosynthesis, respiration, and enzymatic catalysis [24, 29, 46]. A comprehensive understanding of these biological processes requires simulation techniques that properly describe the quantum mechanical nature of the electrons and the light nuclei. With the recent advent of efficient AI-PIMD simulations, these techniques will provide powerful tools to aid in the unravelling of the function of complex biological systems.

5. Acknowledgements

C.M.I. and T.E.M. are supported by U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Award Number DE-SC0014437. L.W. acknowledges a postdoctoral fellowship from the Stanford Center for Molecular Analysis and Design. T.E.M. also acknowledges support from a Cottrell Scholarship from the Research Corporation for Science Advancement and an Alfred P. Sloan Research fellowship.

6. References

References

[1] Altun, A., Shaik, S., Thiel, W., 2006. Systematic qm/mm investigation of factors that affect the cytochrome p450-catalyzed hydrogen abstraction of camphor. J. Comput. Chem. 27, 1324–1337. URL: http://dx.doi.org/10.1002/jcc.20398, doi:10.1002/jcc.20398.

[2] Autschbach, J., Srebro, M., 2014. Delocalization error and “functional tuning” in kohn–sham calculations of molecular properties. Accounts of Chemical Research 47, 2592–2602. URL: http://dx.doi.org/10.1021/ar500171t, doi:10.1021/ar500171t.

[3] Baer, R., Livshits, E., Salzner, U., 2010. Tuned Range-Separated Hybrids in Density Functional Theory, in: Leone, SR and Cremer, PS and Groves, JT and Johnson, MA and Richmond, G (Ed.), ANNUAL REVIEW OF PHYSICAL CHEMISTRY, VOL 61. volume 61 of Ann. Rev. Phys. Chem., pp. 85–109.

[4] Baer, R., Neuhauser, D., 2005. Density functional theory with correct long-range asymptotic behavior. Phys. Rev. Lett. 94, 043002. URL: http://link.aps.org/doi/10.1103/PhysRevLett.94.043002, doi:10.1103/PhysRevLett.94.043002.
[5] Becke, A.D., 1993. Density-functional thermochemistry. iii. the role of exact exchange. J. Chem. Phys. 98, 5648–5652. URL: http://scitation.aip.org/content/aip/journal/jcp/98/7/10.1063/1.464913
[6] Becke, A.D., Johnson, E.R., 2005. Exchange-hole dipole moment and the dispersion interaction. J. Chem. Phys. 122. URL: http://scitation.aip.org/content/aip/journal/jcp/122/15/10.1063/1.1884601
[7] Becke, A.D., Johnson, E.R., 2007. Exchange-hole dipole moment and the dispersion interaction revisited. J. Chem. Phys. 127. URL: http://scitation.aip.org/content/aip/journal/jcp/127/15/10.1063/1.2795701
[8] Benkovic, S.J., Hammes-Schiffer, S., 2003. A perspective on enzyme catalysis. Science 301, 1196–1202. URL: http://science.sciencemag.org/content/301/5637/1196
[9] Berne, B.J., Thirumalai, D., 1986. On the Simulation of Quantum Systems: Path Integral Methods. Ann. Rev. Phys. Chem. 37, 401–424.
[10] Bowler, D.R., Miyazaki, T., 2012. O(n) methods in electronic structure calculations. Reports on Progress in Physics 75, 036503. URL: http://stacks.iop.org/0034-4885/75/i=3/a=036503
[11] Bowler, D.R., Miyazaki, T., Gillan, M.J., 2002. Recent progress in linear scaling ab initio electronic structure techniques. J. Phys.-Condens. Mat. 14, 2781. URL: http://stacks.iop.org/0953-8984/14/i=11/a=303
[12] Burke, K., 2012. Perspective on density functional theory. J. Chem. Phys. 136. URL: http://scitation.aip.org/content/aip/journal/jcp/136/15/10.1063/1.4704546
[13] Cao, J., Berne, B.J., 1989. On energy estimators in path integral Monte Carlo simulations: Dependence of accuracy on algorithm. J. Chem. Phys. 91, 6359.
[14] Cao, J., Voth, G.A., 1994. The formulation of quantum statistical mechanics based on the Feynman path centroid density. IV. Algorithms for centroid molecular dynamics. J. Chem. Phys. 101, 6168–6183.
[15] Case, D.A., Darden, T.A., Cheatham, T.E., Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Walker, R.C., Zhang, W., Merz, K.M., Roberts, B., Hayik, S., Roitberg, A., Seabra, G., Swails, J., Goetz, A.W., Kolossváry, I., Wong, K.F., Paesani, F., Vanicek, J., Wolf, R.M., Liu, J., Wu, X., Brozell, S.R., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Wang, J., Hsieh, M.J., Cui, G., Roe, D.R., Mathews, D.H., Seetin, M.G., Salomon-Ferrer, R., Sagui, C., Babin, V., Luchko, T., Gusrav, S., Kovalenko, A., Kollman, P.A., 2012. Amber 12. URL: http://ambermd.org/
[16] Ceriotti, M., 2010. GLE4MD. epfl-cosmo.github.io/gle4md/.
[17] Ceriotti, M., Bussi, G., Parrinello, M., 2010a. Colored-Noise Thermostats à la Carte. J. Chem. Theory Comput. 6, 1170–1180.
[18] Ceriotti, M., Cuny, J., Parrinello, M., Manolopoulos, D.E., 2013. Nuclear quantum effects and hydrogen bond fluctuations in water. Proc. Natl. Acad. Sci. USA 110, 15591–6.
[19] Ceriotti, M., Manolopoulos, D.E., 2012. Efficient First-Principles Calculation of the Quantum Kinetic Energy and Momentum Distribution of Nuclei. Phys. Rev. Lett. 109, 100604.
[20] Ceriotti, M., Manolopoulos, D.E., Parrinello, M., 2011. Accelerating the convergence of path integral dynamics with a generalized Langevin equation. J. Chem. Phys. 134, 84104.
[21] Ceriotti, M., Markland, T.E., 2013. Efficient methods and practical guidelines for simulating isotope effects. J. Chem. Phys. 138, 014112.
[22] Ceriotti, M., Parrinello, M., Markland, T.E., Manolopoulos, D.E., 2010b. Efficient stochastic thermostatting of path integral molecular dynamics. J. Chem. Phys. 133, 124104.
[23] Chai, J.D., Head-Gordon, M., 2008. Systematic optimization of long-range corrected hybrid density functionals. J. Chem. Phys. 128, 084106.
[24] Chang, C.J., Chang, M.C., Damrauer, N.H., Nocera, D.G., 2004. Proton-coupled electron transfer: a unifying mechanism for biological charge transport, amino acid radical initiation and propagation, and bond making/breaking reactions of water and oxygen. BBA-Bioenergetics 1655, 13–28.
[25] Cleland, W.W., Frey, P.A., Gerlt, J.A., 1998. The low barrier hydrogen bond in enzymatic catalysis.
J. Biol. Chem. 273, 25529–25532. URL: [http://www.jbc.org/content/273/40/25529.short](http://www.jbc.org/content/273/40/25529.short).

[26] Cohen, A.J., Mori-Sánchez, P., Yang, W., 2008a. Fractional charge perspective on the band gap in density-functional theory. Phys. Rev. B 77, 115123.

[27] Cohen, A.J., Mori-Sánchez, P., Yang, W., 2008b. Insights into current limitations of density functional theory. Science 321, 792–794.

[28] Craig, I.R., Manolopoulos, D.E., 2004. Quantum statistics and classical mechanics: Real time correlation functions from ring polymer molecular dynamics. J. Chem. Phys. 121, 3368.

[29] Cukier, R., 2004. Theory and simulation of proton-coupled electron transfer, hydrogen-atom transfer, and proton translocation in proteins. BBA-Bioenergetics 1655, 37–44.

[30] Duan, Y., Wu, C., Chowdhury, S., Lee, M.C., Xiong, W., Yang, R., Cieplak, P., Luo, R., Lee, T., Caldwell, J., Wang, J., Kollman, P., 2003. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J. Comput. Chem. 24, 1999–2012. URL: [http://dx.doi.org/10.1002/jcc.10349](http://dx.doi.org/10.1002/jcc.10349), doi: [10.1002/jcc.10349](http://dx.doi.org/10.1002/jcc.10349).

[31] Elstner, M., Hobza, P., Frauenheim, T., Suhai, S., Kaxiras, E., 2001. Hydrogen bonding and stacking interactions of nucleic acid base pairs: A density-functional-theory based treatment. J. Chem. Phys. 114, 5149–5155. URL: [http://scitation.aip.org/content/aip/journal/jcp/114/12/10.1063/1.1329889](http://scitation.aip.org/content/aip/journal/jcp/114/12/10.1063/1.1329889), doi: [http://dx.doi.org/10.1063/1.1329889](http://dx.doi.org/10.1063/1.1329889).

[32] Fanourgakis, G.S., Markland, T.E., Manolopoulos, D.E., 2009. A fast path integral method for polarizable force fields. J. Chem. Phys. 131. URL: [http://scitation.aip.org/content/aip/journal/jcp/131/9/10.1063/1.3216520](http://scitation.aip.org/content/aip/journal/jcp/131/9/10.1063/1.3216520), doi: [10.1063/1.3216520](http://dx.doi.org/10.1063/1.3216520).

[33] Feynman, R.P., Hibbs, A.R., 1964. Quantum Mechanics and Path Integrals. McGraw-Hill, New York.

[34] Flaig, D., Beer, M., Ochsenfeld, C., 2012. Convergence of electronic structure with the size of the qm region: Example of qm/mm nmr shieldings. J. Chem. Theory Comput. 8, 2260–2271. URL: [http://dx.doi.org/10.1021/ct300036s](http://dx.doi.org/10.1021/ct300036s), doi: [10.1021/ct300036s](http://dx.doi.org/10.1021/ct300036s).

[35] Fox, S.J., Pittock, C., Fox, T., Tautermann, C.S., Malcolm, N., Skylaris, C.K., 2011. Electrostatic embedding in large-scale first principles quantum mechanical calculations on biomolecules. J. Chem. Phys. 135. URL: [http://scitation.aip.org/content/aip/journal/jcp/135/22/10.1063/1.3665893](http://scitation.aip.org/content/aip/journal/jcp/135/22/10.1063/1.3665893), doi: [10.1063/1.3665893](http://dx.doi.org/10.1063/1.3665893).

[36] Friesner, R.A., Guallar, V., 2005. Ab initio quantum chemical and mixed quantum mechanics/molecular mechanics (qm/mm) methods for studying enzymatic catalysis. Annual Review of Physical Chemistry 56, 389–427. URL: [http://dx.doi.org/10.1146/annurev.physchem.55.091602.094410](http://dx.doi.org/10.1146/annurev.physchem.55.091602.094410), doi: [10.1146/annurev.physchem.55.091602.094410](http://dx.doi.org/10.1146/annurev.physchem.55.091602.094410).

[37] Gao, J., Truhlar, D.G., 2002. Quantum mechanical methods for enzyme kinetics. Annu. Rev. Phys. Chem. 53, 467–505. URL: [http://dx.doi.org/10.1146/annurev.physchem.53.091301.150114](http://dx.doi.org/10.1146/annurev.physchem.53.091301.150114), doi: [10.1146/annurev.physchem.53.091301.150114](http://dx.doi.org/10.1146/annurev.physchem.53.091301.150114).

[38] Gaus, M., Cui, Q., Elstner, M., 2014. Density functional tight binding: application to organic and biological molecules. Wiley Interdisciplinary Reviews: Computational Molecular Science 4, 49–61. URL: [http://dx.doi.org/10.1002/wcms.1156](http://dx.doi.org/10.1002/wcms.1156), doi: [10.1002/wcms.1156](http://dx.doi.org/10.1002/wcms.1156).

[39] Gill, P.M.W., Adamson, R.D., Pople, J.A., 1996. Coulomb-attenuated exchange energy density functionals. Mol. Phys. 88, 1005–1009.

[40] Goedecker, S., 1999. Linear scaling electronic structure methods. Rev. Mod. Phys. 71, 1085–1123. URL: [http://link.aps.org/doi/10.1103/RevModPhys.71.1085](http://link.aps.org/doi/10.1103/RevModPhys.71.1085), doi: [10.1103/RevModPhys.71.1085](http://dx.doi.org/10.1103/RevModPhys.71.1085).

[41] Grimme, S., 2006. Semiempirical gga-type density functional constructed with a long-range dispersion correction. J. Comput. Chem. 27, 1787–1799. URL: [http://dx.doi.org/10.1002/jcc.20495](http://dx.doi.org/10.1002/jcc.20495), doi: [10.1002/jcc.20495](http://dx.doi.org/10.1002/jcc.20495).

[42] Grimme, S., 2011. Density functional theory with london dispersion corrections. Wiley Interdisciplinary Reviews: Computational Molecular Science 1, 211–228. URL: [http://dx.doi.org/10.1002/wcms.30](http://dx.doi.org/10.1002/wcms.30), doi: [10.1002/wcms.30](http://dx.doi.org/10.1002/wcms.30).
[43] Grimme, S., Antony, J., Ehrlich, S., Krieg, H., 2010. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. J. Chem. Phys. 132, 154104.

[44] Ha, N.C., Kim, M.S., Lee, W., Choi, K.Y., Oh, B.H., 2000. Detection of large pKa perturbations of an inhibitor and a catalytic group at an enzyme active site, a mechanistic basis for catalytic power of many enzymes. J. Biol. Chem. 275, 41100–41106. URL: http://www.jbc.org/content/275/52/41100.abstract.

[45] Habershon, S., Manolopoulos, D.E., Markland, T.E., III, T.F.M., 2013. Ring-polymer molecular dynamics: Quantum effects in chemical dynamics from classical trajectories in an extended phase space. Annual Review of Physical Chemistry 64, 387–413. URL: http://dx.doi.org/10.1146/annurev-physchem-040412-110122 arXiv:http://dx.doi.org/10.1146/annurev-physchem-040412-110122.

[46] Hammes-Schiffer, S., Stuchebrukhov, A.A., 2010. Theory of coupled electron and proton transfer reactions. Chem. Rev. 110, 6939–6960. doi: 10.1021/cr1001436, arXiv:http://dx.doi.org/10.1021/cr1001436.

[47] Haworth, N.L., Bacskay, G.B., 2002. Heats of formation of phosphorus compounds determined by current methods of computational quantum chemistry. J. Chem. Phys. 117, 11175–11187. URL: http://scitation.aip.org/content/aip/journal/jcp/117/24/10.1063/1.1521760, doi: http://dx.doi.org/10.1063/1.1521760.

[48] Heaton-Burgess, T., Yang, W., 2010. Structural manifestation of the delocalization error of density functional approximations: C4N+2 rings and C-20 bowl, cage, and ring isomers. J. Chem. Phys. 132, 234113.

[49] Henry, D.J., Parkinson, C.J., Radom, L., 2002. An assessment of the performance of high-level theoretical procedures in the computation of the heats of formation of small open-shell molecules. J. Phys. Chem. A 106, 7927–7936. URL: http://dx.doi.org/10.1021/jp0260752, doi:10.1021/jp0260752, arXiv:http://dx.doi.org/10.1021/jp0260752.

[50] Herman, M.F., Bruskin, E.J., Berne, B.J., 1982. On path integral Monte Carlo simulations. J. Chem. Phys. 76, 5150.

[51] Hohenberg, P., Kohn, W., 1964. Inhomogeneous electron gas. Phys. Rev. 136, B864–B871. URL: http://link.aps.org/doi/10.1103/PhysRev.136.B864, doi:10.1103/PhysRev.136.B864.

[52] Hu, L., Söderhjelm, P., Ryde, U., 2011. On the convergence of qm/mm energies. J. Chem. Theory Comput. 7, 761–777. URL: http://dx.doi.org/10.1021/ct100530r, doi:10.1021/ct100530r arXiv:http://dx.doi.org/10.1021/ct100530r.

[53] Iikura, H., Tsuneda, T., Yanai, T., Hirao, K., 2001. A long-range correction scheme for generalized-gradient-exchange functional. J. Chem. Phys. 115, 3540–3544.

[54] Isborn, C.M., Götz, A.W., Clark, M.A., Walker, R.C., Martínez, T.J., 2012. Electronic absorption spectra from mm and ab initio qm/mm molecular dynamics: Environmental effects on the absorption spectrum of photoactive yellow protein. J. Chem. Theory Comput. 8, 5092–5106. URL: http://dx.doi.org/10.1021/ct3006826, doi:10.1021/ct3006826 arXiv:http://dx.doi.org/10.1021/ct3006826.

[55] Isborn, C.M., Luehr, N., Ulitntsev, I.S., Martínez, T.J., 2011. Excited-state electronic structure with configuration interaction singles and tamm-dancoff time-dependent density functional theory on graphical processing units. J. Chem. Theory Comput. 7, 1814–1823. URL: http://dx.doi.org/10.1021/ct200030k, doi:10.1021/ct200030k arXiv:http://dx.doi.org/10.1021/ct200030k.

[56] Isborn, C.M., Mar, B.D., Curchod, B.F.E., Tavernelli, I., Martínez, T.J., 2013. The charge transfer problem in density functional theory calculations of aquously solvated molecules. J. Phys. Chem. B 117, 12189–12201. URL: http://dx.doi.org/10.1021/jp4058274, doi:10.1021/jp4058274 arXiv:http://dx.doi.org/10.1021/jp4058274.

[57] Jang, S., Voth, G.A., 1999. A derivation of centroid molecular dynamics and other approximate time evolution methods for path integral centroid variables. J. Chem. Phys. 111, 2371–2384. URL: http://scitation.aip.org/content/aip/journal/jcp/111/6/10.1063/1.479515, doi:http://dx.doi.org.
[58] Johnson, E.R., Becke, A.D., 2006. A post-hartree-fock model of intermolecular interactions: Inclusion of higher-order corrections. J. Chem. Phys. 124. URL: http://scitation.aip.org/content/aip/journal/jcp/124/17/10.1063/1.2190220, doi:10.1063/1.2190220

[59] Johnson, E.R., Mackie, I.D., DiLabio, G.A., 2009. Dispersion interactions in density-functional theory. J. Phys. Org. Chem. 22, 1127–1135. URL: http://dx.doi.org/10.1002/poc.1606, doi:10.1002/poc.1606

[60] Jorgensen, W.W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., Klein, M.L., 1983. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79, 926.

[61] Kapil, V., VandeVondele, J., Ceriotti, M., 2016. Accurate molecular dynamics and nuclear quantum effects at low cost by multiple steps in real and imaginary time: Using density functional theory to accelerate wavefunction methods. J. Chem. Phys. 144, 054111.

[62] Kim, M.C., Sim, E., Burke, K., 2013. Understanding and reducing errors in density functional calculations. Phys. Rev. Lett. 111, 073003.

[63] Kliman, J.P., Kohen, A., 2013. Hydrogen tunneling links protein dynamics to enzyme catalysis. Annu. Rev. Biochem. 82, 471–496. URL: http://dx.doi.org/10.1146/annurev-biochem-051710-133623, doi:10.1146/annurev-biochem-051710-133623

[64] Kohn, W., Sham, L.J., 1965. Self-consistent equations including exchange and correlation effects. Phys. Rev. 140, A1133–A1138. URL: http://link.aps.org/doi/10.1103/PhysRev.140.A1133, doi:10.1103/PhysRev.140.A1133

[65] Kraissler, E., Kronik, L., 2014. Fundamental gaps with approximate density functionals: The derivative discontinuity revealed from ensemble considerations. J. Chem. Phys. 140, 18A540.

[66] Kronik, L., Stein, T., Refaely-Abramson, S., Baer, R., 2012. Excitation gaps of finite-sized systems from optimally tuned range-separated hybrid functionals. J. Chem. Theory Comput. 8, 1515–1531. URL: http://dx.doi.org/10.1021/ct2009363, doi:10.1021/ct2009363

[67] Lennartz, C., Schäfer, A., Terstegen, F., Thiel, W., 2002. Enzymatic reactions of triosephosphate isomerase: A theoretical calibration study. J. Phys. Chem. B 106, 1758–1767. URL: http://dx.doi.org/10.1021/jp012658k, doi:10.1021/jp012658k

[68] Lin, H., Truhlar, D.G., 2006. Qm/mm: what have we learned, where are we, and where do we go from here? Theoretical Chemistry Accounts 117, 185–199. URL: http://dx.doi.org/10.1007/s00214-006-0143-z, doi:10.1007/s00214-006-0143-z

[69] Luehr, N., Sisto, A., Martinez, T.J., 2016. Gaussian basis set hartree-fock, density functional theory, and beyond on gpus, in: Walker, R., Goetz, A. (Eds.), Electronic Structure Calculations on Graphics Processing Units: From Quantum Chemistry to Condensed Matter Physics. Wiley, West Sussex. chapter 4, pp. 67–100.

[70] Lennartz, C., Schäfer, A., Terstegen, F., Thiel, W., 2002. Enzymatic reactions of triosephosphate isomerase: A theoretical calibration study. J. Phys. Chem. B 106, 1758–1767. URL: http://dx.doi.org/10.1021/jp012658k, doi:10.1021/jp012658k

[71] Markland, T.E., Manolopoulos, D.E., 2008a. A refined ring polymer contraction scheme for systems with electrostatic interactions. Chem. Phys. Lett. 464, 256.

[72] Markland, T.E., Manolopoulos, D.E., 2008b. An efficient ring polymer contraction scheme for imaginary time path integral simulations. J. Chem. Phys. 129, 024105.
[76] Marsalek, O., Chen, P.Y., Dupuis, R., Benoit, M., Meheut, M., Bacic, Z., Tuckerman, M.E., 2014. Efficient calculation of free energy differences associated with isotopic substitution using path-integral molecular dynamics. J. Chem. Theory Comput. 10, 1440–1453. URL: http://dx.doi.org/10.1021/ct400911m

[77] Marsalek, O., Markland, T.E., 2016. Ab initio molecular dynamics with nuclear quantum effects at classical cost: Ring polymer contraction for density functional theory. J. Chem. Phys. 144, 054112.

[78] Marx, D., Parrinello, M., 1996. Ab initio path integral molecular dynamics: Basic ideas. J. Chem. Phys. 104, 4077.

[79] McKenzie, R.H., 2012. A diabatic state model for donor-hydrogen vibrational frequency shifts in hydrogen bonded complexes. Chem. Phys. Lett. 535, 196–200.

[80] Mildvan, A., Massiah, M., Harris, T., Marks, G., Harrison, D., Viragh, C., Reddy, P., Kovach, I., 2002. Short, strong hydrogen bonds on enzymes: {NMR} and mechanistic studies. J. Mol. Struct. 615, 163–175. URL: http://www.sciencedirect.com/science/article/pii/S0022286002002120, doi: http://dx.doi.org/10.1016/S0022-2860(02)00212-0

[81] Mori-Sanchez, P., Cohen, A.J., Yang, W., 2006. Many-electron self-interaction error in approximate density functionals. J. Chem. Phys. 125, 201102.

[82] Mori-Sánchez, P., Cohen, A.J., Yang, W., 2008. Localization and delocalization errors in density functional theory and implications for band-gap prediction. Phys. Rev. Lett. 100, 146401.

[83] Morrone, J.A., Car, R., 2008. Nuclear Quantum Effects in Water. Phys. Rev. Lett. 101, 17801.

[84] Nielsen, J.E., McCammon, J.A., 2003. Calculating pka values in enzyme active sites. Protein Science: A Publication of the Protein Society 12, 1894–1901. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2323987/

[85] Plimpton, S., 1995. Fast Parallel Algorithms for Short-Range Molecular Dynamics. J. Comp. Phys. 117, 1–19.

[86] Ponder, J.W., Case, D.A., 2003. Force fields for protein simulations. Advances in Protein Chemistry 66, 27 – 85.

[87] Perdew, J.P., Yue, W., 1986. Accurate and simple density functional for the electronic exchange energy: Generalized gradient approximation. Phys. Rev. B 33, 8800–8802. URL: http://link.aps.org/doi/10.1103/PhysRevB.33.8800

[88] Perrin, C.L., Nielson, J.B., 1997. "Strong" hydrogen bonds in chemistry and biology. Annual Review of Physical Chemistry 48, 511–544. URL: http://dx.doi.org/10.1146/annurev.physchem.48.1.511, doi: 10.1146/annurev.physchem.48.1.511

[89] Refaely-Abramson, S., Baer, R., Kronik, L., 2011. Fundamental and excitation gaps in molecules of...
relevance for organic photovoltaics from an optimally tuned range-separated hybrid functional. Phys. Rev. B 84, 075144.

[95] Refaely-Abramson, S., Sharifzadeh, S., Govind, N., Autschbach, J., Neaton, J.B., Baer, R., Kronik, L., 2012. Quasiparticle Spectra from a Nonempirical Optimally Tuned Range-Separated Hybrid Density Functional. Phys. Rev. Lett. 109, 226405.

[96] Riccardi, D., Schaefer, P., Yang, Y., Yu, H., Ghosh, N., Prat-Resina, X., König, P., Li, G., Xu, D., Guo, H., Elstner, M., Cui, Q., 2006. Development of effective quantum mechanical/molecular mechanical (qm/mm) methods for complex biological processes. J. Phys. Chem. B 110, 6458–6469. URL: http://dx.doi.org/10.1021/jp056361o, doi: 10.1021/jp056361o, arXiv: http://dx.doi.org/10.1021/jp056361o.

[97] Salzner, U., Baer, R., 2009. Koopmans' springs to life. J. Chem. Phys. 131, 231101.

[98] Schmidt, T., Kraisler, E., Makmal, A., Kronik, L., Kuemmel, S., 2014. A self-interaction-free local hybrid functional: Accurate binding energies vis-a-vis accurate ionization potentials from Kohn-Sham eigenvalues. J. Chem. Phys. 140, 18A510.

[99] Seidl, A., Görling, A., Vogl, P., Majewski, J.A., Levy, M., 1996. Generalized kohn-sham schemes and the band-gap problem. Phys. Rev. B 53, 3764–3774. URL: http://link.aps.org/doi/10.1103/PhysRevB.53.3764, doi: 10.1103/PhysRevB.53.3764.

[100] Simonson, T., Archontis, G., Karplus, M., 2002. Free energy simulations come of age: Protein-ligand recognition. Accounts of Chemical Research 35, 430–437. URL: http://dx.doi.org/10.1021/ar010030m, doi: 10.1021/ar010030m.

[101] Sutcliffe, M.J., Scrutton, N.S., 2002. A new conceptual framework for enzyme catalysis. Eur. J. Biochem. 269, 3096–3102. URL: http://dx.doi.org/10.1046/j.1432-1033.2002.03020.x.

[102] Tao, J., Perdew, J.P., Staroverov, V.N., Scuseria, G.E., 2003. Climbing the density functional ladder: Nonempirical meta-generalized gradient approximation designed for molecules and solids. Phys. Rev. Lett. 91, 146401. URL: http://link.aps.org/doi/10.1103/PhysRevLett.91.146401.
[111] Tawada, Y., Tsumeda, T., Yanagisawa, S., Yanai, T., Hirao, K., 2004. A long-range-corrected time-dependent density functional theory. J. Chem. Phys. 120, 8425–8433.

[112] Toulouse, J., Colonna, F., Savin, A., 2004. Long-range–short-range separation of the electron-electron interaction in density-functional theory. Phys. Rev. A 70, 062505.

[113] Tozer, D.J., 2003. Relationship between long-range charge-transfer excitation energy error and integer discontinuity in kohn-sham theory. J. Chem. Phys. 119, 12697–12699. URL: http://scitation.aip.org/content/aip/journal/jcp/119/24/10.1063/1.1633756, doi:10.1063/1.1633756.

[114] Ufimtsev, I.S., Luehr, N., Martinez, T.J., 2011. Charge transfer and polarization in solvated proteins from ab initio molecular dynamics. J. Phys. Chem. Lett. 2, 1789–1793. URL: http://dx.doi.org/10.1021/jz200697c, arXiv:http://dx.doi.org/10.1021/jz200697c.

[115] Ufimtsev, I.S., Martinez, T.J., 2008. Quantum chemistry on graphical processing units. 1. strategies for two-electron integral evaluation. J. Chem. Theory Comput. 4, 222–231. URL: http://dx.doi.org/10.1021/ct700268q, doi:10.1021/ct700268q, arXiv:http://dx.doi.org/10.1021/ct700268q.

[116] Ufimtsev, I.S., Martinez, T.J., 2009a. Quantum chemistry on graphical processing units. 2. direct self-consistent-field implementation. J. Chem. Theory Comput. 5, 1004–1015. URL: http://dx.doi.org/10.1021/ct9003004, doi:10.1021/ct9003004, arXiv:http://dx.doi.org/10.1021/ct9003004.

[117] Ufimtsev, I.S., Martinez, T.J., 2009b. Quantum chemistry on graphical processing units. 3. analytical energy gradients, geometry optimization, and first principles molecular dynamics. J. Chem. Theory Comput. 5, 2619–2628. URL: http://dx.doi.org/10.1021/ct9003004, doi:10.1021/ct9003004, arXiv:http://dx.doi.org/10.1021/ct9003004.

[118] Vanicek, J., Miller, W.H., 2007. Efficient estimators for quantum instanton evaluation of the kinetic isotope effects: Application to the intramolecular hydrogen transfer in pentadiene. J. Chem. Phys. 127, 114309. URL: http://scitation.aip.org/content/aip/journal/jcp/127/11/10.1063/1.2768930, doi:10.1063/1.2768930.

[119] Walovist, A., Berne, B., 1985. Path-integral simulation of pure water. Chem. Phys. Lett. 117, 214–219. URL: http://dx.doi.org/10.1016/0009-2614(85)80206-2.

[120] Wang, L., Ceriotti, M., Markland, T.E., 2014a. Quantum fluctuations and isotope effects in ab initio descriptions of water. J. Chem. Phys. 141, 104502. URL: http://scitation.aip.org/content/aip/journal/jcp/141/10/10.1063/1.4894287, doi:10.1063/1.4894287.

[121] Wang, L., Fried, S.D., Boxer, S.G., Markland, T.E., 2014b. Quantum delocalization of protons in the hydrogen-bond network of an enzyme active site. Proc. Natl. Acad. Sci. U.S.A. 111, 18454–18459. URL: http://www.pnas.org/content/111/52/18454.

[122] Wang, L.P., Titov, A., McGibbon, R., Liu, F., Pande, V.S., Martínez, T.J., 2014c. Discovering chemistry with an ab initio nanoreactor. Nat Chem 6, 1044–1048. URL: http://dx.doi.org/10.1038/nchem.2099.

[123] Wang, N.X., Wilson, A.K., 2004. The behavior of density functionals with respect to basis set. i. the correlation consistent basis sets. J. Chem. Phys. 121, 7632–7646. URL: http://scitation.aip.org/content/aip/journal/jcp/121/16/10.1063/1.1792071, doi:http://dx.doi.org/10.1063/1.1792071.

[124] Warshel, A., Levitt, M., 1976. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carboxonium ion in the reaction of lysozyme. J. Mol. Biol. 103, 227 – 240. URL: http://www.sciencedirect.com/science/article/pii/0022283679903119, doi:http://dx.doi.org/10.1016/0022-2836(76)90311-9.

[125] Whittleton, S.R., Sosa Vazquez, X.A., Isborn, C.M., Johnson, E.R., 2015. Density-functional errors in ionization potential with increasing system size. J. Chem. Phys. 142, 184106. URL: http://scitation.aip.org/content/aip/journal/jcp/142/18/10.1063/1.4920947, doi:http://dx.doi.org/10.1063/1.4920947.

[126] Wödrich, M.D., Jana, D.F., von Ragüé Schleyer, P., Corminboeuf, C., 2008. Empirical corrections to
density functional theory highlight the importance of nonbonded intramolecular interactions in alkanes. J. Phys. Chem. A 112, 11495–11500. URL: http://dx.doi.org/10.1021/jp806619z, doi:10.1021/jp806619z

[127] Wu, X., Vargas, M.C., Nayak, S., Lotrich, V., Scoles, G., 2001. Towards extending the applicability of density functional theory to weakly bound systems. J. Chem. Phys. 115, 8748–8757. URL: http://scitation.aip.org/content/aip/journal/jcp/115/19/10.1063/1.1412004, doi:10.1063/1.1412004.

[128] Yanai, T., Tew, D., Handy, N., 2004. A new hybrid exchange-correlation functional using the Coulomb-attenuating method (CAM-B3LYP). Chemical Physics Letters 393, 51–57.

[129] Zhang, J., Kulik, H.J., Martinez, T.J., Klinman, J.P., 2015. Mediation of donor–acceptor distance in an enzymatic methyl transfer reaction. Proc. Natl. Acad. Sci. U.S.A. 112, 7954–7959.

[130] Zhang, Y., Yang, W., 1998. A challenge for density functionals: Self-interaction error increases for systems with a noninteger number of electrons. J. Chem. Phys. 109, 2604–2608.

[131] Zhao, Q., Xie, L., Kulik, H.J., 2015. Discovering amorphous indium phosphide nanostructures with high-temperature ab initio molecular dynamics. J. Phys. Chem. C 119, 23238–23249. URL: http://dx.doi.org/10.1021/acs.jpcc.5b07264, doi:10.1021/acs.jpcc.5b07264.

[132] Zheng, X., Liu, M., Johnson, E.R., Contreras-Garcia, J., Yang, W., 2012. Delocalization error of density-functional approximations: A distinct manifestation in hydrogen molecular chains. J. Chem. Phys. 137, 214106.

![Figure 1: Potential energy profile for proton transfer in the active-site hydrogen bond network of KSI$^{D40N}$](image)

Figure 1: Potential energy profile for proton transfer in the active-site hydrogen bond network of KSI$^{D40N}$, as calculated at the B3LYP-D3 level \[^5\]^\[^43\] using five different basis sets. The proton transfer coordinate $\nu_{16} = d_{O16,H16} - d_{O57,H16}$ and $\nu_{16} \geq 0$ represents a proton transfer from residue Tyr16 to Tyr57.
Figure 2: Distributions of bond length for experimental structures (black) compared to those obtained from structural optimization with the classical AMBER force field (orange), ωPBEh/STO-3G (red), ωPBEh/3-21G (green), and ωPBEh/6-31G (blue) for a collection of 58 proteins. (Reproduced from reference [68] with permission from American Chemical Society.)
Figure 3: The DFT error in size-dependent ionization potential is shown by analyzing the neutral and cation SCF density differences for ethene surrounded by 25 water molecules for the LC-BLYP functional with $\omega=0.1$, $0.3$, and $0.5 \, \text{a}_0^{-1}$. The purple isosurface shows where electron density has been removed, and the blue isosurface shows where electron density has been gained. The correct polarization response of the solvent is seen for $\omega=0.5 \, \text{a}_0^{-1}$, where the water molecules polarize their electron density towards the cationic solute. Smaller $\omega$ values show the solvent being incorrectly partially ionized due to the DFT delocalization error present when using approximate exchange based on the local spin density.
Figure 4: Dependence of the PYP excited state energy on the size of the QM region, as obtained from QM/MM calculations for three different PYP configurations. Very large QM regions are required to generate the converged excitation energies. The y-axis is the difference in the computed excited state energy compared to that for QM region 1.

Figure 5: (a) A snapshot of KSI$^{D_{40}N}$ from the AI-PIMD simulation. Its active-site hydrogen bond network enlarged, in which green, red and white represent carbon, oxygen and hydrogen atoms. The blue spheres are the full ring-polymer representation of the protons. For clarity, all the other atoms are shown as their centroids. (b) The free energy surface of the proton movement along $\nu_{16}$, as calculated from AIMD and AI-PIMD simulations.
Figure 6: Potential energy profile for proton transfer in KSI$^{140N}$, as obtained using five density functionals and the 6-31G* basis set.