CHEMICAL REVIEWS

Review

pubs.acs.org/CR

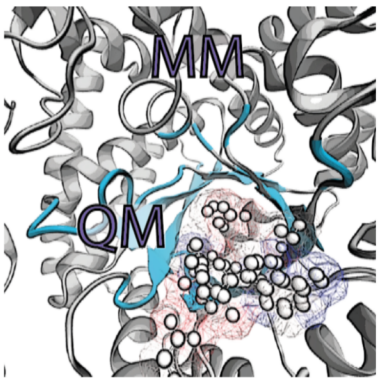
Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States

Elizabeth Brunk[†,‡] and Ursula Rothlisberger*[,†,§]

[†]Laboratory of Computational Chemistry and Biochemistry, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

[‡]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, California 94618, United States

[§]National Competence Center of Research (NCCR) MARVEL—Materials' Revolution: Computational Design and Discovery of Novel Materials, 1015 Lausanne, Switzerland



CONTENTS

1. INTRODUCTION: QUANTUM MECHANICAL PHENOMENA IN BIOLOGICAL SYSTEMS

The quantum nature of electrons and nuclei is manifested in countless biological events including the rearrangements of electrons in biochemical reactions, electron and proton tunneling, coupled proton−electron transfers, photoexcitations,

Special Issue: Calculations on Large Systems

Received: October 30, 2014
Published: April 16, 2015

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

and long-lived quantum coherences and quantum entanglement, observed in biological energy transfer.[1] Quantum mechanical (QM) phenomena are thus at the base of fundamental biological processes, such as light harvesting, photosynthesis, respiration, magnetoreception, and our sensory perceptions of vision, olfaction, and taste. However, describing such events with quantum mechanics seems essentially unreachable, considering the large size of biological macromolecules (which are often comprised of 10,000−100,000 atoms) and the extended time scales (which range from ultrafast electronic processes on the order of atto- to femtoseconds to events that take place on time scales longer than seconds). Furthermore, both the large thermally accessible conformational space and the extremely small relevant energy scale on the order of kT (i.e., <1 kcal/mol) pose serious challenges to the accuracy of a computational method. (Typical melting temperatures of mesophilic proteins are in the range of 25−53 °C; that is, typical thermal energies of $k\Delta T$ < 1 kcal/mol are responsible for the change from native to denatured structures.)

A seminal step toward the quantum mechanical treatment of realistic biological systems has been accomplished in 1976 by Warshel and Levitt[2] through the introduction of mixed quantum mechanical/molecular mechanical (QM/MM) methods that allow one to treat (electronic) quantum phenomena in complex classical environments. Extending the QM/MM approach from adiabatic simulations in the electronic ground state to nonadiabatic dynamics in electronically excited states also compels an account of the quantum nature of the nuclear degrees of freedom adding a further layer of complexity. Here, we review the current state of the art of QM/MM molecular dynamics approaches in ground and electronically excited states and their applications to biological problems. We start from the basic theory of the full time-dependent quantum problem of the combined electronic−nuclear system and present shortly the most widely used trajectory-based methods to treat the dynamics of electrons and nuclei. In section 2, we discuss theoretical as well as practical issues of QM/MM approaches for adiabatic and nonadiabatic dynamics as well as some of the current limitations and ways to overcome them. Section 3 gives some typical examples of applications of QM/MM simulations in ground and excited states. Sections 4 and 5 are dedicated to cross-disciplinary approaches that link the detailed electronic and atomic information gained in a QM/MM simulation of a single biological macromolecule to a wider biological context through the combination with bioinformatics and systems biology approaches. This Review ends with some conclusions and possible future directions.

2. BASIC THEORY

2.1. Starting from the Beginning: The Full Quantum Problem

For a complete description of quantum phenomena, the quantum nature of both electrons and nuclei has to be taken into account. The dynamic evolution of a (nonrelativistic) quantum system is determined by the time-dependent Schrödinger equation (TDSE):

$$i\hbar\frac{\partial}{\partial t}\Phi(\mathbf{r},\mathbf{R},t) = \mathbb{H}(\mathbf{r},\mathbf{R})\Phi(\mathbf{r},\mathbf{R},t) \tag{1}$$

where the time-dependent wave function $\Phi(\mathbf{r},\mathbf{R},t)$ is a function of all electronic $\mathbf{r} = \{\mathbf{r}_i\}$ and nuclear $\mathbf{R} = \{\mathbf{R}_I\}$ positions and the Hamilton operator $\mathbb{H}$ is given by (for the sake of optimal clarity of the derivation, atomic units are only introduced later in the text):

$$\begin{aligned}\mathbb{H} &= -\sum_I \frac{\hbar^2}{2M_I}\nabla_I^2 - \sum_i \frac{\hbar^2}{2m_e}\nabla_i^2 + \frac{1}{4\pi\varepsilon_0}\sum_{i\neq j}\frac{e^2}{|\mathbf{r}_i-\mathbf{r}_j|} \\ &\quad - \frac{1}{4\pi\varepsilon_0}\sum_{I,i}\frac{e^2 Z_I}{|\mathbf{R}_I-\mathbf{r}_i|} + \frac{1}{4\pi\varepsilon_0}\sum_{I\neq J}\frac{e^2 Z_I Z_J}{|\mathbf{R}_I-\mathbf{R}_J|} \\ &= -\sum_I \frac{\hbar^2}{2M_I}\nabla_I^2 - \sum_i \frac{\hbar^2}{2m_e}\nabla_i^2 + \mathbb{V}_{ne}(\mathbf{r},\mathbf{R}) \\ &= -\sum_I \frac{\hbar^2}{2M_I}\nabla_I^2 + \mathbb{H}_e(\mathbf{r},\mathbf{R})\end{aligned} \tag{2}$$

One way of simplifying the coupled electron−nuclear quantum problem is to expand the total wave function in a basis of stationary electronic states ψ_l that are solutions to the time-independent Schrödinger equation (TISE) at fixed nuclear positions (clamped nuclei approximation):

$$\mathbb{H}_e(\mathbf{r},\mathbf{R})\psi_l = E_l(\mathbf{R})\psi_l(\mathbf{r};\mathbf{R}) \tag{3}$$

Using the Born−Huang Ansatz[3] for the total time-dependent wave function:

$$\Phi(\mathbf{r},\mathbf{R},t) = \sum_{l=0}^{\infty}\psi_l(\mathbf{r};\mathbf{R})\chi_l(\mathbf{R},t) \tag{4}$$

the time-dependence can be incorporated in the form of time and nuclear position dependent coefficients $\chi_l(\mathbf{R},t)$ that represent the time-dependent nuclear wave function. Insertion of this Ansatz eq 4 into the TDSE eq 1 followed by multiplication from the left by $\psi_k^*(\mathbf{r};\mathbf{R})$ and integration over the electronic coordinates leads to a set of coupled differential equations for the time-dependent nuclear wave functions $\chi_k(\mathbf{R},t)$:

$$\left[-\sum_I \frac{\hbar^2}{2M_I}\nabla_I^2 + E_k(\mathbf{R})\right]\chi_k + \sum_l C_{kl}\chi_l = i\hbar\frac{\partial}{\partial t}\chi_k \tag{5}$$

The coefficients C_{kl} are due to the nonadiabatic coupling (NAC) between different electronic states:

$$C_{kl} = -\sum_I \frac{\hbar^2}{2M_I}D_{kl}^I(\mathbf{R}) + \sum_I \frac{\hbar^2}{M_I}\mathbf{d}_{kl}^I(\mathbf{R})\nabla_I \tag{6}$$

where $D_{kl}^I(\mathbf{R})$ are the second-order NAC elements (eq 7) and $\mathbf{d}_{kl}^I(\mathbf{R})$ are the first-order NAC vectors (eq 8):

$$D_{kl}^I(\mathbf{R}) = \int \psi_k^*\nabla_I^2\psi_l\,\mathrm{d}\mathbf{r} \tag{7}$$

$$\mathbf{d}_{kl}^I(\mathbf{R}) = \int \psi_k^*\nabla_I\psi_l\,\mathrm{d}\mathbf{r} \tag{8}$$

The electronic wave functions ψ_l and energy eigenvalues E_l of eq 3 can be determined either via precomputing or on-the-fly (direct dynamics[4]). Solving eq 5 thus yields the full array of quantum effects associated with the nuclear motion such as zero point energy, quantization of (vibrational) energy levels, tunneling and interference (coherence/decoherence) of nuclear wavepackets, as well as the NAC between electronic states (for a review, see ref 5). However, solutions of the nuclear quantum

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

dynamics via wavepacket propagation in a fixed basis set or grid representation[6] are costly and, due to the exponential scaling, typically limited to few degrees of freedom (<2−4 atoms). Although approximate quantum dynamics approaches with adaptable basis functions such as the Multiconfigurational Time Dependent Hartree method[7] or variational Gaussian wavepackets are more efficient[8] and allow the treatment of polyatomic systems, they are still limited to the solution of highly simplified (few modes − few states) models of biological problems.[9]

Because both NAC terms in eq 6 are indirectly proportional to the nuclear masses M_I, the NAC between different electronic states becomes small for heavy nuclear masses and the nuclear dynamics proceeds adiabatically on a single potential energy surface (PES) $E_k(\mathbf{R})$; that is, the (slow) nuclear degrees of freedom can be fully separated from the (fast) electronic degrees of freedom (Born−Oppenheimer (BO) approximation).

Even when the dynamics proceeds adiabatically on a single PES, a full treatment of the nuclear quantum dynamics is prohibitive for realistic biological systems, and further approximations have to be introduced.

2.1.1. Approximations to the Nuclear Dynamics. Casting the complex nuclear wave function in polar form with a time-dependent amplitude $A(\mathbf{R},t)$ and phase $S(\mathbf{R},t)$,

$$\chi_k(\mathbf{R}, t) = A_k(\mathbf{R}, t) \exp[iS_k(\mathbf{R}, t)/\hbar] \tag{9}$$

inserting this Ansatz into the TDSE eq 1 and separating real and imaginary parts leads to the equations:

$$\frac{\partial S_k}{\partial t} + \sum_I \frac{1}{2M_I}(\nabla_I S_k)^2 + E_k = \hbar^2 \sum_I \frac{1}{2M_I}\frac{\nabla_I^2 A_k}{A_k} \tag{10}$$

$$\frac{\partial A_k}{\partial t} + \sum_I \frac{1}{M_I}(\nabla_I A_k)(\nabla_I S_k) + \sum_I \frac{1}{2M_I} A_k(\nabla_I S_k) = 0 \tag{11}$$

Solution of the coupled eqs 10 and 11 by a trajectory-based approach (Bohmian dynamics[10]) constitutes an exact answer to the electronic−nuclear quantum problem and can also be extended to the nonadiabatic case and an on-the-fly approach.[11] Taking the classical limit $\hbar \to 0$ of the phase equation eq 10 leads to

$$\frac{\partial S_k}{\partial t} + \sum_I \frac{1}{2M_I}(\nabla_I S_k)^2 + E_k = 0 \tag{12}$$

which is isomorphic to the equation of motion in the Hamilton−Jacobi formulation of classical mechanics; that is, in this limit the nuclear quantum dynamics can be described by trajectories of classical point particles. A systematic expansion in terms of $\hbar$ leads instead to a variety of semiclassical methods (for a review, see ref 12).

2.1.1.1. Adiabatic On-the-Fly Dynamics with Classical Nuclei. As just shown, in the limit $\hbar \to 0$, the nuclei move according to classical mechanics in the effective potential E_k^{BO} given by the Born−Oppenheimer PES $E_k(\mathbf{R})$ of the kth electronic state:

$$M_I \ddot{\mathbf{R}}_I(t) = -\nabla_I E_k^{\mathrm{BO}}(\mathbf{R}(t)) \tag{13}$$

An efficient approach of obtaining $E_k(\mathbf{R})$ is to calculate the potential energy surface on-the-fly by solving the TISE (eq 3) at the fixed nuclear configurations $\mathbf{R}(t)$ that are visited during the dynamics. The most straightforward way of doing this is to solve eq 3 for an initial set of nuclear coordinates $\mathbf{R}_I(t = 0)$, to calculate the forces acting on the nuclei from the Hellman−Feynman theorem, and to use these to propagate the system according to classical mechanics (eq 13) to a new configuration $\mathbf{R}_I(t + \Delta t)$ (Born−Oppenheimer Dynamics, BOMD).[13] Because time steps for the integration of the classical equations of motion in ab initio BOMD are typically <1 fs, ten thousands of electronic structure calculations are required to cover time scales of tens of picoseconds. Efficient extrapolation schemes that can make optimal use of the optimized wave functions from previous steps are thus essential.[14]

As an alternative to the above approach, in 1985, Car and Parrinello introduced an elegant approach that makes use of fictitious classical dynamics of the electronic degrees of freedom to propagate them along the nuclear motion (Car−Parrinello Molecular Dynamics, CPMD).[15] The Car−Parrinello method is based on the extended Lagrangian $\mathcal{L}_{\mathrm{ex}}$:

$$\mathcal{L}_{\mathrm{ex}} = \mathcal{T}_{\mathrm{N}} + \mathcal{T}_{\mathrm{e}} - E_{\mathrm{pot}} \tag{14}$$

where $\mathcal{T}_{\mathrm{N}}$ is the kinetic energy of the nuclei, $\mathcal{T}_{\mathrm{e}}$ is the analogous term for the electronic degrees of freedom, and E_{pot} is the potential energy that depends on both nuclear positions $\mathbf{R}_I$ and electronic variables ϕ_i. $\mathcal{L}_{\mathrm{ex}}$ can be written as

$$\mathcal{L}_{\mathrm{ex}} = \sum_I \frac{1}{2} M \dot{\mathbf{R}}_I^2 + \sum_i \frac{1}{2}\mu |\dot{\phi}_i|^2 - \langle \psi_0 | \mathbb{H} | \psi_0 \rangle + \sum_{i,j} \Lambda_{ij}\left[\left\{\int \phi_i^*(\mathbf{r})\phi_j(\mathbf{r})\, \mathrm{d}\mathbf{r}\right\} - \delta_{ij}\right] \tag{15}$$

where the Lagrange multipliers Λ_{ij} ensure orthonormality of the wave functions ϕ_i, μ is a fictitious mass associated with the electronic degrees of freedom, and the potential energy is given by the expectation value of the total (ground state) energy $E = \langle \psi_0 | \mathbb{H} | \psi_0 \rangle$ of the system. The Lagrangian in eq 15 determines the time evolution of a fictitious classical system in which nuclear positions as well as electronic degrees of freedom are treated as dynamic variables. The classical equations of motion are given by the Euler−Lagrange equations:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}_i^*}\right) = \frac{\partial \mathcal{L}}{\partial q_i^*} \tag{16}$$

where q_i stands for a generalized coordinate. With the Lagrangian of eq 15, the equations of motion for the nuclear and electronic degrees of freedom become

$$M_I \ddot{\mathbf{R}}_I = -\frac{\partial E}{\partial \mathbf{R}_I} \tag{17}$$

$$\mu \ddot{\phi}_i = -\mathbb{H}\phi_i + \sum_j \Lambda_{ij}\phi_j \tag{18}$$

where the term with the Lagrange multipliers accounts for the constraint forces that are needed to keep the wave functions orthonormal during the dynamics. Equations 17 and 18 can be integrated simultaneously with a standard MD integration algorithm. The parameter μ is a purely fictitious variable and can be assigned an arbitrary value. In analogy to the nuclear degrees of freedom, μ determines the rate at which the electronic variables evolve in time. In particular, the ratio of M_I to μ characterizes the relative speed, in which the electronic variables are propagated with respect to the nuclear positions.

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

For $\mu \ll \mathcal{M}_I$, the electronic degrees of freedom adjust instantaneously to changes in the nuclear coordinates, and the resulting dynamics is adiabatic. Under this condition $\mathcal{T}_e \ll \mathcal{T}_N$, and the extended Lagrangian in eq 15 becomes identical to the physical Lagrangian $\mathcal{L}$ of the system:

$$\mathcal{L} = \mathcal{T}_N - E_{pot} \tag{19}$$

For finite values of μ, the system moves within a limited width, given by the fictitious electronic kinetic energy, over the Born–Oppenheimer surface. Adiabaticity is ensured when the highest frequency of the nuclear motion ω_I^{max} is well separated from the lowest frequency associated with the fictitious motion of the electronic degrees of freedom ω_e^{min}. It can be shown that ω_e^{min} is proportional to the electronic gap E_g, that is, the difference between highest occupied and lowest unoccupied energy levels:[16]

$$\omega_e^{min} \propto \sqrt{\frac{E_g}{\mu}} \tag{20}$$

For systems with finite E_g, the parameter μ can be used to shift the electronic frequency spectrum so that $\omega_e^{min} \gg \omega_I^{max}$ and no energy transfer between nuclear and electronic subsystems occurs. In practice, it is easy to check if adiabatic conditions are fulfilled by monitoring the energy conservation of the physical Lagrangian in eq 19.

2.1.1.2. Semiclassical Nonadiabatic Dynamics. Most of the current QM/MM applications are based on adiabatic dynamics in the electronic ground state. However, for dynamics in electronically excited states, the coupling between states, which is mediated via the nuclear motion, can be sizable, and nonadiabatic effects have to be taken into account. The two most popular approaches that can incorporate nonadiabatic effects beyond the BO approximation and that have been combined with QM/MM simulations are Ehrenfest Dynamics (ED) and Trajectory Surface Hopping (SH). Both methods are mixed quantum-classical approaches, in which the nuclear dynamics is described with purely classical trajectories. The two approaches differ in the way that the forces that the quantum electronic system exerts on the classical nuclei (and vice versa) are described (Figure 1).

2.1.1.2.1. Ehrenfest Meanfield Dynamics. Ehrenfest dynamics[17] is based on a meanfield description of the electron–nuclear dynamics. The time-dependent electron–nuclear wave function is described with the Ansatz:

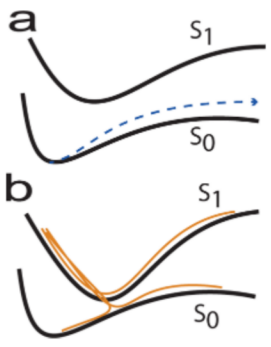


Figure 1. Schematic representation of the forces that the quantum systems exert on the classical nuclei in Ehrenfest dynamics (a) and trajectory surface hopping (b).

$$\Phi(\mathbf{r}, \mathbf{R}, t) = \Psi(\mathbf{r}, t)\chi(\mathbf{R}, t)e^{i[\int_{t_0}^{t} dt' E_e(t')]/\hbar} \tag{21}$$

E_e in the phase term is given by

$$E_e(t) = \iint d\mathbf{r}\, d\mathbf{R}\, \Psi^*(\mathbf{r}, t)\chi^*(\mathbf{R}, t)\mathbb{H}_e\Psi(\mathbf{r}, t)\chi(\mathbf{R}, t) \tag{22}$$

Inserting eq 21 into the TDSE eq 1, multiplying from the left by $\chi^*(\mathbf{R}, t)$ and integrating over $\mathbf{R}$, respectively, multiplying from the left by $\Psi^*(\mathbf{r}, t)$ and integrating over $\mathbf{r}$, leads after some additional transformations to the Time-Dependent Self-Consistent Field (TDSCF) equations:[18]

$$i\hbar\frac{\partial}{\partial t}\Psi(\mathbf{r}, t) = (\mathbb{T}_e + \langle\chi(\mathbf{R}, t)|V_{ne}(\mathbf{r}, \mathbf{R})|\chi(\mathbf{R}, t)\rangle)\Psi(\mathbf{r}, t) \tag{23}$$

$$i\hbar\frac{\partial}{\partial t}\chi(\mathbf{R}, t) = (\mathbb{T}_N + \langle\Psi(\mathbf{r}, t)|\mathbb{H}_e(\mathbf{r}, \mathbf{R})|\Psi(\mathbf{r}, t)\rangle)\chi(\mathbf{R}, t) \tag{24}$$

where $\mathbb{T}_N$ and $\mathbb{T}_e$ are the nuclear and electronic kinetic energy operators, respectively, and V_{ne} and $\mathbb{H}_e$ are the total electron–nuclear potential and the electronic Hamiltonian defined in eq 2. Equations 23 and 24 describe the time evolution of the electronic degrees of freedom in the average field created by the nuclei (and vice versa). Taking the classical point particle limit for $\chi(\mathbf{R}, t)$ recovers the time-dependent electronic Schrodinger eq 25 in which both the Hamiltonian and the wave function depend parametrically on the time-dependent position $\mathbf{R}(t)$ and the classical equations of motion for the nuclei eq 26:

$$i\hbar\frac{\partial}{\partial t}\Psi(\mathbf{r}; \mathbf{R}(t), t) = \mathbb{H}_e(\mathbf{r}; \mathbf{R}(t), t)\Psi(\mathbf{r}; \mathbf{R}(t), t) \tag{25}$$

$$\mathbf{F} = M\ddot{\mathbf{R}} = -\nabla_{\mathbf{R}}\langle\Psi(\mathbf{r}, t)|\mathbb{H}_e(\mathbf{r}; \mathbf{R})|\Psi(\mathbf{r}, t)\rangle \tag{26}$$

The time-dependent wave function $\Psi(\mathbf{r}, t)$ corresponds to a superposition of adiabatic states (eq 3), and the nuclei move according to eq 26 with forces that are derived from the gradient of the average energy from different electronic states (Figure 1a). Such a meanfield description is a good approximation as long as the classical trajectories corresponding to different electronic states are not too different from one another.[19] This is often the case for phenomena in which the electronic structure relaxes on much shorter time scales than the nuclear rearrangements such as in ultrafast electronic processes in intense laser fields.[20] However, ED can lead to qualitatively incorrect behavior in other situations.[21] A further drawback of ED is that it violates detailed balance.[22]

2.1.1.2.2. Trajectory Surface Hopping. In the surface hopping approach introduced by Tully,[18b,23] the nuclear wavepacket is replaced by a swarm of independent classical trajectories (indexed α). The time-dependent wave function is expanded in a complete set of adiabatic states ψ_I with time-dependent amplitudes $C_I^\alpha(t)$:

$$\Phi(\mathbf{r}, \mathbf{R}, t) = \sum_{I=0}^{\infty} C_I^\alpha(t)\psi_I(\mathbf{r}; \mathbf{R}) \tag{27}$$

Note that in contrast to the nuclear wave function coefficients $\chi_I(\mathbf{R}, t)$ in eq 4, the time-dependent coefficients $C_I^\alpha(t)$ do not depend explicitly on the nuclear coordinates. Inserting eq 27 into the TDSE eq 1, multiplying with ψ_k^*, and integrating over the electronic degrees of freedom results after some reformulation in a set of coupled first-order differential

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217–6263

equations for the amplitudes $C_k^\alpha(t)$ that in contrast to the exact quantum dynamics equations (eq 5) only depend on the nonadiabatic derivative coupling vectors $\mathbf{d}_{kl}^{\alpha}(\mathbf{R})$ defined in eq 8.

$$i\dot{C}_k^{\alpha}(t) = \sum_{l=0}^{\infty} C_l^{\alpha}(t)[\mathbb{H}_{kl}\delta_{kl} - i\dot{\mathbf{R}}^{\alpha}\mathbf{d}_{kl}^{\alpha}] \tag{28}$$

where δ_{kl} is the Kronecker delta. Equation 28 describes the time evolution of the state populations. In contrast to Ehrenfest dynamics, the nuclear dynamics follows at any time the PES of a single state (the "running state") (Figure 1b), while the time-dependent coefficients of other electronic states are propagated along the nuclear dynamics according to eq 28. The system at time t can hop from one electronic state k to another state l in the time interval Δt according to the hopping probability $g_{k,l}$:

$$g_{k,l}(t, \Delta t) = -2\int_t^{t+\Delta t} \mathrm{d}\tau \frac{\mathcal{R}e[C_l(\tau)C_k^*(\tau)\sigma_{kl}(\tau)]}{C_k(\tau)C_k^*(\tau)} \tag{29}$$

where $\sigma_{kl}(\tau)$ are the scalar nonadiabatic couplings:

$$\sigma_{kl}(\tau) = \dot{\mathbf{R}}\mathbf{d}_{kl} \tag{30}$$

In Tully's fewest switches SH, a Metropolis algorithm is applied to accept or reject hops with a probability given in eq 29. In the limit of a sufficient number of trajectories, the average state populations should reproduce the correct quantum probability density. The fewest switches algorithm achieves this convergence with the fewest possible number of stochastic hops.

The facts that, at any given time, the nuclear dynamics evolves on a single electronic state and that the different SH trajectories are independent from one another (independent trajectory approximation (ITA)) are at variance to the exact quantum dynamics where the nuclear wave packets on different electronic states can evolve along different trajectories and interfere with each other. Several approaches have been suggested to correct for the lack of decoherence[24] such as the multiple spawning approach introduced by Martinez and co-workers.[25]

All of the methods discussed in this section use classical trajectories to describe the nuclear motion. ED and SH can describe the nuclear quantum effects that lead to the NAC between electronic states. In contrast, due to the underlying purely classical dynamics of the nuclei, nuclear quantum effects such as zero point energy, quantization of (vibrational) energy levels, and nuclear tunneling cannot be captured. Semiclassical methods like variational transition state theory with multi-dimensional tunneling corrections,[26] quantum-classical path integral approaches,[27] and quantum-classical path methods[28] can be used to account for nuclear quantum effects in, for example, enzyme kinetics[29] also in a QM/MM context, and excellent reviews are available on this subject.[30] Most of these methods are based on previously determined classical paths, and quantum effects can be added in the form of a posteriori corrections.

2.1.2. Approximate Solutions of the Electronic Problem. The potential energy surfaces $E(\mathbf{R})$ can be obtained by solving the TISE eq 3 or, as described in section 3.6.2, from analytic parametrization. While only a few methods have been developed that are able to solve the combined nuclear−electronic problem such as multicomponent density functional theory[31] or the nuclear−electronic orbital approach,[32] the majority of methods find approximate solutions to the TISE for many-body systems based on the Born−Oppenheimer approximation. A wide range of well-established quantum chemical methods exist for this purpose, which are described in detail in a number of textbooks,[33] and comprehensive reviews are also available for the time-dependent extensions and the treatment of electronically excited states.[34] Here, we briefly mention the electronic structure methods for ground and electronically excited states that are frequently encountered in QM/MM simulations.

2.1.2.1. Ab Initio Methods. The majority of ab initio (wave function-based)/MM simulations of biological systems are based on the Hartree−Fock[35] (HF) method,[36] and electron correlation effects are often added via Møller−Plesset perturbation theory at second order[37] (MP2) (according to a Web of Knowledge search with "method name" AND QM/MM).[38] Roughly a quarter of all ab initio/MM applications make use of higher correlated coupled cluster methods,[39] sometimes in combination with a polarizable environment.[39a] One of the currently largest ab initio/MM applications for biological systems based on linear scaling local MP2 (LMP2)[40] and local coupled cluster singles and doubles with perturbative triples (LCCSD(T))[41] on the enzyme *para*-hydroxybenzoate hydroxylase contains a quantum part of 49 atoms.[42] Despite local linear scaling approaches, the high computational cost associated with highly correlated methods limits their applications to few selected geometries (e.g., along a reaction path), while the sampling of configurational space has to be performed with a lower level method.

The direct extension of the Hartree−Fock method to the time-dependent case results in the time-dependent Hartree−Fock (TDHF) equations (random phase approximation) for the calculation of excited states. Assuming that the system is initially in the ground state and that a small perturbation $\delta v(\mathbf{r},t)$ is applied at $t > t_0$, the TDHF equations can be solved in a suitable basis within a linear response approximation yielding the non Hermitian eigenvalue problem:[34]

$$\begin{bmatrix} A & B \\ B^* & A^* \end{bmatrix}\begin{bmatrix} \mathbf{X}_n \\ \mathbf{Y}_n \end{bmatrix} = \omega_n \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} \mathbf{X}_n \\ \mathbf{Y}_n \end{bmatrix} \tag{31}$$

where ω_n are the excitation energies, and the matrix elements are defined as

$$A_{ia,jb} = \delta_{ij}\delta_{ab}(\varepsilon_a - \varepsilon_i) + (ia|bj) - (ij|ba)$$
$$B_{ia,jb} = (ia|bj) - (ib|aj) \tag{32}$$

where $(ia|bj)$ is a shorthand for the two electron integral between orbitals ϕ_i^*, ϕ_a, ϕ_b^*, and ϕ_j:

$$(ia|bj) = \iint \mathrm{d}\mathbf{r}_1\, \mathrm{d}\mathbf{r}_2\, \phi_i^*(\mathbf{r}_1)\phi_a(\mathbf{r}_1)\frac{1}{r_{12}}\phi_b^*(\mathbf{r}_2)\phi_j(\mathbf{r}_2) \tag{33}$$

and i and j refer to occupied and a and b to unoccupied Hartree−Fock orbitals, respectively. The linear response vectors $\mathbf{X}$ and $\mathbf{Y}$ correspond to a description of excited states in singly excited determinants. In fact, using the Tamm−Dancoff approximation (setting $B = 0$) leads to the Hermitian eigenvalue problem:

$$A\mathbf{X}_n = \omega_n\mathbf{X}_n \tag{34}$$

and the TDHF equations become equivalent to Configuration Interaction Singles (CIS).[34]

CIS is a popular method in QM/MM applications due to its computational expedience (with a scaling of $O(N^3 - N^4)$) that allows for the treatment of several hundred atoms. However, due to the lack of dynamic correlation, CIS can lead to

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

qualitatively wrong results.[43] CIS does not yield reliable intensities and fails for electronic transitions with significant double excitation character. Furthermore, due to the Brillouin theorem, CIS does not predict the correct topology of conical intersections between excited states and the ground state. For these reasons, multireference ab initio methods such as Multiconfigurational Self Consistent Field (MCSCF) in a Complete Active Space SCF (CASSCF) formulation augmented by CASPT2 or Multiconfigurational QuasiDegenerate Perturbation Theory (MCQDPT)[44] to account for dynamic correlation effects are currently dominating QM/MM simulations of excited states.[45] QM/MM applications using approximate coupled cluster treatments for excited states such as the approximate coupled cluster singles-and-doubles model CC2[46] and the second-order algebraic diagrammatic construction method ADC(2)[47] are starting to emerge,[39a] and the Spectroscopy Oriented Configuration (SORCI) method[48] that combines multireference CI with multireference perturbation theory is also gaining popularity for the prediction of excited-state properties in QM/MM simulations.[49] Recently, combinations of many-body Green's function methods (GW and Bethe−Salpeter equation) and Quantum Monte Carlo methods with a classical environment have been developed and applied to calculate excitation energies of biological chromophores.[50]

CASSCF/MM methods have also been implemented in the context of nonadiabatic simulations in combination with surface hopping[51] and multiple spawning.[52] In addition, a CC2-based surface hopping approach based on Landau−Zener transition probabilities has also been developed.[53]

2.1.2.2. Semiempirical Methods. Because of their low computational cost, semiempirical electronic structure methods are extensively used in QM/MM applications. We limit our discussion here to explicit semiempirical electronic structure methods. QM/MM methods based on the Empirical Valence Bond (EVB) approach have been reviewed extensively elsewhere.[54] A wide range of different semiempirical methods have been developed (a recent review is given in ref 55) that are either simplifications of the Density Functional Theory such as the self-consistent charge density tight binding (SCC-DFTB) method[56] or approximations to the Hartree−Fock method. Computational efficiency is gained through the fact that only valence electrons (usually s and p electrons) are treated within minimal basis sets and the core interaction is parametrized. The HF-based methods make use of the Zero Differential Overlap (ZDO) or Neglect of Diatomic Differential Overlap (NDDO) approximations[57] in which it is assumed that different basis orbitals have zero overlap with one another so that all products of basis functions $\varphi_a\varphi_b$ can be set to zero to reduce the computational effort in calculating four-center and three-center integrals. Applying the ZDO/NDDO approximation to varying extents gives rise to a hierarchy of methods: the Complete Neglect of Differential Overlap (CNDO) method in which the ZDO approximation is applied to all orbital products, the Intermediate Neglect of Differential Overlap (INDO) method where the ZDO is not applied when all 4 basis functions are located on the same atom, and the Modified Neglect of Differential Overlap (MNDO)[58] and Austin Method 1 (AM1)[59] where atomic orbital products are retained when the basis functions for the first electron are both located on atom *A* and the basis functions of the second one on *B*. AM1 has been reparameterized for a larger set of reference data leading to the Parameterized Model 3 (PM3).[60] In addition, the Pairwise Distance Directed Gaussian (PDDG)/PM3 method[61] uses a functional group specific reparameterization of the core repulsion function.

Since the 1990s, the MNDO method has been further extended to a d orbital basis. This MNDO/d extension allows for the treatment of heavier elements.[62] Recent developments also include PM6, an updated parametrization of PM3 that is based on a large experimental data set of over 9000 compounds, including 83 elements.[60] In all semiempirical methods, the overlap matrix is replaced by the unit matrix; that is, the basis is not orthogonalized. The series of OM*x* orthogonalization methods (OM1, OM2, and OM3) remedies this by the inclusion of orthogonalization corrections to better account for the effects of Pauli repulsion.[63]

The whole panoply of semiempirical methods has been applied in QM/MM simulations of biological systems. Because of the fact that the most commonly applied integral approximations scale as $O(N^2)$, and $O(N)$ implementations are also available,[64] semiempirical methods are roughly a factor 1000× faster than first-principles approaches, enabling the simulation of large systems (up to 20 000 atoms[65]) or extended time scales in the range of nanoseconds.[66]

Popular semiempirical methods for excited-state properties are the ZINDO approach originally developed by Zerner and co-workers,[67] time-dependent extensions of DFTB in a linear response formulation (LR-TDDFTB derived from LR-TDDFT),[68] and the Floating Occupation Molecular Orbital Configuration Interaction (FOMO-CI).[69] The OM*x* family of methods has also demonstrated high accuracy and reliability for excited-state applications and can be used as a semiempirical basis for Multireference CI calculations (OM2/MRCI).[70]

Several semiempirical methods have recently been combined with surface hopping (FOMO-CI,[71] OM2,[72] OM2/MRCI,[73] and TDDFTB) and multiple spawning (FOMO-CI)[52]).

2.1.2.3. Density Functional Theory (DFT). Density Functional Theory (DFT) is based on two theorems introduced by Hohenberg and Kohn in 1964[74] that show that the total ground-state energy of a nondegenerate *n* electron system in the external potential V^{ex} can be expressed solely as a functional of the electronic density $\rho(\mathbf{r})$:

$$E = T[\rho] + \int V^{ex}(\mathbf{r})\rho(\mathbf{r})\,d\mathbf{r} + \frac{1}{2}\iint \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,d\mathbf{r}\,d\mathbf{r}' + E_{xc}[\rho] + \frac{1}{2}\sum_I \sum_J \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} \tag{35}$$

with

$$\rho(\mathbf{r}) = \int \dots \int d\mathbf{r}_2 \dots d\mathbf{r}_n\, \psi^*(\mathbf{r}, \mathbf{r}_2 \dots \mathbf{r}_n)\psi(\mathbf{r}, \mathbf{r}_2 \dots \mathbf{r}_n) \tag{36}$$

and T and E_{xc} are the kinetic and exchange-correlation (xc) energy functionals, respectively, V^{ex} represents the external potential created by the positively charged nuclei, and the nuclear charges of atoms *I* and *J* are given, respectively, by Z_I and Z_J.

Mapping of the many-electron system to a noninteracting effective one-particle system with the same density[75] allows one to express the electron density and the kinetic energy functional of the noninteracting system T_s in terms of one-particle orbitals $\phi_i(\mathbf{r}_i)$ leading to the Kohn−Sham (KS) energy functional E^{KS}:

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

$$E^{KS} = \sum_i \left\langle \phi_i \middle| -\frac{1}{2}\nabla^2 \middle| \phi_i \right\rangle + \int V^{ex}(\mathbf{r})\rho(\mathbf{r})\,d\mathbf{r} + \frac{1}{2}\iint \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,d\mathbf{r}\,d\mathbf{r}' + E_{xc}[\rho] + \frac{1}{2}\sum_I \sum_J \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} \quad (37)$$

where

$$\rho(\mathbf{r}) = \sum_i |\phi_i(\mathbf{r})|^2 \quad (38)$$

and $E_{xc}[\rho]$ also includes the difference between the kinetic energy of the many-body and the noninteracting system $T[\rho] - T_s[\{\phi_i\}]$. Minimizing eq 37 with respect to the Kohn–Sham orbitals $\phi_i(\mathbf{r}_i)$ under orthonormality constraints leads to the one-particle Kohn–Sham equations:

$$\left\{-\frac{1}{2}\nabla^2 + V^{KS}(\mathbf{r}, \mathbf{R})\right\}\phi_i(\mathbf{r}) = \varepsilon_i\phi_i(\mathbf{r}) \quad (39)$$

where the Kohn–Sham potential $V^{KS}(\mathbf{r}; \mathbf{R})$ is given by

$$V^{KS}(\mathbf{r}, \mathbf{R}) = V^{ex}(\mathbf{r}, \mathbf{R}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,d\mathbf{r}' + V_{xc}[\rho] \quad (40)$$

with the exchange-correlation potential:

$$V_{xc}[\rho] = \frac{\partial E_{xc}}{\partial \rho(\mathbf{r})} \quad (41)$$

A whole range of approximations at increasing levels of sophistication (that can be classified according to Perdew's Jacob's ladder[76]) have been developed for accurate descriptions of the exchange correlation energy and potential, and many textbooks and reviews are available on DFT.[33a,77]

Because of the favorable complexity $<O(N^3)$ and the good accuracy of DFT methods, the majority of QM/MM simulations in the electronic ground state uses a DFT[74] treatment of the QM part. This is a reflection of the general popularity of DFT methods often referred to as the "de facto work horse" for large-scale correlated quantum chemical studies. A large number of DFT/MM simulations have been performed covering essentially all rungs of the Jacob's ladder[76] with respect to the employed exchange-correlation functionals. Typical DFT/MM applications can contain a few hundreds of atoms with simulation times in the tens to hundreds of picosecond range. In addition, the development of $O(N)$ DFT methods (for a recent review, see ref 78) enables the treatment of large systems (or in combination with QM/MM, large quantum regions) of several thousands of atoms. DFT/MM methods and applications to biological ground-state problems have been reviewed in detail in several publications.[30a,79−81]

The formal extension of DFT to the time-dependent case by Runge and Gross[82] and later by van Leeuwen[83] leads to the time-dependent Kohn–Sham equations and enables a treatment of excited states by TDDFT. However, in contrast to ground-state applications, wave function-based methods have by far been more frequently applied for QM/MM simulations of electronically excited states, while DFT-based approaches have only emerged relatively recently.[84]

The first DFT-based QM/MM implementation for excited states was using the Restricted Open Shell Kohn–Sham (ROKS) method with a spin-adapted multideterminantal representation of the wave function in the first excited singlet state[84−86] (in analogy with the sum method for multiplet states[87]).

2.1.2.3.1. Linear Response Time-Dependent Density Functional Theory (LR-TDDFT). In analogy to TDHF, the time-dependent Kohn–Sham equations can be solved in a linear response approximation by assuming an initially unperturbed state, a small time-dependent perturbation $\delta v(\mathbf{r}, t)$ at a later time, and a suitable basis, resulting in matrix equations (Casida equations[88]) analogous to eq 31 with the matrix elements (for nonhybrid functionals):[34]

$$A_{ia,jb} = \delta_{ij}\delta_{ab}(\varepsilon_a - \varepsilon_i) + (ia|bj) + (ia|f_{xc}|jb)$$
$$B_{ia,jb} = (ia|bj) + (ia|f_{xc}|bj) \quad (42)$$

The last terms in eq 42 replace the (exact) exchange contributions in the TDHF case and are due to the response of the time-dependent xc potential. In most TDDFT calculations, the nonlocal time-dependence of the xc potential is neglected by adopting time-independent approximations (adiabatic approximation) without any memory or retardation effects.

$$V_{xc}[\rho](\mathbf{r}, t) \approx \frac{\delta E_{xc}[\rho_t]}{\delta \rho_t(\mathbf{r})} = V_{xc}[\rho_t](\mathbf{r}) \quad (43)$$

where ρ_t denotes the electron density at a given time t. In this case, f_{xc} in eq 42 is given by the second functional derivative of the xc energy, the xc kernel:

$$(ia|f_{xc}|jb) = \int d\mathbf{r}\,d\mathbf{r}\,\phi_i^*(\mathbf{r})\phi_a(\mathbf{r})\frac{\delta^2 E_{xc}}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')}\phi_j^*(\mathbf{r}')\phi_b(\mathbf{r}') \quad (44)$$

and can be calculated with the standard exchange-correlation approximations developed for ground-state DFT. As in the case of TDHF, applying the Tamm–Dancoff approximation yields a Hermitian eigenvalue problem. Alternatively, the LR-TDDFT equations can be solved using the Sternheimer formalism[89] that does not involve virtual orbitals (which is advantageous for basis sets like plane waves with a very large space of virtual orbitals). LR-TDDFT also enables the calculation of excited-state nuclear forces, and analytic gradients are now available in several programs.[89,90]

TDDFT/MM applications have focused on calculations of environment-induced changes of excitations energies such as shifts in the vertical absorption spectra of biological chromophores in solution[80,84,91] and different protein environments.[92] In many applications, the classical surrounding is only represented by a (static) point charge model, and the optical properties are evaluated for a single optimized (zero Kelvin) structure, albeit for biological systems finite temperature and sampling effects can be sizable.[50a,93] Few applications have also performed adiabatic excited-state dynamics or excited-state optimization of chromophore structures to calculate emission spectra and Stoke shifts[91a] or to follow excited-state reactions.[84]

Early implementations of DFT surface hopping were based on approximate descriptions of the excited states using a ROKS formalism[90b] or Kohn–Sham (KS) potential energy surfaces,[94] and first implementations within LR-TDDFT made use of approximate hopping probabilities based on Landau–Zener estimates of the transition probabilities.[95,96]

Direct implementations of LR-TDDFT-based Tully Fewest Switches surface hopping (eq 28) face the problem that the

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

many-electron wave functions $\psi_I(\mathbf{r};\mathbf{R})$ of the adiabatic states that are needed for a calculation of the nonadiabatic derivative coupling vectors (eq 8) are not available within DFT where all quantities have to be expressed as a functional of the density (or equivalently the KS orbitals). Chernyak and Mukamel were the first to provide exact expressions for nonadiabatic couplings based on a density-matrix formulation.[97] In addition, a method to calculate NAC vectors between ground and excited states from real-time propagation has been proposed.[98] Alternatively, NACs can also be calculated by using auxiliary many-electron wave functions in the form of singly excited KS Slater determinants. For the wave function of the kth excited state:

$$|\psi_k\rangle = \sum_{i=1}^{n_{occ}} \sum_{a=n_{occ}+1}^{n_{occ}+n_{vir}} c_{ia}^{k} |\psi_i^a\rangle \tag{45}$$

where, as in eq 33, i and a refer to occupied, respectively, unoccupied KS orbitals and the $|\psi_i^a\rangle$ are generated from single excitations of the determinant $|\psi_o\rangle$ constructed with the occupied KS orbitals of the ground state:

$$|\psi_i^a\rangle = \sum_{i=1}^{n_{occ}} \sum_{a=n_{occ}+1}^{n_{occ}+n_{vir}} \hat{a}_a^\dagger \hat{a}_i |\psi_o\rangle \tag{46}$$

where $\hat{a}_a^\dagger$ and $\hat{a}_i$ are creation and annihilation operators, respectively. Using Casida's assignment Ansatz[88] for the characterization of electronic transitions, the expansion coefficients can be determined by projecting the linear response orbitals ϕ_i^- on a subset of unoccupied KS orbitals:

$$c_{ia}^{k} = \sqrt{\frac{\langle \phi_i^- | \phi_a \rangle}{\omega_{0k}}}\, \varepsilon_{ia}^{k} \tag{47}$$

where ω_{0k} are the excitation energies and $\varepsilon_{ia}^{k} = \varepsilon_a - \varepsilon_i$ is the KS eigenvalue difference between virtual orbital a and occupied orbital i. Using eq 45, the scalar NACs (eq 30) that are needed for the propagation of the state amplitudes can be calculated via a finite difference scheme,[99,100] while the full derivative NAC vectors that are used for a rescaling of the nuclear velocities after hops can be obtained directly from LR-TDDFT quantities.[99,100] It can be shown that the nonadiabatic couplings between the ground and any excited state obtained in this way are formally correct[101] and that the ones between excited states[102] (which correspond to a second-order response quantity) coincide within the Tamm–Dancoff approximation with the electronic oscillator approach by Mukamel and co-workers.[103] A numerical comparison of the NACs derived in this way with wave function-based methods (MR-CISD) shows a good general agreement.[100,102] A similar (and equivalent) approach has been developed by Sugino et al. for both first-order[104] and second-order NACs.[105] Alternatively, first-order NACs can be derived directly from linear response TDDFT without explicit calculations of excited-state wave functions.[106]

Several LR-TDDFT-based SH methodologies have been developed and implemented in a QM/MM framework[107] with typical applications to photoactive proteins, photodamage,[108] and electron transfer.[109] Recently, LR-TDDFT SH/MM simulations have also been extended to the presence of an external electromagnetic field $\mathbf{E}(t)$[110] and pulse shaping via local control[111] with modified equations for the amplitudes (eq 28):

$$i\dot{C}_k^{\alpha}(t) = \sum_{l=0}^{\infty} C_l^{\alpha}(t)[\mathbb{H}_{kl}\delta_{kl} - i\dot{\mathbf{R}}^{\alpha}\mathbf{d}_{kl}^{\alpha} - \boldsymbol{\mu}_{kl}^{\alpha}\mathbf{E}(t)] \tag{48}$$

and transition probabilities:

$$g_{k,l}(t,\Delta t) \approx -2\int_t^{t+\Delta t} \mathrm{d}\tau \frac{\mathcal{R}e[C_l(\tau)C_k^*(\tau)\{\sigma_{kl}(\tau) - i\boldsymbol{\mu}_{kl}^{\alpha}\mathbf{E}(t)\}]}{C_k(\tau)C_k^*(\tau)} \tag{49}$$

with $\boldsymbol{\mu}_{kl}^{\alpha}$ being the transition dipole vector between states k and l.

2.1.2.3.2. Real Time Propagation TDDFT (P-TDDFT). The linear response formulation of TDDFT has the advantage that all excited-state properties can be calculated from ground-state quantities and LR-TDDFT is used in the majority of current TDDFT calculations. However, if the time-dependent external perturbation is not small (such as in a strong laser field), the linear response approximation may break down. Alternatively, eq 1 can be solved directly in the framework of TDDFT by direct propagation of the time-dependent Kohn–Sham orbitals:

$$\left\{-\frac{1}{2}\nabla^2 + V^{KS}(\mathbf{r},t)\right\}\phi_i(\mathbf{r},t) = i\frac{\partial}{\partial t}\phi_i(\mathbf{r},t) \tag{50}$$

with suitable approximations to the time propagator.[112] Real-time propagation time-dependent density functional theory (P-TDDFT) was originally introduced by Teilhaber[113] and has been implemented by several groups.[112−114] P-TDDFT can be combined with Ehrenfest dynamics for the nuclei that evolve classically according to eq 26. In ref 115, for instance, eq 50 is propagated in time via a time stepping method using a second-order Crank–Nicholson expansion[112] of the propagator in combination with a two-step Runge–Kutta scheme to estimate the effective Kohn–Sham potential at mid time step. The set of integral equations is solved iteratively until convergence. Integrals are calculated via Chebyshev interpolation in the time domain.

P-TDDFT is computationally demanding because a very small integration time step of ∼0.1−1 has to be used. As a consequence, only very short time windows of few tens of femtoseconds can be investigated. P-TDDFT has mostly been applied to study systems in strong laser fields,[116] but few applications to biological problems have recently emerged such as the cis–trans photoisomerization of a minimal retinal model,[115] electron transfer in DNA photolyase based on an approximate DFT model,[117] and radiation damage of biological matter.[118] Some specific QM/MM applications of P-TDDFT will be discussed in section 4.2.3.

2.1.2.3.3. Improving the Accuracy of DFT/TDDFT. Despite the wide success of DFT methods, DFT is also plagued by a number of serious failures due to the limited accuracy of xc functionals and kernels, which have been discussed in several reviews.[77a,105,119] For ground-state applications, one of the most serious failures in view of biological applications is the lack of dispersion interactions when local (or semilocal) approximations to the xc functional are used. During the past decade, many methods have been developed to cure this deficiency ranging from adding dispersion corrections in the form of empirical or first-principles-based R^{-6} interatomic pair potentials (DFT-D[120] to DFT-D3,[121] dDsC[122] XDM,[123] TS-vdW,[124] LRD[125]), incorporating dispersion interactions via nonlocal density functionals (vDW-DF,[126] vDW-DF2,[127]

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

VV10[128]), the parametrization of new generation XC functionals (M05[129] to N12[130]), and the inclusion of many-body van der Waals interactions (DFT-MBD[131]) to the correction of xc functionals through parametrized nonlocal atom-centered potentials (DCACP).[132]

The situation deteriorates for excited states where the self-interaction error, the missing integer discontinuity, and the wrong asymptotic behavior of xc functionals are at the origin of the failure of TDDFT for charge-transfer excitations. Furthermore, due to the adiabatic approximation of the xc kernel, transitions with large double excitation or multireference character cannot be described. It has also been pointed out that TDDFT is not able to reproduce the correct dimensionality of conical intersections (CIs)[133] and has difficulties to describe CIs between excited states and the ground state. The development of improved xc functionals and kernels is a highly active field of research, and many advances have been made through the last years. The introduction of range-separated functionals[119] has contributed a lot to alleviate the problem with charge-transfer transitions. Recent developments in spin-flip DFT[134] are able to improve the description of double excitations,[135] and new approaches to improve the description of CIs have also been introduced lately.[136] Hopefully, the intense activities in the development of improved xc descriptions will further increase the number and scope of TDDFT/MM applications for excited states.

3. THE QM/MM APPROACH

While electronic structure calculations of entire proteins with linear scaling semiempirical[64b,65,66,137] or DFT[138] methods are possible nowadays and allow the treatment of large systems of up to 150 000 atoms,[139] the large system size comes at the price of a very limited exploration of phase space that is typically reduced to a minimal number of configurations.

On the other extreme, a full reduction of the QM treatment to minimal cluster models is not always applicable, except in certain successful cases, for example; to deduce information about the chemical reactivity of some transition metal enzymes[140] for which the electronic effects are strongly dominated by the highly localized orbital structure of the transition metal center and the active site is efficiently screened by the surrounding protein. Active sites are often embedded in a complex hydrogen-bond network, and a complete neglect of all environment effects can lead to drastic alteration of the properties of the QM region, in particular pK_a's or redox properties,[141] and induce artifacts that are not easily recovered at the level of a cluster model even when a continuum representation of the environment is included. In addition, the protein surrounding imposes steric constraints on the active site residues that have to be approximated by somewhat arbitrary position constraints in minimal models. These approaches are thus limited to the optimization of zero Kelvin structures and neglect thermal sampling.

Investigations of QM phenomena in realistic biological systems such as proteins or protein−protein and protein−DNA adducts, solvated in biologically representative environments (explicit water with counterions or biological membranes), typically comprise 10,000−100,000 atoms that have to be sampled for extended time scales. This can be achieved using a multiscale hybrid approach, in which only a portion of the system is treated at the QM level while the rest of the system is represented with a computationally more expedient lower level method, for example, an empirical force field, leading to a mixed Quantum Mechanical/Molecular Mechanical (QM/MM) partitioning of the system (Figure 2).

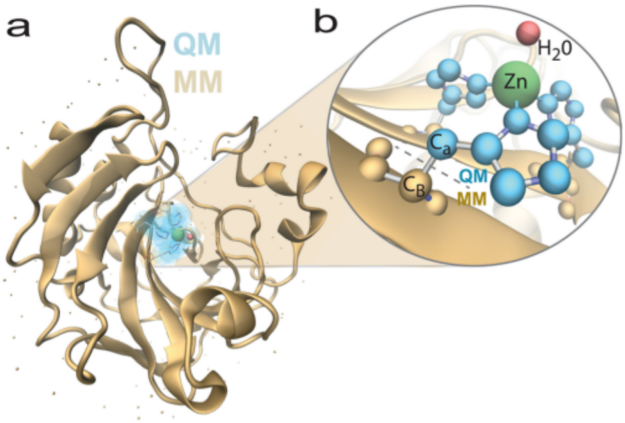


Figure 2. Typical QM/MM partitioning of a complex system. QM, region treated with a quantum mechanical method; MM, region treated with a molecular mechanics force field; IL, optional intermediate layer between QM and MM parts. Note that in most QM/MM schemes, the initial partitioning of the system is maintained throughout the simulation; that is, atoms belonging to the QM part remain QM atoms, and MM atoms are always treated on the classical level (PDB entry 2BCA).

3.1. The General Idea

The wide success of QM/MM schemes (see Figure 3) for the description of quantum phenomena such as chemical reactions

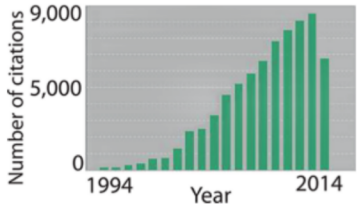


Figure 3. The number of citations of papers related to QM/MM approaches has grown rapidly since 1994, according to a Web of Knowledge search with "method name" AND QM/MM. The final number of citations for 2014 is expected to be higher as it was taken before the end of the year.

and photoexcitations in extended systems is based on the fact that electronic structure effects are often localized to a fairly small part of the system (a consequence of the "nearsightedness" of electrons discussed by Kohn[142]). In this perspective, a full QM treatment of the system (e.g., including fully solvated amino acid residues on the protein surface that are located far away from the active site) might be a waste of computer time. Using a QM/MM approach, the computational effort can be concentrated where it is needed, that is, for the portion of the system that is subject to strong electronic rearrangements, and the rest of the system is treated in a more economic way.

As there are a number of reviews available on QM/MM topics,[79a,143] we choose to focus here primarily on dynamical QM/MM simulations in the adiabatic and nonadiabatic regimes. The aim of this section is to introduce the theoretical and practical aspects involved in QM/MM simulations and to discuss their capabilities and limitations.

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

3.2. Some Historical Aspects

Almost 40 years ago, Warshel and Levitt published a seminal paper that introduced, for the first time, a semiempirical QM/MM approach to describe the enzymatic reaction of lysozyme.[2] This contribution was a significant extension of some early ideas by Warshel and Karplus[144] and Birge et al.[145] Revolutionary and far ahead of its time, it not only suggested a QM treatment of biological systems at a time when computer power was still very limited and most quantum chemical calculations were performed on small gas-phase molecules, but it also advocated the use of a polarizable MM environment and introduced hybrid atomic orbitals for treating covalent QM/MM boundaries, features that even now have not become standard QM/MM practice. The idea was indeed so radical that it took 10 years before it was taken up by Singh and Kollman[146] who published the first ab initio/MM implementation at the HF level and Bash, Field, and Karplus[147] who introduced a QM/MM free energy perturbation approach both applied to the same S_N2 reaction in aqueous solution. Since then, many implementations of QM/MM schemes have been developed, and QM/MM applications have grown in scope mainly as a result of the rapid advancement of computer power that allows for the coupling of molecular mechanics methods to a wide range of different QM methods as well as large QM regions and significant configurational sampling. Up to today, roughly 8000 QM/MM papers have been published (currently ca. 350 articles per year) that receive around 9500 citations per year (Figure 3). In the following sections, we will first introduce the two main flavors of QM/MM approaches (subtractive versus additive schemes) and then give a detailed theoretical motivation of the QM/MM approach and its intrinsic approximations.

3.3. Subtractive and Additive QM/MM Schemes

Most of the current QM/MM methods use either a subtractive or an additive QM/MM scheme. In a subtractive scheme (Figure 4a), the QM calculation is performed on an isolated system. The influence of the environment is estimated at the lower (MM) level as the difference between two calculations in which either the entire system (QM + MM) or the QM part (QM) only is described with an MM method, assuming that the change in energy is fully transferable from the lower to the higher level treatment. Using this approach, the total energy of the embedded system can be represented by

$$E = E^{\mathrm{QM}}(\mathrm{QM}) + E^{\mathrm{MM}}(\mathrm{QM}+\mathrm{MM}) - E^{\mathrm{MM}}(\mathrm{QM}) \tag{51}$$

where E^{QM} represents the quantum mechanical energy and E^{MM} the energy calculated with molecular mechanics. The corresponding nuclear forces can be calculated by direct differentiation:

$$\mathbf{F}_I = -\frac{\partial E}{\partial \mathbf{R}_I} = -\frac{\partial E^{\mathrm{QM}}(\mathrm{QM})}{\partial \mathbf{R}_I} - \frac{\partial E^{\mathrm{MM}}(\mathrm{QM}+\mathrm{MM})}{\partial \mathbf{R}_I} + \frac{\partial E^{\mathrm{MM}}(\mathrm{QM})}{\partial \mathbf{R}_I} \tag{52}$$

where $\mathbf{F}_I$ is the force acting on atom I at position $\mathbf{R}_I$.

A standard subtractive scheme involves a so-called "mechanical coupling" (a term introduced originally in ref 148) between the QM and MM parts, in which the electrons of the QM system do not "feel" the classical electrostatic field of the MM environment. The electrostatic interactions between the QM and MM systems are instead described fully at the MM level as Coulomb interaction between effective classical point charges. This is a good approximation if the difference between the higher level and lower level calculations is relatively small.

A subtractive scheme has the advantage that it is straightforward to implement and that there is no QM/MM interface to account for. Furthermore, multiple layers and descriptions can be added easily, such as in the ONIOM (our own *n*-layered integrated molecular orbital and molecular mechanics) approach developed by Morokuma et al.[149] as a generalization of the integrated molecular-orbital molecular mechanics (IMOMM) method.[150] A typical ONIOM calculation entails a high-level first-principles (e.g., CCSD(T)) calculation for the inner region, a lower level QC method (e.g., Hartree−Fock, DFT, or semiempirical) calculation for the adjacent region, and a semiempirical or molecular mechanics calculation for the outer layer.

The majority of current QM/MM approaches are instead based on an additive scheme, in which the system is described by a single hybrid Hamiltonian:

$$\mathbb{H} = \mathbb{H}_{\mathrm{QM}} + \mathbb{H}_{\mathrm{MM}} + \mathbb{H}_{\mathrm{QM/MM}} \tag{53}$$

where $\mathbb{H}_{\mathrm{QM}}$ is the quantum Hamiltonian, $\mathbb{H}_{\mathrm{MM}}$ is the Hamiltonian describing the MM system, and $\mathbb{H}_{\mathrm{QM/MM}}$ is the interaction Hamiltonian that describes the coupling between QM and MM regions. Using this representation, the total energy of the mixed quantum/classical system can be computed from the lowest eigenvalue of the Hamiltonian in eq 53.

$$E = E_{\mathrm{QM}} + E_{\mathrm{MM}} + E_{\mathrm{QM/MM}} \tag{54}$$

In contrast to the subtractive scheme, in an additive scheme, the QM calculation is performed in the presence of the MM environment. The electrostatic coupling between QM and MM part can be taken into account via classical point charge interactions ("mechanical embedding"[148]) or at the level of the QM method by including the MM point charges as an additional part of the external potential of the QM Hamiltonian ("electrostatic embedding"[148]) so that the electron density of

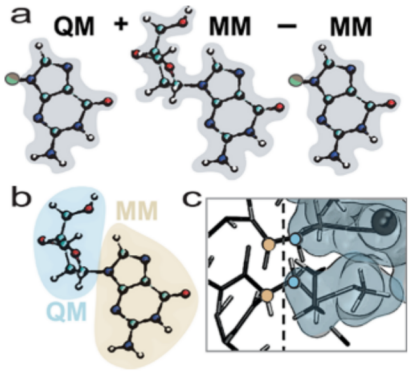


Figure 4. (a) Example of a subtractive QM/MM scheme; (b) additive QM/MM scheme; and (c) zoom in to the QM/MM interface region of an additive coupling scheme. In the case shown here, the QM/MM boundary (indicated by the gray dashed line) runs across covalent bonds. A contour plot of the electron density distribution of the QM part is shown in gray. Atomic positions of QM and MM part are represented by sticks. The last QM atoms at the boundary are indicated by a red circle, and the first MM atoms are indicated with a dark gray circle.

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

the QM system is polarized by the electrostatic field created by the surrounding. Most of the current additive QM/MM schemes use this latter type of coupling. A further extension is possible when a polarizable MM environment is used and the mutual polarization of QM and MM part is taken into account self-consistently ("polarizable embedding"[148]). One advantage of an additive formalism is that also very different levels of theory, such as a first-principles electronic structure and a classical force field description, can be used to describe the QM and MM parts of the system and their coupling. The disadvantage of an additive approach is the creation of an abrupt quantum/classical boundary when crossing from the QM to the MM part of the system, the electrons cease to exist. This means that the electronic description is discontinuous, a fact that must be dealt with in an appropriate way. A hybrid description can at best constitute a somewhat crude representation of the true system. In the next subsection, we will introduce the theoretical framework of an additive QM/MM scheme for the prototypical case of a DFT-based QM/MM approach to better illustrate the underlying assumptions.

3.4. Theoretical Formalism and Main Approximations

Let us consider first a case where the entire system (QM + MM) is described at the QM (DFT) level. The total energy is given by the energy functional of eq 35 (respectively, eq 37).

First approximations arise when the total system is arbitrarily partitioned into different parts, A and B, in analogy to the inherent approximations in frozen density embedding[151] or fragment-based[152] approaches. Each part of the system will have a respective density, ρ_A and ρ_B, such that the total density, ρ, can be represented as

$$\rho(\mathbf{r}) = \rho_A(\mathbf{r}) + \rho_B(\mathbf{r}) \tag{55}$$

The total energy, analogous to eq 54, is given by

$$E = E_A + E_B + E_{A-B} \tag{56}$$

or using the energy functional in eq 35:

$$\begin{aligned} E = {} & T[\rho_A] + T[\rho_B] + T^{\mathrm{NL}} + \int_\Omega V^{\mathrm{ex}}(\mathbf{r})\rho_A(\mathbf{r})\,\mathrm{d}\mathbf{r} \\ & + \int_\Omega V^{\mathrm{ex}}(\mathbf{r})\rho_B(\mathbf{r})\,\mathrm{d}\mathbf{r} + \frac{1}{2}\iint \frac{\rho_A(\mathbf{r})\rho_A(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}' \\ & + \frac{1}{2}\iint \frac{\rho_B(\mathbf{r})\rho_B(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}' \\ & + \frac{1}{2}\iint \frac{\rho_A(\mathbf{r})\rho_B(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}' + E_{\mathrm{xc}}[\rho_A] + E_{\mathrm{xc}}[\rho_B] \\ & + E_{\mathrm{xc}}^{\mathrm{NL}} + \frac{1}{2}\sum_I \sum_J \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} \end{aligned} \tag{57}$$

where

$$\begin{aligned} T^{\mathrm{NL}} &= T[\rho_A + \rho_B] - (T[\rho_A] + T[\rho_B]); \\ E_{\mathrm{xc}}^{\mathrm{NL}} &= E_{\mathrm{xc}}[\rho_A + \rho_B] - (E_{\mathrm{xc}}[\rho_A] + E_{\mathrm{xc}}[\rho_B]) \end{aligned} \tag{58}$$

The terms T^{NL} and $E_{\mathrm{xc}}^{\mathrm{NL}}$ account for the nonlinearity of the kinetic energy and the exchange-correlation functionals, respectively. They are zero or negligible when ρ_A and ρ_B are spatially far apart from one another and nonoverlapping.

Similar nonlinearity corrections in the exchange-correlation term, $E_{\mathrm{xc}}^{\mathrm{NL}}$, also arise in the construction of ab initio atomic pseudopotentials when the electron density of core and valence electrons have significant overlap. In this case, $E_{\mathrm{xc}}^{\mathrm{NL}}$ is referred to as nonlinear core correction.[153]

As we expect to describe the A and B parts with different levels of theory, it is convenient to also separate the external potential, V^{ex}, into contributions that arise from nuclei located on the A and B parts, V_A^{ex} and V_B^{ex}, respectively.

$$V^{\mathrm{ex}}(\mathbf{r}) = V_A^{\mathrm{ex}}(\mathbf{r}) + V_B^{\mathrm{ex}}(\mathbf{r}) \tag{59}$$

E_A and E_B in eq 56 are thus given by

$$\begin{aligned} E_\gamma = {} & T[\rho_\gamma] + \int_\Omega V_\gamma^{\mathrm{ex}}(\mathbf{r})\rho_\gamma(\mathbf{r})\,\mathrm{d}\mathbf{r} \\ & + \frac{1}{2}\iint \frac{\rho_\gamma(\mathbf{r})\rho_\gamma(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}' + E_{\mathrm{xc}}[\rho_\gamma] \\ & + \frac{1}{2}\sum_{I\in\gamma}\sum_{J\in\gamma} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} \end{aligned} \tag{60}$$

where $\gamma = A, B$, respectively.

The nuclear charges, Z_I and Z_J, can also be expanded into Gaussian charge distributions of the form:

$$Z_I = \int \rho_I^{\mathrm{nuclear}}(\mathbf{r} - \mathbf{R}_I)\,\mathrm{d}\mathbf{r} = \int_\Omega \frac{Z_I}{R_c^3}\pi^{-2/3}\exp\left[\frac{|\mathbf{r}-\mathbf{R}_I|^2}{R_c^2}\right] \tag{61}$$

where R_c describes the width of the Gaussian. Using eq 61 for the nuclear charge distribution, the three classical Coulomb terms can be summarized into one equation, which depends on the total (electronic and nuclear) charge density, $\rho^{\mathrm{el+nuc}} = \rho^{\mathrm{el}} + \rho^{\mathrm{nuc}}$:

$$\begin{aligned} & \int_\Omega V_\gamma^{\mathrm{ex}}(\mathbf{r})\rho_\gamma(\mathbf{r})\,\mathrm{d}\mathbf{r} + \frac{1}{2}\iint \frac{\rho_\gamma(\mathbf{r})\rho_\gamma(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}' \\ & + \frac{1}{2}\sum_{I\in\gamma}\sum_{J\in\gamma} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} \\ & = \frac{1}{2}\iint \frac{\rho_\gamma^{\mathrm{el+nuc}}(\mathbf{r})\rho_\gamma^{\mathrm{el+nuc}}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}' \end{aligned} \tag{62}$$

The interface term, E_{A-B}, describes the interactions between A and B and therefore contains the remaining terms of eq 57:

$$\begin{aligned} E_{A-B} = {} & T^{\mathrm{NL}} + \int_\Omega v_B^{\mathrm{ex}}(\mathbf{r})\rho_A(\mathbf{r})\,\mathrm{d}\mathbf{r} + \int_\Omega v_A^{\mathrm{ex}}(\mathbf{r})\rho_B(\mathbf{r})\,\mathrm{d}\mathbf{r} \\ & + \frac{1}{2}\iint \frac{\rho_A(\mathbf{r})\rho_B(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}' + E_{\mathrm{xc}}^{\mathrm{NL}} \\ & + \frac{1}{2}\sum_{I\in A}\sum_{J\in B} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} \end{aligned} \tag{63}$$

Let us now consider the particular case that A is treated with a QM method and B with an MM method so that in

$$\begin{aligned} E_\gamma = {} & T[\rho_\gamma] + \frac{1}{2}\iint \frac{\rho_\gamma^{\mathrm{el+nuc}}(\mathbf{r})\rho_\gamma^{\mathrm{el+nuc}}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}' \\ & + E_{\mathrm{xc}}[\rho_\gamma] \end{aligned} \tag{64}$$

$E_\gamma = E_{\mathrm{QM}}$ and $E_\gamma = E_{\mathrm{MM}}$, respectively, and E_{MM} is substituted by a classical mechanics description. None of the force fields terms will provide an exact match for the expression in eq 64. Nevertheless, force fields (such as the biomolecular force fields

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

discussed in section 3.6.2) use analytical expressions that were designed to mimic most of the physical effects in eq 64.

Electronic effects are only included implicitly in a force field that is based on a molecular fragment approach in which one or several representative small molecule configurations are used for parametrization. From an electronic point of view, this corresponds to a frozen density decomposition, and under this condition the kinetic energy term $T[\rho_{MM}]$ is a (composition-dependent) additive constant and is not directly taken into account. The effect of the exchange-correlation term $E_{xc}[\rho_{MM}]$ is typically approximated by a pair-additive van der Waals term, such as a 12-6 Lennard-Jones potential:

$$E_{xc}[\rho_{MM}] \approx E_{vdW} = \frac{1}{2}\sum_{I\in MM}\sum_{J\in MM} 4\varepsilon_{IJ}\left(\left(\frac{\sigma_{IJ}}{R_{IJ}}\right)^{12} - \left(\frac{\sigma_{IJ}}{R_{IJ}}\right)^{6}\right) \quad (65)$$

where $R_{IJ} = |\mathbf{R}_I - \mathbf{R}_J|$. In most biomolecular force fields, the electrostatic interaction energy of the combined electronic and ionic charge distribution in eq 64 is approximated by effective (fixed) point charges, which are typically located at atomic positions:

$$\frac{1}{2}\iint \frac{\rho_{MM}^{el+nuc}(\mathbf{r})\rho_{MM}^{el+nuc}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,d\mathbf{r}\,d\mathbf{r}' \approx \frac{1}{2}\sum_{I\in MM}\sum_{J\in MM}\frac{q_I q_J}{|\mathbf{R}_I - \mathbf{R}_J|} \quad (66)$$

More sophisticated electrostatic MM representations are based on distributed multipoles,[154] which however exhibit a strong conformational dependence and low transferability. Alternatively, the MM environment can be described with the help of effective fragment potentials.[152]

There is no guarantee that the effective point charges used in (biomolecular) force fields are a good approximation to the quantum mechanical charge distribution. This could be especially questionable if one considers the fact that in a QM/MM approach the QM part might be described with a different electronic structure method than the one that was used for the parametrization of the classical force field. Furthermore, not all current biomolecular force fields use a QC-based approach to derive charges, and the ones that do are based on small gas-phase molecules for a determination of (fixed) point charges. In addition, the point charge approximation might break down. These issues about possible inconsistencies between QM and MM electrostatics are further discussed in section 3.5.3 as one of the possible pitfalls of QM/MM simulations.

While the approximation of a constant kinetic energy contribution and the description of the electrostatic interactions via point charges might be reasonable for nonbonded parts of the MM system, these assumptions are sure to break down in the description of covalent chemical bonds that are characterized by highly inhomogeneous, highly directional, and highly overlapping distributions of the electron density. Clearly, simple van der Waals/electrostatic descriptions such as those in eqs 65 and 66 cannot reproduce the intricacies of chemical bonding. In most force fields, all interactions with atoms separated by 1, 2, and 3 bonding interactions are thus described by explicit mechanical bond, angle, and torsional terms given by

$$E_{MM}^{bonded} = \sum_d \frac{1}{2}k_d(d_{IJ} - d_0)^2 + \sum_\theta \frac{1}{2}k_\theta(\theta_{IJK} - \theta_0)^2 + \sum_\phi\sum_n k_n[1 + \cos(n\phi_{IJKL} + \phi_0)] \quad (67)$$

where the first term runs over all bonds d with harmonic force constant k_d and equilibrium bond length d_0. The second term includes all bonding angles θ with harmonic force constant k_θ and equilibrium bond angle θ_0, and the last term is a sum over all dihedral angles ϕ with multiplicity n, and corresponding force constants k_n and phases ϕ_0. E_{MM} is thus represented by the sum of the three terms in eqs 65−67, and the interaction energy E_{QM-MM} in eq 63 becomes

$$E_{QM-MM} = \sum_I \int \frac{q_I}{|\mathbf{R}_I - \mathbf{r}|^2}\rho_{QM}^{el+nuc}(\mathbf{r})\,d\mathbf{r} + \sum_I\sum_{I'} 4\varepsilon_{I'I}\left(\left(\frac{\sigma_{I'I}}{R_{I'I}}\right)^{12} - \left(\frac{\sigma_{I'I}}{R_{I'I}}\right)^{6}\right) + \sum_d \frac{1}{2}k_d(d_{I'I} - d_0)^2 + \sum_\theta \frac{1}{2}k_\theta(\theta_{I''J''K''} - \theta_0)^2 + \sum_\phi\sum_n k_n[1 + \cos(n\phi_{I''J''K''L''} + \phi_0)] \quad (68)$$

where I' runs over all QM atoms, I over MM atoms, and (I'', J'', K'') and (I'', J'', K'', L'') refer to the triple or quadruple bonded sets of atoms in which at least one atom is QM and the rest are MM. In this formulation, the van der Waals term and bonding terms are acting only on atomic positions and are not directly felt by the electronic system.

On the QM level, the errors introduced by the MM representation and consequently by the E_{QM-MM} interaction can in principle be compensated by a correction term ΔV to the total potential V_{tot} that the electrons of the QM part experience:

$$V_{tot} = V_{QM} + V_{QM/MM} + \Delta V \quad (69)$$

$$\Delta V = V^{NL} + \Delta V^{MM} \quad (70)$$

where ΔV^{MM} accounts for the errors in the MM representation that include factors due to the deviation of the classical electrostatic potential from the full electronic structure description and the reduction of the electronic density to a point charge representation and the differences in the van der Waals interactions and bonded terms. The other term, V^{NL}, describes the nonlinearity corrections due to the density partitioning in eq 58.

To minimize the latter, the QM part must be chosen in such a way that the electronic wave functions are well localized within the QM region. As discussed in section 3.5.2, when this criterion is not met, V^{NL} gains importance and must be addressed explicitly.

In practice, the importance of ΔV can be evaluated by determining how much the electron density in the QM region obtained from a QM/MM calculation differs from that of a full QM description. Ideally, the density, ρ_{QM}, of the QM subsystem should match as closely as possible the density of the true (or full) QM system, ρ_{true}, within the region of interest. According to the Hohenberg−Kohn theorem,[74] if the two densities are identical, all of the calculated properties for the QM region are identical to those of the real system. In other

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

words, if we determine the correction potential ΔV in such a way that the total electronic potential in a QM/MM simulation V_{tot} of eq 69 minimizes the density difference:

$$\int_{\Omega} (\rho_{true}(\boldsymbol{r}) - \rho_{QM}(\boldsymbol{r}))^2 \, d\boldsymbol{r} \tag{71}$$

where Ω represents a suitably chosen volume for the QM region, the QM calculation within a QM/MM framework approaches the full QM reference result in an optimal way. However, in practice, QM/MM simulations are not always performed in such ideal conditions that minimize ΔV, and in the next section, we will discuss the most frequently encountered sources of error, while practical issues such as the selection of the QM region and QM/MM boundary that can help reducing ΔV are elaborated in section 3.6.

3.5. Pitfalls

Most QM/MM errors are caused by an insufficient match between the approximate MM representation of E_{MM} and E_{QM-MM} in eq 68 with respect to the full quantum mechanical formulation in eq 64.

3.5.1. Electron Spill-Out. The artificial phenomenon referred to as "electron spill-out" is a common pitfall of QM/MM simulations, especially when using an extended, highly flexible basis set, such as plane waves. This problem arises from the approximation in eq 65, where the Pauli exchange interactions between QM and MM systems are only considered on the level of atomic pair potentials. As such terms do not explicitly affect the electrons, an electronic correction term, ΔV_{xc}^{NL}, must be included. As we have seen, this term is especially important for regions with overlapping or nearly overlapping densities between QM and MM parts, which is particularly the case for the nearby MM atoms surrounding the QM region. The reason this phenomenon is called "electron spill-out" is linked to the observation that in the absence of electronic wave functions in the MM part, the electrons in the QM system do not experience any Pauli repulsion from the closed-shell cores of the MM region. This can lead to a localization of QM density on nearby positively charged classical point charges, that is, a spill-out of the electron density into the MM region.

Such an artificial behavior can be prevented by applying Gaussian smeared classical charges or by screening the electrostatic interactions with suitably constructed analytic pseudopotentials:[155]

$$E_{QM/MM}^{ele} = \sum_{I \in MM} q_I \int d\mathbf{r}\, \rho(\mathbf{r}) \nu_I(|\mathbf{r} - \mathbf{R}_I|)$$

$$\nu_I(|\mathbf{r} - \mathbf{R}_I|) = \nu_I(r_I) \frac{r_c^4 - r_I^4}{r_c^5 - r_I^5} \tag{72}$$

where q_I is the classical point charge found at $\mathbf{R}_I$ and r_c is the covalent radius of atom I. In this way, the Coulombic interaction potential is modified at short-range and "spill-out" is prevented.

3.5.2. QM/MM Boundaries through Covalent Bonds. Most of the approximations discussed in section 3.4 are strongest at the QM/MM boundary and break down in situations where the real electron density distribution of the MM part is highly anisotropic or where the densities of the QM and MM atoms are strongly overlapping. Both situations occur in the case of covalent chemical bonds. Force fields resolve this situation by describing covalent bonds with special bonding terms, as shown in eq 67. However, for the E_{QM-MM} term, this presents problems as the bonding terms are functions of the atomic coordinates only and cannot influence the electronic potential in a direct way.

Unfortunately, in the case of biological macromolecules, oftentimes the QM/MM partitioning has to be defined in such a way that it intersects a covalent bond between a QM and MM atom. In this special case, the QM system contains unsaturated valence electrons, which, in nature, would be chemically reactive and induce strong artifacts in QM/MM simulations. The problem of treating mixed QM/MM bonds has plagued QM/MM simulations and many different approaches have been developed to address these issues that are reviewed in refs 79a and 143. One way that this situation can be dealt with is by using boundary pseudopotentials. In this case, the correction potential ΔV in eqs 69 and 70 can, for example, be represented by a monovalent pseudopotential, constructed in such a manner that the electrons of the QM region are correctly distributed and scattered properly by the MM region. This monovalent pseudopotential is located at the first covalently bound MM atom at the QM/MM boundary ("link atom"). In plane wave/pseudopotential-based QM/MM simulations, link atoms can be represented with empirically (re)parametrized pseudopotentials of, for example, the von Barth−Car-type:[156]

$$V_{ps}(r) = -\frac{Z_V}{r} \operatorname{erf}\left(\frac{r}{r_c}\right) + \sum_{l=0}^{n} V_l(r) \hat{P}_l \tag{73}$$

where

$$V_l(r) = (a_l + b_l r^2)\, e^{-(r/r_l)^2} \tag{74}$$

The parameters $\{\sigma_i\} = \{r_c, r_l, a_l, b_l\}$ are adjustable and can be determined by minimizing a density penalty:

$$\{\sigma_i\} = \arg\min F[\rho_{QM}(\mathbf{r}), \{\sigma_i\}]$$

$$= \int_{\Omega} d\mathbf{r} \, |\rho_{ref}(\mathbf{r}) - \rho_{QM}(\mathbf{r}), \{\sigma_i\}|^2 \tag{75}$$

ρ_{ref} is a reference density that approximates ρ_{true} of eq 71. In practice, ρ_{ref} is first determined by performing a QM/MM simulation with an extended QM region[157] (Figure 5).

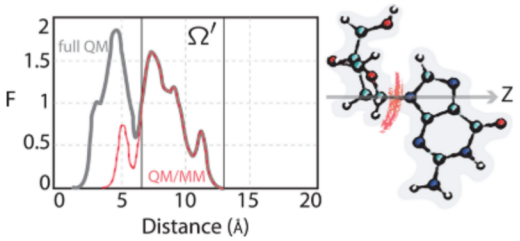


Figure 5. Density profile along the gray arrow indicated for the molecule on the right. Black, density from a full QM calculation; red, corresponding density from a QM/MM calculation with link atom pseudopotentials that have been optimized via density fitting.

This or analogous procedures can be used to construct empirical monovalent or heptavalent pseudopotentials for boundary cuts[158] and, in a semiempirical context, the parametrization of special link atoms.[159] Other approaches aimed to remedy the issue surrounding unsaturated valencies include adding hydrogen or fluorine atoms to cap the QM atoms at the boundary or a representation of the boundary QM atoms with frozen frontier orbitals.[160] In the case of hydrogen

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

(or fluorine) capping, additional degrees of freedom are introduced, and the artificial interactions of "non-existing" dummy atoms with the classical environment must be excluded from the E_{QM-MM} term. With the frozen orbital approach, on the other hand, no additional physical interactions or degrees of freedom are introduced; however, a different drawback makes this approach less appealing: the frozen orbitals must be determined on small model systems, and they cannot adapt (i.e., they are frozen) to a changing environment, while link-atoms (like pseudopotentials) are, to some extent, adaptable and transferable to different environments. More recently developed approaches are the polarized-boundary redistributed charge and the polarized-boundary redistributed charge and dipole schemes that allow for a self-consistent polarization of the MM regions close to covalent bond cuts[161] and an approach for optimal link orbitals based on density matrix embedding.[162]

3.5.3. Compatibility between QM and MM Electrostatics. A reduction of the quantum mechanical charge distribution to a set of effective (sometimes empirical) point charges, such as the ones used in current biomolecular force fields, cannot be expected to closely reproduce the left-hand side of eq 66, that is, to be fully consistent with the electronic structure method used for the QM part. Fortunately, tests for a number of prototypical systems show that, although the magnitude of effective point charges used in different force fields can vary to a large degree, the average electrostatic potentials seem to be in surprisingly good agreement with each other as well as with a full DFT description.[163] Full QM and QM/MM Car–Parrinello MD simulations of a Gly-Ala dipeptide (the QM part) in its zwitterionic form in aqueous solution (the MM part) show that standard, nonpolarizable force fields like AMBER (parm95)[164] and GROMOS96[165] have the capacity to reproduce the average electrostatic field of the quantum simulations within roughly 10% and yield solvation pattern in close agreement with the QM-based descriptions. On the other hand, for certain systems, larger deviations in the description of the electrostatic properties can occur due to a breakdown of the point charge approximation.[166]

Other issues pertaining to incompatibilities between QM and MM parts include imbalances in the strength of electrostatic interactions (QM–QM versus QM–MM versus MM–MM) that can lead to artificial preferences for specific intermolecular interactions (e.g., preferred h-bonding between QM water molecules as compared to QM–MM or MM–MM equivalents), which will ultimately lead to segregation effects. Another possible source of inconsistencies arises from the exclusion rules, that is, the selective exclusion or scaling of nonbonded interactions for pairs, triples, or quadruples of atoms connected via chemical bonds, which are applied in most force fields. A rigorous equivalent of these rules in context of a QM (or QM/MM) treatment is not obvious. The transferability of such a classical description to a many-body QM framework is far from trivial but can be achieved by a mapping of the many-body electronic Hamiltonian to a pair additive point charge representation.[167] Fortunately, it turns out that use of such a fully consistent QM/MM analogue of classical exclusion rules has no significant impact on most QM/MM simulations.

The overall consistency between QM and MM representations of the system can be further improved by applying a force-matching approach (discussed in more detail in section 3.10.1.7) to generate in situ force fields with maximum QM compatibility.

3.6. Practical Issues

In practice, the accuracy and reliability of QM/MM simulations can be influenced significantly through judicious choices of the QM and MM method, the QM region, and the specific location of the QM/MM boundary. This section is thus devoted to the discussion of some of the practical issues that have a direct impact on QM/MM performance.

3.6.1. The QM Part. *3.6.1.1. Choice of QM Method.* As was already discussed in section 2.1.2, the QM/MM formulation houses all types of QM methods spanning the whole spectrum of accuracy and computational cost. Whereas semiempirical/MM approaches allow the treatment of large QM regions of the order of thousands of atoms for extended time scales (from 100 ps to 1 ns), they bear the disadvantage that the performance of the semiempirical methods for the problem of interest has to be carefully evaluated beforehand. DFT-based QM/MM simulations typically contain QM regions of few hundreds atoms and comprise sampling times of the order of 10–100 ps. Ab initio (wave function-based)/MM schemes with MC-CASSCF/CASPT2 or higher level coupled cluster approaches, on the other hand, are often used in static QM/MM applications for a limited number of (optimized) configurations along, for example, a minimum energy path.

3.6.1.2. Choice of QM Region. The choice of the QM region is essentially always based on previous (bio)chemical information, and larger QM parts alleviate the problem that segments of the system that could be of potential electronic importance might accidentally be omitted. What makes the selection of a suitable size of the QM region challenging is to find the QM to MM ratio that leads to an optimal accuracy/speed trade-off in particular in view of the fact that increasing the size of the QM region will come at the prize of a more limited phase space sampling. In practice, the density difference in eq 71 can be used to check convergence with respect to QM size by comparing the electron density in the region of interest in QM/MM simulations with increasing sizes of the QM region. Recent studies provide concrete examples of the above ideology by performing QM/MM simulations with up to 1600 atoms in direct comparison with full QM calculations.[168] Two interesting findings arise from these studies: (i) the convergence to the full QM limit is much faster using a QM/MM technique than a QM method on isolated cluster models; and (ii) quantitatively accurate results necessitate fairly large QM regions. The authors indicate that the QM subset should comprise a minimum of 300 atoms to achieve accuracies on the order of 10 kJ mol^{-1} and 300–1000 atoms for an accuracy of 2 kJ mol^{-1}. Furthermore, Solt et al. have shown that the error in nuclear forces at the center of the QM region can be strongly dependent on QM size. Their work reports that forces may not fall within acceptable limits even for quantum regions with radii of 9 Å.[168a]

3.6.2. The MM Part. The accuracy and predictive power of QM/MM simulations relies heavily on an accurate representation of the MM region. In standard biomolecular force fields, the MM part is described with bonded and nonbonded terms of the form of eqs 65–67. Force fields of this type include OPLS,[164,169] CHARMM,[170] AMBER,[171] and GROMOS,[172] which have all been highly successful in the study of biological systems but also have some well-known intrinsic and practical limitations. There are various conditions for which standard

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217–6263

(unpolarizable) force fields fail to correctly depict intermolecular interactions, examples of which include polarization and charge transfer phenomena. In the context of QM/MM simulations, these problems can, in principle, be easily circumvented if the available computational resources allow for extended QM regions, in which all of the electronically relevant components are included.

Alternatively, if the polarization effects extend over large areas, polarizable force fields might be advantageous in combination with a QM/MM scheme. In this case, the mutual polarization of QM and MM part has to be treated self-consistently ("polarizable embedding") either with matrix inversion, iteration, or via an extended Lagrangian scheme, in which the induced polarization in the form of fluctuating charges or dipoles is treated as dynamical variable.[173]

Most of the popular fixed point charge force fields have also been extended to polarizable versions (AMBER,[174] CHARMM,[175] and OPLS/PFF[176]), although they have not yet found a widespread use similar to their nonpolarizable counterparts. These extensions of fixed point charge models to the polarizable case are based on three main categories: (1) the shell model (Drude oscillator, charge-on-spring models)[175c,177] in which atoms are represented by two oppositely charged shells connected by a harmonic potential with a force constant proportional to the atomic polarizability and the intershell distance varies in response to the surrounding electric field; (2) fluctuating charge (and dipole) schemes based on electronegativity equalization (charge equilibration, chemical potential equalization, fluctuating charge models);[175a,b,176] and (3) induced point dipoles (respectively, multipoles) centered at atomic sites.[174,176]

In addition, significant progress in the development of second generation polarizable force fields, such as the AMOEBA (Atomic Multipole Optimized Energetics for Biomolecular Applications) force field, developed by Ponder and co-workers,[178] has been made in recent years.[179] AMOEBA differs from standard biomolecular force fields in a number of additional terms and the replacement of fixed partial charges with polarizable atomic multipoles up to quadrupole moments. With respect to computational cost, AMOEBA and other transferable polarizable force fields, such as the sum of interactions between fragments ab initio (SIBFA)[180] and nonempirical molecular orbital (NEMO)[181] methods, are between polarizable biomolecular force fields and QM/MM approaches such as the Direct Reaction Field (DRF) method.[182] For a more detailed account on polarizable force fields, we would like to direct readers to helpful reviews.[183]

Despite the recent progress made in this field, there have been relatively few QM/MM applications that treat the MM part with a polarizable force field. Some of the rare examples can be found mostly in the area of spectral shift calculations such as the early pioneering work of Thompson et al.[184] who used a INDO/S/MMpol model to study microsolvation effects on the ground- and excited-state properties of the bacteriochlorophyll b dimer of the photosynthetic reaction center of *Rhodopseudomonas viridis* and the recent implementation of a fully consistent polarizable embedding scheme implemented in the context of HF and DFT/MM simulations by Olsen et al.[185]

3.6.3. The QM/MM Boundary. All of the inherent approximations in QM/MM approaches discussed in section 3.4 affect particularly the region around the QM/MM boundary. It is therefore advisible to move this border as far as possible away from the core region of interest, which should be located in the center of the QM part. Furthermore, to minimize errors, the QM/MM boundary should preferably be placed in regions where the density overlap is small (i.e., nonbonded QM/MM border) or if that is not possible within regions of homogeneous density (e.g., cutting through apolar bonds). In most QM/MM approaches, the initial partitioning into QM and MM atoms is maintained throughout the simulations, an assumption that can break down if, for example, an active site is highly water accessible and new water molecules are approaching the catalytically relevant part of the system. In this case, the definition of QM and MM regions has to be manually redefined. Several methods have also been developed that allow for adaptable QM/MM boundaries[186] or smooth morphing schemes.[187]

3.7. QM/MM Static versus Dynamic Approaches

The term "molecular mechanics" in the QM/MM terminology is mainly used from a historical perspective because the combined use of QM/MM was a natural extension of early studies of enzymatic reactions that used molecular mechanics optimizations to generate initial coordinates for subsequent QM calculations.[188] The acronym QM−MM was first coined by Gao et al.[189] and has been used ever since, although many QM/MM approaches to date are actually performing molecular dynamics (instead of molecular mechanics) calculations.

In a number of static QM/MM calculations, the MM environment is reduced to an electrostatic point charge representation, which is used to optimize the wave functions of the QM part in the presence of an external field (electrostatic embedding). This has been extensively applied in the calculation of environment-induced shifts of excitation energies but might be insufficient in other cases.[190] For the characterization of certain enzymatic reactions, performing energy minimizations, transition state searches or single point energy calculations may already provide valuable information.[143c] However, most enzymes function at body temperature in contrast to conditions at zero Kelvin (as in geometry optimized structures), in which water molecules are frozen into an amorphous state, which might not be representative. Instead, using a full MM representation of the type in eqs 65−67 allows one to perform mixed QM/MM MD studies in which either the full system is evolved dynamically or only the adjoining MM region is flexible while more distant MM parts are kept frozen. This enables thermal sampling of statistically relevant conformations that all contribute to the overall properties of the system. In addition, the (low) probability of the occurrence of, for example, reactive (near attack) configurations can make important contributions to the entropic part of the free energy barrier of an enzymatic reaction.

A natural question that follows is "How much sampling is considered to be enough?" Current first-principles QM/MM MD methods sample the conformational landscape for durations on the order of tenths of nanoseconds. Using a semiempirical description of the QM region, this can be extended to the nanosecond time scale. Most biochemical reactions, on the other hand, take place on microsecond to milliseconds or seconds. QM/MM simulations thus often rely on a presampling of the configurational space (and subsequent cluster analysis) using purely force field-based dynamics that can reach into the hundreds of nanoseconds and microsecond time scales.[191] With the recent advances in computer resources, purely force field-based MD can even extend into the millisecond range.[192] The restricted sampling in QM/MM

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

simulations is an important limitation that will be discussed in more detail in section 3.10.1 on the combination of QM/MM simulations with enhanced sampling methods.

Looking at various QM/MM studies, it becomes clear that the required amount of conformational sampling is highly dependent on the system and the problem of interest. McCammon and co-workers generated enzyme−substrate conformations with classical molecular dynamics simulations and subsequently mapped out a minimum reaction energy path for each conformational snapshot with combined QM/MM calculations.[193] The authors studied the convergence of potential energy calculations of acetylcholinesterase by inspecting different frames of a 1 ns MM trajectory. In this particular case, while the orientation and behavior of the catalytic triad are consistent over this time scale, they report significant fluctuations in the energy barrier along the reaction coordinate with the inclusion of only 10 snapshots (equally spaced over 1 ns). Mulholland and co-workers have performed a clustering analysis of a 1 ns classical trajectory for fatty acid amidehydrolase.[194] The 4000 snapshots chosen in this way lead to geometrical fluctuations of the active site, which significantly affect the overall energetic barrier. On the other hand, an extensive sampling study performed by Carloni and co-workers on HIV-1 protease indicates that some active sites may not be affected by slow large-scale domain motions.[195]

Unfortunately, it is difficult to predict a priori whether a system requires sampling over longer time scales. Various sampling schemes and practical techniques have been developed to provide a wide range of strategies that help to address this key point (see also section 3.10.1).

Large-scale sampling coupled to a QM/MM approach is not limited to molecular dynamics methods. Monte Carlo techniques in conjunction with QM/MM methods are increasingly popular.[196] While not a conventional electronic-structure QM method, the empirical valence bond (EVB) method has also been applied with much success, notably by Warshel and colleagues.[197]

For the remainder of this review, however, we will focus on DFT/MM methods coupled to all-atom molecular dynamics, with a special emphasis on Car−Parrinello QM/MM simulations in the electronic ground state and TDDFT-based hybrid QM/MM simulations for excited states.

3.8. QM/MM Car−Parrinello MD

3.8.1. The QM/MM Extended Lagrangian. The Car−Parrinello Lagrangian in eq 15 and the corresponding equations of motion 17 and 18 can easily be generalized into a QM/MM context:

$$\mathcal{L}_{ex}^{QM/MM} = \frac{1}{2}\mu \sum_i \int d\mathbf{r}\, \dot{\phi}_i^*(\mathbf{r})\dot{\phi}_i(\mathbf{r}) + \frac{1}{2}\sum_I M_I \dot{\mathbf{R}}_\mathbf{I}^2 - E_{MM} - E_{QM/MM} - E_{QM} + \sum_{i,j}\Lambda_{i,j}\left(\int \{d\mathbf{r}\, \phi_i^*(\mathbf{r})\phi_j(\mathbf{r})\} - \delta_{i,j}\right) \quad (76)$$

where E_{QM} is given by the Kohn−Sham energy density functional (eq 37):

$$E_{QM} = E_{KS}[\phi_i, \mathbf{R}_I] = -\frac{1}{2}\sum_i \int d\mathbf{r}\, \phi_i^*(\mathbf{r})\nabla^2\phi_i(\mathbf{r}) + \int d\mathbf{r}\, V^{ex}(\mathbf{r})\rho_{QM}(\mathbf{r}) + \frac{1}{2}\int d\mathbf{r}\, d\mathbf{r}'\, \rho_{QM}(\mathbf{r})\frac{1}{|\mathbf{r}-\mathbf{r}'|}\rho_{QM}(\mathbf{r}') + E_{xc}[\rho_{QM}(\mathbf{r})] \quad (77)$$

$$\rho_{QM}(\mathbf{r}) = 2\sum_i \phi_i^*(\mathbf{r})\phi_i(\mathbf{r}) \quad (78)$$

and, for the spin unpolarized case, the electron density, $\rho_{QM}(\mathbf{r})$, is given by the sum of the densities of the doubly occupied one-particle states (eq 78). E_{MM} is described by a standard (nonpolarizable) biomolecular force field (eqs 65−67). The QM/MM CPMD simulations described in section 4 of this contribution have been performed using the classical force fields GROMOS96[165] and AMBER[171a] in combination with a particle−particle−particle mesh (P3M) treatment of the long-range electrostatic interactions.[198] The QM/MM coupling energy E_{QM-MM} in eq 76 is given by eq 68. Current implementations of DFT-based first-principles molecular dynamics are based on plane wave (PW) or mixed atom-centered/PW[199] basis sets. Using a PW scheme, the calculation of the first term of eq 68, which evaluates the interaction of the electron density at every point in space (defined by the real space mesh of the PW expansion), with every classical point charge is computationally demanding because it involves of the order of ($N_r \times N_{MM}$) operations, where N_r is the number of grid points (typically ca. 100^3) and N_{MM} is the number of point charges (atoms) in the MM subset (i.e., of the order of 10 000−100 000). Therefore, the interactions between QM atoms and the more distant MM atoms are dealt with by a Hamiltonian term that explicitly couples the multipole moments of the quantum charge distribution with the classical point charges. Furthermore, a third, intermediate layer is typically created that makes use of a coupling via variational D-RESP[167,200] charges. Recently, dual grid[201] and multi grid approaches with Gaussian expansions[202] have also been implemented.

Using a PW basis, special care must be taken when dealing with periodic boundary conditions (PBC). A typical QM/MM Car−Parrinello simulation is performed with two simulation cells (or multiple grids), a small subcell for the QM part and a larger one comprising the total system. For the QM part, the Poisson equation is solved for an isolated system, that is, by decoupling the periodic images.[203] In this way, a smaller periodic box is used for the PW expansion, while the periodic boundary condition is only applied for the entire system.

Because of the fact that in PW-based QM/MM schemes the wave function has a particularly high spatial flexibility as compared to the minimal basis set representations used in semiempirical QM/MM simulations, special care must be taken to represent the QM/MM boundary in the most accurate and consistent way possible to minimize the QM/MM correction term, ΔV (see section 3.4), as discussed in sections 3.5.1 and 3.5.2.

QM/MM CPMD simulations provide an efficient and quantitative platform for performing dynamical simulations of complex extended systems, including 100−1,000 QM atoms and 10,000 to 100,000 MM atoms for time scales of the order of 10−100 ps. Using this framework, chemical processes can be studied in light of an explicit environment, such that the steric and electrostatic effects of their surroundings are accounted for.

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

This QM/MM implementation is freely distributed with the CPMD package (www.cpmd.org).

3.9. TDDFT/MM Implementations for Adiabatic and Nonadiabatic Excited-State Dynamics

The development of time-dependent density functional theory has opened a new avenue for first-principles studies of the properties and dynamics of fairly large molecular systems in electronically excited states. The implementation of a QM/MM scheme for the calculation of excitation energies within LR-TDDFT and a nonpolarizable MM environment is straightforward. Because all excited-state properties are calculated solely as functionals of the ground-state electron density (or equivalently of the corresponding Kohn−Sham orbitals), all that is needed is a QM/MM method for the electronic ground state such as the one described with eq 68 where the effective point charges of the MM environment are taken into account as an additional contribution to the external potential:

$$V'_{\mathrm{ex}}(\mathbf{r}, \mathbf{R}^{\mathrm{QM}}, \mathbf{R}^{\mathrm{MM}}) = V_{\mathrm{ex}}(\mathbf{r}, \mathbf{R}^{\mathrm{QM}}) + V_{\mathrm{ex}}(\mathbf{r}, \mathbf{R}^{\mathrm{MM}})$$

$$V'_{\mathrm{ex}}(\mathbf{r}, \mathbf{R}^{\mathrm{QM}}, \mathbf{R}^{\mathrm{MM}}) = \sum_{I\in\mathrm{QM}} \frac{Z_I}{|\mathbf{R}_I - \mathbf{r}|} + \sum_{m\in\mathrm{MM}} \frac{q_M}{|\mathbf{R}_m - \mathbf{r}|} \tag{79}$$

The new external potential $V'_{\mathrm{ex}}(\mathbf{r}, \mathbf{R}^{\mathrm{QM}}, \mathbf{R}^{\mathrm{MM}})$ generates a new (polarized) ground-state density $\rho'_0(\mathbf{r})$, and its linear response to a frequency-dependent perturbation yields the new excitations energies ω'_n via eqs 31 and 34, respectively .

For adiabatic LR-TDDFT/MM dynamics in an electronically excited state, the classical environment has to be coupled to the excited electron density. The electron density in the nth excited state $\rho'_n(\mathbf{r})$ can be reconstructed from the ground-state density and its linear response $\Delta\rho'_n(\mathbf{r})$:

$$\rho'_n(\mathbf{r}) = \rho'_0(\mathbf{r}) + \Delta\rho'_n(\mathbf{r}) \tag{80}$$

and the QM/MM interaction energy $E_{\mathrm{QM-MM}}$ of eq 68 becomes

$$\begin{aligned} E_{\mathrm{QM/MM}} = & \sum_I \int_\Omega \frac{q_I}{|\mathbf{R}_I - \mathbf{r}|^2} (\rho_{\mathrm{QM}}^{\mathrm{el+nuc}}(\mathbf{r}) + \Delta\rho'(\mathbf{r}))\, \mathrm{d}\mathbf{r} \\ & + \sum_I \sum_{I'} 4\varepsilon_{I'I} \left(\left(\frac{\sigma_{I'I}}{R_{I'I}}\right)^{12} - \left(\frac{\sigma_{I'I}}{R_{I'I}}\right)^{6} \right) \\ & + \sum_d \frac{1}{2} k_d (R_{I'I} - d_0)^2 + \sum_\theta \frac{1}{2} k_\theta (\theta_{I''J''K''} - \theta_0)^2 \\ & + \sum_\phi \sum_n k_n [1 + \cos(n\phi_{I''J''K''L''} + \phi_0)] \end{aligned} \tag{81}$$

Equation 81 can also be used for a further extension to nonadiabatic LR-TDDFT dynamics within a Tully Fewest Switches surface hopping approach. In this case, $\Delta\rho'_n(\mathbf{r})$ corresponds to the running state for which the nuclear forces are derived.

An extension to P-TDDFT/MM dynamics is also possible by replacing the external potential in the total Kohn−Sham potential V^{KS} in eq 40 by the one defined in eq 79. The propagation equation then reads:

$$\begin{aligned} & \left\{ -\frac{1}{2}\nabla^2 + V_{\mathrm{H}}(\mathbf{r}, t) + V_{\mathrm{xc}}(\mathbf{r}, t) + V_{\mathrm{ex}}{}'(\mathbf{r}, t) \right\} \phi_i(\mathbf{r}, t) \\ & = i\frac{\partial}{\partial t}\phi_i(\mathbf{r}, t) \end{aligned} \tag{82}$$

where $V_{\mathrm{H}}(\mathbf{r}, t)$ is the time-dependent Hartree and $V_{\mathrm{xc}}(\mathbf{r}, t)$ the time-dependent exchange-correlation potential.

3.10. Overcoming Limitations of QM/MM Simulations

Despite the steady advances in computer power, numerous limitations continue to encumber the use of QM/MM MD simulations in the study of biological phenomena. The main challenges are (i) remaining restrictions due to the tractable system size, (ii) the short accessible time window that limits the configurational sampling and thus the statistical accuracy of the calculated thermodynamic properties, as well as (iii) the limited intrinsic accuracy of QM, MM and QM/MM descriptions. Among these three issues, the time scale problem is especially severe, in particular for first-principles-based QM/MM MD, and most simulations have to be coupled with enhanced sampling techniques to be able to reach the characteristic times scales involved in biological events.

The time scale problem is a longstanding issue for all MD simulations, and a vast array of methods have been developed to enhance sampling efficiency (for reviews, see refs 204 and 205) that are also applicable in a QM/MM context. The main strategies range from a simple rescaling of atomic masses, to using temperature to accelerate the sampling of rare events (e.g., Temperature-Accelerated MD (TAMD);[206c,d] Canonical Adiabatic Free Energy Sampling (CAFES),[206b] parallel tempering[207]), partitioning into slow and fast degrees of freedom (e.g., multiple time scale MD,[208] constraints,[209] essential dynamics,[210] subspace integration MD[211]), imposing restrictions to specific phase space regions (e.g., thermodynamic integration in the Blue Moon ensemble,[212] reaction path method,[213] path optimization,[214] transition path sampling[215]), to well-controlled deformations of the underlying potential energy surface designed in such a way as to lower activation free energies and increase the rate for activated events (e.g., umbrella sampling,[216] multidimensional adaptive umbrella sampling,[217] hypersurface deformation[218] local elevation,[219] conformational flooding,[220] hyperdynamics,[221] accelerated MD,[222] metadynamics,[223] and paradynamics[224]). In this section, we give a brief overview of the enhanced sampling techniques and related free energy methods that have most frequently been applied in QM/MM simulations.

3.10.1. Enhanced Sampling: Configuration and Systems Space. *3.10.1.1. Multiple Time Step QM/MM.* One of the most direct ways of enhancing sampling in mixed QM/MM MD simulations is to integrate the classical equations of motion of the QM and the MM parts with different time steps[225] using a multiple time step algorithm.[208] By assigning a large time step to the QM region, the computationally expensive electronic structure evaluation has to be performed only at longer time intervals, leading to an increased efficiency and extended simulation times. Because there is usually no natural separation between the time scales of the atomic motion in the QM and MM parts, this division is somewhat arbitrary, although a dynamical separation can be achieved to a certain extent by rescaling masses. However, such an approach is in general not aimed at reproducing the physical dynamics of the QM/MM system but has rather to be seen as a more efficient exploration of phase space, in which every QM configuration "sees" many different MM environments; that is, the MM part that is usually larger and spans a higher dimensional configurational space is "oversampled" with respect to the QM region, which is exposed to the potential of mean force created by the MM surrounding. Like for many other enhanced sampling techniques, multiple

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

time step QM/MM is a proficient tool for a more efficient evaluation of thermodynamic ensemble averages at the cost of giving up on the actual dynamics of the system.

3.10.1.2. Using Temperature To Enhance Sampling. Because the probability p of crossing energetic barriers of height E increases exponentially as a function of temperature T ($p \propto e^{-\beta E}$ where $\beta = (k_B T)^{-1}$, and k_B is the Boltzmann constant), increasing the system's temperature can be a very efficient means to enhance the sampling of rare events. For this reason, first-principles MD and QM/MM simulations at elevated temperatures are routinely used to extend phase space exploration. Furthermore, whereas essentially all of the enhanced sampling techniques are based on an identification of "critical" slow motions, the temperature of the system is a global variable that can be tuned without a priori knowledge about the system's characteristics. However, the configurations explored at high temperatures might not necessarily be relevant for the low temperature ensemble; for example, extended MD of a protein at high temperature will lead to denaturation and unfolding. To solve this problem, methods such as replica exchange (or parallel tempering)[207b] have been introduced in which several replicas of the system are simulated simultaneously at different temperatures and configurations are exchanged at given intervals, applying a Metropolis criterion to ensure that all replicas preserve proper Boltzmann distributions. Another method that allows efficient, temperature-accelerated sampling of the physical free energy surface of a subsystem (e.g., the QM region) in a (MM) environment is Canonical Adiabatic Free Energy Sampling (CAFES).[206] The two systems are dynamically decoupled by introducing fictitious masses so that the QM part evolves slowly and adiabatically on the potential of mean force generated by the MM surrounding. In this way, different temperatures can be applied for the two parts of the system without introducing an irreversible heat flow. CAFES-QM/MM simulations have, for instance, been applied to characterize the reaction mechanism of HIV-1 protease.[226]

3.10.1.3. Atomic and Electronic Bias Potentials. A further category of enhanced sampling approaches modifies the underlying potential energy surface through the introduction of suitably chosen bias potentials V_{bias} that either lower the energy at the top of barriers or increase the energy of minima, leading to exponentially increased sampling rates proportional to $\sim e^{\beta V_{bias}}$. The thermodynamic average of a function $\langle f(\mathbf{p}, \mathbf{q})\rangle$ in the original unbiased ensemble can be obtained from reweighted averages of the biased ensemble:

$$\langle f(\mathbf{p}, \mathbf{q})\rangle_{\beta H} = \frac{\langle f(\mathbf{p}, \mathbf{q})\, e^{\beta' H' - \beta H}\rangle_{\beta' H'}}{\langle e^{\beta' H' - \beta H}\rangle_{\beta' H'}} \tag{83}$$

where H and β are, respectively, the Hamiltonian and inverse temperature of the original ensemble, and H' and β' are those of the biased ensemble. If the modified Hamiltonian H' is chosen appropriately, the ensemble averages in the biased ensemble are converging faster than in the original one. A further requirement is that the overlap of the two ensembles has to be sufficiently high, that is, that a large number of configurations sampled for H' are also relevant for the original ensemble so that the reweighting factor $e^{\beta' H' - \beta H}$ is nonzero.

A large number of different bias potentials have been proposed. In fact, the temperature-accelerated MD techniques discussed in the previous section can also be considered as a bias potential approach, in which the inverse temperature β' is used to bias the system and in analogy to eq 83:

$$\langle f(\mathbf{p}, \mathbf{q})\rangle_{\beta H} = \frac{\langle f(\mathbf{p}, \mathbf{q})\, e^{(\beta' - \beta) H}\rangle_{\beta'}}{\langle e^{(\beta' - \beta) H}\rangle_{\beta'}} \tag{84}$$

For QM/MM simulations, a particularly interesting option is the use of classical force fields as bias potentials for the enhanced sampling of conformational space in the QM region (e.g., for the acceleration of rare dihedral transitions). The MM description of the QM part can be taken as a first approximation of the shape of the quantum mechanical potential energy surface so that the resulting potential $V' = V - V_{bias}$ is almost flat allowing for efficient crossing of barriers.[227]

The observation of rare reactive events, on the other hand, can be accelerated through the use of electronic bias potentials: a finite electronic temperature[206] or bias potentials constructed from electronic structure information such as the sum of orbital eigenvalues of the highest occupied (HOMO) and lowest unoccupied (LUMO) set of orbitals that can be used to induce an artificial rise in the energy of the HOMOs or a decrease of the LUMO energies, rendering the system more reactive.[228]

3.10.1.4. Thermodynamic Integration. Thermodynamic integration, in combination with the Blue Moon ensemble method,[229] can be used to determine free energy profiles along a given reaction coordinate $\zeta(\mathbf{r})$ and to determine, for example, activation free energies (and within transition state theory, associated rate constants) for enzymatic reactions. To explore a reaction step of the catalytic mechanism, a series of constraints are applied to force the exploration of configurations along a specified reaction coordinate.

From the average constraint force at each point along the reaction coordinate, the free energy difference between two points ζ_1 and ζ_2 along the reaction coordinate can be calculated:

$$F(\zeta_2) - F(\zeta_1) = \int_{\zeta_1}^{\zeta_2} d\zeta' \left\langle \frac{\partial H}{\partial \zeta} \right\rangle_{\zeta'}^{\mathrm{cond}} \tag{85}$$

where F is the free energy, H is the Hamiltonian, and $\langle ... \rangle_{\zeta'}^{\mathrm{cond}}$ represents the conditional average evaluated at $\zeta(\mathbf{r}) = \zeta'$. After applying the blue moon ensemble to unbias the system, the conditional average of eq 85 can be evaluated from the Lagrange multiplier λ associated with the constraint:

$$\frac{dF}{d\zeta'} = \frac{\langle Z^{-1/2}[-\lambda + k_B T G]\rangle_{\zeta'}}{\langle Z^{-1/2}\rangle_{\zeta'}} \tag{86}$$

in which

$$Z = \sum_i \frac{1}{m_i} \left(\frac{\delta \zeta}{\delta \mathbf{r}_i} \right)^2 \tag{87}$$

and

$$G = \frac{1}{Z^2} \sum_{i,j} \frac{1}{m_i m_j} \frac{\delta \zeta}{\delta \mathbf{r}_i} \frac{\delta^2 \zeta}{\delta \mathbf{r}_i \delta \mathbf{r}_j} \frac{\delta \zeta}{\delta \mathbf{r}_j} \tag{88}$$

In the case of a simple distance constraint, $\zeta = |\mathbf{r}_i - \mathbf{r}_j|$, Z is a constant, $G = 0$, and the free energy difference in eq 85 can be evaluated as

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

$$F(\zeta_2) - F(\zeta_1) = \int_{\zeta_1}^{\zeta_2} d\zeta' \frac{dF}{d\zeta'} = \int_{\zeta_1}^{\zeta_2} d\zeta' \langle \lambda \rangle_{\zeta'} \quad (89)$$

3.10.1.5. Alchemical Transformation and Free Energy Perturbation Methods. Alchemical free energy (AFE) calculations,[230] in which parts of a system can be "transmuted" or perturbed into a different chemical state, provide a rigorous estimation of the free energy difference between two chemical states by evaluating the free energy change along an alchemical path. AFE simulations allow a quantitative determination of, for example, relative binding free energies,[231] pK_a's,[232] and redox potentials.[233]

AFE has also been applied in the context of QM/MM simulations[234] using a smooth alchemical switching transformation from the starting state $A_{QM/MM}$ to the end state, $B_{QM/MM}$. The hybrid potential energy function is written as

$$U_0 = U_S^{A,QM/MM}(1 - \lambda) + \lambda U_S^{B,QM/MM} + U_e \quad (90)$$

where $U_S^{A,((QM)/(MM))}$ and $U_S^{B,((QM)/(MM))}$ are the QM/MM energy terms that describe states A and B, respectively, and U_e is the common environmental term shared by the two states. In this case, alchemical mixing is implemented on the basis of two independent QM/MM electronic structure calculations. Molecular dynamics or Monte Carlo simulations are required for sampling the hybrid system at different λ values to estimate the free energy difference between the two end states $A(\lambda = 0)$ and $B(\lambda = 1)$. Many different procedures are employed in AFE simulations, which vary with the chosen free energy simulation method (e.g., "direct" versus "indirect" schemes).[189,233a,235] The limited time scale of QM/MM simulations (in particular for first-principles-based approaches) can hamper proper convergence of QM/MM AFE simulation, and improvements to enhance the efficiency of AFE simulation techniques have been proposed.[236]

Using free-energy perturbation (FEP), the free energy can be estimated on the basis of an unperturbed Hamiltonian and a perturbation term ΔE_{pert}. By sampling ΔE_{pert}, it is possible to calculate the free energy change due to the perturbation by exponential averaging:

$$\Delta A = -1/\beta \ln\langle \exp(-\beta \Delta E_{pert}) \rangle \quad (91)$$

where $\langle\rangle$ denotes a canonical ensemble average. Free energy derivatives can be computed using various implementations, such as the "dual-topology" or the "single topology" approach. For more information on QM/MM free energy perturbation schemes, we refer the reader to the review of ref 233a.

3.10.1.6. Metadynamics. Thermodynamic integration and AFE can mostly be applied for processes that can be described with a single (collective) reaction coordinate or when the system can be transformed into the target state via a relatively small perturbation. However, for many biological and biochemical processes, the free energy surface might consist of numerous minima that are separated by barriers much larger than the thermal energies and multiple multidimensional reaction paths might contribute. Metadynamics[223,237] is a technique that was introduced by Laio and Parrinello in 2002 and is used to overcome barriers on multidimensional free energy landscapes. This method has been successfully applied to a variety of problems in biophysics, chemistry, and material science.[238]

In metadynamics, the free energy surface spanned by a limited number d of predefined collective variables (CVs), $S_\alpha(x)$, $\alpha = 1, d$ is reconstructed in a recursive manner by adding a repulsive Gaussian potential at every τ_G MD steps to the currently visited region in CV space. The total history-dependent bias potential V_G acting on the system at time t is thus given by

$$V_G(S(x), t) = \omega \sum_{t'=\tau_G,2\tau_G,\ldots t'<t} \exp\left(-\sum_{\alpha=1}^{d} \frac{(S_\alpha(x) - s_\alpha(t'))^2}{2\delta {s_\alpha}^2} \right) \quad (92)$$

where $s_\alpha(t')$ refers to the value of the CV α at time t', and ω and δs are the height and width of the Gaussian, respectively. The three parameters τ_G, ω, and δs are chosen by the user and can be adapted to the specific problem. In this way, the free energy surface is constructed successively, from the bottom of minima to the top via a history-dependent random walker. After sufficient sampling, an estimate of the reconstructed free energy, $\lim_{t\to\infty} V_G(s, t) \approx -F(s)$, is obtained.[239] The choice of the CVs strongly influences the convergence of the metadynamics approach. If the free energy surface is filled up "smoothly", it is probable that the set of chosen variables was complete.[240]

Recently, much progress has been made and many variants and extensions of metadynamics have emerged, such as well-tempered metadynamics,[241] multiple walkers metadynamics,[242] and histogram reweighted metadynamics.[243] For more information, we refer the interested reader to the many helpful reviews on this subject.[237,244] Different variants of metadynamics are frequently applied in conjunction with QM/MM molecular dynamics.[245]

3.10.1.7. QM/MM Consistent Force Fields via Force Matching. Another way of enhancing sampling is to use a lower-level, computationally more economical method for an extended phase space exploration. Ideally, the configurations visited by the lower level method coincide with the ones that would have been sampled with a higher level method. In the context of QM/MM simulations, presampling is often performed with force-field-based dynamics, but as discussed in section 3.5.3, the description of the system at the full MM level is not necessarily fully consistent with the one at the QM/MM level.

Opportunely, the data obtained in QM/MM simulations can be used in turn for a generation of consistent in situ force fields via force-matching procedures.[166,246,247] Because QM/MM force-matched force fields are constructed to reproduce the QM/MM results as closely as possible, they can be used for propagating the system on long time scales eventually applying periodic corrections through short QM/MM runs (as in the optimal potential method[248] and the "learn-on-the-fly" approach[249]). Alternatively, the classical force field description can be considered as a bias potential, and the thermodynamic properties of the corresponding QM/MM ensemble can be recalculated from the force field dynamics using the bias potential approach described in section 3.10.1.3. In addition, as the development of a reliable force field for biological systems can be a cumbersome and time-consuming task, especially for systems containing metal ions or nonstandard amino acids, QM/MM force matching can be used as an automatized way of generating force fields for parts of the system for which no parameters are available.

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

In the QM/MM force matching method,[166,247b] one copy of the system is described at the full MM level and another one at the QM/MM level with all parts of the system that have to be parametrized included in the QM part (Figure 6). The

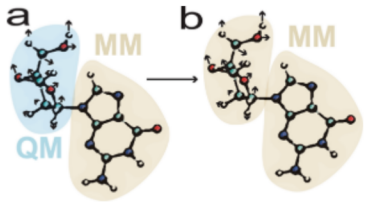


Figure 6. QM/MM force matching schemes describe one copy of the system at the full MM level and another at the QM/MM level. All parts of the system that have to be parametrized are included in the QM part. As shown in the figure, the forces on the QM part of the guanosine moiety are matched to parametrize an MM force field for this part of the molecule.

parameters of the force field $\{\sigma_i\}$ are adjusted to minimize the penalty function $\Lambda(\{\sigma_i\})$:

$$\{\sigma_i\} = \arg\min \Lambda(\{\sigma_i\}) = \frac{1}{L \cdot N} \sum_{l=1}^{L} \sum_{\alpha=1}^{N} |F_{l\alpha}^{\mathrm{MM}}(\{\sigma_i\}) - F_{l\alpha}^{\mathrm{QM/MM}}|^2 \quad (93)$$

where L is a number of atomic configurations taken from a QM/MM trajectory, N is the number of atoms in the QM part, $F_{l\alpha}^{\mathrm{MM}}(\{\sigma_i\})$ is the MM force on atom α in configuration l, and $F_{l\alpha}^{\mathrm{QM/MM}}$ is the corresponding reference force in the QM/MM setup. In contrast to the standard procedure for the development of biomolecular force fields that relies on QM calculations of model compounds in the gas phase,[250] force fields that are generated via eq 93 are built in situ in a specific electrostatic and steric environment.

QM/MM force matching has been applied to the parametrization of new anticancer azole-bridged platinum and organometallic ruthenium compounds binding to DNA[251] to describe the drug-induced DNA distortions on the nanosecond−millisecond time scale or for generating high-accuracy force fields for the protonated Schiffbase chromophore in rhodopsin.[252]

An analogous force matching approach can also be applied for an in situ refitting of semiempirical methods.[247b] At difference to the force matching adaption of classical force fields, such a semiempirical force matching scheme can also be applied for the exploration of chemical reactions.[253]

3.10.1.8. Sampling Chemical Space with Genetic Algorithms. QM/MM simulations are a powerful tool to characterize the underlying molecular mechanisms of a wealth of biological processes and thus provide us with a unique fundamental understanding of the inner workings of natural processes. In many cases, a next step that ensues is to make use of this knowledge to re-engineer biological systems in such as way as to alter their function in specific ways or to design biomimetic compounds that copy the action of natural counterparts. This leads to an inverse optimization problem, in which starting from a known target property, the corresponding system whose assets best match the desired features is sought. However, even the computational design of small biomimetic compounds or miniproteins with tailored properties requires not only efficient methods for configurational sampling but also extensive sampling of the chemical or sequence space. In fact, the sampling problem in compound space is incomparably more severe than the one we face in phase space sampling: a full exploration of all possible amino acid sequences involving even a very small number of residues is not feasible, due to the curse of dimensionality; for instance, for only eight amino acids, there are already $20^8 = \sim 10^{18}$ possible peptide sequences. Therefore, the optimization of the chemical composition of such systems through systematic one-by-one evaluations is intractable, and the problem has to be approached from an entirely different angle.

One possible strategy to tackle this challenge is to resort to a "bioinspired" optimization procedure. Genetic algorithms (GA)[254] belong to the larger class of evolutionary algorithms that provide a heuristic search inspired by natural evolution. One of the main advantages of using genetic algorithms is their supreme ability to find near-optimal solutions in very large property spaces.[254]

In a GA, a set (population P) of N trial solutions (phenotypes, individuals I) to an optimization problem is iteratively improved toward better solutions by applying operations that mimic natural evolution strategies such as selection, inheritance, crossover, and mutation. The population of each new iteration is called a generation. In the following, we will use the sequence optimization of an 8-amino acid peptide for the property A as an illustrative example.

To apply a GA, the optimization problem has to be encoded in a genetic representation (genotype, chromosome C), usually in binary or integer format, that consists of an ordered list or sequence of elements (n-tuple), in which every position represents an individual gene that encodes for a given feature. For the concrete case of the octapeptide, every solution C:

$$C = (f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8) \quad (94)$$

where the indices 1−8 represent amino acid position 1−8 and $f_i = \{1...20\}$ are integers encoding the different possible amino acids (e.g., $f_i = 1 \Rightarrow$ Gly, etc.). An objective function $\hat{A}$ of the optimization then has to be defined. The value of the objective function for each solution C is called the fitness F, which measures how close the specific property of solution C approaches the target property A. A typical workflow of GA optimization (Figure 7) consists of the following:

(1) Generation of an initial population P_0 with N random solutions C_I (the population size N is problem specific but is typical of the order of 100−1000).

(2) Evaluation of the fitness F_{C_I} for each individual solution I represented by the chromosome C_I:

$$\hat{A}(C_I) = F_{C_I} \quad (95)$$

(3) Selection of individual solutions C_I based on their fitness F_{C_I} into a mating pool M, by applying a selection operator $\hat{S}$:

$$\hat{S}(P_n) = M_n \subset P_n \quad (96)$$

Different algorithms can be used for the selection operator $\hat{S}$: truncation selection, tournament selection with and without replacement, roulette wheel selection, stochastic universal selection, and/or elitist selection, in which a certain percentage of the best organisms of the current population are directly transferred to the new generation, which guarantees that the GA never leads to a "worse" (i.e., less fit) overall solution.

(4) Creation ("breeding") of a new population P_{n+1} from individuals of the parent solutions in M by applying genetic

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

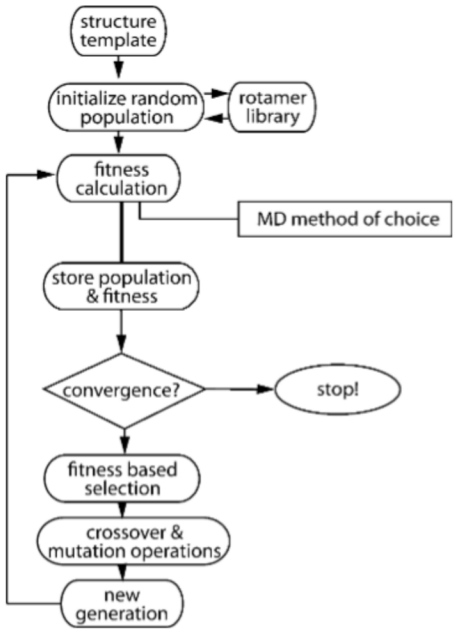


Figure 7. General scheme of a genetic algorithm to optimize properties of proteins and peptidic scaffolds. The workflow starts with a randomized population (structure). The fitness of the minimized protein-based structure is computed and convergence is tested. If the convergence threshold is met, the program stops. If the convergence criterion is not met, an individual is selected out of a pool of individuals with a probability proportional to the computed fitness. Crossover and mutation operations are performed on the individual with given probabilities, and a new generation is formed to repeat the process in an iterative manner.

operations of crossover (recombination) $\hat{R}$, in which part of the genetic information on one parent solution is combined with a part from another parent, for example, for a specific one point crossover in the illustrative example:

$$C_{M_K} = (8, 3, 20, 11, 2, 17, 18, 14)$$
$$C_{M_L} = (5, 10, 1, 15, 16, 4, 19, 6)$$
$$\hat{R}(C_{M_K}, C_{M_L}) = (C_{n+1_K}, C_{n+1_L})$$
$$C_{n+1_K} = (8, 3, 20, 15, 16, 4, 19, 6)$$
$$C_{n+1_L} = (5, 10, 1, 11, 2, 17, 18, 14) \quad (97)$$

Crossover operations are performed with a given probability p_R and with different possible algorithms (e.g., 1 or 2 point crossover, uniform crossover, simulated binary crossover). In addition to crossover, mutation operations $\hat{T}$ can be performed in which one or more genes are randomly changed, for instance:

$$\hat{T}(C_{n+1_K}) = (8, 3, 20, 17, 16, 4, 19, 6) \quad (98)$$

with the mutation probability p_T and different mutation procedures (selective mutation, gene-wise mutation, polynomial mutation). In this way, a new (child) generation is created with the same population size (N), and the procedure is repeated starting from step 2 until a given number of generations or a predefined convergence criterion is reached (Figure 7).

Genetic algorithms are an innovative way of performing rational design. However, the fitness function has to be chosen appropriately and has to be quickly computable, as the GA must be iterated many times to produce a practical result. Balancing the speed of execution with the objective of the designer's goals can be a nontrivial challenge, but, once accomplished, provides a powerful stochastic search and optimization tool that works for a wide array of problems, which are difficult to solve by conventional approaches.

4. APPLICATIONS OF QM/MM MD

In this section, we attempt to illustrate the wide scope of possible QM/MM applications for ground- and excited-state problems with representative examples. Each case was chosen to exemplify a specific ability or special niche of QM/MM simulations. Some of the unique abilities of QM/MM simulations that we focus on include the capacity of this methodology to study enzymatic reactions and to capture systems containing transition metal ions or photoactive biological chromophores.

4.1. Electronic Ground State

4.1.1. Enzymes in Action. Studying enzyme catalysis has been one of the original motivations for the development of QM/MM methodologies, and this topic has remained the most intensely studied subject related to QM/MM simulations. Here, we present some examples of how QM/MM simulations were used for the characterization of the mechanisms of action of metalloenzymes, the re-engineering of enzymes, and the design of artificial biomimetic enzymes (see refs 226 and 255 for more studies of enzymes with the same QM/MM methodology).

4.1.1.1. Human Carbonic Anhydrase. Human Carbonic Anhydrases (CA, EC 4.2.1.1) are a family of enzymes that catalyze the rapid interconversion of CO_2 and bicarbonate (Figure 8a). This enzymatic hydration reaction is reversible and known to be one of the fastest enzymatic reactions, with reaction rates approaching the diffusion-controlled limit for the dehydration reaction (10^{10} s^{-1}),[256] and rates for the hydration reaction ranging between 10^4 and 10^6 s^{-1}.[257] Rapid interconversion of CO_2 and bicarbonate is required by most biological systems, in plants, for photosynthesis, and in mammals, for the selective transport of H^+/HCO_3^- across lipid membranes to maintain acid–base balance in blood.

The putative catalytic mechanism of CA-catalyzed CO_2 hydration involves a rate-limiting transfer of a proton from a water molecule, which is bound to the zinc moiety, through a chain of hydrogen-bonded water molecules, eventually passing the proton off to a histidine residue (His64), which acts as a shuttle to bulk solvent. The zinc-bound OH^- moiety subsequently performs a nucleophilic attack on the CO_2 molecule, ultimately forming a Zn-bound bicarbonate ion. Bicarbonate is displaced by a water molecule, regenerating the resting state of the enzyme. Since its initial discovery, this proton-transfer system has been extensively studied and is now one of the best-understood mechanisms for the transfer of protons in a protein environment.

Considerable attention has been focused on elucidating the catalytic mechanism using both experimental and computational studies including QM/MM simulations.[193,258,259] These simulations have aided the discovery that a "proton hole" pathway,[260] a canonical system used to characterize the "Grotthus mechanism"[261] in biology, may be relevant to the catalytic mechanism of type II carbonic anhydrases (CAII). For

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217–6263

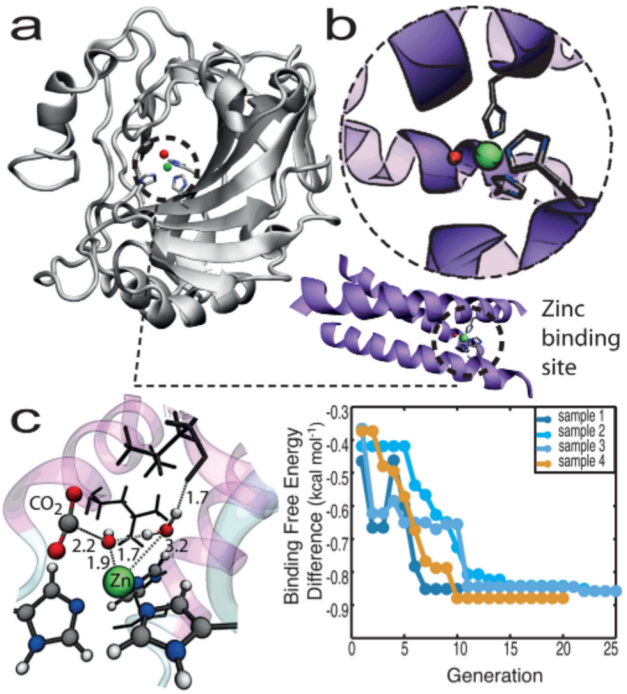


Figure 8. (a) Comparison of the Zn_3NO catalytic sites between HCA (PDB entry 2CBA) and a three stranded coiled coil (3SCC, PDB entry 3PBJ), with reported CO_2 conversion rates at high pH (b). Nucleophilic attack of the zinc-bound hydroxyl on CO_2 (c, left panel). Left panel (c), heuristic-based amino acid sequence optimization of 3SCC. Nine amino acids at the C-terminal end of the protein were mutated to tune the CO_2 binding affinity.

a proton transfer in water, the bond exchange occurs spontaneously, rendering the process diffusion-controlled. Other computational studies have led to several conclusions about the role of His64 as proton shuttle.[259c]

Athough initial deprotonation of the zinc-bound water molecule is considered to be rate-limiting, the H_2O substitution reaction, which regenerates the zinc-bound water molecule in the product release step, is also likely to affect the turnover rate of CA. For example, understanding how the bicarbonate molecule reorients itself for product-release is of great interest as it is likely to influence both the regeneration of the zinc-bound water molecule and the product release step. Despite this fact, these later stages of the reaction mechanism have received less attention and have remained somewhat controversial because two different conversion mechanisms, suggested by Limbscomb and Lindskog, respectively, had originally been suggested. Some QM studies on small mimetic zinc complexes point to the fact that both pathways are feasible.[262] QM/MM simulations of HCA have been performed such that the substrate (CO_2) together with the key catalytic residues (the Zn moiety, the histidine residues bound to the zinc ion, and the Zn-bound water molecule) were described by DFT and the rest of the protein and solvent with a MM force field (GROMOS96).[255d] The QM/MM simulations were able to reconcile the differing viewpoints on the conversion step by demonstrating that the actual mechanism constitutes a mixture of the two pathways.

4.1.1.2. B_{12}-Dependent Mutase. A different type of enzymatic role involves the integration of highly reactive species into a catalytic mechanism. Controlling radical species is an intensely demanding task, requiring clever tactics to find a delicate balance between upholding reactivity for the sake of catalysis and maintaining reaction selectivity by preventing unwanted and destructive side reactions (this has sometimes been termed "negative catalysis").[263] When such a balance is achieved, the enzyme can effectively overcome challenges brought forward by some of the most difficult chemical reactions, such as, in the case of Vitamin B_{12}, carbon skeleton rearrangements.[264] For 5′-deoxyadenosyl-cobalamin (AdoCbl)-dependent isomerases (Figure 9a), homolysis of the carbon−

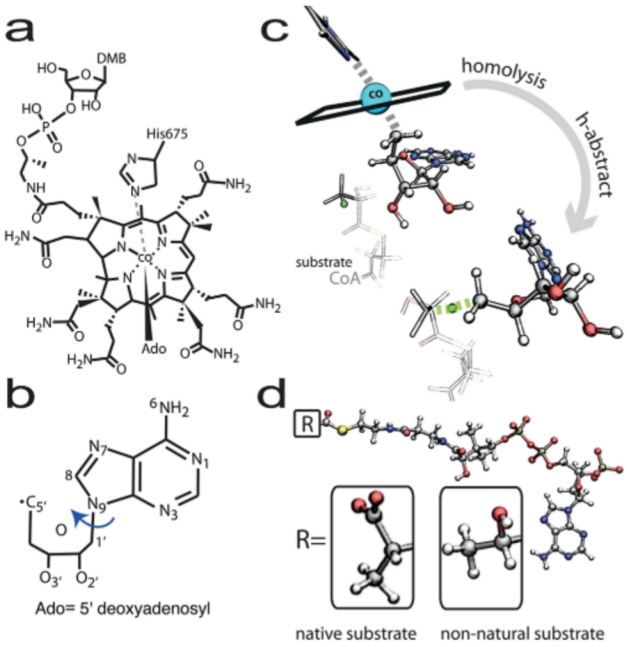


Figure 9. (a) B_{12} cofactor doubly coordinated with a histidine residue (DMB-off conformation) and the 5′-deoxyadenosyl moiety (Ado*), shown in (b). (c) Hemolysis and hydrogen abstraction, in which Ado* diffuses out of its cavity and undergoes a spontaneous conformational change. (d) Ball-and-stick representation of the natural substrate, methylmalonyl-CoA, in comparison with the non-natural substrate, lactoyl-CoA. Lactoyl-CoA is identical to methyl malonyl-CoA aside from a hydroxyl group, which replaces a carboxyl group on the terminal acetyl moiety.

cobalt bond, the so-called "radical reservoir",[264] generates a highly reactive, primary carbon radical species (Ado*, Figure 9b and c). This radical species has the capability of abstracting a hydrogen atom from an unactivated carbon atom of the substrate (Figure 9c) and initiates a rearrangement reaction. To this end, migration of Ado* takes place over several angstroms from its original position, proximal to co(II)balamin.[265]

Many efforts have been made to better understand the elusive nature of the adenosyl radical moiety. From crystallographic structures, it becomes apparent that various conformations of Ado* might be important to catalysis (e.g., a pseudorotation of the glycosyl moiety in diol dehydratase,[265b] changes in puckering of the ribose moiety in GM,[265a] and an adenine-ribose rotation about the glycosidic bond in MCM[266]). QM studies have been conducted to further explore the potential energy landscape of the radical moiety in the gas phase, reporting certain conformational changes of Ado*, such as glycosyl rotation[267] consistent with some crystallographic

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

structures.[265a] However, it is unclear whether such gas-phase observations remain relevant in an enzyme environment.

The considerable protein effect on homolytic cleavage was pointed out by ONIOM (DFT/MM) calculations,[268] and for certain AdoCbl-dependent proteins, large-scale domain motions appear to be coupled to the activation of the AdoCbl bond in certain AdoCbl-dependent enzymes.[269] QM/MM simulations have provided key support for a stepwise mechanism, as well as for the identification of low energy conformation of the adenosyl radical intermediate after Co−C cleavage.[270] Moreover, simulations not only point out how but, more importantly, why the enzyme manipulates the conformation of Ado* during the catalytic cycle as a possible mechanism to ensure selectivity.[271] In fact, the enzyme acts as a mechanochemical switch to manage conformational changes in Ado* by imposing or releasing strain at distinct stages of the catalytic cycle to effectively channel the reaction in a desired direction. Interestingly, this switch can be activated or deactivated through intramolecular interactions between the adenine and ribose moieties, using changes in geometry as a transformational trigger, which suggests that such a mechanism may be inherent in other AdoCbl-dependent enzymes.[271]

4.1.1.3. DNA Repair Enzyme MutY. Cells have evolved elaborate repair systems to defend themselves against unwarranted and detrimental DNA mutations. To counter the constant threats by environment-induced stress, Base Excision Repair (BER) proteins act as one of the central machineries that maintain the integrity of cellular DNA. The most common form of oxidative damage repaired by BER proteins is the 8-oxo-guanine (OG) lesion, resulting from the attack of highly reactive hydroxyl radicals on the C8 atom. Because of this DNA damage, OG preferentially binds to adenine (OG:A),[272] resulting in a high rate of G−C to T−A transversion mutations during DNA replication.[273] Such lesions and the resulting mispairs are managed by the "GO" system, which involves the MutM, MutT, and MutY enzymes.[272a,274] Among this suite of interacting proteins, MutY plays a key role in the recognition of OG:A, the excision of the mispaired adenine base, and the protection of the "empty" apurinic site after excision has taken place[275] (Figure 10a).

The mechanism adopted by MutY to excise the mispaired adenine base has been extensively studied by means of crystallographic structures and transition state (TS) analysis using kinetic isotope effects (KIEs). The crystallographic structure of *Bacillus stearothermophilus* MutY[276] illustrates the way in which MutY flips the adenine moiety outside of the double helix into an extrahelical cavity of the protein. Once inside this cavity, the glycosidic bond of adenine is cleaved, resulting in the formation of an oxacarbenium ion intermediate, which can then be hydrolyzed to generate the empty apurinic site. From these structural insights, a putative S_N1 mechanism was proposed in which a nearby glutamate residue (E43) initiates cleavage by transferring its proton to the adenine base at N7 through a nearby water molecule. The capture step, catalyzed by a second water molecule, hydrolyzes the oxacarbenium ion intermediate.

Many *N*-glycosylases utilize water-assisted catalytic mechanisms to direct the course of excision of mispaired DNA bases. In fact, it has been demonstrated that the involvement of water molecules during the cleavage reaction potentially enables MutY to lower the energetic barrier of a proton transfer, which contributes a catalytic rate enhancement equivalent to nearly one-half of the required activation energy barrier.[277] This

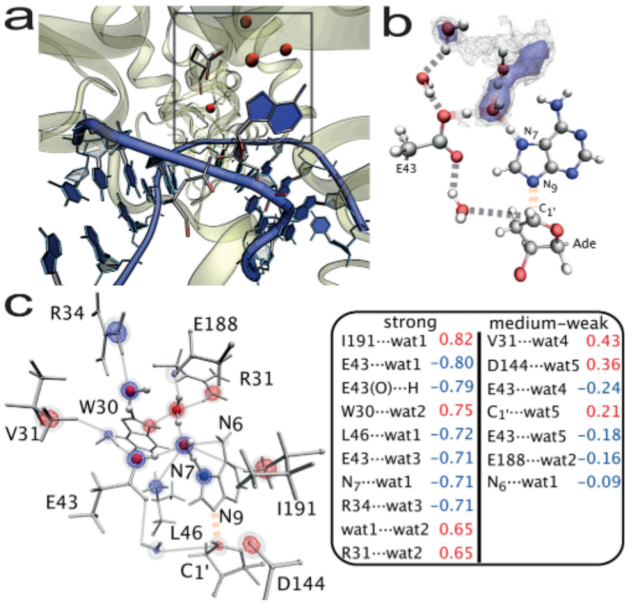


Figure 10. (a) A mispaired adenine is extruded into an extrahelical cavity. As shown in the inset of (a), a concerted proton transfer occurs from E43 to N7, mediated by a water molecule in the vicinity of the catalytic pocket, during cleavage. (b) Solvent distribution (maximally occupied positions of water molecules) within 5 Å of adenine in the active site. The distribution is represented by a contour plot, depicting the probability to find water molecules at a given position during the classical molecular dynamics simulations. The three maxima, illustrated in dark blue (contour level of 0.5), correspond to positions of three out of five structured water molecules in the MutY crystal structure (PDB entry 1RRQ). (c) Graphical representation of a Pearson-like correlation coefficient used to measure the strength of the coupling between the change in reaction coordinate and the interaction distances of the five different water molecules with various residues lining the active site. A contour plot was generated to visualize the strongest correlating residues by placing spherical Gaussians on the respective interacting atom center. Stronger correlation is represented by larger densities in the contour plot, depicting the interactions that are the most cooperative to the increase in reaction coordinate. Surprisingly, the residue−solvent interactions that correlate most strongly to reaction coordinate correspond to hydrophobic residues, positioned close to the site of N7 protonation that interact with wat1 during proton transfer.

astonishing observation naturally leads to the question, "What is the exact involvement of water during enzymatic catalysis and how is it secured?" QM/MM simulations that were used to probe environmental factors that may influence catalysis demonstrate a unique ability of this enzyme to organize a subset of water molecules into a special configuration to protonate the adenine base, which consequently turns it into a better leaving group.[191a] A long-lived solvation pattern is established (Figure 10b), which facilitates the involvement of three structured water molecules in the active site. How the enzyme secures this configuration is interesting because both hydrophilic and hydrophobic interactions play a role in convening, ordering, and constraining specified orientations of water molecules during the key points in the reaction[191a] (see also section 5.1).

4.1.1.4. Enzyme Re-engineering. A great deal of progress in understanding how enzymes function and what molecular strategies they use during catalysis is due, in part, to the ever increasing number of high-resolution structures from X-ray

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

crystallography and NMR as well as to detailed mechanistic information from kinetics, modeling, and in particular QM/MM simulations. Although valuable in its own right, an additional advantage of knowing the structure and function of enzymes is transfer of this knowledge to the design of synthetic catalysts that are capable of performing novel chemical tasks. Understanding the unique attributes of native enzymes reveals the key chemical principles that underlie their efficacy as catalysts. These "bioinspired" catalysts have many advantages, which include the ability to be tuned in terms of selectivity or specificity to act on an expanded or reduced scope of substrates.

Developments in metabolic engineering and synthetic biology bring promise of engineering complex reaction pathways within a single microorganism.[278] Coupling the engineering of new biosynthetic routes with the discovery and redesign of enzymes as biocatalysts would deliver an innovative strategy for exploiting microorganisms as renewable resources for the production of specialty and industrial chemicals. Using microorganisms to produce desired compounds requires discovering and re-engineering biosynthetic routes, because the biosynthesis of many compounds has not been reported. For example, some production routes may call for known reactions, in which an existing enzyme catalyzes its customary reaction to transform a natural substrate into a desired product or intermediate. On the other hand, they may require novel reactions, in which an existing enzyme catalyzes the transformation of a non-natural substrate into a desired product or intermediate.

The knowledge gained performing QM/MM molecular dynamics simulations brings insights into how a native chemical process can be tailored into the design of a bioinspired system with novel functionalites. One application of this concept involves the redesign of methyl-malonyl-CoA mutase (MCM) in a biosynthetic pathway for the production of desired commodity chemicals from metabolic precursors.[279] Unlike de novo protein design,[196b,280] an existing enzyme mechanism was re-engineered to catalyze a non-natural reaction for the bioysynthetic production of 3-hydroxypropanoate (3HP), a chemical whose derivatives have extensive applications in the area of polymer plastics.[281,282] Engineering catalytic promiscuity in MCM enables the transformation of lactoyl-CoA into 3HP-CoA, a precursor in the biosynthesis of the final product (Figure 9d).

Classical and QM/MM molecular dynamics simulations were used to assess feasibility of the non-natural reaction in terms of whether MCM can efficiently bind and catalyze the desired transformation of a non-natural substrate. The information gained from this study was then used to redesign and optimize the novel biocatalyst based on binding affinity and catalytic efficacy via site-directed mutagenesis.[279]

4.1.1.5. Biomimetic Systems with Tailored Functionalities. Using biomimicry for the production of desired chemicals takes inspiration from nature's unique ability to create exceptional catalysts. Exploring the natural machinery of enzymes lends insight into understanding how novel biocatalysts can be engineered to operate under mild conditions and, at the same time, be environmentally benign, highly selective, and efficient. By understanding strategies of biological processes, which have been optimized over millions of years of evolution, many of their sustainable qualities can be exploited for practical applications that are desired by humankind. As shown in the previous section, such a biomimetic strategy can be used for an in-cell production of chemical compounds. However, in some cases, in-cell production of desired products may be economically infeasible due to high concentrations of the target compound, which could either be toxic to cells or cannot be produced by single cells with acceptable yields. Alternatively, the same philosophy can be applied for out-of-cell catalysis based on a biomimetic design of protein-like scaffolds with both steric and electronic features that reinforce the proper orientation of catalytic residues.

4.1.1.5.1. Designer Peptidic Scaffolds. Determining the structure and aggregation characteristics of helical peptidic bundles has enabled the design of metal-binding peptidic scaffolds.[283] A coiled coil structure consists of two or more amphipathic helices wrapped around one another and embodies many of the characteristics of native protein structures. For example, these helical bundles are stabilized by an interplay of hydrophobic and hydrophilic residues, in which specific positions are occupied by hydrophobic residues at the interhelical interface.[284] One advantage of the de novo design of peptide scaffolds is that their structures are far less complex when compared to entire proteins where site-directed mutagenesis often leads to nonlinear changes in the complex interaction networks which are impossible to deconvolute.

A catalytically active synthetic protein scaffold[285] has recently been reported to reproduce the catalytic reaction of Human Carbonic Anhydrase (HCA; see section 4.1.1.1 and Figure 8b). This peptidic scaffold belongs to the TRI family of peptides [Ac-G(LKALEEK)4G-CONH2][286] and bears strong resemblance to HCA in that a pseudotetrahedral Zn(II) is coordinated by three imidazoles and one water molecule. The tris(histidine) coordinated Zn−OH^- site reproduces the metal binding site in HCA (Figure 8b and c), while an additional tris(thiol) coordinated Hg(II) confers a significant amount of extra stability. This artificial protein[285] successfully imitates several of the important characteristics of CA; the CO_2 hydration rate, however, is several orders of magnitude lower than its wild-type counterpart.

The knowledge gained by studying the catalytic mechanism of CA was used to redesign the three-stranded coiled coil mimic (Figure 8b).[255d] Classical and QM/MM simulations were performed to understand the solvation characteristics, long-term stability, CO_2 binding affinity, and catalytic efficacy for the hydration reaction in the mimic. Two routes were then chosen to optimize the mimic: (i) increasing CO_2 binding affinity and (ii) tuning the pK_a of the Zn−OH_2 site. While the first route increases the ability of the mimic to bind CO_2 in the active site, the latter optimizes the rate-limiting step of catalysis, the initial deprotonation of the zinc-bound water molecule.

To this end, a genetic algorithm (GA) was employed to search sequence space. Taking this approach to optimize the peptidic scaffold, we selected nine amino acids in the vicinity of the CO_2 and zinc binding sites to be screened by the GA. The objective function to evaluate the relative fitness was chosen to be the binding free energy of CO_2 computed via Molecular Mechanics Poisson−Boltzmann Surface Area (MMPBSA)[287] calculations as implemented in the AMBER suite of programs and the pK_a of the zinc bound water estimated using the Adaptive Poisson−Boltzmann Solver (APBS).[288] The resulting mutants offered an increased binding affinity to CO_2 by 1 kcal mol^{-1} (Figure 8d), and the pK_a for the zinc-bound water molecule was down-shifted by 2 pK_a units with respect to the original system (from a pK_a of 10.4 to 8). This shift indicates that the zinc-bound water molecule is much more likely to be deprotonated at a lower pH, hence increasing not only the

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

efficacy of the mimic but also making it more environmentally benign, because the reaction can be performed around neutral pH.

4.1.1.5.2. Biomimetic Solar Cells. Dye-sensitized solar cells (DSSCs) have gained widespread attention in recent years because of their low production costs, ease of fabrication, and tunable optical properties such as color and transparency. Numerous attempts have been made to optimize these devices toward their theoretical maximal performance. For solar cells with traditional ruthenium-based sensitizers, the conversion efficiency is hampered by the low molar extinction coefficients of the dyes (e.g., $\epsilon < 10\,000\ M^{-1}\ cm^{-1}$ for the metal to ligand charge transfer (MLCT) band of the black dye[289]). Furthermore, the limited availability of precious ruthenium metal impedes its practical applications in the solar industry field.

In nature, porphyrin-based chromophores have evolved as highly efficient systems to capture solar light and convert it into chemical energy. Given their primary role in photosynthesis, the use of porphyrins to harvest solar energy and thus the use of porphyrin molecules as dye-sensitizer in DSSCs emerge naturally.[290] Following such a biomimetic strategy, it was possible to computationally design new dye-sensitizers that lead to a DSSC with the currently record efficiency of 13% (with an open-circuit voltage 0.91 V, a short-circuit current density of 18.1 mA cm^{-2}, and a fill factor of 0.78).[291]

4.1.2. Transition Metal Ion Interactions with DNA, Proteins, and Nucleosome. The interaction of biological systems with transition metals whose d orbital-induced chemical interactions can be highly anisotropic is often difficult to describe at the level of classical force fields, and is thus another domain in which QM/MM simulations have found widespread applications.[292]

4.1.2.1. Ru- and Pt-Based Anticancer Drugs. The transition metal compound cisplatin $[Pt(Cl)_2(NH_3)_2]$ is currently the most widely used chemotherapeutic agent.[293] Its primary mode of action is the formation of DNA lesions that interfere with transcription and result in cellular apoptosis. Despite this wide use, the efficacy of Pt-based cancer therapies is limited due to acquired drug resistance[294] caused by geometric constraints of DNA−drug interactions[295] or the overexpression of certain proteins in detoxification pathways. In addition, traditional platinum agents show severe toxicity and limited selectivity. This prompted the further development of new chemotherapeutic strategies, based, for example, on alternative transition metals such as ruthenium.[296] Ru-based compounds show great promise of being effective anticancer agents due to their selective activity against specific types of cancer cells in conjunction with low toxicity.[297] From a drug resistance perspective, the octahedral bonding of Ru(II) and Ru(III) species provides a higher degree of conformational versatility, as compared to the square planar geometry of the platinum species, which enables the tailoring of site selectivity, size discrimination, and the use of favorable pharmacological attributes.

In contrast to platinum-based drugs, very little is known about the molecular mechanism of drug interactions in the case of ruthenium compounds. In particular, open issues concern an identification of the primary targets (DNA versus protein), the specific binding sites, and the molecular response of the target upon drug binding. To address these questions, QM/MM simulations are among the only computational tools with the capacity to characterize these types of transition metal interactions with biological matter. On one hand, electronic structure methods are required to describe the binding properties of the transition metals, given that Ru- and Pt-based compounds bind covalently to the target receptor site. On the other hand, it is essential to study the compound in the presence of a protein or DNA environment to take electrostatic and steric effects into account that directly influence drug binding. In a number of publications, the interactions of metal-based anticancer drugs with (i) DNA as a possible first target,[251b,298] (ii) proteins involved in detoxification pathways,[299] and (iii) specific histone sites on the nucleosome core particle[298a] have been studied with the help of QM/MM simulations.

The binding processes of both monofunctional and chelating Ru-based compounds to double-stranded DNA and the resulting structural perturbations were characterized by means of DFT/MM and classical molecular dynamics simulations. The high flexibility of DNA allows for fast accommodation of ruthenium complexes into the major groove, with the ruthenium center binding to the N7 atom of guanine bases. Once bound to the host, different types of Ru-based complexes induce DNA structural distortions in distinct ways, either by releasing complexation-induced strain in the DNA backbone by local disruption of a Watson−Crick base-pair or by a bending of the DNA helix toward its major groove, resembling the characteristic DNA distortion induced by the classic anticancer drug cisplatin. In both cases, force-matched force fields have been used to characterize the long-time scale DNA distortions.[251b]

Although it is traditionally assumed that DNA is the most likely pharmacological target, recent evidence suggests that, for certain compounds, proteins are the main players involved in the therapeutic effect.[300] In fact, nonclassical drugs that focus on specific cellular pathways and interact with targets other than DNA display increased selectivity.[301] The recognition mechanism between two ethacrynic acid (EA) derivatives based on Ru(II) and Pt(IV) called ethaRAPTA and ethacraplatin, respectively, with glutathione *S*-transferase (GST P1-1), an enzyme responsible for the detoxification of cisplatin and related metallodrugs, has been studied with QM/MM simulations. Performing metadynamics within the framework of classical MD simulations, the free energy profiles for the binding of anticancer drugs to the active site of the GST were estimated (Figure 11a). The QM/MM simulations indicate that particular residues, such as Tyr108, play an important role accommodating the drug into the binding cavity.[299]

Very recently, it has been shown that ruthenium anticancer drugs bind directly to the nucleosome core particle (Figure 11b).[298a] Furthermore, the monofunctional compound $[(\eta^6\text{-}p\text{-cymene})Ru(ethylenediamine)Cl]PF_6$ (RAED-C)[302] was found to preferentially bind to the DNA component of chromatin in cells, whereas the $[(\eta^6\text{-}p\text{-cymene})Ru(1,3,5\text{-triaza-}7\text{-phosphaadamantane})Cl_2]$ (RAPTA-C)[303] compound associates with the protein (histone) component (Figure 11c). The latter is particularly interesting, because histone-binding could directly influence gene expression and open new avenues for possible epigenetic cancer therapies. QM/MM simulations with the explicitly solvated NCP were able to rationalize the surprising binding selectivity of these two structurally related Ru-compounds by demonstrating that RAED-C-DNA adducts are thermodynamically preferred and that, as a consequence of steric clash, the activation barrier for RAPTA-C binding to the DNA is considerably higher.

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

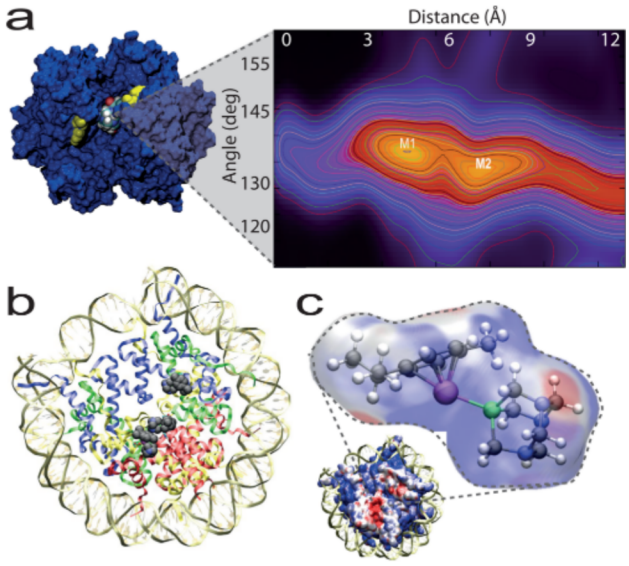


Figure 11. Interactions with Ru-based anticancer drugs. (a) Classical metadynamics free energy profile of the binding of anticancer drugs to the active site of the glutathione *S*-transferase (GST). (b) Crystal structure of nucleosome core particle with bound Ru anticancer drugs. Certain compounds associate with the protein (histone) component, as shown in (b). (c) Zoom of the Ru-based compound, [(η^6-*p*-cymene)Ru(1,3,5-triaza-7-phosphaadamantane)Cl_2] (RAPTA-C).

Such atomistic insights gained from QM/MM studies might serve as a guide for further experimental investigations and assist the rational development of more effective and targeted anticancer drugs.

4.2. Electronically Excited States

An increasing number of QM/MM applications have been devoted to the investigation of electronically excited states in complex environments. The most intensely explored subject is photoactive proteins such as photoactive yellow protein,[51a] green fluorescent protein,[69,304] (flavin) photoreceptors,[136,305] bacterial photosynthetic reaction center,[184,306] photosystem II,[307] bacteriorhodopsin,[308] and rhodopsin (for a recent review of QM/MM simulations of photoactive proteins, see refs 45b, 190, and 309). Rhodopsin, the visual photoreceptor, is the textbook example of a light-sensitive protein and has been explored particularly extensively with a full spectrum of quantum chemical approaches from full nuclear quantum dynamics on simplified models[9a,310] to a large variety of QM/MM studies. The latter are briefly discussed in the following section.

4.2.1. Photoactive Proteins: Rhodopsin. The ability of the eye to detect light is a complex biological process involving different cells in the retina. The rod cells, responsible for peripheral and night vision, accomplish this function using an integral membrane protein, rhodopsin, as a fundamental detector. Light-activated rhodopsin undergoes a structural change and binds to the intracellular G-protein, provoking a signaling transmission cascade that leads to a neuronal signal and, thus, to vision.[311] The structure of rhodopsin is composed of a seven transmembrane helical bundle surrounding the covalently bound chromophore, the protonated Schiff base of retinal (RPSB). Light absorption by the pigment triggers the isomerization of the 11-*cis* to the *all-trans* form of the RPSB. This reaction occurs with high efficiency (quantum yield 0.67),[312] and the primary photoproduct, photorhodopsin, is formed within a very short time (200 fs).[313] Subsequently, a number of different intermediates can be detected spectroscopically; the first one, photo rodopsin, thermally relaxes within a few picoseconds to a distorted *all-trans* configuration, batho rhodopsin.[314] On a nanosecond time scale, other photoproducts are formed, batho rhodopsin forms an equilibrium with a blue-shifted intermediate (BSI), before the mixture decays to form lumi rhodopsin. Lumi rhodopsin is then transformed into meta rhodopsin I (meta I) on a microsecond time scale and finally into meta rhodopsin II (meta II), the active conformation for G-protein coupling[315] (Figure 12a).

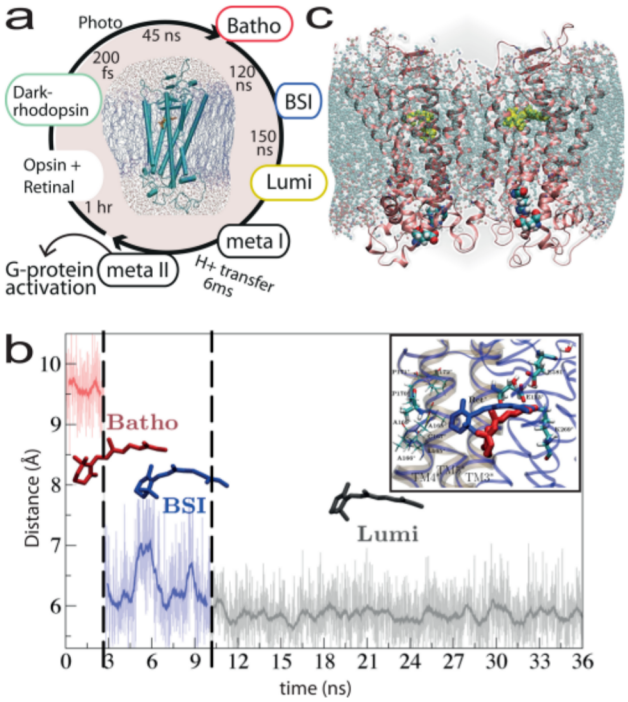


Figure 12. (a) Intermediates in the rhodopsin photocycle. (b) The time evolution of the distance between the ionone ring of the chromophore and Ala169 after photoinduced isomerization. (c) Rhodopsin dimer model in an explicit membrane.

The early intermediates in the photocycle can be distinguished by their UV/vis absorption maxima; batho (529 nm, 2.34 eV) displays a red-shifted maximum with respect to the one of dark state rhodopsin (498 nm, 2.49 eV), BSI has a maximum at 477 nm (2.60 eV), lumi rhodopsin absorbs at a wavelength (490 nm, 2.53 eV) similar to that of the dark state, while at low temperature meta I absorbs at 478 nm (2.59 eV) and meta II at 380 nm (3.26 eV), respectively. A three-dimensional structure of bovine rhodopsin in its resting dark state has been solved by X-ray crystallography in 2000[316] as the first member of the large family of G protein-coupled receptors (GPCRs). GPCRs transduce extracellular signals into intracellular stimuli and are involved in all types of stimulus-response pathways from intercellular communication to physiological senses. Later, the structures of some of the intermediates (batho and lumi) in the rhodopsin photocycle have also been determined, and, very recently, the structure of the activated meta II form has been resolved.[317,318] In addition, time-resolved Fourier transformed infrared,[319] circular dichroism,[320] nuclear magnetic resonance,[321] and resonance Raman

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

spectroscopy[322] have been utilized to characterize some of the conformational changes that the chromophore and the protein environment undergo during the early steps of signal transduction.

QM/MM studies have been applied extensively to clarify different aspects of the rhodopsin photocycle. DFTB/MM calculations have been applied to refine the crystal structure of dark state[316b] and batho rhodopsin,[323] and DFT/MM simulations in combination with long time-scale force field-based molecular dynamics simulations were able to predict the structures of the early intermediates of the photo cycle (batho, BSI, and lumi) with high accuracy before any crystal structure was available[93e,324] (Figure 12b). DFT/MM simulations have also been used in an attempt to help elucidate the long-debated protonation state of Glu181, an active site residue close to the isomerizing double bond, through calculation of the IR, Raman, UV−vis properties, and the NMR chemical shifts.[325] In addition, data from extended DFT/MM simulations were employed to generate a highly accurate force field for the RPBS, which was able to capture subtle structural changes of the chromophore that have a pronounced influence on the optical properties.[252] By using combinations of classical and QM/MM simulations, it was also possible to investigate the proton transfer between the RPBS and its counterion that occurs in the late stages of the photocycle.

The structure and optical properties of the rhodopsin chromophore in different environments (gas phase, methanol solution, protein environment) have also been investigated thoroughly with semiempirical,[190,326] DFT/TDDFT,[93d] CC2,[106b,327] multireference,[328] Green's function-based many-body perturbation theory (GW, BSE),[50b] and Quantum Monte Carlo[93d] methods. In fact, the absorption spectrum of rhodopsin in the dark state and the early intermediates have served as benchmarks for the performance of different electronic structure methods[93d,329] and have illustrated one of the most prominent failures of LR-TDDFT when using purely local approximations to the exchange-correlation functional and kernel. Recent systematic studies have also assessed the influence of different QM/MM trajectories, QM/MM coupling schemes, and the effect of finite temperature sampling.[93b,d,e]

Numerous studies have been devoted to the subject of spectral tuning among the different photointermediates of the rhodopsin photocycle and the color tuning in cone rhodopsin and retinal binding proteins.[325b,330−332] Various properties such as the changes in the bond length alternation (BLA) of single and double bonds along the conjugated chain of the chromophore, the distance to the negatively charged counterion Glu113, structural distortions of the chromophore, or polarization effects of the environment have been proposed as decisive factors for spectral tuning, but no consensus has been reached. The problem is complicated by the fact that the observed shifts are small (typically smaller than the thermal width of the absorption band) and that various protein components can give small but nonnegligible contributions. Recently, a statistically relevant number of thermally sampled configurations extracted from QM/MM trajectories was used to establish a one-to-one correspondence between the structures of the different early intermediates (dark, batho, BSI, lumi) involved in the initial steps of the rhodopsin photoactivation mechanism and their optical spectra. By applying machine-learning techniques (further discussed in section 5.2) to analyze the complex structure−function relationship, it was possible to identify five factors that determine the observed spectral shifts, the most prominent one being the BLA.[93e]

Both static and dynamic investigations have been performed to characterize the cis−trans photoisomerization reaction. Conte et al.[333] used Green's Function Perturbation Theory to calculate the excitation energy along a cis−trans isomerization path for a minimal gas-phase model for which Szymczak and co-workers performed CASSCF/surface hopping.[334] QM/MM simulations of the excited-state reaction of the full systems were performed at various levels of theory including ROKS,[84] SA-CASSCF[335] in combination with surface hopping,[336] and multiple spawning based on CASSCF/FOMO-CI.[52] All simulations agree with the ultrafast (essentially barrierless) isomerization dynamics observed experimentally, in which the system in the first excited singlet state relaxes through a conical intersection to the ground state. However, it has recently been suggested that the lowest lying triplet state might also be involved.[337]

Force-matched force field-based MD can be used to simulate the further relaxation of the photoisomerized chromophore in the ground state up to the microsecond time scale and provides insights into the influence of the protein environment and the effects of receptor dimerization[324c] (Figure 12c).

4.2.2. Photodamage in Proteins. From an evolutionary point of view, the integrity of nucleobases is essential to the faithful transcription and replication of the genetic code. Thus, mechanisms have evolved that either prevent photodamage of DNA[338] or repair photolesions[339] in an efficient manner. For proteins, however, the presence of aromatic side chains (e.g., tryptophan) makes them also susceptible to UV light-damage. Less attention has been given to the mechanisms that nature has put in place to prevent proteins from light-induced harm. Theoretical studies, conducted in gas phase, were devoted to the study of the photoexcitation of single neutral or zwitterionic amino acid water clusters to the first optically allowed state. Their findings suggest that UV absorption can lead to NH_3 detachment, in agreement with high quantum yields for ammonia in UV photolysis.[340] However, recent cold-ion spectroscopy experiments[96] have demonstrated that the excited-state properties of small amino acid complexes are potentially influenced by the environment. In fact, the addition of as few as two water molecules can drastically alter the excited-state properties, and it is thus not clear to what extent the findings on small gas-phase molecules are directly transferable to more realistic systems.

To further elucidate the role that the environment plays in the photoexcitation properties of tryptophan residues, we have recently performed a comparative study of the excited-state properties of tryptophan in both small gas-phase complexes involving bare and microsolvated protonated tryptophan $[H^+$-Trp$]$-$(H_2O)_n$ (n = 0, 1, or 2) and the dipeptide H^+Gly-Trp as well as in a realistic protein environment. For the latter, QM/MM excited-state simulations were performed for Human Serum Albumin (HSA), which contains a single tryptophan (W214) located in the vicinity of a positively charged lysine (K199) (Figure 13). W214 is sequestered inside a deep pocket, which contains on average 8 water molecules within hydrogen-bond distance.[341]

The findings for the gas-phase systems indicate that the vicinity of a bare positively charged group to the indole ring of tryptophan can render the system potentially susceptible to photodamage. Both theory and experiment[96] demonstrate that rapid dissociation occurs for both bare H^+-Trp and H^+Gly-Trp.

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

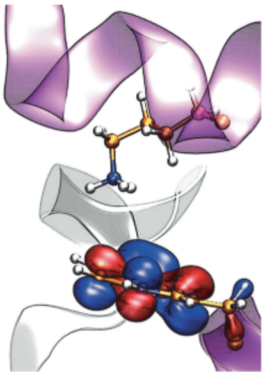


Figure 13. Properties of the photoexcited desolvated Lys199$^+$-Trp214 motif in the region described at the DFT/TDDFT level of theory.

Instead, in solvated H$^+$-Trp, the presence of two water molecules drastically changes the electronic structure of the low energy excited states; that is, the transitions to σ^*_{N-H} orbitals are shifted up by ∼1.3 eV due to the presence of hydrogen bonds preventing photodissociation.

To study the photodynamics of HSA under fully hydrated conditions, we performed surface hopping LR-TDDFT/MM TDDFT simulations with a simulation time of 250 fs. After a short relaxation (∼60 fs), the system evolves in a long-lived nondissociative state as a direct result of the presence of a strong hydrogen-bond interaction between the protein and a water molecule in the vicinity of the Trp214 cavity. Analogous excited-state simulations performed for HSA after removing three water molecules hydrogen bonded to K199 indicate instead that rapid changes in the electronic state induce major structural alterations. After a relatively short relaxation phase (∼20 fs), the σ^*_{N-H} and π^* contributions decrease progressively, while the σ^*_{C-N} antibonding character increases. The initial stretching of one N−H bond (t < 20 fs) is followed by the elongation of the C−N bond leading to the rapid dissociation of NH_3 from the lysine residue, in analogy to the findings for unsolvated protonated tryptophan complexes in the gas phase.[95a]

These studies suggest that UV-sensitive tryptophan residues in the vicinity of a positive charge (such as Lys or Arg residues) can act as potential sources of photoinduced damage in proteins, in particular for buried residues. In the solvated case, the only state accessible under UV irradiation is a non-dissociative $\pi\pi^*$ state from where radiationless de-excitation might occur via a conical intersection to the ground state. These results suggest that, indeed, water is the "solvent of life" that protects proteins and other biological components from immediate photodamage.

4.2.3. Biological Applications of P-TDDFT. Recently, first applications of real-time propagation TDDFT/MM to biological problems have emerged to describe long-distance electron transfer and radiation damage in biological matter.

4.2.3.1. Electron Transfer in Azurin. Azurin is a relatively small (128 amino acids) copper-binding protein (Figure 14a) that can transport electrons between itself and other proteins (e.g., nitrite reductase and cytochrome C-551 oxidase). The electron transfer (ET) mechanism is extremely efficient across the protein for distances up to 18 Å. The remarkable properties of the blue copper site in azurin and related proteins have been both experimentally and theoretically studied for nearly five decades. Additionally, accurate crystal structures have been resolved for azurin, both in oxidized and in reduced forms. For these reasons, azurin is an ideal test case for the evaluation of new simulation methodologies.

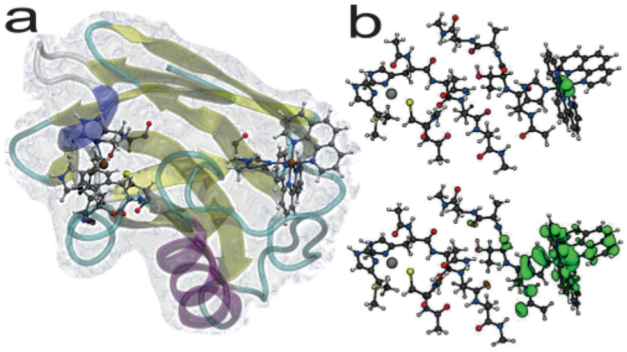


Figure 14. Zoom of the QM part of the molecular setup, consisting of the azurin active site with the addition of a ruthenium complex. The spin density contour in (b) depicts the location of the initial hole, generated by oxidation of the ruthenium ion. After a few femtoseconds (bottom of (b)), the hole is displaced toward the copper center (silver) of azurin.

The optical[342] and redox[343] properties of azurin have also been characterized with DFT/MM simulations. P-TDDFT/MM simulations have been performed of the electron transfer (ET) process occurring between the azurin copper binding site and a ruthenium donor complex attached to a histidine residue located at the surface of the protein (Figure 14a). The size of the quantum part ranges from ca. 250 to 500 atoms spanning the entire protein matrix between the donor and the acceptor centers. The system was first equilibrated with a hole (depletion of one electron) created at the copper site. The simulations were then initiated with the hole moved to the ruthenium complex to create a nonequilibrium situation. The latter simulation favors the ET from the ruthenium to the copper center. Simulating in this way allows accelerating the ET process by several orders of magnitude, because the reorganization of the solvent and of the protein matrix is the rate-limiting step of the entire reaction (Figure 14b).

A similar study of long-range electron transfer using Ehrenfest QM/MM dynamics based on an approximate DFT-based description was recently performed for the DNA repair enzyme photolyase.[117]

4.2.3.2. Radiation Damage and Hadron Therapy. Bombardment with charged ultrafast cations (such as H$^+$ and C^{4+}) (Figure 15) is used in a number of cancer therapies (hadron therapy). As compared to electron radiation, hadron therapy is characterized by a deeper penetration depth of tens of centimeters and a sharp Bragg profile that allows for highly focused deposition. The high energy (∼100 MeV) cations remove electrons from encountered molecules inducing Coulomb explosion. This radiation-induced damage leads to strand breaks in the DNA of bombarded cancer cells that have an impaired ability for repair and undergo apoptosis. Although DNA is the main biological target, a large part of the radiation induced damage also affects water which leads to the production of highly reactive oxygen radical species that attack other biological molecules in secondary reactions causing long-lasting side effects over periods of days and even years. Very little is known about the effects that this type of radiation has on biological matter, and a better atomistic understanding is

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

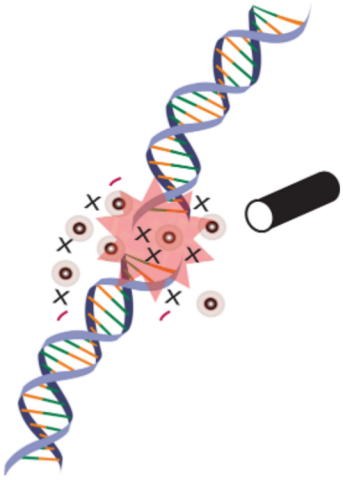


Figure 15. Cartoon depiction of the bombardment of DNA with charged and ultrafast cations (such as H^+, C^{4+}), which is used in hadron therapy.

essential for developing strategies that minimize unwanted damage.

P-TDDFT in combination with Ehrenfest dynamics has been used to perform real-time simulations of the interaction of hadron particles with prototypical biological systems (2-deoxy-D-ribose,[344] doubly charged uracil in the gas phase,[345] and in aqueous solution[118,346]) to characterize radiation products. These simulations show that bombardement with C^{4+} ions leads to ultrafast extraction of electrons from water molecules and DNA bases on the femtosecond time scale. Ionized water molecules decompose by forming $H^•$ atom and $OH^•$ radicals, while in doubly ionized uracil, electron extraction is followed by Coulomb explosion that leads to the ejection of the carbonyl oxygen atom.[118,346]

5. ANALYSIS OF HIGH-DIMENSIONAL DATA SETS

QM/MM simulations provide a wealth of data, rich in information about the structural, dynamical, and electronic properties of the system. In principle, computer experiments enable the probing of the energetic contribution and role of every residue for a given process. However, the information generated from simulations quickly becomes unwieldy, as systems are typically on the order of 10 000−100 000 atoms and sampled trajectories can reach into the microsecond time scales. Moreover, using QM/MM simulations adds electronic degrees of freedom, which increases the complexity even further.

Consequently, it can be a great challenge to derive meaning from such high-dimensional data sets. For example, we are typically interested to ascertain whether specific parts of the protein play a crucial role in the biological mechanism under investigation or whether these parts serve an entirely different purpose. One of the main hurdles to be overcome in this case is the high cost of performing large numbers of QM/MM simulations at the first-principles level, which is often a critical experiment to test the influence of different parts of the system in an unbiased way. Thus, finding ways to efficiently analyze, decipher, and find meaning in the data that are generated from simulations is a very important issue to address.

In this section, we will describe two strategies to find meaning in the highly complex QM/MM simulation data by establishing correlations between structural/electronic features and a given biological process and identifying causal relationships between electronic and geometric changes in protein residues and molecular mechanisms.

Feature selection methods and causality inference[347] techniques and machine learning algorithms[348] are well-developed strategies to reduce the dimensionality of the problem and thus are commonly used in areas such as text mining, bioinformatics, combinatorial chemistry, and multivariate image processing. Chemical and biological applications that heavily depend on such analyses are typically dealing with high-dimensional sequence and microarray data, such as in the case of transcriptomics and proteomics, and other deep sequencing technologies, or are used to facilitate the discovery of more selective drugs from large chemical libraries in virtual screening studies to improve the predictability of QSAR methodologies.[349]

We will discuss the application of these techniques in the context of QM/MM simulations for two case studies: (i) to assess the role that the environment plays during catalysis for the DNA repair enzyme, MutY; and (ii) to identify the molecular determinants responsible for the spectral shifts during the photoactivation of the visual pigment rhodopsin.

5.1. Correlation Analyses

The putative stepwise reaction of MutY obtained from QM/MM simulations (discussed in section 4.1.1.3) is in good agreement with kinetic isotope experiments (KIEs).[277] However, missing from this description are the details of the enzymatic reaction mechanism and, in particular, the respective role that the environment plays in catalysis. KIEs have already provided many clues to suggest that protonation occurs in a pre-equilibrium step before cleavage,[277b,350] which indicates that N7 protonation may be a critical event to enable departure of adenine. Moreover, other interactions between the adenine moiety (Ade) and the environment appear to be important during the cleavage reaction, such as Ade:N6, which undergoes a loss of hydrogen bonding and/or solvation at the transition state,[277b] and Ade:N1 and Ade:N3, which may also facilitate the departure of the adenine leaving group.[277a,351]

Direct involvement of solvent interactions with adenine can be deduced from various crystallographic structures,[276,352] which reveal a conserved structuring of five to six water molecules in the vicinity of the adenine moiety (see Figure 10a and b). One proposed role for water molecules in this region is to form a "water bridge", connecting the nearby catalytic glutamate (E43) and adenine moieties to relay the proton in the initial protonation event at N7.[276,277] Another proposed role for a nearby water molecule is to hydrolyze the oxacarbenium ion intermediate. Therefore, at least one water molecule is directly involved in both cleavage and hydrolysis steps, but whether all five structured water molecules are truly conserved and involved in catalysis is a question that requires analysis beyond a comparison of crystal structures. We were interested in establishing the characteristic functions of the catalytically active water molecules in addition to other environmental factors that influence catalysis and to probe nature's strategy for securing the involvement of solvent molecules in water-assisted chemical reactions.

Classical molecular dynamics simulations show that the structured active site water molecules observed in the crystal structure in fact persist in the fully solvated system assuming specific characteristic positions and orientations[191a] (see Figure

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

10b). This raises the question whether or not the structured water molecules actually contribute to catalysis. Starting from the classically equilibrated structure, the catalytic cycle of MutY was studied using QM/MM Car−Parrinello simulations. As was already discussed in section 4.1.1.2, the putative stepwise S_N1 mechanism of the overall reaction was examined in two steps: (i) cleavage of the glycosidic bond and (ii) hydrolysis of the oxacarbenium intermediate. We have chosen to describe cleavage with a single reaction coordinate, the N9−C1′ bond distance.

A Pearson-like correlation coefficient, r:

$$r = \frac{\sum_i^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_i^n (X_i - \bar{X})^2}\sqrt{\sum_i^n (Y_i - \bar{Y})^2}} \quad (99)$$

where $\bar{X}$ and $\bar{Y}$ are averages of the time series, was used to measure the strength of relationships between the reaction coordinate (Y) and various intermolecular distances of residues (X) and the structured water molecules in the 5 Å vicinity of Ade:N7. In this particular case, Y is not a random variable but a linear function that describes the increase in the reaction coordinate, which has been predefined and fixed during the simulation. Thus, the correlation we measure using these parameters describes whether the intermolecular distance is correlated to linear changes along the reaction coordinate (Figure 10c). The correlation coefficients giving the highest correlation are displayed in the right panel of Figure 10c. The contour plot, shown in the left panel, was generated to visualize the residues most affected by the change in reaction coordinate by representing the correlation coefficients with spherical Gaussians centered on their respective interacting atoms.

Our findings reveal that interactions involving three of the five water molecules (see water molecules in Figure 10b with the largest positional densities in the vicinity of Ade:N7) are more cooperative with cleavage than those involving the other two water molecules. As expected, hydrophilic residues (R31 and R34) strongly interact with two out of the three water molecules that are strongly tied to the reaction coordinate, and these interactions may help in preserving their unique configuration. More surprising are the interactions that correlate most strongly, which involve two hydrophobic residues (I191 and L46) sitting opposite to one another at the site of N7 protonation. Residues strategically placed at this site appear to aid the proton transfer by constraining the orientation of the bridging water molecule. Furthermore, cooperative interactions involving W30 and one of the three structured water molecules may similarly manipulate the orientation of the solvent molecule to assist the bridging water molecule during proton transfer. It is interesting to point out that a similar structural motif is found in the MutY structure of *E. coli*, in which hydrophobic residues, M185 and L40, guard the site of N7 protonation.

Using this correlation analysis to reduce the dimensionality of the data obtained in QM/MM molecular dynamics simulations led to the discovery that both the water network and the specific amino acid residues around the active site play specific roles in catalysis. In particular, transient water molecules become structured in a persistent arrangement, which coordinates the catalytically active water molecules into prime positions to assist the proton transfer. Such a configuration enables the solvent molecules to take on three distinct roles, depending on their positions: (i) a water molecule frequently bridges the catalytic residues and (ii) the bridging water molecule is assisted by 1−2 other "supporting" water molecules. To maintain this configuration, MutY, surprisingly, uses hydrophobic residues in combination with hydrophilic residues to tune the microenvironment into a "water trap". Hydrophilic residues prolong solvent residence times by maintaining hydrogen-bonding networks, whereas the hydrophobic residues constrain the positioning of the catalytic water molecules that assist the proton transfer event. In this way, the enzyme uses both entropic and enthalpic effects to guide the water-assisted reaction.

5.2. Machine Learning Techniques

Techniques from machine learning such as feature selection and causal inference are successfully applied in many fields, for example, for risk assessment. By using feature selection algorithms[348] to identify the most appropriate subset of relevant variables that describe a certain phenomenon, the high-dimensionality of QM/MM data can be reduced and used for further analysis with causal inference algorithms[347] to establish unique cause−effect relationships.

Such a combination of feature selection/causality inference analysis was recently applied to identify the molecular factors that determine the spectral shifts in rhodopsin.[93e] As mentioned in section 4.2.1, many possible factors have been suggested as being responsible for color tuning in rhodopsins, but most studies have been based on the investigation of single (0 K) structures neglecting the large thermal span of the absorption bands.

On the basis of a large set of classical and QM/MM data on the early intermediates of the rhodopsin photocycle, it was possible to perform a comprehensive and systematic analysis that allowed one to extract the statistically relevant factors in a high-dimensional space. An initial comprehensive set of 4×10^6 descriptors (distances, angles, and dihedrals) was drastically reduced by applying a Correlation Based Feature Selection (CBFS) algorithm to perform attribute selection. CBFS is implemented in the Weka machine learning package[353] and identifies the best subset of features from a given data set such that variables are selected if they are highly correlated with the target property, in this case the spectral shifts in the absorption spectra, while being minimally correlated with one another.[353] CBFS ranks feature subsets according to a correlation-based evaluation function:

$$\mathrm{CBFS}_S = \frac{k\langle r_{\mathrm{ft}}\rangle}{\sqrt{k + k(k-1)\langle r_{\mathrm{ff}}\rangle}} \quad (100)$$

where CBFS_S is the heuristic merit of the subset S containing k features f, $\langle r_{\mathrm{ft}}\rangle$ is the mean feature−target correlation, and $\langle r_{\mathrm{ff}}\rangle$ is the average feature−feature correlation of subset S ($f \in S$). Therefore, the numerator of this equation indicates how predictable a given subset of features is for the target property, while the denominator gives a measure of the redundancy among that group of features. In this way, irrelevant, redundant, and noisy points in the data are screened out, and the identification of a representative subset of features that globally maintains a predictive ability similar to that presented by the original set can be attained. Applying the CBFS algorithm, it was possible to reduce the initial set of variables from 4×10^6 descriptors to as little as 13.

In this reduced set, it is possible to apply causal inference methods such as the PC-LiNGAM algorithm[354] that is able to estimate the dependency structures in networks with arbitrary (both Gaussian and non-Gaussian) distributions. Using this

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

algorithm in conjunction with a directed acyclic graph (DAG) provides information on whether a particular variable influences another and whether relationships exist between variables. Applying this protocol, it was possible to describe all spectral shifts among the early intermediates of the rhodopsin photocycle with a set of five structural primary variables.[93e] These findings suggest that the spectral tuning between the early intermediates in the rhodopsin photocycle is mainly governed by modifications in intrinsic properties of the chromophore structure itself, which are altered by changes in the intermolecular interactions between the RPSB and the surrounding residues as a consequence of the structural relaxation of the chromophore upon photoinduced isomerization. Surprisingly, the distance to the counterion Glu113 that has often been proposed as an important ingredient in color tuning does not have any significant influence on the relative shifts among the early intermediates.

6. COMBINING BIOINFORMATICS AND QM/MM

QM/MM simulations can provide valuable and highly detailed information about a given biological process in one specific system. Often, it would be highly desirable to assess if the information gained for a single system can be generalized to an entire protein class or if specific characteristics of the enzymatic reaction mechanism could be shared among diverse proteins with similar active site features. In turn, bioinformatics techniques can be used to identify prominent statistical patterns such as conserved sequences or residues that can be rationalized subsequently with the help of a more detailed atomistic simulation method. Therefore, more and more studies have emerged recently that try to bridge the gap between (QM/MM) molecular dynamics simulations and bioinformatics approaches.[355]

6.1. Location and Refinement of Metal Binding Sites

The Prion Protein PrP is a widespread cell surface tethered protein expressed in the tissues of the central nervous system, and a conformational isomer of PrP (the "scrapie" form) has been attributed to playing an active role in a group of neurodegenerative disorders.[356] As a common denominator with many other neurodegenerative diseases, it has been suggested that metal ions such as copper could be implicated in the normal cellular function of PrP[357] or its disease related transformation.[357i,j,358,359]

While it is well established that several copper ions can bind to the unstructured part of PrP (residues 1−122), it has remained unclear whether or not copper binding sites also exist in the structured part of the protein (residues 123−231) (Figure 16a). Many studies have demonstrated that all of the disease related mutations lie within the region 90−231 of the protein, which further suggests that the structured part could play an important part in PrP related diseases, although experimental evidence is still controversial with regard to the binding affinity of copper ions to the C-terminal domain.[357g,360,361]

In an attempt to resolve these discrepancies and to locate possible copper binding sites, a statistical bioinformatics analysis was combined with explicit QM/MM simulations. First, an analysis of all known copper protein structures was performed to determine the probability of different amino acids to be implicated in copper coordination. As can be expected, His was identified as most likely ligand, followed by Cys and Met residues, together constituting 92% of the copper ligands,

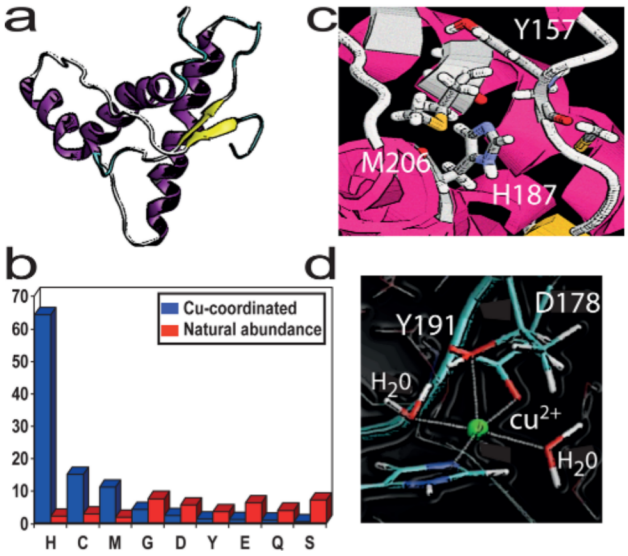


Figure 16. (a) C-terminal structured domain of the mouse PrP(124−226). (b) Statistical analyses for copper binding sites in proteins in the PDB database. (c) Identified high probability region for copper ion binding. (d) QM/MM refined copper binding site.

whereas their combined natural abundance is less than 7% (Figure 16b). This probability map was used to scan the PrP structure and identify regions with high propensity for copper ion binding. In total, six binding sites (one of them shown in Figure 16c) were identified in this way, which constitute of variations in residues in the vicinities of H140, H177, and H187. All candidate sites were subsequently refined with QM/MM Car−Parrinello simulations, ultimately leading to three putative binding sites. For all three sites, EPR spectra were calculated as averages over a number of snapshots (10 snapshots at time intervals of 0.35 ps) along the QM/MM dynamics and compared to the available experimental data.[360] This identified the site involving H187, E196, T191, and two water molecules (Figure 16d) as the most likely copper binding site at neutral pH. For the experimentally observed complex 1 at low pH,[360] the possible binding site of H140, D144, and D147, in which both of the imidazole nitrogen atoms are protonated, was suggested.

7. COMBINING SYSTEMS BIOLOGY AND QM/MM

Understanding and modeling biological systems is multifaceted, requiring a combination of approaches from several perspectives, including both reductionist and holistic views. The main aim of this section is to briefly introduce several examples of how the union of QM/MM simulations with computational systems biology methods enables the characterization of biochemical processes at graded levels of detail. As further discussed in this section, by combining these two different, but essentially complementary, perspectives, conclusions can be gained that would otherwise have been out-of-reach when pursued by either one of these disciplines alone.

7.1. Computational Systems Biology

Computational systems biology is an emerging discipline that lies at the intersection of graph theory and linear optimization. From the graph theory perspective, compounds or metabolites within a network are nodes, and the chemical reactions represent the edges of a graph. Genome-scale models of

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

metabolism are becoming increasingly more available for various organisms,[362] and the procedure is well established.[363] These metabolic network reconstructions together with constraint-based modeling techniques, such as Flux Balance Analysis (FBA),[364] lead to the possibility of estimating cellular phenotypes, such as the growth rates of cells, on the basis of metabolite uptake and secretion rates. These models and methods have also been applied to probe behavior and functionality through the response of the biochemical network to genomic alterations, such as gene deletions.

The knowledge of the relationship between cellular genotype and phenotype provides a great deal of information for engineering microbial species for human benefit. Recent developments in the genomics era have led to the development of metabolic engineering, which has a great potential to become an enabling strategy for microbial production of commodity chemicals. Metabolic engineering is a discipline that focuses on the improvement of cellular properties (e.g., cellular robustness or increase in the range of substrates that can be utilized by the cell) that can enhance the capacity of the organism to produce new chemicals and existing compounds at higher yields or degrade toxic compounds.

To be competitive to classical chemical synthesis protocols, a metabolic engineer must first understand how to utilize the full capacity of the host organism to produce target compounds with high yields and turnover rates. In most cases, organisms have not evolved to meet the demands of human needs, and further pathway engineering is required. In silico predictions of native and non-natural biosynthetic reactions (i.e., reactions that do not exist in the host organism) for further engineering have been predicted by pathway predictor algorithms, such as the Biochemical Network Integrated Computational Explorer (BNICE) algorithm.[365] BNICE explores all of the reaction space with respect to standard steady-state conditions by considering the general thermodynamic properties pertaining to each metabolite transformation and obtains a discriminative ranking of the nondescript list of thousands of predicted pathways.

To understand the capacity of the organism demands methods capable of handling complexity on many levels, such as managing the complexity of large-scale biochemical networks while respecting the critical chemical phenomena at the atomic scale. At the network scale, optimal pathways must be selected from thousands of possible interacting networks. At the atomistic scale, finding optimal enzymatic catalysts may require "fine-tuning" the reaction mechanism for the desired chemical task, such as increasing the binding affinity of the catalytic active site of an enzyme to a target molecule. As we will discuss in the following section, BNICE is one computational systems-scale method, which has been successfully used in combination with QM/MM simulations in the exploration of novel chemical processes.[279]

7.2. Combining a Systems Scale Analysis with a Molecular Scale Analysis

Combining BNICE with QM/MM molecular dynamics simulations allowed for addressing the feasibility of an engineered pathway at various levels of chemical detail. By integrating these computational methods, metabolic pathways were engineered in silico to respect the feasible ranges of intracellular metabolites and enzyme concentrations, while the molecular simulations (i.e., QM/MM) assessed the capacity of the enzyme to catalyze the desired reaction at the atomistic and electronic scales. The information gained from these in silico studies was then used to redesign and optimize the novel biocatalyst via site-directed mutagenesis. One of the main goals of this study was to assemble and rank the network of metabolic pathways, comprised of both natural and non-natural reactions, to produce a target compound.

As shown in Figure 17, a novel biosynthetic pathway was proposed by BNICE, and all of the possible enzyme candidates for one step of the pathway are given in (a). Using substrate-similarity-based comparison, the most likely candidates of this initial pool of enzymes were chosen in (b). From this filtered set, classical molecular dynamics simulations together with QM/MM simulations further assessed the binding free energy differences and catalytic cycle of the novel reaction in (c). Finally, the outcome of this assessment together with optimization of the binding affinity for the non-natural substrate led to the final enzyme biocatalyst in (d). This multiscale approach enabled the modeling of cellular processes covering a wide range of biological detail and made it possible

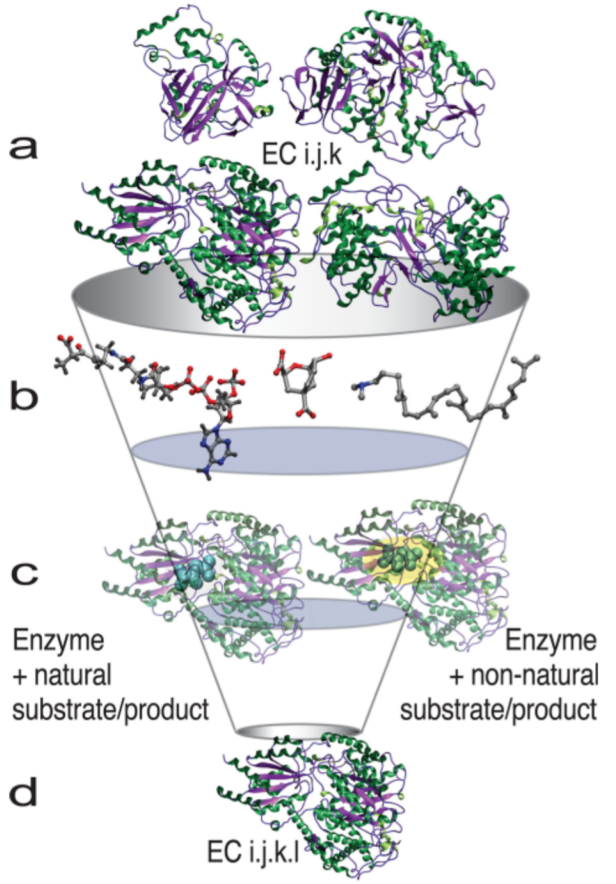


Figure 17. Scheme illustrating a process that filters through potential enzyme candidates (a) to find the most suitable and optimal catalyst for the desired reaction. The generalized scheme first generates and assembles the novel biosynthetic routes from a metabolite (e.g., pyruvate) to a desired product (e.g., 3HP), selects an enzyme, (e.g., 5.4.99.2), to act as the novel biocatalyst by comparing the structural similarity between natural and non-natural substrates (b), assesses the binding affinity and catalytic efficacy of the novel catalyst (c), and further optimizes the binding affinity or catalytic turnover of the novel reaction to give the final biocatalyst with novel functionalities (d).

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

to probe biological questions at varying levels of chemical detail.[279]

7.3. Bridging the Gap between Systems and Molecular Scale Analyses

A challenge that has recently emerged is the development of special measures to effectively bridge the gap between structural biology, computational chemistry methods, and systems biology methods in a useful and meaningful way.[279,298a,366] Reconstructed metabolic networks provide information on the connectivity of reactions within a biological network, such as metabolism; however, the molecular details of these interactions are sparse. The complementarity of computational systems biology and structural biology led to the first genome-scale model integrated with protein structures (GEM-PRO) for two different organisms, *Thermotoga maritima*[367] and *Escherichia coli*.[366b] Numerous studies have emerged since these two domains have been brought together: the evolution of fold families in metabolic proteins,[367b] identification of causal off-target actions of drugs,[368] identification of protein−protein interactions,[369] and determination of causal mutations for disease susceptibility.[369b,370]

A key challenge that remains, however, is establishing a link between computational chemistry methods and computational systems biology approaches. One application that could potentially benefit from such a combination of methods is prediction of protein promiscuity. Protein promiscuity is an increasingly important problem, transferrable to both protein engineering and pharmacological applications. Recent studies[371] have emerged that characterize the extent of underground metabolism using a top-down data driven approach that includes weak underground reactions in developed reconstructions of *E. coli*.[371a,b] Such studies are likely to facilitate a model-driven approach to discover new reactions and further the understanding of the role that promiscuity plays in the context of metabolism.

Our current understanding of protein promiscuity is limited. From the assemblies of large numbers of interacting reactions emerge organized, functional networks as a result of evolution, because maintaining modular structure is one force that helps systems adapt to new environmental conditions.[372] Networks were initially assembled in "patches", through the recruitment of primordial, nonspecific enzymes that reacted with a wide range of chemically similar compounds.[373] The patchwork assembly of generalized reactions enabled early cells to overcome the limiting coding capabilities of small genomes.[374] The versatility of these primitive elements has since evolved into specialization of orthologous enzymes. However, certain enzymes maintain a degree of substrate promiscuity, possibly as an evolutionary mechanism that enables a high degree of catalytic versatility. Understanding the physiological role of this so-called "underground" metabolic network or the latent activities of certain proteins in a network would benefit greatly from a combined systems biology/QM/MM approach.

8. OUTLOOK: QM/MM QUO VADIS?

In the almost 40 years since their introduction, QM/MM simulations have come of age in the technical details of their implementations and the scope of their applications that now routinely include not only problems in electronic ground state but also electronically excited states. Despite the versatility and general reliability of modern QM/MM simulations, intense research efforts are devoted to advance the current frontiers of (i) system size, (ii) time scale, and (iii) accuracy even further. To extend the system size, generalized mixed multishell approaches involving QM/MM/CG (coarse grained MM) and QM/MM/continuum approaches have been implemented.[187b,375] While enhanced sampling methods are already standardly applied in many QM/MM simulations, the combined challenge of high accuracy and sufficient sampling is still an open issue. Several methods have emerged recently to address this problem by introducing elegant ways of sampling with the high efficiency of computationally expedient lower level methods while retaining the accuracy of a more expensive high level treatment of the electronic structure.[224] More accurate treatments of the nuclear quantum motion are also being implemented in a QM/MM framework in combination with optimal control.[376] The most severe unsolved challenge is probably linked to the issue of how to go beyond the "single macromolecule" level and embed the QM/MM information into a wider biological context. Several attempts in this direction are currently ongoing, for example, as part of the Human Brain project.[377]

AUTHOR INFORMATION

Corresponding Author

*E-mail: ursula.roethlisberger@epfl.ch.

Notes

The authors declare no competing financial interest.

Biographies

Elizabeth Brunk was born in Santa Cruz, CA in 1985. She received her B.Sc. in biochemistry from the University of Michigan, Ann Arbor in 2007. She completed her M.Sc. in chemistry from both the Ecole Normale Superieure de Lyon and Universiteit Van Amsterdam in 2008. She joined the group of Prof. Rothlisberger in 2008 and completed her Ph.D. in computational biochemistry in 2013. She is currently a recipient of the Swiss National Science Foundation Early Postdoctoral Mobility Fellowship and collaborates with the University of California, Berkeley and the Joint BioEnergy Institute and the Systems Biology Research Group at the University of California, San Diego. Her current research interests involve merging systems biology approaches with molecular dynamics simulations.

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

Ursula Rothlisberger was born and raised in Switzerland. In 1991 she completed her Ph.D. thesis in Physical Chemistry at the University of Bern in collaboration with the IBM Zurich Research Laboratory in Rüschlikon. From 1992−1995, she was a postdoctoral research fellow in the group of Prof. Michael Klein at the University of Pennsylvania. After that, she spent one year in the group of Prof. Michele Parrinello at the Max-Planck Institute in Stuttgart, Germany. In 1997, she became Assistant Professor for Computer-Aided Inorganic Chemistry at the ETH Zurich until she moved to EPF Lausanne in 2002 as Associate Prof. of Computational Chemistry and Biochemistry. Since 2009 she is Full Professor at EPFL. She has published more than 200 articles and is the recipient of the Ruzicka medal and the 2005 Dirac medal of the World Organization of Theoretically Oriented Chemists WATOC. Her research interests are the development of density functional-based mixed QM/MM simulations for ground and excited states and their applications to problems in biology and materials science.

ACKNOWLEDGMENTS

We thank all of the current and former members of the Laboratory of Computational Chemistry and Biochemistry (LCBC) for the valuable contributions. In particular, we would like to thank Martin Bircher and Elisa Liberatore for a careful reading of the manuscript and Ali Ebrahim for technical support. We also thank Professor Vassily Hatzimanikatis and the Laboratory of Computational Systems Biotechnology. Funding from the Swiss National Science Foundation (Grant No. 200020-146645 and Mobility Postdoctorate fellowship: P2ELP2_148961) and the interdisciplinary research programs NCCR MUST and MARVEL is gratefully acknowledged. We thank the IT domain (DIT) of EPFL, the Swiss National Computing Center (CSCS), and CADMOS project for computing resources.

REFERENCES

(1) Panitchayangkoon, G.; Hayes, D.; Fransted, K. A.; Caram, J. R.; Harel, E.; Wen, J.; Blankenship, R. E.; Engel, G. S. Long-Lived Quantum Coherence in Photosynthetic Complexes at Physiological Temperature. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 12766−12770.

(2) Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103*, 227−249.

(3) Born, M.; Huang, K. *Dynamical Theory of Crystal Lattices*; Clarendon Press, Oxford University Press: Oxford, New York, 1988.

(4) Worth, G. A.; Robb, M. A.; Burghardt, I. A Novel Algorithm for Non-Adiabatic Direct Dynamics Using Variational Gaussian Wavepackets. *Faraday Discuss.* **2004**, *127*, 307−323.

(5) Meyer, H. D.; Gatti, F.; Worth, G. A. *Multidimensional Quantum Dynamics*; Wiley: New York, 2009.

(6) Kosloff, R. Time-Dependent Quantum-Mechanical Methods for Molecular Dynamics. *J. Phys. Chem.* **1988**, *92*, 2087−2100.

(7) (a) Beck, M. The Multiconfiguration Time-Dependent Hartree (MCTDH) Method: A Highly Efficient Algorithm for Propagating Wavepackets. *Phys. Rep.* **2000**, *324*, 1−105. (b) Meyer, H. D.; Manthe, U.; Cederbaum, L. S. The Multi-Configurational Time-Dependent Hartree Approach. *Chem. Phys. Lett.* **1990**, *165*, 73−78.

(8) (a) Burghardt, I.; Meyer, H. D.; Cederbaum, L. S. Approaches to the Approximate Treatment of Complex Molecular Systems by the Multiconfiguration Time-Dependent Hartree Method. *J. Chem. Phys.* **1999**, *111*, 2927−2939. (b) Hsu, D.; Coker, D. F. Quantum Dynamics Via Mobile Basis Sets: The Dirac Variational Principle. *J. Chem. Phys.* **1992**, *96*, 4266−4271.

(9) (a) Arango, C. A.; Brumer, P. Communication: One-Photon Phase Control of Cis-Trans Isomerization in Retinal. *J. Chem. Phys.* **2013**, *138*, 071104. (b) Makri, N. Quantum Dissipative Dynamics: A Numerically Exact Methodology. *J. Phys. Chem. A* **1998**, *102*, 4414−4427.

(10) Lopreore, C. L.; Wyatt, R. E. Quantum Wave Packet Dynamics with Trajectories. *Phys. Rev. Lett.* **1999**, *82*, 5190.

(11) Curchod, B. F. E.; Tavernelli, I.; Rothlisberger, U. Trajectory-Based Solution of the Nonadiabatic Quantum Dynamics Equations: An on-the-Fly Approach for Molecular Dynamics Simulations. *Phys. Chem. Chem. Phys.* **2011**, *13*, 3231−3236.

(12) Herman, M. F. Dynamics by Semiclassical Methods. *Annu. Rev. Phys. Chem.* **1994**, *45*, 83−111.

(13) Payne, M. C.; Teter, M. P.; Allan, D. C.; Arias, T. A.; Joannopoulos, J. D. Iterative Minimization Techniques for Ab Initio Total-Energy Calculations: Molecular Dynamics and Conjugate Gradients. *Rev. Mod. Phys.* **1992**, *64*, 1045.

(14) Marx, D.; Hutter, J. *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*; Cambridge University Press: New York, 2009.

(15) Car, R.; Parrinello, M. Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.* **1985**, *55*, 2471−2474.

(16) (a) Pastore, G.; Smargiassi, E.; Buda, F. Theory of Ab Initio Molecular-Dynamics Calculations. *Phys. Rev. A* **1991**, *44*, 6334−6347. (b) Bornemann, F. A.; Schütte, C. A Mathematical Investigation of the Car-Parrinello Method. *Numerische Mathematik* **1998**, *78*, 359−376.

(17) (a) Kirson, Z.; Gerber, R. B.; Nitzan, A.; Ratner, M. A. Dynamics of Metal Electron Excitation in Atom-Surface Collisions: A Quantum Wave Packet Approach. *Surf. Sci.* **1984**, *137*, 527−550. (b) McLachlan, A. D. A Variational Solution of the Time-Dependent Schrodinger Equation. *Mol. Phys.* **1964**, *8*, 39−44.

(18) (a) Drukker, K. Basics of Surface Hopping in Mixed Quantum/Classical Simulations. *J. Comput. Phys.* **1999**, *153*, 225−272. (b) Tully, J. Mixed Quantum−Classical Dynamics. *Faraday Discuss.* **1998**, *110*, 407−419.

(19) Billing, G. D. Classical Path Method in Inelastic and Reactive Scattering. *Int. Rev. Phys. Chem.* **1994**, *13*, 309−335.

(20) Li, X.; Tully, J. C.; Schlegel, H. B.; Frisch, M. J. Ab Initio Ehrenfest Dynamics. *J. Chem. Phys.* **2005**, *123*, 084106.

(21) Fang, J.-Y.; Hammes-Schiffer, S. Comparison of Surface Hopping and Mean Field Approaches for Model Proton Transfer Reactions. *J. Chem. Phys.* **1999**, *110*, 11166−11175.

(22) Parandekar, P. V.; Tully, J. C. Detailed Balance in Ehrenfest Mixed Quantum-Classical Dynamics. *J. Chem. Theory Comput.* **2006**, *2*, 229−235.

(23) (a) Hammes-Schiffer, S.; Tully, J. C. Proton Transfer in Solution: Molecular Dynamics with Quantum Transitions. *J. Chem. Phys.* **1994**, *101*, 4657−4667. (b) Tully, J. C.; Preston, R. K. Trajectory Surface Hopping Approach to Nonadiabatic Molecular Collisions: The Reaction of H+ with D2. *J. Chem. Phys.* **1971**, *55*, 562−572.

(24) (a) Granucci, G.; Persico, M.; Zoccante, A. Including Quantum Decoherence in Surface Hopping. *J. Chem. Phys.* **2010**, *133*, 134111. (b) Jaeger, H. M.; Fischer, S.; Prezhdo, O. V. Decoherence-Induced Surface Hopping. *J. Chem. Phys.* **2012**, *137*, 22A545. (c) Landry, B. R.; Subotnik, J. E. Communication: Standard Surface Hopping Predicts Incorrect Scaling for Marcus' Golden-Rule Rate: The Decoherence

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

Problem Cannot Be Ignored. *J. Chem. Phys.* **2011**, *135*, 191101. (d) Zhu, C.; Jasper, A. W.; Truhlar, D. G. Non-Born-Oppenheimer Liouville-Von Neumann Dynamics. Evolution of a Subsystem Controlled by Linear and Population-Driven Decay of Mixing with Decoherent and Coherent Switching. *J. Chem. Theory Comput.* **2005**, *1*, 527−540. (e) Zhu, C.; Nangia, S.; Jasper, A. W.; Truhlar, D. G. Coherent Switching with Decay of Mixing: An Improved Treatment of Electronic Coherence for Non-Born-Oppenheimer Trajectories. *J. Chem. Phys.* **2004**, *121*, 7658−7670.

(25) (a) Ben-Nun, M.; Martínez, T. J. Semiclassical Tunneling Rates from Ab Initio Molecular Dynamics. *J. Phys. Chem. A* **1999**, *103*, 6055−6059. (b) Martinez, T. J.; Ben-Nun, M.; Levine, R. D. Multi-Electronic-State Molecular Dynamics: A Wave Function Approach with Applications. *J. Phys. Chem.* **1996**, *100*, 7884−7895.

(26) Pu, J.; Gao, J.; Truhlar, D. G. Multidimensional Tunneling, Recrossing, and the Transmission Coefficient for Enzymatic Reactions. *Chem. Rev.* **2006**, *106*, 3140−3169.

(27) Makri, N. Accurate Quantum Mechanical Calculation of Thermally Averaged Reaction Rate Constants for Polyatomic Systems. *J. Chem. Phys.* **1991**, *94*, 4949−4958.

(28) Hwang, J.-K.; Warshel, A. How Important Are Quantum Mechanical Nuclear Motions in Enzyme Catalysis? *J. Am. Chem. Soc.* **1996**, *118*, 11745−11751.

(29) Hammes-Schiffer, S. Hydrogen Tunneling and Protein Motion in Enzyme Reactions. *Acc. Chem. Res.* **2006**, *39*, 93−100.

(30) (a) Gao, J.; Truhlar, D. G. Quantum Mechanical Methods for Enzyme Kinetics. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467−505. (b) Sutcliffe, M. J.; Scrutton, N. S. Computational Studies of Enzyme Mechanism: Linking Theory with Experiment in the Analysis of Enzymic H-Tunnelling. *Phys. Chem. Chem. Phys.* **2006**, *8*, 4510−4516. (c) Wang, H.; Sun, X.; Miller, W. H. Semiclassical Approximations for the Calculation of Thermal Rate Constants for Chemical Reactions in Complex Molecular Systems. *J. Chem. Phys.* **1998**, *108*, 9726−9736.

(31) (a) Capitani, J. F.; Nalewajski, R. F.; Parr, R. G. Non-Born−Oppenheimer Density Functional Theory of Molecular Systems. *J. Chem. Phys.* **1982**, *76*, 568−573. (b) Kreibich, T.; Gross, E. K. Multicomponent Density-Functional Theory for Electrons and Nuclei. *Phys. Rev. Lett.* **2001**, *86*, 2984−2987.

(32) Webb, S. P.; Iordanov, T.; Hammes-Schiffer, S. Multiconfigurational Nuclear-Electronic Orbital Approach: Incorporation of Nuclear Quantum Effects in Electronic Structure Calculations. *J. Chem. Phys.* **2002**, *117*, 4106−4118.

(33) (a) Helgaker, T.; Jorgensen, P.; Olsen, J. *Molecular Electronic-Structure Theory*; Wiley: New York, 2013. (b) Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; Dover Publications: New York, 2012.

(34) Dreuw, A.; Head-Gordon, M. Single-Reference Ab Initio Methods for the Calculation of Excited States of Large Molecules. *Chem. Rev.* **2005**, *105*, 4009−4037.

(35) (a) Fock, V. Näherungsmethode Zur Lösung Des Quantenmechanischen Mehrkörperproblems. *Z. Phys.* **1930**, *61*, 126−148. (b) Hartree, D. R. *The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods*; Cambridge University Press: New York, 1928; Vol. *24*. (c) Slater, J. C. Note on Hartree's Method. *Phys. Rev.* **1930**, *35*, 210.

(36) Pu, J.; Gao, J.; Truhlar, D. G. Generalized Hybrid Orbital (GHO) Method for Combining Ab Initio Hartree-Fock Wave Functions with Molecular Mechanics. *J. Phys. Chem. A* **2004**, *108*, 632−650.

(37) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618.

(38) (a) Cummins, P. L.; Gready, J. E. Energetically Most Likely Substrate and Active-Site Protonation Sites and Pathways in the Catalytic Mechanism of Dihydrofolate Reductase. *J. Am. Chem. Soc.* **2001**, *123*, 3418−3428. (b) Ridder, L.; Harvey, J. N.; Rietjens, I. M.; Vervoort, J.; Mulholland, A. J. Ab Initio QM/MM Modeling of the Hydroxylation Step in P-Hydroxybenzoate Hydroxylase. *J. Phys. Chem. B* **2003**, *107*, 2118−2126.

(39) (a) Kongsted, J.; Osted, A.; Mikkelsen, K. V.; Christiansen, O. Linear Response Functions for Coupled Cluster/Molecular Mechanics Including Polarization Interactions. *J. Chem. Phys.* **2003**, *118*, 1620−1633. (b) Woodcock, H. L., III; Hodoscek, M.; Gilbert, A. T. B.; Gill, P. M. W.; Schaefer, H. F., III; Brooks, B. R. Interfacing Q-Chem and Charmm to Perform QM/MM Reaction Path Calculations. *J. Comput. Chem.* **2007**, *28*, 1485−1502.

(40) Schütz, M.; Hetzer, G.; Werner, H.-J. Low-Order Scaling Local Electron Correlation Methods. I. Linear Scaling Local MP2. *J. Chem. Phys.* **1999**, *111*, 5691−5705.

(41) (a) Schütz, M. Low-Order Scaling Local Electron Correlation Methods. III. Linear Scaling Local Perturbative Triples Correction (T). *J. Chem. Phys.* **2000**, *113*, 9986−10001. (b) Manby, F. R.; Werner, H.-J.; Adler, T. B.; May, A. J. Explicitly Correlated Local Second-Order Perturbation Theory with a Frozen Geminal Correlation Factor. *J. Chem. Phys.* **2006**, *124*, 94103.

(42) (a) Claeyssens, F.; Harvey, J. N.; Manby, F. R.; Mata, R. A.; Mulholland, A. J.; Ranaghan, K. E.; Schütz, M.; Thiel, S.; Thiel, W.; Werner, H. j. High Accuracy Computation of Reaction Barriers in Enzymes. *Angew. Chem., Int. Ed.* **2006**, *118*, 7010−7013. (b) Mata, R. A.; Werner, H.-J.; Thiel, S.; Thiel, W. Toward Accurate Barriers for Enzymatic Reactions: QM/MM Case Study on P-Hydroxybenzoate Hydroxylase. *J. Chem. Phys.* **2008**, *128*, 025104. (c) Mulholland, A. J. Chemical Accuracy in QM/MM Calculations on Enzyme-Catalysed Reactions. *Chem. Cent. J.* **2007**, *1*, 19.

(43) Serrano-Andrés, L.; Merchán, M. Quantum Chemistry of the Excited State: 2005 Overview. *J. Mol. Struct. (THEOCHEM)* **2005**, *729*, 99−108.

(44) Nakano, H. Quasidegenerate Perturbation Theory with Multiconfigurational Self-Consistent-Field Reference Functions. *J. Chem. Phys.* **1993**, *99*, 7983−7992.

(45) (a) Bearpark, M. J.; Ogliaro, F.; Vreven, T.; Boggio-Pasqua, M.; Frisch, M. J.; Larkin, S. M.; Morrison, M.; Robb, M. A. Casscf Calculations for Photoinduced Processes in Large Molecules: Choosing When to Use the RASSCF, ONIOM and MMVB Approximations. *J. Photochem. Photobiol., A* **2007**, *190*, 207−227. (b) Schapiro, I.; Ryazantsev, M. N.; Ding, W. J.; Huntress, M. M.; Melaccio, F.; Andruniow, T.; Olivucci, M. Computational Photobiology and Beyond. *Aust. J. Chem.* **2010**, *63*, 413−429.

(46) Christiansen, O.; Koch, H.; Jørgensen, P. The Second-Order Approximate Coupled Cluster Singles and Doubles Model CC2. *Chem. Phys. Lett.* **1995**, *243*, 409−418.

(47) Trofimov, A. B.; Schirmer, J. An Efficient Polarization Propagator Approach to Valence Electron Excitation Spectra. *J. Phys. B: At., Mol. Opt. Phys.* **1995**, *28*, 2299.

(48) Neese, F. A Spectroscopy Oriented Configuration Interaction Procedure. *J. Chem. Phys.* **2003**, *119*, 9428−9443.

(49) Hoffmann, M.; Wanko, M.; Strodel, P.; König, P. H.; Frauenheim, T.; Schulten, K.; Thiel, W.; Tajkhorshid, E.; Elstner, M. Color Tuning in Rhodopsins: The Mechanism for the Spectral Shift between Bacteriorhodopsin and Sensory Rhodopsin Ii. *J. Am. Chem. Soc.* **2006**, *128*, 10808−10818.

(50) (a) Filippi, C.; Buda, F.; Guidoni, L.; Sinicropi, A. Bathochromic Shift in Green Fluorescent Protein: A Puzzle for QM/MM Approaches. *J. Chem. Theory Comput.* **2011**, *8*, 112−124. (b) Varsano, D.; Coccia, E.; Pulci, O.; Conte, A. M.; Guidoni, L. Ground State Structures and Electronic Excitations of Biological Chromophores at Quantum Monte Carlo/Many Body Green's Function Theory Level. *Comput. Theor. Chem.* **2014**, *10*, 1048−1061.

(51) (a) Boggio-Pasqua, M.; Burmeister, C. F.; Robb, M. A.; Groenhof, G. Photochemical Reactions in Biological Systems: Probing the Effect of the Environment by Means of Hybrid Quantum Chemistry/Molecular Mechanics Simulations. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7912−7928. (b) Ruckenbauer, M.; Barbatti, M.; Sellner, B.; Muller, T.; Lischka, H. Azomethane: Nonadiabatic Photodynamical Simulations in Solution. *J. Phys. Chem. A* **2010**, *114*, 12585−12590. (c) Weingart, O.; Altoè, P.; Stenta, M.; Bottoni, A.; Orlandi, G.; Garavelli, M. Product Formation in Rhodopsin by Fast Hydrogen Motions. *Phys. Chem. Chem. Phys.* **2011**, *13*, 3645−3648.

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

(52) Virshup, A. M.; Punwong, C.; Pogorelov, T. V.; Lindquist, B. A.; Ko, C.; Martínez, T. J. Photodynamics in Complex Environments: Ab Initio Multiple Spawning Quantum Mechanical/Molecular Mechanical Dynamics. *J. Phys. Chem. B* **2009**, *113*, 3280−3291.
(53) Tapavicza, E.; Tavernelli, I.; Rothlisberger, U. Ab Initio Excited State Properties and Dynamics of a Prototype Sigma-Bridged-Donor-Acceptor Molecule. *J. Phys. Chem. A* **2009**, *113*, 9595−9602.
(54) Kamerlin, S. C. L.; Warshel, A. The EVB as a Quantitative Tool for Formulating Simulations and Analyzing Biological and Chemical Reactions. *Faraday Discuss.* **2010**, *145*, 71−106.
(55) Thiel, W. Semiempirical Quantum−Chemical Methods. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2014**, *4*, 145−157.
(56) (a) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. A QM/MM Implementation of the Self-Consistent Charge Density Functional Tight Binding (SCC-DFTB) Method. *J. Phys. Chem. B* **2001**, *105*, 569−585. (b) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. Self-Consistent-Charge Density-Functional Tight-Binding Method for Simulations of Complex Materials Properties. *Phys. Rev. B: Condens. Matter* **1998**, *58*, 7260−7268.
(57) Pople, J. A.; Santry, D. P.; Segal, G. A. Approximate Self-Consistent Molecular Orbital Theory. I. Invariant Procedures. *J. Chem. Phys.* **1965**, *43*, S129−S135.
(58) Dewar, M. J. S.; Thiel, W. Ground States of Molecules. 38. The MNDO Method. Approximations and Parameters. *J. Am. Chem. Soc.* **1977**, *99*, 4899−4907.
(59) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and Use of Quantum Mechanical Molecular Models. 76. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.
(60) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods V: Modification of NDDO Approximations and Application to 70 Elements. *J. Mol. Model.* **2007**, *13*, 1173−1213.
(61) Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. PDDG/PM3 and PDDG/MNDO: Improved Semiempirical Methods. *J. Comput. Chem.* **2002**, *23*, 1601−1622.
(62) Thiel, W.; Voityuk, A. A. Extension of Mndo to D Orbitals: Parameters and Results for the Second-Row Elements and for the Zinc Group. *J. Phys. Chem.* **1996**, *100*, 616−626.
(63) Weber, W.; Thiel, W. Orthogonalization Corrections for Semiempirical Methods. *Theor. Chem. Acc.* **2000**, *103*, 495−506.
(64) (a) Billeter, S. R.; Turner, A. J.; Thiel, W. Linear Scaling Geometry Optimisation and Transition State Search in Hybrid Delocalised Internal Coordinates. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2177−2186. (b) Dixon, S. L.; Merz, K. M., Jr. Semiempirical Molecular Orbital Calculations with Linear System Size Scaling. *J. Chem. Phys.* **1996**, *104*, 6643−6649. (c) Lee, T. s.; York, D. M.; Yang, W. Linear-Scaling Semiempirical Quantum Calculations for Macromolecules. *J. Chem. Phys.* **1996**, *105*, 2744−2750.
(65) Daniels, A. D.; Millam, J. M.; Scuseria, G. E. Semiempirical Methods with Conjugate-Gradient Density-Matrix Search to Replace Diagonalization for Molecular-Systems Containing Thousands of Atoms. *J. Chem. Phys.* **1997**, *107*, 425−431.
(66) Elstner, M.; Frauenheim, T.; Suhai, S. An Approximate DFT Method for QM/MM Simulations of Biological Structures and Processes. *J. Mol. Struct. (THEOCHEM)* **2003**, *632*, 29−41.
(67) Ridley, J.; Zerner, M. An Intermediate Neglect of Differential Overlap Technique for Spectroscopy: Pyrrole and the Azines. *Theor. Chim. Acta* **1973**, *32*, 111−134.
(68) Frauenheim, T.; Seifert, G.; Elstner, M.; Niehaus, T.; Kohler, C.; Amkreutz, M.; Sternberg, M.; Hajnal, Z. n.; Carlo, A. D.; Suhai, S. n. Atomistic Simulations of Complex Materials: Ground-State and Excited-State Properties. *J. Phys.: Condens. Matter* **2002**, *14*, 3015.
(69) Toniolo, A.; Granucci, G.; Martínez, T. J. Conical Intersections in Solution: A QM/MM Study Using Floating Occupation Semi-empirical Configuration Interaction Wave Functions. *J. Phys. Chem. A* **2003**, *107*, 3822−3830.
(70) Silva-Junior, M. R.; Thiel, W. Benchmark of Electronically Excited States for Semiempirical Methods: MNDO, AM1, PM3, OM1, OM2, OM3, INDO/S, and INDO/S2. *J. Chem. Theory Comput.* **2010**, *6*, 1546−1564.
(71) Persico, M.; Granucci, G.; Inglese, S.; Laino, T.; Toniolo, A. Semiclassical Simulation of Photochemical Reactions in Condensed Phase. *J. Mol. Struct. (THEOCHEM)* **2003**, *621*, 119−126.
(72) Petersen, J.; Wohlgemuth, M.; Sellner, B.; Bonacić-Koutecký, V.; Lischka, H.; Mitrić, R. Laser Pulse Trains for Controlling Excited State Dynamics of Adenine in Water. *Phys. Chem. Chem. Phys.* **2012**, *14*, 4687−4694.
(73) Lan, Z.; Lu, Y.; Fabiano, E.; Thiel, W. QM/MM Nonadiabatic Decay Dynamics of 9H-Adenine in Aqueous Solution. *ChemPhysChem* **2011**, *12*, 1989−1998.
(74) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864−B871.
(75) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133.
(76) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the Density Functional Ladder: Nonempirical Meta−Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.* **2003**, *91*, 146401.
(77) (a) Burke, K. Perspective on Density Functional Theory. *J. Chem. Phys.* **2012**, *136*, 150901. (b) Sholl, D.; Steckel, J. A. *Density Functional Theory: A Practical Introduction*; John Wiley & Sons: New York, 2011.
(78) Kussmann, J.; Beer, M.; et al. Linear-Scaling Self-Consistent Field Methods for Large Molecules. *Computat. Mol. Biosci.* **2013**, *3*, 614−636.
(79) (a) Friesner, R. A.; Guallar, V. Ab Initio Quantum Chemical and Mixed Quantum Mechanics/Molecular Mechanics (QM/MM) Methods for Studying Enzymatic Catalysis. *Annu. Rev. Phys. Chem.* **2005**, *56*, 389−427. (b) Hu, H.; Yang, W. Free Energies of Chemical Reactions in Solution and in Enzymes with Ab Initio Quantum Mechanics/Molecular Mechanics Methods. *Annu. Rev. Phys. Chem.* **2008**, *59*, 573−601.
(80) Sulpizi, M.; Folkers, G.; Rothlisberger, U.; Carloni, P.; Scapozza, L. Applications of Density Functional Theory-Based Methods in Medicinal Chemistry. *Quant. Struct.-Act. Relat.* **2002**, *21*, 173−181.
(81) Thar, J.; Reckien, W.; Kirchner, B. Car−Parrinello Molecular Dynamics Simulations and Biological Systems. *Atomistic Approaches in Modern Biology*; Springer: New York, 2007; pp 133−171.
(82) Runge, E.; Gross, E. K. U. Density-Functional Theory for Time-Dependent Systems. *Phys. Rev. Lett.* **1984**, *52*, 997.
(83) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W.; Tironi, I. G.; van Leeuwen, R. Biomolecular Simulation: The Gromos96 Manual and User Guide-Causality and Symmetry in Time-Dependent Density-Functional Theory. *Verlag der Fachvereine, Zürich* **1996**, *80*, 1−1024.
(84) Röhrig, U. F.; Guidoni, L.; Rothlisberger, U. Solvent and Protein Effects on the Structure and Dynamics of the Rhodopsin Chromophore. *ChemPhysChem* **2005**, *6*, 1836−1847.
(85) Frank, I.; Hutter, J.; Marx, D.; Parrinello, M. Molecular Dynamics in Low-Spin Excited States. *J. Chem. Phys.* **1998**, *108*, 4060−4069.
(86) Röhrig, U. F.; Guidoni, L.; Laio, A.; Frank, I.; Rothlisberger, U. A Molecular Spring for Vision. *J. Am. Chem. Soc.* **2004**, *126*, 15328−15329.
(87) Ziegler, T.; Rauk, A.; Baerends, E. J. On the Calculation of Multiplet Energies by the Hartree-Fock-Slater Method. *Theor. Chim. Acta* **1977**, *43*, 261−271.
(88) Casida, M. E. Time-Dependent Density Functional Response Theory for Molecules. *Recent Adv. Density Funct. Methods* **1995**, *1*, 155.
(89) Furche, F.; Ahlrichs, R. Adiabatic Time-Dependent Density Functional Methods for Excited State Properties. *J. Chem. Phys.* **2002**, *117*, 7433−7447.
(90) (a) Hutter, J. Excited State Nuclear Forces from the Tamm−Dancoff Approximation to Time-Dependent Density Functional Theory within the Plane Wave Basis Set Framework. *J. Chem. Phys.* **2003**, *118*, 3928−3934. (b) Doltsinis, N. L.; Marx, D. Nonadiabatic Car-Parrinello Molecular Dynamics. *Phys. Rev. Lett.* **2002**, *88*, 166402.

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

(c) Chiba, M.; Tsuneda, T.; Hirao, K. Excited State Geometry Optimizations by Analytical Energy Gradient of Long-Range Corrected Time-Dependent Density Functional Theory. *J. Chem. Phys.* **2006**, *124*, 144106. (d) Liu, F.; Gan, Z.; Shao, Y.; Hsu, C.-P.; Dreuw, A.; Head-Gordon, M.; Miller, B. T.; Brooks, B. R.; Yu, J.-G.; Furlani, T. R. A Parallel Implementation of the Analytic Nuclear Gradient for Time-Dependent Density Functional Theory within the Tamm−Dancoff Approximation. *Mol. Phys.* **2010**, *108*, 2791−2800.

(91) (a) Röhrig, U. F.; Frank, I.; Hutter, J.; Laio, A.; VandeVondele, J.; Rothlisberger, U. QM/MM Car-Parrinello Molecular Dynamics Study of the Solvent Effects on the Ground State and on the First Excited Singlet State of Acetone in Water. *ChemPhysChem* **2003**, *4*, 1177−1182. (b) Sulpizi, M.; Carloni, P.; Hutter, J.; Rothlisberger, U. A Hybrid TDDFT/MM Investigation of the Optical Properties of Aminocoumarins in Water and Acetonitrile Solution. *Phys. Chem. Chem. Phys.* **2003**, *5*, 4798−4805. (c) Yoo, S.; Zahariev, F.; Sok, S.; Gordon, M. S. Solvent Effects on Optical Properties of Molecules: A Combined Time-Dependent Density Functional Theory/Effective Fragment Potential Approach. *J. Chem. Phys.* **2008**, *129*, 144112.

(92) Govind, N.; Valiev, M.; Jensen, L.; Kowalski, K. Excitation Energies of Zinc Porphyrin in Aqueous Solution Using Long-Range Corrected Time-Dependent Density Functional Theory. *J. Phys. Chem. A* **2009**, *113*, 6041−6043.

(93) (a) Moret, M.-E.; Tapavicza, E.; Guidoni, L.; Röhrig, U. F.; Sulpizi, M.; Tavernelli, I.; Rothlisberger, U. Quantum Mechanical/Molecular Mechanical (QM/MM) Car-Parrinello Simulations in Excited States. *Chimia* **2005**, *59*, 493−498. (b) Wanko, M.; Hoffmann, M.; Strodel, P.; Koslowski, A.; Thiel, W.; Neese, F.; Frauenheim, T.; Elstner, M. Calculating Absorption Shifts for Retinal Proteins: Computational Challenges. *J. Phys. Chem. B* **2005**, *109*, 3606−3615. (c) Parac, M.; Doerr, M.; Marian, C. M.; Thiel, W. QM/MM Calculation of Solvent Effects on Absorption Spectra of Guanine. *J. Comput. Chem.* **2010**, *31*, 90−106. (d) Valsson, O.; Campomanes, P.; Tavernelli, I.; Rothlisberger, U.; Filippi, C. Rhodopsin Absorption from First Principles: Bypassing Common Pitfalls. *J. Chem. Theory Comput.* **2013**, *9*, 2441−2454. (e) Campomanes, P.; Neri, M.; Horta, B. A. C.; Röhrig, U. F.; Vanni, S.; Tavernelli, I.; Rothlisberger, U. Origin of the Spectral Shifts among the Early Intermediates of the Rhodopsin Photocycle. *J. Am. Chem. Soc.* **2014**, *136*, 3842−3851.

(94) Craig, C. F.; Duncan, W. R.; Prezhdo, O. V. Trajectory Surface Hopping in the Time-Dependent Kohn-Sham Approach for Electron-Nuclear Dynamics. *Phys. Rev. Lett.* **2005**, *95*, 163001.

(95) (a) Guglielmi, M.; Doemer, M.; Tavernelli, I.; Rothlisberger, U. Photodynamics of Lys+-Trp Protein Motifs: Hydrogen Bonds Ensure Photostability. *Faraday Discuss.* **2013**, *163*, 189−203 discussion 243−75. (b) Guglielmi, M.; Tavernelli, I.; Rothlisberger, U. On the Proton Transfer Mechanism in Ammonia-Bridged 7-Hydroxyquinoline: A TDDFT Molecular Dynamics Study. *Phys. Chem. Chem. Phys.* **2009**, *11*, 4549−4555.

(96) Mercier, S. R.; Boyarkin, O. V.; Kamariotis, A.; Guglielmi, M.; Tavernelli, I.; Cascella, M.; Rothlisberger, U.; Rizzo, T. R. Microsolvation Effects on the Excited-State Dynamics of Protonated Tryptophan. *J. Am. Chem. Soc.* **2006**, *128*, 16938−16943.

(97) Chernyak, V.; Mukamel, S. Density-Matrix Representation of Nonadiabatic Couplings in Time-Dependent Density Functional (TDDFT) Theories. *J. Chem. Phys.* **2000**, *112*, 3572−3579.

(98) Baer, R. Non-Adiabatic Couplings by Time-Dependent Density Functional Theory. *Chem. Phys. Lett.* **2002**, *364*, 75−79.

(99) Tapavicza, E.; Tavernelli, I.; Rothlisberger, U. Trajectory Surface Hopping within Linear Response Time-Dependent Density-Functional Theory. *Phys. Rev. Lett.* **2007**, *98*, 023001.

(100) (a) Tavernelli, I.; Tapavicza, E.; Rothlisberger, U. Non-Adiabatic Dynamics Using Time-Dependent Density Functional Theory: Assessing the Coupling Strengths. *J. Mol. Struct. (THEOCHEM)* **2009**, *914*, 22−29. (b) Tavernelli, I.; Tapavicza, E.; Rothlisberger, U. Nonadiabatic Coupling Vectors within Linear Response Time-Dependent Density Functional Theory. *J. Chem. Phys.* **2009**, *130*, 124107.

(101) Tavernelli, I.; Curchod, B. F. E.; Rothlisberger, U. On Nonadiabatic Coupling Vectors in Time-Dependent Density Functional Theory. *J. Chem. Phys.* **2009**, *131*, 196101.

(102) Tavernelli, I.; Curchod, B. F. E.; Laktionov, A.; Rothlisberger, U. Nonadiabatic Coupling Vectors for Excited States within Time-Dependent Density Functional Theory in the Tamm-Dancoff Approximation and Beyond. *J. Chem. Phys.* **2010**, *133*, 194104.

(103) Tretiak, S.; Mukamel, S. Density Matrix Analysis and Simulation of Electronic Excitations in Conjugated and Aggregated Molecules. *Chem. Rev.* **2002**, *102*, 3171−3212.

(104) Hu, C.; Hirai, H.; Sugino, O. Nonadiabatic Couplings from Time-Dependent Density Functional Theory: Formulation in the Casida Formalism and Practical Scheme within Modified Linear Response. *J. Chem. Phys.* **2007**, *127*, 064103.

(105) Hu, C.; Sugino, O.; Watanabe, K. Second-Order Nonadiabatic Couplings from Time-Dependent Density Functional Theory: Evaluation in the Immediate Vicinity of Jahn-Teller/Renner-Teller Intersections. *J. Chem. Phys.* **2011**, *135*, 074101.

(106) (a) Doltsinis, N. L.; Kosov, D. S. Plane Wave/Pseudopotential Implementation of Excited State Gradients in Density Functional Linear Response Theory: A New Routevia Implicit Differentiation. *J. Chem. Phys.* **2005**, *122*, 144101. (b) Send, R.; Furche, F. First-Order Nonadiabatic Couplings from Time-Dependent Hybrid Density Functional Response Theory: Consistent Formalism, Implementation, and Performance. *J. Chem. Phys.* **2010**, *132*, 044107.

(107) Tomasello, G.; Wohlgemuth, M.; Petersen, J.; Mitrić, R. Photodynamics of Free and Solvated Tyrosine. *J. Phys. Chem. B* **2012**, *116*, 8762−8770.

(108) Masson, F.; Laino, T.; Rothlisberger, U.; Hutter, J. A QM/MM Investigation of Thymine Dimer Radical Anion Splitting Catalyzed by DNA Photolyase. *ChemPhysChem* **2009**, *10*, 400−410.

(109) Kubař, T.; Elstner, M. Efficient Algorithms for the Simulation of Non-Adiabatic Electron Transfer in Complex Molecular Systems: Application to DNA. *Phys. Chem. Chem. Phys.* **2013**, *15*, 5794−5813.

(110) Tavernelli, I.; Curchod, B. F. E.; Rothlisberger, U. Mixed Quantum-Classical Dynamics with Time-Dependent External Fields: A Time-Dependent Density-Functional-Theory Approach. *Phys. Rev. A* **2010**, *81*, 052508.

(111) Curchod, B. F. E.; Penfold, T. J.; Rothlisberger, U.; Tavernelli, I. Local Control Theory in Trajectory-Based Nonadiabatic Dynamics. *Phys. Rev. A* **2011**, *84*, 042507.

(112) Castro, A.; Marques, M. A. L.; Rubio, A. Propagators for the Time-Dependent Kohn-Sham Equations. *J. Chem. Phys.* **2004**, *121*, 3425−3433.

(113) Theilhaber, J. Ab Initio Simulations of Sodium Using Time-Dependent Density-Functional Theory. *Phys. Rev. B: Condens. Matter* **1992**, *46*, 12990−13003.

(114) (a) Yabana, K.; Bertsch, G. F. Time-Dependent Local-Density Approximation in Real Time. *Phys. Rev. B: Condens. Matter* **1996**, *54*, 4484−4487. (b) Ojanperä, A.; Havu, V.; Lehtovaara, L.; Puska, M. Nonadiabatic Ehrenfest Molecular Dynamics within the Projector Augmented-Wave Method. *J. Chem. Phys.* **2012**, *136*, 144103.

(115) Tavernelli, I.; Röhrig, U. F.; Rothlisberger, U. Molecular Dynamics in Electronically Excited States Using Time-Dependent Density Functional Theory. *Mol. Phys.* **2005**, *103*, 963−981.

(116) (a) De Giovannini, U.; Brunetto, G.; Castro, A.; Walkenhorst, J.; Rubio, A. Simulating Pump-Probe Photoelectron and Absorption Spectroscopy on the Attosecond Timescale with Time-Dependent Density Functional Theory. *ChemPhysChem* **2013**, *14*, 1363−1376. (b) Miyamoto, Y.; Zhang, H.; Rubio, A. Pulse-Induced Non-equilibrium Dynamics of Acetylene inside Carbon Nanotube Studied by an Ab Initio Approach. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 8861−8865. (c) Driscoll, J. A.; Varga, K. Time-Dependent Density-Functional Study of Field Emission from Tipped Carbon Nanotubes. *Phys. Rev. B: Condens. Matter* **2009**, *80*, 245431. (d) Chu, X.; McIntyre, M. Comparison of the Strong-Field Ionization of N 2 and F 2: A Time-Dependent Density-Functional-Theory Study. *Phys. Rev. A* **2011**, *83*, 013409. (e) Bubin, S.; Atkinson, M.; Varga, K.; Xie, X.; Roither, S.; Kartashov, D.; Baltuška, A.; Kitzler, M. Strong Laser-Pulse-

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

Driven Ionization and Coulomb Explosion of Hydrocarbon Molecules. *Phys. Rev. A* **2012**, *86*, 043407.
(117) Woiczikowski, P. B.; Steinbrecher, T.; Kubař, T.; Elstner, M. Nonadiabatic QM/MM Simulations of Fast Charge Transfer in Escherichia Coli DNA Photolyase. *J. Phys. Chem. B* **2011**, *115*, 9846−9863.
(118) López-Tarifa, P.; Gaigeot, M. p.; Vuilleumier, R.; Tavernelli, I.; Alcamí, M.; Martín, F.; Hervé du Penhoat, M. a.; Politis, M. f. Ultrafast Damage Following Radiation-Induced Oxidation of Uracil in Aqueous Solution. *Angew. Chem., Int. Ed.* **2013**, *52*, 3160−3163.
(119) Baer, R.; Livshits, E.; Salzner, U. Tuned Range-Separated Hybrids in Density Functional Theory. *Annu. Rev. Phys. Chem.* **2010**, *61*, 85−109.
(120) Grimme, S. Semiempirical GGA-Type Density Functional Constructed with a Long-Range Dispersion Correction. *J. Comput. Chem.* **2006**, *27*, 1787−1799.
(121) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32*, 1456−1465.
(122) Steinmann, S. N.; Corminboeuf, C. A System-Dependent Density-Based Dispersion Correction. *J. Chem. Theory Comput.* **2010**, *6*, 1990−2001.
(123) Becke, A. D.; Johnson, E. R. A Density-Functional Model of the Dispersion Interaction. *J. Chem. Phys.* **2005**, *123*, 154101.
(124) Tkatchenko, A.; Scheffler, M. Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. *Phys. Rev. Lett.* **2009**, *102*, 073005.
(125) Sato, T.; Nakai, H. Density Functional Method Including Weak Interactions: Dispersion Coefficients Based on the Local Response Approximation. *J. Chem. Phys.* **2009**, *131*, 224104.
(126) Langreth, D. C.; Dion, M.; Rydberg, H.; Schröder, E.; Hyldgaard, P.; Lundqvist, B. I. Van Der Waals Density Functional Theory with Applications. *Int. J. Quantum Chem.* **2005**, *101*, 599−610.
(127) Lee, K.; Murray, É. D.; Kong, L.; Lundqvist, B. I.; Langreth, D. C. Higher-Accuracy Van Der Waals Density Functional. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2010**, *82*, 081101.
(128) Vydrov, O. A.; Van Voorhis, T. Nonlocal Van Der Waals Density Functional: The Simpler the Better. *J. Chem. Phys.* **2010**, *133*, 244103.
(129) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. Design of Density Functionals by Combining the Method of Constraint Satisfaction with Parametrization for Thermochemistry, Thermochemical Kinetics, and Noncovalent Interactions. *J. Chem. Theory Comput.* **2006**, *2*, 364−382.
(130) Peverati, R.; Truhlar, D. G. Exchange−Correlation Functional with Good Accuracy for Both Structural and Energetic Properties While Depending Only on the Density and Its Gradient. *J. Chem. Theory Comput.* **2012**, *8*, 2310−2319.
(131) Tkatchenko, A.; DiStasio, R. A., Jr.; Car, R.; Scheffler, M. Accurate and Efficient Method for Many-Body Van Der Waals Interactions. *Phys. Rev. Lett.* **2012**, *108*, 236402.
(132) von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. Optimization of Effective Atom Centered Potentials for London Dispersion Forces in Density Functional Theory. *Phys. Rev. Lett.* **2004**, *93*, 153004.
(133) Levine, B. G.; Ko, C.; Quenneville, J.; Martínez, T. J. Conical Intersections and Double Excitations in Time-Dependent Density Functional Theory. *Mol. Phys.* **2006**, *104*, 1039−1051.
(134) Huix-Rotllant, M.; Natarajan, B.; Ipatov, A.; Muhavini Wawire, C.; Deutsch, T.; Casida, M. E. Assessment of Noncollinear Spin-Flip Tamm−Dancoff Approximation Time-Dependent Density-Functional Theory for the Photochemical Ring-Opening of Oxirane. *Phys. Chem. Chem. Phys.* **2010**, *12*, 12811−12825.
(135) Xu, X.; Yang, K. R.; Truhlar, D. G. Testing Noncollinear Spin-Flip, Collinear Spin-Flip, and Conventional Time-Dependent Density Functional Theory for Predicting Electronic Excitation Energies of Closed-Shell Atoms. *J. Chem. Theory Comput.* **2014**, *10*, 2070−2084.
(136) Li, S. L.; Marenich, A. V.; Xu, X.; Truhlar, D. G. Configuration Interaction-Corrected Tamm−Dancoff Approximation: A Time-Dependent Density Functional Method with the Correct Dimensionality of Conical Intersections. *J. Phys. Chem. Lett.* **2014**, *5*, 322−328.
(137) York, D. M.; Lee, T.-S.; Yang, W. Quantum Mechanical Treatment of Biological Macromolecules in Solution Using Linear-Scaling Electronic Structure Methods. *Phys. Rev. Lett.* **1998**, *80*, 5011.
(138) Sulpizi, M.; Raugei, S.; VandeVondele, J.; Carloni, P.; Sprik, M. Calculation of Redox Properties: Understanding Short- and Long-Range Effects in Rubredoxin. *J. Phys. Chem. B* **2007**, *111*, 3969−3976.
(139) Canfield, P.; Dahlbom, M. G.; Hush, N. S.; Reimers, J. R. Density-Functional Geometry Optimization of the 150,000-Atom Photosystem-I Trimer. *J. Chem. Phys.* **2006**, *124*, 024301.
(140) Siegbahn, P. E.; Blomberg, M. R. Density Functional Theory of Biologically Relevant Metal Centers. *Annu. Rev. Phys. Chem.* **1999**, *50*, 221−249.
(141) Sham, Y. Y.; Chu, Z. T.; Warshel, A. Consistent Calculations of pKA's of Ionizable Residues in Proteins: Semi-Microscopic and Microscopic Approaches. *J. Phys. Chem. B* **1997**, *101*, 4458−4472.
(142) Kohn, W. Density Functional and Density Matrix Method Scaling Linearly with the Number of Atoms. *Phys. Rev. Lett.* **1996**, *76*, 3168−3171.
(143) (a) Warshel, A. Computer Simulations of Enzyme Catalysis: Methods, Progress, and Insights. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425−443. (b) Lin, H.; Truhlar, D. G. QM/MM: What Have We Learned, Where Are We, and Where Do We Go from Here? *Theor. Chem. Acc.* **2007**, *117*, 185−199. (c) Senn, H. M.; Thiel, W. QM/MM Methods for Biological Systems. *Atomistic Approaches in Modern Biology*; Springer: New York, 2007; pp 173−290. (d) van der Kamp, M. W.; Mulholland, A. J. Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology. *Biochemistry* **2013**, *52*, 2708−2728. (e) Vidossich, P.; Magistrato, A. QM/MM Molecular Dynamics Studies of Metal Binding Proteins. *Biomolecules* **2014**, *4*, 616−645.
(144) Warshel, A.; Karplus, M. Calculation of Ground and Excited State Potential Surfaces of Conjugated Molecules. I. Formulation and Parametrization. *J. Am. Chem. Soc.* **1972**, *94*, 5612−5625.
(145) Birge, R. R.; Sullivan, M. J.; Kohler, B. E. The Effect of Temperature and Solvent Environment on the Conformational Stability of 11-Cis-Retinal. *J. Am. Chem. Soc.* **1976**, *98*, 358−367.
(146) Singh, U. C.; Kollman, P. A. A Combined Ab Initio Quantum Mechanical and Molecular Mechanical Method for Carrying out Simulations on Complex Molecular Systems: Applications to the $CH3Cl^{+}$ Cl^{-} Exchange Reaction and Gas Phase Protonation of Polyethers. *J. Comput. Chem.* **1986**, *7*, 718−730.
(147) Bash, P. A.; Field, M. J.; Karplus, M. Free Energy Perturbation Method for Chemical Reactions in the Condensed Phase: A Dynamic Approach Based on a Combined Quantum and Molecular Mechanics Potential. *J. Am. Chem. Soc.* **1987**, *109*, 8092−8094.
(148) Bakowies, D.; Thiel, W. Hybrid Models for Combined Quantum Mechanical and Molecular Mechanical Approaches. *J. Phys. Chem.* **1996**, *100*, 10580−10594.
(149) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. Oniom: A Multilayered Integrated Mo+ MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels-Alder Reactions and $Pt(P(t\text{-}Bu)_3)^{2+}$ H_2 Oxidative Addition. *J. Phys. Chem.* **1996**, *100*, 19357−19363.
(150) Maseras, F.; Morokuma, K. Imomm: A New Integrated Ab Initio+ Molecular Mechanics Geometry Optimization Scheme of Equilibrium Structures and Transition States. *J. Comput. Chem.* **1995**, *16*, 1170−1179.
(151) Wesolowski, T. A.; Warshel, A. Frozen Density Functional Approach for Ab Initio Calculations of Solvated Molecules. *J. Phys. Chem.* **1993**, *97*, 8050−8053.
(152) (a) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. The Effective Fragment Potential Method: A QM-Based MM Approach to Modeling Environmental Effects in Chemistry. *J. Phys. Chem. A* **2001**, *105*, 293−307. (b) Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. An Effective Fragment Method

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

for Modeling Solvent Effects in Quantum Mechanical Calculations. *J. Chem. Phys.* **1996**, *105*, 1968−1986.
(153) Louie, S. G.; Froyen, S.; Cohen, M. L. Nonlinear Ionic Pseudopotentials in Spin-Density-Functional Calculations. *Phys. Rev. B: Condens. Matter* **1982**, *26*, 1738−1742.
(154) Price, S. L.; Stone, A. J. Electrostatic Models for Polypeptides: Can We Assume Transferability? *J. Chem. Soc., Faraday Trans.* **1992**, *88*, 1755−1763.
(155) Laio, A.; VandeVondele, J.; Rothlisberger, U. A Hamiltonian Electrostatic Coupling Scheme for Hybrid Car−Parrinello Molecular Dynamics Simulations. *J. Chem. Phys.* **2002**, *116*, 6941−6947.
(156) Dal Corso, A.; Baroni, S.; Resta, R.; de Gironcoli, S. Ab Initio Calculation of Phonon Dispersions in Ii-Vi Semiconductors. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1993**, *47*, 3588.
(157) von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. Variational Optimization of Effective Atom Centered Potentials for Molecular Properties. *J. Chem. Phys.* **2005**, *122*, 14113.
(158) Zhang, Y.; Lee, T.-S.; Yang, W. A Pseudobond Approach to Combining Quantum Mechanical and Molecular Mechanical Methods. *J. Chem. Phys.* **1999**, *110*, 46−54.
(159) Antes, I.; Thiel, W. Adjusted Connection Atoms for Combined Quantum Mechanical and Molecular Mechanical Methods. *J. Phys. Chem. A* **1999**, *103*, 9290−9295.
(160) Assfeld, X.; Rivail, J.-L. Quantum Chemical Computations on Parts of Large Molecules: The Ab Initio Local Self Consistent Field Method. *Chem. Phys. Lett.* **1996**, *263*, 100−106.
(161) Zhang, Y.; Lin, H.; Truhlar, D. G. Self-Consistent Polarization of the Boundary in the Redistributed Charge and Dipole Scheme for Combined Quantum-Mechanical and Molecular-Mechanical Calculations. *J. Chem. Theory Comput.* **2007**, *3*, 1378−1398.
(162) Sun, Q.; Chan, G. K.-L. Exact and Optimal Quantum Mechanics/Molecular Mechanics Boundaries. *J. Chem. Theory Comput.* **2014**, *10*, 3784−3790.
(163) (a) Sebastiani, D.; Rothlisberger, U. Advances in Density-Functional-Based Modeling Techniques-Recent Extensions of the Car-Parrinello Approach. *Quantum Med. Chem.* **2003**, 1. (b) Hugosson, H. W.; Laio, A.; Maurer, P.; Rothlisberger, U. A Comparative Theoretical Study of Dipeptide Solvation in Water. *J. Comput. Chem.* **2006**, *27*, 672−684.
(164) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(165) Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F. The Gromos Biomolecular Simulation Program Package. *J. Phys. Chem. A* **1999**, *103*, 3596−3607.
(166) Maurer, P.; Laio, A.; Hugosson, H. W.; Colombo, M. C.; Rothlisberger, U. Automated Parametrization of Biomolecular Force Fields from Quantum Mechanics/Molecular Mechanics (QM/MM) Simulations through Force Matching. *J. Chem. Theory Comput.* **2007**, *3*, 628−639.
(167) Laio, A.; Gervasio, F. L.; VandeVondele, J.; Sulpizi, M.; Rothlisberger, U. A Variational Definition of Electrostatic Potential Derived Charges. *J. Phys. Chem. B* **2004**, *108*, 7963−7968.
(168) (a) Solt, I.; Kulhánek, P.; Simon, I.; Winfield, S.; Payne, M. C.; Csányi, G.; Fuxreiter, M. Evaluating Boundary Dependent Errors in QM/MM Simulations. *J. Phys. Chem. B* **2009**, *113*, 5728−5735. (b) Sumowski, C. V.; Ochsenfeld, C. A Convergence Study of QM/MM Isomerization Energies with the Selected Size of the QM Region for Peptidic Systems. *J. Phys. Chem. A* **2009**, *113*, 11734−11741.
(169) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the Opls All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(170) (a) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. Charmm: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187−217. (b) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr. Charmm General Force Field: A Force Field for Drug-Like Molecules Compatible with the Charmm All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2010**, *31*, 671−690.
(171) (a) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E., III; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. Amber, a Package of Computer Programs for Applying Molecular Mechanics, Normal Mode Analysis, Molecular Dynamics and Free Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Comput. Phys. Commun.* **1995**, *91*, 1−41. (b) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber FF99sb Protein Force Field. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1950−1958.
(172) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The Gromos Force-Field Parameter Sets 53a5 and 53a6. *J. Comput. Chem.* **2004**, *25*, 1656−1676.
(173) (a) Rick, S. W.; Stuart, S. J.; Berne, B. J. Dynamical Fluctuating Charge Force Fields: Application to Liquid Water. *arXiv [chem-ph]*; AIP Publishing: New York, 1994; Vol. *101*, pp 6141−6156. (b) Sprik, M.; Klein, M. L. A Polarizable Model for Water Using Distributed Charge Sites. *J. Chem. Phys.* **1988**, *89*, 7556−7560.
(174) Wang, Z.-X.; Zhang, W.; Wu, C.; Lei, H.; Cieplak, P.; Duan, Y. Strike a Balance: Optimization of Backbone Torsion Parameters of Amber Polarizable Force Field for Simulations of Proteins and Peptides. *J. Comput. Chem.* **2006**, *27*, 781−790.
(175) (a) Patel, S.; Brooks, C. L., III. Charmm Fluctuating Charge Force Field for Proteins: I Parameterization and Application to Bulk Organic Liquid Simulations. *J. Comput. Chem.* **2004**, *25*, 1−15. (b) Patel, S.; Mackerell, A. D., Jr.; Brooks, C. L., III. Charmm Fluctuating Charge Force Field for Proteins: Ii Protein/Solvent Properties from Molecular Dynamics Simulations Using a Nonadditive Electrostatic Model. *J. Comput. Chem.* **2004**, *25*, 1504−1514. (c) Lamoureux, G.; Harder, E.; Vorobyov, I. V.; Roux, B.; MacKerell, A. D., Jr. A Polarizable Model of Water for Molecular Dynamics Simulations of Biomolecules. *Chem. Phys. Lett.* **2006**, *418*, 245−249.
(176) (a) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. Development of an Accurate and Robust Polarizable Molecular Mechanics Force Field from Ab Initio Quantum Chemistry. *J. Phys. Chem. A* **2004**, *108*, 621−627. (b) Harder, E.; Kim, B.; Friesner, R. A.; Berne, B. J. Efficient Simulation Method for Polarizable Protein Force Fields: Application to the Simulation of Bpti in Liquid Water. *J. Chem. Theory Comput.* **2005**, *1*, 169−180.
(177) (a) Lamoureux, G.; Roux, B. t. Modeling Induced Polarization with Classical Drude Oscillators: Theory and Molecular Dynamics Simulation Algorithm. *J. Chem. Phys.* **2003**, *119*, 3025−3039. (b) Geerke, D. P.; van Gunsteren, W. F. On the Calculation of Atomic Forces in Classical Simulation Using the Charge-on-Spring Method to Explicitly Treat Electronic Polarization. *J. Chem. Theory Comput.* **2007**, *3*, 2128−2137.
(178) (a) Piquemal, J.-P.; Perera, L.; Cisneros, G. A.; Ren, P.; Pedersen, L. G.; Darden, T. A. Towards Accurate Solvation Dynamics of Divalent Cations in Water Using the Polarizable Amoeba Force Field: From Energetics to Structure. *J. Chem. Phys.* **2006**, *125*, 054511. (b) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A., Jr.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current Status of the Amoeba Polarizable Force Field. *J. Phys. Chem. B* **2010**, *114*, 2549−2564.
(179) Donchev, A. G.; Ozrin, V. D.; Subbotin, M. V.; Tarasov, O. V.; Tarasov, V. I. A Quantum Mechanical Polarizable Force Field for Biomolecular Interactions. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 7829−7834.
(180) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. Anisotropic, Polarizable Molecular Mechanics Studies of Inter- and

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

Intramolecular Interactions and Ligand-Macromolecule Complexes. A Bottom-up Strategy. *J. Chem. Theory Comput.* **2007**, *3*, 1960−1986.
(181) Åstrand, P. O.; Linse, P.; Karlström, G. Molecular Dynamics Study of Water Adopting a Potential Function with Explicit Atomic Dipole Moments and Anisotropic Polarizabilities. *Chem. Phys.* **1995**, *191*, 195−202.
(182) Swart, M.; van Duijnen, P. T. Drf90: A Polarizable Force Field. *Mol. Simul.* **2006**, *32*, 471−484.
(183) Halgren, T. A.; Damm, W. Polarizable Force Fields. *Curr. Opin. Struct. Biol.* **2001**, *11*, 236−242.
(184) Thompson, M. A.; Schenter, G. K. Excited States of the Bacteriochlorophyll B Dimer of Rhodopseudomonas Viridis: A QM/MM Study of the Photosynthetic Reaction Center That Includes MM Polarization. *J. Phys. Chem.* **1995**, *99*, 6374−6386.
(185) Olsen, J. M.; Aidas, K.; Kongsted, J. Excited States in Solution through Polarizable Embedding. *J. Chem. Theory Comput.* **2010**, *6*, 3721−3734.
(186) (a) Bulo, R. E.; Ensing, B.; Sikkema, J.; Visscher, L. Toward a Practical Method for Adaptive QM/MM Simulations. *J. Chem. Theory Comput.* **2009**, *5*, 2212−2221. (b) Heyden, A.; Lin, H.; Truhlar, D. G. Adaptive Partitioning in Combined Quantum Mechanical and Molecular Mechanical Calculations of Potential Energy Functions for Multiscale Simulations. *J. Phys. Chem. B* **2007**, *111*, 2231−2241. (c) Waller, M. P.; Kumbhar, S.; Yang, J. A Density-Based Adaptive Quantum Mechanical/Molecular Mechanical Method. *ChemPhysChem* **2014**, *15*, 3218−3225.
(187) (a) Pezeshki, S.; Lin, H. Adaptive-Partitioning Redistributed Charge and Dipole Schemes for QM/MM Dynamics Simulations: On-the-Fly Relocation of Boundaries That Pass through Covalent Bonds. *J. Chem. Theory Comput.* **2011**, *7*, 3625−3634. (b) Böckmann, M.; Marx, D.; Peter, C.; Site, L. D.; Kremer, K.; Doltsinis, N. L. Multiscale Modelling of Mesoscopic Phenomena Triggered by Quantum Events: Light-Driven Azo-Materials and Beyond. *Phys. Chem. Chem. Phys.* **2011**, *13*, 7604−7621.
(188) Hayes, D. M.; Kollman, P. A. Electrostatic Potentials of Proteins. 2. Role of Electrostatics in a Possible Catalytic Mechanism for Carboxypeptidase A. *J. Am. Chem. Soc.* **1976**, *98*, 7811−7814.
(189) Gao, J.; Xia, X.; Priori, A. Evaluation of Aqueous Polarization Effects through Monte Carlo QM-MM Simulations. *Science* **1992**, *258*, 631−635.
(190) Wanko, M.; Hoffmann, M.; Frauenheim, T.; Elstner, M. Computational Photochemistry of Retinal Proteins. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 511−518.
(191) (a) Brunk, E.; Arey, J. S.; Rothlisberger, U. Role of Environment for Catalysis of the DNA Repair Enzyme Muty. *J. Am. Chem. Soc.* **2012**, *134*, 8608−8616. (b) Hu, P.; Wang, S.; Zhang, Y. Highly Dissociative and Concerted Mechanism for the Nicotinamide Cleavage Reaction in Sir2tm Enzyme Suggested by Ab Initio QM/MM Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **2008**, *130*, 16721−16728.
(192) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science* **2010**, *330*, 341−346.
(193) Zhang, Y.; Kua, J.; McCammon, J. A. Influence of Structural Fluctuation on Enzyme Reaction Energy Barriers in Combined Quantum Mechanical/Molecular Mechanical Studies. *J. Phys. Chem. B* **2003**, *107*, 4459−4463.
(194) Lodola, A.; Mor, M.; Zurek, J.; Tarzia, G.; Piomelli, D.; Harvey, J. N.; Mulholland, A. J. Conformational Effects in Enzyme Catalysis: Reaction Via a High Energy Conformation in Fatty Acid Amide Hydrolase. *Biophys. J.* **2007**, *92*, L20−2.
(195) Carnevale, V.; Raugei, S.; Micheletti, C.; Carloni, P. Large-Scale Motions and Electrostatic Properties of Furin and HIV-1 Protease. *J. Phys. Chem. A* **2007**, *111*, 12327−12332.
(196) (a) Alcantara, R. E.; Xu, C.; Spiro, T. G.; Guallar, V. A Quantum-Chemical Picture of Hemoglobin Affinity. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 18451−18455. (b) Alexandrova, A. N.; Röthlisberger, D.; Baker, D.; Jorgensen, W. L. Catalytic Mechanism and Performance of Computationally Designed Enzymes for Kemp Elimination. *J. Am. Chem. Soc.* **2008**, *130*, 15907−15915. (c) Tubert-Brohman, I.; Acevedo, O.; Jorgensen, W. L. Elucidation of Hydrolysis Mechanisms for Fatty Acid Amide Hydrolase and Its Lys142ala Variant Via QM/MM Simulations. *J. Am. Chem. Soc.* **2006**, *128*, 16904−16913.
(197) Klähn, M.; Braun-Sand, S.; Rosta, E.; Warshel, A. On Possible Pitfalls in Ab Initio Quantum Mechanics/Molecular Mechanics Minimization Approaches for Studies of Enzymatic Reactions. *J. Phys. Chem. B* **2005**, *109*, 15645−15650.
(198) Hünenberger, P. H. Optimal Charge-Shaping Functions for the Particle−Particle—Particle−Mesh (P3M) Method for Computing Electrostatic Interactions in Molecular Simulations. *J. Chem. Phys.* **2000**, *113*, 10464−10476.
(199) 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.
(200) Laio, A.; VandeVondele, J.; Rothlisberger, U. D-Resp: Dynamically Generated Electrostatic Potential Derived Charges from Quantum Mechanics/Molecular Mechanics Simulations. *J. Phys. Chem. B* **2002**, *106*, 7300−7307.
(201) Yarne, D. A.; Tuckerman, M. E.; Martyna, G. J. A Dual Length Scale Method for Plane-Wave-Based, Simulation Studies of Chemical Systems Modeled Using Mixed Ab Initio/Empirical Force Field Descriptions. *J. Chem. Phys.* **2001**, *115*, 3531−3539.
(202) Laino, T.; Mohamed, F.; Laio, A.; Parrinello, M. An Efficient Real Space Multigrid QM/MM Electrostatic Coupling. *J. Chem. Theory Comput.* **2005**, *1*, 1176−1184.
(203) (a) Hockney, R. W. *The Potential Calculation and Some Applications*, 1970; ntrs.nasa.gov/search.jsp?R=19700060674. (b) Martyna, G. J.; Tuckerman, M. E. A Reciprocal Space Based Method for Treating Long Range Interactions in Ab Initio and Force-Field-Based Calculations in Clusters. *J. Chem. Phys.* **1999**, *110*, 2810−2821.
(204) Berne, B. J.; Straub, J. E. Novel Methods of Sampling Phase Space in the Simulation of Biological Systems. *Curr. Opin. Struct. Biol.* **1997**, *7*, 181−189.
(205) (a) Röthlisberger, U.; Sprik, M.; Hutter, J. Time and Length Scales in Ab Initio Molecular Dynamics. *Bridging Time Scales: Molecular Simulations for the Next Decade*; Springer: New York, 2002; pp 413−442. (b) Dellago, C.; Bolhuis, P. G. Transition Path Sampling and Other Advanced Simulation Techniques for Rare Events. *Advanced Computer Simulation Approaches for Soft Matter Sciences Iii*; Springer: Berlin, Heidelberg, 2009; pp 167−233. (c) Bonella, S.; Meloni, S.; Ciccotti, G. Theory and Methods for Rare Events. *Eur. Phys. J. B* **2012**, *85*, 1−19.
(206) (a) VandeVondele, J.; Rothlisberger, U. Accelerating Rare Reactive Events by Means of a Finite Electronic Temperature. *J. Am. Chem. Soc.* **2002**, *124*, 8163−8171. (b) VandeVondele, J.; Rothlisberger, U. Canonical Adiabatic Free Energy Sampling (CAFES): A Novel Method for the Exploration of Free Energy Surfaces. *J. Phys. Chem. B* **2002**, *106*, 203−208. (c) Abrams, J. B.; Tuckerman, M. E. Efficient and Direct Generation of Multidimensional Free energy Surfaces via Adiabatic Dynamics without Coordinate Transformations. *J. Phys. Chem. B* **2008**, *112*, 15742−15757. (d) Maragliano, L.; Vanden-Eijnden, E. Single-sweep methods for free energy calculations. *J. Chem. Phys.* **2008**, *128*, 184110.
(207) (a) Okamoto, Y., Metropolis Algorithms in Generalized Ensemble. *arXiv preprint cond-mat/0308119*, 2003. (b) Okamoto, Y. Generalized-Ensemble Algorithms: Enhanced Sampling Techniques for Monte Carlo and Molecular Dynamics Simulations. *J. Mol. Graphics Modell.* **2004**, *22*, 425−439.
(208) Tuckerman, M. E.; Martyna, G. J.; Berne, B. J. Molecular Dynamics Algorithm for Condensed Systems with Multiple Time Scales. *J. Chem. Phys.* **1990**, *93*, 1287−1291.
(209) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

(210) Amadei, A.; Linssen, A. B.; Berendsen, H. J. Essential Dynamics of Proteins. *Proteins* **1993**, *17*, 412−425.
(211) Space, B.; Rabitz, H.; Askar, A. Long Time Scale Molecular Dynamics Subspace Integration Method Applied to Anharmonic Crystals and Glasses. *J. Chem. Phys.* **1993**, *99*, 9070−9079.
(212) Sprik, M.; Ciccotti, G. Free Energy from Constrained Molecular Dynamics. *J. Chem. Phys.* **1998**, *109*, 7737−7744.
(213) Elber, R.; Karplus, M. A Method for Determining Reaction Paths in Large Molecules: Application to Myoglobin. *Chem. Phys. Lett.* **1987**, *139*, 375−380.
(214) Olender, R.; Elber, R. Calculation of Classical Trajectories with a Very Large Time Step: Formalism and Numerical Examples. *J. Chem. Phys.* **1996**, *105*, 9299.
(215) Dellago, C.; Bolhuis, P. G.; Csajka, F. S.; Chandler, D. Transition Path Sampling and the Calculation of Rate Constants. *J. Chem. Phys.* **1998**, *108*, 1964−1977.
(216) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.
(217) Bartels, C.; Karplus, M. Multidimensional Adaptive Umbrella Sampling: Applications to Main Chain and Side Chain Peptide Conformations. *J. Comput. Chem.* **1997**, *18*, 1450−1462.
(218) (a) Piela, L.; Kostrowicki, J.; Scheraga, H. A. On the Multiple-Minima Problem in the Conformational Analysis of Molecules: Deformation of the Potential Energy Hypersurface by the Diffusion Equation Method. *J. Phys. Chem.* **1989**, *93*, 3339−3346. (b) Wales, D. J.; Scheraga, H. A. Global Optimization of Clusters, Crystals, and Biomolecules. *Science* **1999**, *285*, 1368−1372.
(219) Huber, T.; Torda, A. E.; van Gunsteren, W. F. Local Elevation: A Method for Improving the Searching Properties of Molecular Dynamics Simulation. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 695−708.
(220) Grubmüller, H. Predicting Slow Structural Transitions in Macromolecular Systems: Conformational Flooding. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **1995**, *52*, 2893−2906.
(221) Voter, A. F. Hyperdynamics: Accelerated Molecular Dynamics of Infrequent Events. *Phys. Rev. Lett.* **1997**, *78*, 3908−3911.
(222) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *J. Chem. Phys.* **2004**, *120*, 11919−11929.
(223) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562−12566.
(224) Plotnikov, N. V.; Kamerlin, S. C. L.; Warshel, A. Paradynamics: An Effective and Reliable Model for Ab Initio QM/MM Free-Energy Calculations and Related Tasks. *J. Phys. Chem. B* **2011**, *115*, 7950−7962.
(225) Woo, T. K.; Margl, P.; Blöchl, P. E.; Ziegler, T. Sampling Phase Space by a Combined QM/MM Ab Initio Car-Parrinello Molecular Dynamics Method with Different (Multiple) Time Steps in the Quantum Mechanical (QM) and Molecular Mechanical (MM) Domains. *J. Phys. Chem. A* **2002**, *106*, 1173−1182.
(226) Piana, S.; Bucher, D.; Carloni, P.; Rothlisberger, U. Reaction Mechanism of HIV-1 Protease by Hybrid Car-Parrinello/Classical MD Simulations. *J. Phys. Chem. B* **2004**, *108*, 11139−11149.
(227) VandeVondele, J.; Rothlisberger, U. Efficient Multidimensional Free Energy Calculations for Ab Initio Molecular Dynamics Using Classical Bias Potentials. *J. Chem. Phys.* **2000**, *113*, 4863−4868.
(228) Guidoni, L.; Rothlisberger, U. Scanning Reactive Pathways with Orbital Biased Molecular Dynamics. *J. Chem. Theory Comput.* **2005**, *1*, 554−560.
(229) Ciccotti, G.; Ferrario, M. Rare Events by Constrained Molecular Dynamics. *J. Mol. Liq.* **2000**, *89*, 1−18.
(230) Chipot, C.; Pohorille, A. *Free Energy Calculations: Theory and Applications in Chemistry and Biology*; Springer: New York, 2007; Vol. *86*.
(231) (a) Kollman, P. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395−2417. (b) Kollman, P. A. Advances and Continuing Challenges in Achieving Realistic and Predictive Simulations of the Properties of Organic and Biological Molecules. *Acc. Chem. Res.* **1996**, *29*, 461−469. (c) Simonson, T.; Archontis, G.; Karplus, M. Free Energy Simulations Come of Age: Protein-Ligand Recognition. *Acc. Chem. Res.* **2002**, *35*, 430−437.
(232) Li, G.; Cui, Q. P K a Calculations with QM/MM Free Energy Perturbations. *J. Phys. Chem. B* **2003**, *107*, 14521−14528.
(233) (a) Li, G.; Zhang, X.; Cui, Q. Free Energy Perturbation Calculations with Combined QM/MM Potentials Complications, Simplifications, and Applications to Redox Potential Calculations. *J. Phys. Chem. B* **2003**, *107*, 8643−8653. (b) Min, D.; Zheng, L.; Harris, W.; Chen, M.; Lv, C.; Yang, W. Practically Efficient QM/MM Alchemical Free Energy Simulations: The Orthogonal Space Random Walk Strategy. *J. Chem. Theory Comput.* **2010**, *6*, 2253−2266.
(234) Hu, H.; Yang, W. Dual-Topology/Dual-Coordinate Free-Energy Simulation Using QM/MM Force Field. *J. Chem. Phys.* **2005**, *123*, 041102.
(235) Wesolowski, T.; Warshel, A. Ab Initio Free Energy Perturbation Calculations of Solvation Free Energy Using the Frozen Density Functional Approach. *J. Phys. Chem.* **1994**, *98*, 5183−5187.
(236) (a) Zheng, L.; Chen, M.; Yang, W. Random Walk in Orthogonal Space to Achieve Efficient Free-Energy Simulation of Complex Systems. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 20227−20232. (b) Abrams, J. B.; Rosso, L.; Tuckerman, M. E. Efficient and Precise Solvation Free Energies Via Alchemical Adiabatic Molecular Dynamics. *J. Chem. Phys.* **2006**, *125*, 074115. (c) Christ, C. D.; van Gunsteren, W. F. Enveloping Distribution Sampling: A Method to Calculate Free Energy Differences from a Single Simulation. *J. Chem. Phys.* **2007**, *126*, 184110. (d) Bitetti-Putzer, R.; Yang, W.; Karplus, M. Generalized Ensembles Serve to Improve the Convergence of Free Energy Simulations. *Chem. Phys. Lett.* **2003**, *377*, 633−641. (e) Li, H.; Fajer, M.; Yang, W. Simulated Scaling Method for Localized Enhanced Sampling and Simultaneous "Alchemical" Free Energy Simulations: A General Method for Molecular Mechanical, Quantum Mechanical, and Quantum Mechanical/Molecular Mechanical Simulations. *J. Chem. Phys.* **2007**, *126*, 024106. (f) Knight, J. L.; Brooks, C. L., III. Lambda-Dynamics Free Energy Simulation Methods. *J. Comput. Chem.* **2009**, *30*, 1692−1700.
(237) Laio, A.; Gervasio, F. L. Metadynamics: A Method to Simulate Rare Events and Reconstruct the Free Energy in Biophysics, Chemistry and Material Science. *Rep. Prog. Phys.* **2008**, *71*, 126601.
(238) (a) Blumberger, J.; Ensing, B.; Klein, M. L. Formamide Hydrolysis in Alkaline Aqueous Solution: Insight from Ab Initio Metadynamics Calculations. *Angew. Chem., Int. Ed.* **2006**, *45*, 2893−2897. (b) Fiorin, G.; Pastore, A.; Carloni, P.; Parrinello, M. Using Metadynamics to Understand the Mechanism of Calmodulin/Target Recognition at Atomic Detail. *Biophys. J.* **2006**, *91*, 2768−2777. (c) Oganov, A. R.; Martoňák, R.; Laio, A.; Raiteri, P.; Parrinello, M. Anisotropy of Earth's D-Layer and Stacking Faults in the $MgSiO_3$ Post-Perovskite Phase. *Nature* **2005**, *438*, 1142−1144. (d) Rodríguez-Fortea, A.; Iannuzzi, M.; Parrinello, M. Ab Initio Molecular Dynamics Study of Heterogeneous Oxidation of Graphite by Means of Gas-Phase Nitric Acid. *J. Phys. Chem. B* **2006**, *110*, 3477−3484.
(239) Micheletti, C.; Laio, A.; Parrinello, M. Reconstructing the Density of States by History-Dependent Metadynamics. *Phys. Rev. Lett.* **2004**, *92*, 170601.
(240) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A. Plumed: A Portable Plugin for Free-Energy Calculations with Molecular Dynamics. *Comput. Phys. Commun.* **2009**, *180*, 1961−1972.
(241) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100*, 020603.
(242) Raiteri, P.; Laio, A.; Gervasio, F. L.; Micheletti, C.; Parrinello, M. Efficient Reconstruction of Complex Free Energy Landscapes by Multiple Walkers Metadynamics. *J. Phys. Chem. B* **2006**, *110*, 3533−3539.
(243) Smiatek, J.; Heuer, A. Calculation of Free Energy Landscapes: A Histogram Reweighted Metadynamics Approach. *J. Comput. Chem.* **2011**, *32*, 2084−2096.

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

(244) (a) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *WIREs Comput. Mol. Sci.* **2011**, *1*, 826−843. (b) Ensing, B.; De Vivo, M.; Liu, Z.; Moore, P.; Klein, M. L. Metadynamics as a Tool for Exploring Free Energy Landscapes of Chemical Reactions. *Acc. Chem. Res.* **2006**, *39*, 73−81.

(245) (a) Zhang, Y.; Yan, S.; Yao, L. Mechanism of the Humicola Insolens Cel7b E197S Mutant Catalyzed Flavonoid Glycosides Synthesis: A QM/MM Metadynamics Simulation Study. *Theor. Chem. Acc.* **2013**, *132*, 1−10. (b) Petersen, L.; Ardèvol, A.; Rovira, C.; Reilly, P. J. Molecular Mechanism of the Glycosylation Step Catalyzed by Golgi Alpha-Mannosidase Ii: A QM/MM Metadynamics Investigation. *J. Am. Chem. Soc.* **2010**, *132*, 8291−8300. (c) Cantu, D. C. *Fatty Acid and Polyketide Synthesis Enzymes: Sequences, Structures, and Mechanisms*; Iowa State University: IA, 2013. (d) Biarnés, X.; Ardèvol, A.; Iglesias-Fernández, J.; Planas, A.; Rovira, C. Catalytic Itinerary in 1,3−1,4-B-Glucanase Unraveled by QM/MM Metadynamics. Charge Is Not yet Fully Developed at the Oxocarbenium Ion-Like Transition State. *J. Am. Chem. Soc.* **2011**, *133*, 20301−20309. (e) Bucher, D.; Sandala, G. M.; Durbeej, B.; Radom, L.; Smith, D. M. The Elusive 5′-Deoxyadenosyl Radical in Coenzyme-B12-Mediated Reactions. *J. Am. Chem. Soc.* **2012**, *134*, 1591−1599.

(246) Izvekov, S.; Parrinello, M.; Burnham, C. J.; Voth, G. A. Effective Force Fields for Condensed Phase Systems from Ab Initio Molecular Dynamics Simulation: A New Method for Force-Matching. *J. Chem. Phys.* **2004**, *120*, 10896−10913.

(247) (a) Ercolessi, F.; Adams, J. B. Interatomic Potentials from First-Principles Calculations: The Force-Matching Method. *arXiv [cond-mat]*; IOP Publishing: PA, 1993; Vol. *26*, p 583. (b) Doemer, M.; Liberatore, E.; Knaup, J. M.; Tavernelli, I.; Rothlisberger, U. In Situ Parameterisation of SCC-DFTB Repulsive Potentials by Iterative Boltzmann Inversion. *Mol. Phys.* **2013**, *111*, 3595−3607.

(248) Laio, A.; Bernard, S.; Chiarotti, G. L.; Scandolo, S.; Tosatti, E. Physics of Iron at Earth's Core Conditions. *Science* **2000**, *287*, 1027−1030.

(249) Csányi, G.; Albaret, T.; Payne, M. C.; De Vita, A. "Learn on the Fly": A Hybrid Classical and Quantum-Mechanical Molecular Dynamics Simulation. *Phys. Rev. Lett.* **2004**, *93*, 175503.

(250) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The Resp Model. *J. Phys. Chem.* **1993**, *97*, 10269−10280.

(251) (a) Spiegel, K.; Magistrato, A.; Carloni, P.; Reedijk, J.; Klein, M. L. Azole-Bridged Diplatinum Anticancer Compounds. Modulating DNA Flexibility to Escape Repair Mechanism and Avoid Cross Resistance. *J. Phys. Chem. B* **2007**, *111*, 11873−11876. (b) Gossens, C.; Tavernelli, I.; Rothlisberger, U. DNA Structural Distortions Induced by Ruthenium-Arene Anticancer Compounds. *J. Am. Chem. Soc.* **2008**, *130*, 10921−10928.

(252) Doemer, M.; Maurer, P.; Campomanes, P.; Tavernelli, I.; Rothlisberger, U. Generalized QM/MM Force Matching Approach Applied to the 11-Cis Protonated Schiff Base Chromophore of Rhodopsin. *J. Chem. Theory Comput.* **2013**, *10*, 412−422.

(253) Zhou, Y.; Pu, J. Reaction Path-Force Matching: A New Strategy of Fitting Specific Reaction Parameters for Semiempirical Methods in Combined QM/MM Simulations. *J. Chem. Theory Comput.* **2014**, *10*, 3038−3054.

(254) Simpson, A. R.; Dandy, G. C.; Murphy, L. J. Genetic Algorithms Compared to Other Techniques for Pipe Optimization. *J. Water Resour. Planning Manage.* **1994**, *120*, 423−443.

(255) (a) Biarnés, X.; Ardevol, A.; Planas, A.; Rovira, C.; Laio, A.; Parrinello, M. The Conformational Free Energy Landscape of B-D-Glucopyranose. Implications for Substrate Preactivation in B-Glucoside Hydrolases. *J. Am. Chem. Soc.* **2007**, *129*, 10686−10693. (b) De Vivo, M.; Dal Peraro, M.; Klein, M. L. Phosphodiester Cleavage in Ribonuclease H Occurs Via an Associative Two-Metal-Aided Catalytic Mechanism. *J. Am. Chem. Soc.* **2008**, *130*, 10955−10962. (c) Boero, M.; Ikeda, T.; Ito, E.; Terakura, K. Hsc70 ATPase: An Insight into Water Dissociation and Joint Catalytic Role of K^+ and Mg^{2+} Metal Cations in the Hydrolysis Reaction. *J. Am. Chem. Soc.* **2006**, *128*, 16798−16807. (d) Bozkurt, E.; Ashari-Astani, N.; Browning, N.; Brunk, E.; Campomanes, P.; Perez, M. A. S.; Rothlisberger, U. Lessons from Nature: Computational Design of Biomimetic Compounds and Processes. *Chimia* **2014**, *68*, 642−647.

(256) Koenig, S. H.; Brown, R. D. H_2CO_3 as Substrate for Carbonic Anhydrase in the Dehydration of HCO3. *Proc. Natl. Acad. Sci. U.S.A.* **1972**, *69*, 2422−2425.

(257) (a) Khalifah, R. G. The Carbon Dioxide Hydration Activity of Carbonic Anhydrase. I. Stop-Flow Kinetic Studies on the Native Human Isoenzymes B and C. *J. Biol. Chem.* **1971**, *246*, 2561−2573. (b) Lindskog, S. Structure and Mechanism of Carbonic Anhydrase. *Pharmacol. Ther.* **1997**, *74*, 1−20. (c) Badger, M. R.; Price, G. D. The Role of Carbonic Anhydrase in Photosynthesis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1994**, *45*, 369−392.

(258) (a) Aqvist, J.; Warshel, A. Computer Simulation of the Initial Proton Transfer Step in Human Carbonic Anhydrase I. *J. Mol. Biol.* **1992**, *224*, 7−14. (b) Fisher, S. Z.; Maupin, C. M.; Budayova-Spano, M.; Govindasamy, L.; Tu, C.; Agbandje-McKenna, M.; Silverman, D. N.; Voth, G. A.; McKenna, R. Atomic Crystal and Molecular Dynamics Simulation Structures of Human Carbonic Anhydrase Ii: Insights into the Proton Transfer Mechanism. *Biochemistry* **2007**, *46*, 2930−2937. (c) Lu, D.; Voth, G. A. Proton Transfer in the Enzyme Carbonic Anhydrase: An Ab Initio Study. *J. Am. Chem. Soc.* **1998**, *120*, 4006−4014.

(259) (a) Nair, S. K.; Christianson, D. W. Unexpected Ph-Dependent Conformation of His-64, the Proton Shuttle of Carbonic Anhydrase Ii. *J. Am. Chem. Soc.* **1991**, *113*, 9455−9458. (b) Riccardi, D.; König, P.; Guo, H.; Cui, Q. Proton Transfer in Carbonic Anhydrase Is Controlled by Electrostatics Rather Than the Orientation of the Acceptor. *Biochemistry* **2008**, *47*, 2369−2378. (c) Toba, S.; Colombo, G.; Merz, K. M. Solvent Dynamics and Mechanism of Proton Transfer in Human Carbonic Anhydrase Ii. *J. Am. Chem. Soc.* **1999**, *121*, 2290−2302.

(260) Riccardi, D.; Yang, S.; Cui, Q. Proton Transfer Function of Carbonic Anhydrase: Insights from QM/MM Simulations. *Biochim. Biophys. Acta* **2010**, *1804*, 342−351.

(261) Agmon, N. The Grotthuss Mechanism. *Chem. Phys. Lett.* **1995**, *244*, 456−462.

(262) Bräuer, M.; Pérez-Lustres, J. L.; Weston, J.; Anders, E. Quantitative Reactivity Model for the Hydration of Carbon Dioxide by Biomimetic Zinc Complexes. *Inorg. Chem.* **2002**, *41*, 1454−1463.

(263) Rétey, J. Enzymic Reaction Selectivity by Negative Catalysis or How Do Enzymes Deal with Highly Reactive Intermediates? *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 355−361.

(264) Banerjee, R.; Ragsdale, S. W. The Many Faces of Vitamin B12: Catalysis by Cobalamin-Dependent Enzymes. *Annu. Rev. Biochem.* **2003**, *72*, 209−247.

(265) (a) Gruber, K.; Reitzer, R.; Kratky, C. Radical Shuttling in a Protein: Ribose Pseudorotation Controls Alkyl-Radical Transfer in the Coenzyme B12 Dependent Enzyme Glutamate Mutase. *Angew. Chem., Int. Ed.* **2001**, *40*, 3377−3380. (b) Masuda, J.; Shibata, N.; Morimoto, Y.; Toraya, T.; Yasuoka, N. How a Protein Generates a Catalytic Radical from Coenzyme B(12): X-Ray Structure of a Diol-Dehydratase-Adeninylpentylcobalamin Complex. *Structure* **2000**, *8*, 775−788.

(266) Banerjee, R. *Chemistry and Biochemistry of B12*; John Wiley & Sons: New York, 1999.

(267) Khoroshun, D. V.; Warncke, K.; Ke, S.-C.; Musaev, D. G.; Morokuma, K. Internal Degrees of Freedom, Structural Motifs, and Conformational Energetics of the 5′-Deoxyadenosyl Radical: Implications for Function in Adenosylcobalamin-Dependent Enzymes. A Computational Study. *J. Am. Chem. Soc.* **2003**, *125*, 570−579.

(268) Li, X.; Chung, L. W.; Paneth, P.; Morokuma, K. Dft and Oniom(Dft:MM) Studies on Co-C Bond Cleavage and Hydrogen Transfer in B12-Dependent Methylmalonyl-Coa Mutase. Stepwise or Concerted Mechanism? *J. Am. Chem. Soc.* **2009**, *131*, 5115−5125.

(269) Pang, J.; Li, X.; Morokuma, K.; Scrutton, N. S.; Sutcliffe, M. J. Large-Scale Domain Conformational Change Is Coupled to the Activation of the Co-C Bond in the B12-Dependent Enzyme

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

Ornithine 4,5-Aminomutase: A Computational Study. *J. Am. Chem. Soc.* **2012**, *134*, 2367−2377.
(270) Kwiecien, R. A.; Khavrutskii, I. V.; Musaev, D. G.; Morokuma, K.; Banerjee, R.; Paneth, P. Computational Insights into the Mechanism of Radical Generation in B12-Dependent Methylmalonyl-Coa Mutase. *J. Am. Chem. Soc.* **2006**, *128*, 1287−1292.
(271) Brunk, E. C.; Kellett, W. F.; Richards, N. G. J.; Rothlisberger, U. A Mechano-Chemical Switch to Control Radical Intermediates. *Biochemistry* **2014**, *53*, 3830−3838.
(272) (a) Tchou, J.; Grollman, A. P. Repair of DNA Containing the Oxidatively-Damaged Base, 8-Oxoguanine. *Mutat. Res.* **1993**, *299*, 277−287. (b) Tchou, J.; Kasai, H.; Shibutani, S.; Chung, M. H.; Laval, J.; Grollman, A. P.; Nishimura, S. 8-Oxoguanine (8-Hydroxyguanine) DNA Glycosylase and Its Substrate Specificity. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 4690−4694.
(273) Cabrera, M.; Nghiem, Y.; Miller, J. H. Mutm, a Second Mutator Locus in Escherichia Coli That Generates G.C−−T.A Transversions. *J. Bacteriol.* **1988**, *170*, 5405−5407.
(274) Michaels, M. L.; Miller, J. H. The Go System Protects Organisms from the Mutagenic Effect of the Spontaneous Lesion 8-Hydroxyguanine (7, 8-Dihydro-8-Oxoguanine). *J. Bacteriol.* **1992**, *174*, 6321.
(275) Michaels, M. L.; Tchou, J.; Grollman, A. P.; Miller, J. H. A Repair System for 8-Oxo-7, 8-Dihydrodeoxyguanine. *Biochemistry* **1992**, *31*, 10964−10968.
(276) Fromme, J. C.; Banerjee, A.; Huang, S. J.; Verdine, G. L. Structural Basis for Removal of Adenine Mispaired with 8-Oxoguanine by Muty Adenine DNA Glycosylase. *Nature* **2004**, *427*, 652−656.
(277) (a) Berti, P. J.; McCann, J. A. B. Toward a Detailed Understanding of Base Excision Repair Enzymes: Transition State and Mechanistic Analyses of N-Glycoside Hydrolysis and N-Glycoside Transfer. *Chem. Rev.* **2006**, *106*, 506−555. (b) McCann, J. A. B.; Berti, P. J. Transition-State Analysis of the DNA Repair Enzyme Muty. *J. Am. Chem. Soc.* **2008**, *130*, 5789−5797.
(278) (a) Hofvendahl, K.; Hahn-Hägerdal, B. Factors Affecting the Fermentative Lactic Acid Production from Renewable Resources(1). *Enzyme Microb. Technol.* **2000**, *26*, 87−107. (b) Young, A. L. Biotechnology for Food, Energy, and Industrial Products New Opportunities for Bio-Based Products. *Environ. Sci. Pollut. Res.* **2003**, *10*, 273−276.
(279) Brunk, E.; Neri, M.; Tavernelli, I.; Hatzimanikatis, V.; Rothlisberger, U. Integrating Computational Methods to Retrofit Enzymes to Synthetic Pathways. *Biotechnol. Bioeng.* **2012**, *109*, 572−582.
(280) (a) Röthlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S.; Baker, D. Kemp Elimination Catalysts by Computational Enzyme Design. *Nature* **2008**, *453*, 190−195. (b) Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Röthlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Barbas, C. F., III; Hilvert, D.; Houk, K. N.; Stoddard, B. L.; Baker, D. De Novo Computational Design of Retro-Aldol Enzymes. *Science* **2008**, *319*, 1387−1391. (c) Siegel, J. B.; Zanghellini, A.; Lovick, H. M.; Kiss, G.; Lambert, A. R.; St Clair, J. L.; Gallaher, J. L.; Hilvert, D.; Gelb, M. H.; Stoddard, B. L.; Houk, K. N.; Michael, F. E.; Baker, D. Computational Design of an Enzyme Catalyst for a Stereoselective Bimolecular Diels-Alder Reaction. *Science* **2010**, *329*, 309−313.
(281) Andreessen, B.; Steinbüchel, A. Biosynthesis and Biodegradation of 3-Hydroxypropionate-Containing Polyesters. *Appl. Environ. Microbiol.* **2010**, *76*, 4919−4925.
(282) (a) Cameron, D. C.; Suthers, P. F. Production of 3-Hydroxypropionic Acid in Recombinant Organisms. Patent 2001016346 A1, 2001/3/8, 2001. (b) Willke, T.; Vorlop, K. D. Industrial Bioconversion of Renewable Resources as an Alternative to Conventional Chemistry. *Appl. Microbiol. Biotechnol.* **2004**, *66*, 131−142.
(283) (a) Dieckmann, G. R.; McRorie, D. K.; Lear, J. D.; Sharp, K. A.; DeGrado, W. F.; Pecoraro, V. L. The Role of Protonation and Metal Chelation Preferences in Defining the Properties of Mercury-Binding Coiled Coils. *J. Mol. Biol.* **1998**, *280*, 897−912. (b) Farrer, B. T.; Pecoraro, V. L. Hg (Ii) Binding to a Weakly Associated Coiled Coil Nucleates an Encoded Metalloprotein Fold: A Kinetic Analysis. *Proc. Natl. Acad. Sci. U.S.a.* **2003**, *100*, 3760−3765. (c) Tegoni, M.; Yu, F.; Bersellini, M.; Penner-Hahn, J. E.; Pecoraro, V. L. Designing a Functional Type 2 Copper Center That Has Nitrite Reductase Activity within A-Helical Coiled Coils. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 21234−21239.
(284) (a) Betz, S.; Fairman, R.; O'Neil, K.; Lear, J.; Degrado, W. Design of Two-Stranded and Three-Stranded Coiled-Coil Peptides. *Philos. Trans. R. Soc., B.* **1995**, *348*, 81−88. (b) Handel, T. M.; Williams, S. A.; DeGrado, W. F. Metal Ion-Dependent Modulation of the Dynamics of a Designed Protein. *Science* **1993**, *261*, 879−885. (c) Hill, R. B.; DeGrado, W. F. Solution Structure of A2d, a Nativelike De Novo Designed Protein. *J. Am. Chem. Soc.* **1998**, *120*, 1138−1145. (d) Kammerer, R. A.; Steinmetz, M. O. De Novo Design of a Two-Stranded Coiled-Coil Switch Peptide. *J. Struct. Biol.* **2006**, *155*, 146−153. (e) Lupas, A. Coiled Coils: New Structures and New Functions. *Trends Biochem. Sci.* **1996**, *21*, 375−382. (f) O'Neil, K. T.; DeGrado, W. F. A Thermodynamic Scale for the Helix-Forming Tendencies of the Commonly Occurring Amino Acids. *Science* **1990**, *250*, 646−651.
(285) Zastrow, M. L.; Peacock, A. F. A.; Stuckey, J. A.; Pecoraro, V. L. Hydrolytic Catalysis and Structural Stabilization in a Designed Metalloprotein. *Nat. Chem.* **2012**, *4*, 118−123.
(286) Lee, K. h.; Cabello, C.; Hemmingsen, L.; Marsh, E. N. G.; Pecoraro, V. L. Using Nonnatural Amino Acids to Control Metal-Coordination Number in Three-Stranded Coiled Coils. *Angew. Chem., Int. Ed.* **2006**, *45*, 2864−2868.
(287) Rastelli, G.; Del Rio, A.; Degliesposti, G.; Sgobba, M. Fast and Accurate Predictions of Binding Free Energies Using MM-PBSA and MM-GBSA. *J. Comput. Chem.* **2010**, *31*, 797−810.
(288) Baker, N.; Holst, M.; Wang, F. Adaptive Multilevel Finite Element Solution of the Poisson–Boltzmann Equation Ii. Refinement at Solvent-Accessible Surfaces in Biomolecular Systems. *J. Comput. Chem.* **2000**, *21*, 1343−1352.
(289) Kimura, M.; Nomoto, H.; Masaki, N.; Mori, S. Dye Molecules for Simple Co-Sensitization Process: Fabrication of Mixed-Dye-Sensitized Solar Cells. *Angew. Chem., Int. Ed.* **2012**, *51*, 4371−4374.
(290) O'regan, B.; Grfitzeli, M. A Low-Cost, High-Efficiency Solar Cell Based on Dye-Sensitized. *Nature* **1991**, *353*, 737−740.
(291) Mathew, S.; Yella, A.; Gao, P.; Humphry-Baker, R.; Curchod, B. F. E.; Ashari-Astani, N.; Tavernelli, I.; Rothlisberger, U.; Nazeeruddin, M. K.; Grätzel, M. Dye-Sensitized Solar Cells with 13% Efficiency Achieved through the Molecular Engineering of Porphyrin Sensitizers. *Nat. Chem.* **2014**, *6*, 242−247.
(292) (a) Cavalli, A.; Carloni, P.; Recanatini, M. Target-Related Applications of First Principles Quantum Chemical Methods in Drug Design. *Chem. Rev.* **2006**, *106*, 3497−3519. (b) Binolfi, A.; Rodriguez, E. E.; Valensin, D.; D'Amelio, N.; Ippoliti, E.; Obal, G.; Duran, R.; Magistrato, A.; Pritsch, O.; Zweckstetter, M. Bioinorganic Chemistry of Parkinson's Disease: Structural Determinants for the Copper-Mediated Amyloid Formation of Alpha-Synuclein. *Inorg. Chem.* **2010**, *49*, 10668−10679. (c) Spiegel, K.; Magistrato, A. Modeling Anticancer Drug−DNA Interactions Via Mixed QM/MM Molecular Dynamics Simulations. *Org. Biomol. Chem.* **2006**, *4*, 2507−2517. (d) Dal Peraro, M.; Ruggerone, P.; Raugei, S.; Gervasio, F. L.; Carloni, P. Investigating Biological Systems Using First Principles Car−Parrinello Molecular Dynamics Simulations. *Curr. Opin. Struct. Biol.* **2007**, *17*, 149−156.
(293) (a) Jung, Y.; Lippard, S. J. Direct Cellular Responses to Platinum-Induced DNA Damage. *Chem. Rev.* **2007**, *107*, 1387−1407. (b) Todd, R. C.; Lippard, S. J. Inhibition of Transcription by Platinum Antitumor Compounds. *Metallomics* **2009**, *1*, 280−291.
(294) Kelland, L. The Resurgence of Platinum-Based Cancer Chemotherapy. *Nat. Rev. Cancer* **2007**, *7*, 573−584.
(295) Wu, B.; Davey, G. E.; Nazarov, A. A.; Dyson, P. J.; Davey, C. A. Specific DNA Structural Attributes Modulate Platinum Anticancer Drug Site Selection and Cross-Link Generation. *Nucleic Acids Res.* **2011**, *39*, 8200−8212.

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

(296) Sava, G.; Bergamo, A.; Dyson, P. J. , Metal-Based Antitumour Drugs in the Post-Genomic Era: What Comes Next? *Dalton Trans.* **2011**, *40*, 9069−9075.
(297) Ang, W. H.; Casini, A.; Sava, G.; Dyson, P. J. Organometallic Ruthenium-Based Antitumor Compounds with Novel Modes of Action. *J. Organomet. Chem.* **2011**, *696*, 989−998.
(298) (a) Adhireksan, Z.; Davey, G. E.; Campomanes, P.; Groessl, M.; Clavel, C. M.; Yu, H.; Nazarov, A. A.; Yeo, C. H. F.; Ang, W. H.; Dröge, P. Ligand Substitutions between Ruthenium−Cymene Compounds Can Control Protein Versus DNA Targeting and Anticancer Activity. *Nat. Commun.* **2014**, *5*. (b) Dorcier, A.; Dyson, P. J.; Gossens, C.; Rothlisberger, U.; Scopelliti, R.; Tavernelli, I. Binding of Organometallic Ruthenium(Ii) and Osmium(Ii) Complexes to an Oligonucleotide: A Combined Mass Spectrometric and Theoretical Study. *Organometallics* **2005**, *24*, 2114−2123. (c) Gossens, C.; Tavernelli, I.; Rothlisberger, U. Structural and Energetic Properties of Organometallic Ruthenium (Ii) Diamine Anticancer Compounds and Their Interaction with Nucleobases. *J. Chem. Theory Comput.* **2007**, *3*, 1212−1222. (d) Gossens, C.; Tavernelli, I.; Rothlisberger, U. Binding of Organometallic Ruthenium(Ii) Anticancer Compounds to Nucleobases: A Computational Study. *J. Phys. Chem. A* **2009**, *113*, 11888−11897. (e) Gossens, C.; Dorcier, A.; Dyson, P. J.; et al. pKa Estimation of Ruthenium (Ii)-Arene Pta Complexes and Their Hydrolysis Products Via a Dft/Continuum Electrostatics Approach. *Organometallics* **2007**, *26*, 3969−3975. (f) Gossens, C.; Tavernelli, I.; Rothlisberger, U. Rational Design of Organo-Ruthenium Anticancer Compounds. *Chimia* **2005**, *59*, 81−84. (g) Renfrew, A. K.; Phillips, A. D.; Tapavicza, E.; Scopelliti, R.; Rothlisberger, U.; Dyson, P. J. Tuning the Efficacy of Ruthenium (II)-Arene (RAPTA) Antitumor Compounds with Fluorinated Arene Ligands. *Organometallics* **2009**, *28*, 5061−5071. (h) Scolaro, C.; Geldbach, T. J.; Rochat, S.; Dorcier, A.; Gossens, C.; Bergamo, A.; Cocchietto, M.; Tavernelli, I.; Sava, G.; Rothlisberger, U. Influence of Hydrogen-Bonding Substituents on the Cytotoxicity of Rapta Compounds. *Organometallics* **2006**, *25*, 756−765.
(299) Parker, L. J.; Italiano, L. C.; Morton, C. J.; Hancock, N. C.; Ascher, D. B.; Aitken, J. B.; Harris, H. H.; Campomanes, P.; Rothlisberger, U.; De Luca, A. Studies of Glutathione Transferase P1−1 Bound to a Platinum (IV)-Based Anticancer Compound Reveal the Molecular Basis of Its Activation. *Chem.—Eur. J.* **2011**, *17*, 7806−7816.
(300) (a) Groessl, M.; Zava, O.; Dyson, P. J. Cellular Uptake and Subcellular Distribution of Ruthenium-Based Metallodrugs under Clinical Investigation Versus Cisplatin. *Metallomics* **2011**, *3*, 591−599. (b) Heffeter, P.; Bock, K.; Atil, B.; Reza Hoda, M. A.; Korner, W.; Bartel, C.; Jungwirth, U.; Keppler, B. K.; Micksche, M.; Berger, W.; Koellensperger, G. Intracellular Protein Binding Patterns of the Anticancer Ruthenium Drugs Kp1019 and Kp1339. *J. Biol. Inorg. Chem.* **2010**, *15*, 737−748. (c) Wolters, D. A.; Stefanopoulou, M.; Dyson, P. J.; Groessl, M. Combination of Metallomics and Proteomics to Study the Effects of the Metallodrug Rapta-T on Human Cancer Cells. *Metallomics* **2012**, *4*, 1185−1196. (d) Wu, B.; Ong, M. S.; Groessl, M.; Adhireksan, Z.; Hartinger, C. G.; Dyson, P. J.; Davey, C. A. A Ruthenium Antimetastasis Agent Forms Specific Histone Protein Adducts in the Nucleosome Core. *Chemistry* **2011**, *17*, 3562−3566.
(301) Sebolt-Leopold, J. S.; English, J. M. Mechanisms of Drug Inhibition of Signalling Molecules. *Nature* **2006**, *441*, 457−462.
(302) (a) Aird, R. E.; Cummings, J.; Ritchie, A. A.; Muir, M.; Morris, R. E.; Chen, H.; Sadler, P. J.; Jodrell, D. I. In Vitro and in Vivo Activity and Cross Resistance Profiles of Novel Ruthenium (Ii) Organometallic Arene Complexes in Human Ovarian Cancer. *Br. J. Cancer* **2002**, *86*, 1652−1657. (b) Morris, R. E.; Aird, R. E.; Murdoch, P. d. S.; Chen, H.; Cummings, J.; Hughes, N. D.; Parsons, S.; Parkin, A.; Boyd, G.; Jodrell, D. I.; Sadler, P. J. Inhibition of Cancer Cell Growth by Ruthenium(Ii) Arene Complexes. *J. Med. Chem.* **2001**, *44*, 3616−3621.
(303) Scolaro, C.; Bergamo, A.; Brescacin, L.; Delfino, R.; Cocchietto, M.; Laurenczy, G.; Geldbach, T. J.; Sava, G.; Dyson, P. J. In Vitro and in Vivo Evaluation of Ruthenium(Ii)-Arene Pta Complexes. *J. Med. Chem.* **2005**, *48*, 4161−4171.
(304) Bravaya, K. B.; Khrenova, M. G.; Grigorenko, B. L.; Nemukhin, A. V.; Krylov, A. I. Effect of Protein Environment on Electronically Excited and Ionized States of the Green Fluorescent Protein Chromophore. *J. Phys. Chem. B* **2011**, *115*, 8296−8303.
(305) (a) Sadeghian, K.; Bocola, M.; Schütz, M. A Conclusive Mechanism of the Photoinduced Reaction Cascade in Blue Light Using Flavin Photoreceptors. *J. Am. Chem. Soc.* **2008**, *130*, 12501−12513. (b) Salzmann, S.; Silva-Junior, M. R.; Thiel, W.; Marian, C. M. Influence of the Lov Domain on Low-Lying Excited States of Flavin: A Combined Quantum-Mechanics/Molecular-Mechanics Investigation. *J. Phys. Chem. B* **2009**, *113*, 15610−15618.
(306) Hutter, M. C.; Hughes, J. M.; Reimers, J. R.; Hush, N. S. Modeling the Bacterial Photosynthetic Reaction Center. 2. A Combined Quantum Mechanical/Molecular Mechanical Study of the Structure of the Cofactors in the Reaction Centers of Purple Bacteria. *J. Phys. Chem. B* **1999**, *103*, 4906−4915.
(307) Sproviero, E. M.; Gascón, J. A.; McEvoy, J. P.; Brudvig, G. W.; Batista, V. S. Quantum Mechanics/Molecular Mechanics Study of the Catalytic Cycle of Water Splitting in Photosystem Ii. *J. Am. Chem. Soc.* **2008**, *130*, 3428−3442.
(308) (a) Warshel, A.; Chu, Z. T. Nature of the Surface Crossing Process in Bacteriorhodopsin: Computer Simulations of the Quantum Dynamics of the Primary Photochemical Event. *J. Phys. Chem. B* **2001**, *105*, 9857−9871. (b) Altoè, P.; Cembran, A.; Olivucci, M.; Garavelli, M. Aborted Double Bicycle-Pedal Isomerization with Hydrogen Bond Breaking Is the Primary Event of Bacteriorhodopsin Proton Pumping. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 20172−20177.
(309) Neugebauer, J. Subsystem-Based Theoretical Spectroscopy of Biomolecules and Biomolecular Assemblies. *ChemPhysChem* **2009**, *10*, 3148−3173.
(310) (a) Hahn, S.; Stock, G. Femtosecond Secondary Emission Arising from the Nonadiabatic Photoisomerization in Rhodopsin. *Chem. Phys.* **2000**, *259*, 297−312. (b) Hahn, S.; Stock, G. Quantum-Mechanical Modeling of the Femtosecond Isomerization in Rhodopsin. *J. Phys. Chem. B* **2000**, *104*, 1146−1149. (c) Flores, S. C.; Batista, V. S. Model Study of Coherent-Control of the Femtosecond Primary Event of Vision. *J. Phys. Chem. B* **2004**, *108*, 6745−6749. (d) Lucas, F.; Hornberger, K. Incoherent Control of the Retinal Isomerization in Rhodopsin. *Phys. Rev. Lett.* **2014**, *113*, 058301.
(311) (a) Menon, S. T.; Han, M.; Sakmar, T. P. Rhodopsin: Structural Basis of Molecular Physiology. *Physiol. Rev.* **2001**, *81*, 1659−1688. (b) Okada, T.; Ernst, O. P.; Palczewski, K.; Hofmann, K. P. Activation of Rhodopsin: New Insights from Structural and Biochemical Studies. *Trends Biochem. Sci.* **2001**, *26*, 318−324.
(312) Kim, J. E.; Tauber, M. J.; Mathies, R. A. Wavelength Dependent Cis-Trans Isomerization in Vision. *Biochemistry* **2001**, *40*, 13774−13778.
(313) Schoenlein, R. W.; Peteanu, L. A.; Mathies, R. A.; Shank, C. V. The First Step in Vision: Femtosecond Isomerization of Rhodopsin. *Science* **1991**, *254*, 412−415.
(314) Palings, I.; van den Berg, E. M.; Lugtenburg, J.; Mathies, R. A. Complete Assignment of the Hydrogen out-of-Plane Wagging Vibrations of Bathorhodopsin: Chromophore Structure and Energy Storage in the Primary Photoproduct of Vision. *Biochemistry* **1989**, *28*, 1498−1507.
(315) (a) Kliger, D. S.; Lewis, J. W. Spectral and Kinetic Characterization of Visual Pigment Photointermediates. *Isr. J. Chem.* **1995**, *35*, 289−307. (b) Jäger, S.; Lewis, J. W.; Zvyaga, T. A.; Szundi, I.; Sakmar, T. P.; Kliger, D. S. Chromophore Structural Changes in Rhodopsin from Nanoseconds to Microseconds Following Pigment Photolysis. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 8557−8562.
(316) (a) Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C. A.; Motoshima, H.; Fox, B. A.; Le Trong, I.; Teller, D. C.; Okada, T.; Stenkamp, R. E.; Yamamoto, M.; Miyano, M. Crystal Structure of Rhodopsin: A G Protein-Coupled Receptor. *Science* **2000**, *289*, 739−745. (b) Okada, T.; Sugihara, M.; Bondar, A.-N.; Elstner, M.; Entel, P.; Buss, V. The Retinal Conformation and Its Environment in Rhodopsin in Light of a New 2.2 Å Crystal Structure. *J. Mol. Biol.* **2004**, *342*, 571−583.

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

(317) Choe, H.-W.; Kim, Y. J.; Park, J. H.; Morizumi, T.; Pai, E. F.; Krauss, N.; Hofmann, K. P.; Scheerer, P.; Ernst, O. P. Crystal Structure of Metarhodopsin Ii. *Nature* **2011**, *471*, 651−655.
(318) Deupi, X.; Edwards, P.; Singhal, A.; Nickle, B.; Oprian, D.; Schertler, G.; Standfuss, J. Stabilized G Protein Binding Site in the Structure of Constitutively Active Metarhodopsin-Ii. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 119−124.
(319) (a) Ganter, U. M.; Gärtner, W.; Siebert, F. Rhodopsin-Lumirhodopsin Phototransition of Bovine Rhodopsin Investigated by Fourier Transform Infrared Difference Spectroscopy. *Biochemistry* **1988**, *27*, 7480−7488. (b) Ganter, U. M.; Kashima, T.; Sheves, M.; Siebert, F. Ftir Evidence of an Altered Chromophore-Protein Interaction in the Artificial Visual Pigment Cis-5, 6-Dihydroisorhodopsin and Its Photoproducts Bsi, Lumirhodopsin, and Metarhodopsin-I. *J. Am. Chem. Soc.* **1991**, *113*, 4087−4092. (c) Kandori, H.; Maeda, A. Ftir Spectroscopy Reveals Microscopic Structural Changes of the Protein around the Rhodopsin Chromophore Upon Photoisomerization. *Biochemistry* **1995**, *34*, 14220−14229.
(320) Thomas, Y. G.; Szundi, I.; Lewis, J. W.; Kliger, D. S. Microsecond Time-Resolved Circular Dichroism of Rhodopsin Photointermediates. *Biochemistry* **2009**, *48*, 12283−12289.
(321) (a) Smith, S. O.; Courtin, J.; De Groot, H.; Gebhard, R.; Lugtenburg, J. Carbon-13 Magic-Angle Spinning Nmr Studies of Bathorhodopsin, the Primary Photoproduct of Rhodopsin. *Biochemistry* **1991**, *30*, 7409−7415. (b) Carravetta, M.; Zhao, X.; Johannessen, O. G.; Lai, W. C.; Verhoeven, M. A.; Bovee-Geurts, P. H. M.; Verdegem, P. J. E.; Kiihne, S.; Luthman, H.; de Groot, H. J. M.; deGrip, W. J.; Lugtenburg, J.; Levitt, M. H. Protein-Induced Bonding Perturbation of the Rhodopsin Chromophore Detected by Double-Quantum Solid-State Nmr. *J. Am. Chem. Soc.* **2004**, *126*, 3948−3953. (c) Concistrè, M.; Gansmüller, A.; McLean, N.; Johannessen, O. G.; Marín Montesinos, I.; Bovee-Geurts, P. H. M.; Verdegem, P.; Lugtenburg, J.; Brown, R. C. D.; DeGrip, W. J.; Levitt, M. H. Double-Quantum 13c Nuclear Magnetic Resonance of Bathorhodopsin, the First Photointermediate in Mammalian Vision. *J. Am. Chem. Soc.* **2008**, *130*, 10490−10491.
(322) (a) Palings, I.; Pardoen, J. A.; van den Berg, E.; Winkel, C.; Lugtenburg, J.; Mathies, R. A. Assignment of Fingerprint Vibrations in the Resonance Raman Spectra of Rhodopsin, Isorhodopsin, and Bathorhodopsin: Implications for Chromophore Structure and Environment. *Biochemistry* **1987**, *26*, 2544−2556. (b) Pan, D.; Ganim, Z.; Kim, J. E.; Verhoeven, M. A.; Lugtenburg, J.; Mathies, R. A. Time-Resolved Resonance Raman Analysis of Chromophore Structural Changes in the Formation and Decay of Rhodopsin's Bsi Intermediate. *J. Am. Chem. Soc.* **2002**, *124*, 4857−4864.
(323) Schreiber, M.; Sugihara, M.; Okada, T.; Buss, V. Quantum Mechanical Studies on the Crystallographic Model of Bathorhodopsin. *Angew. Chem., Int. Ed.* **2006**, *45*, 4274−4277.
(324) (a) Röhrig, U. F.; Guidoni, L.; Rothlisberger, U. Early Steps of the Intramolecular Signal Transduction in Rhodopsin Explored by Molecular Dynamics Simulations. *Biochemistry* **2002**, *41*, 10799−10809. (b) Saam, J.; Tajkhorshid, E.; Hayashi, S.; Schulten, K. Molecular Dynamics Investigation of Primary Photoinduced Eventsin the Activation of Rhodopsin. *Biophys. J.* **2002**, *83*, 3097−3112. (c) Neri, M.; Vanni, S.; Tavernelli, I.; Rothlisberger, U. Role of Aggregation in Rhodopsin Signal Transduction. *Biochemistry* **2010**, *49*, 4827−4832.
(325) (a) Röhrig, U. F.; Sebastiani, D. Nmr Chemical Shifts of the Rhodopsin Chromophore in the Dark State and in Bathorhodopsin: A Hybrid QM/MM Molecular Dynamics Study. *J. Phys. Chem. B* **2008**, *112*, 1267−1274. (b) Frähmcke, J. S.; Wanko, M.; Phatak, P.; Mroginski, M. A.; Elstner, M. The Protonation State of Glu181 in Rhodopsin Revisited: Interpretation of Experimental Data on the Basis of QM/MM Calculations. *J. Phys. Chem. B* **2010**, *114*, 11338−11352.
(326) Montero-Cabrera, L. A.; Röhrig, U.; Padrón-Garcia, J. A.; Crespo-Otero, R.; Montero-Alejo, A. L.; Garcia de la Vega, J. M.; Chergui, M.; Rothlisberger, U. CNDOL: A Fast and Reliable Method for the Calculation of Electronic Properties of Very Large Systems. Applications to Retinal Binding Pocket in Rhodopsin and Gas Phase Porphine. *J. Chem. Phys.* **2007**, *127*, 145102.
(327) Send, R.; Kaila, V. R. I.; Sundholm, D. Benchmarking the Approximate Second-Order Coupled-Cluster Method on Biochromophores. *J. Chem. Theory Comput.* **2011**, *7*, 2473−2484.
(328) Ferré, N.; Olivucci, M. Probing the Rhodopsin Cavity with Reduced Retinal Models at the CASPT2//CASSCF/AMBER Level of Theory. *J. Am. Chem. Soc.* **2003**, *125*, 6868−6869.
(329) Coccia, E.; Varsano, D.; Guidoni, L. Protein Field Effect on the Dark State of 11-Cis Retinal in Rhodopsin by Quantum Monte Carlo/Molecular Mechanics. *J. Chem. Theory Comput.* **2012**, *9*, 8−12.
(330) (a) Fujimoto, K.; Hasegawa, J.-Y.; Hayashi, S.; Kato, S.; Nakatsuji, H. Mechanism of Color Tuning in Retinal Protein: Sac-Ci and QM/MM Study. *Chem. Phys. Lett.* **2005**, *414*, 239−242. (b) Coto, P. B.; Strambi, A.; Ferré, N.; Olivucci, M. The Color of Rhodopsins at the Ab Initio Multiconfigurational Perturbation Theory Resolution. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 17154−17159. (c) Altun, A.; Yokoyama, S.; Morokuma, K. Color Tuning in Short Wavelength-Sensitive Human and Mouse Visual Pigments: Ab Initio Quantum Mechanics/Molecular Mechanics Studies. *J. Phys. Chem. A* **2009**, *113*, 11685−11692.
(331) (a) Altun, A.; Yokoyama, S.; Morokuma, K. Spectral Tuning in Visual Pigments: An Oniom(QM:MM) Study on Bovine Rhodopsin and Its Mutants. *J. Phys. Chem. B* **2008**, *112*, 6814−6827. (b) Hernández-Rodríguez, E. W.; Sánchez-García, E.; Crespo-Otero, R.; Montero-Alejo, A. L.; Montero, L. A.; Thiel, W. Understanding Rhodopsin Mutations Linked to the Retinitis Pigmentosa Disease: A QM/MM and DFT/MRCI Study. *J. Phys. Chem. B* **2012**, *116*, 1060−1076.
(332) Rajamani, R.; Lin, Y.-L.; Gao, J. The Opsin Shift and Mechanism of Spectral Tuning in Rhodopsin. *J. Comput. Chem.* **2011**, *32*, 854−865.
(333) Conte, A. M.; Guidoni, L.; Del Sole, R.; Pulci, O. Many-Body Study of the Photoisomerization of the Minimal Model of the Retinal Protonated Schiff Base. *Chem. Phys. Lett.* **2011**, *515*, 290−295.
(334) Szymczak, J. J.; Barbatti, M.; Lischka, H. Is the Photoinduced Isomerization in Retinal Protonated Schiff Bases a Single- or Double-Torsional Process? *J. Phys. Chem. A* **2009**, *113*, 11907−11918.
(335) Hayashi, S.; Tajkhorshid, E.; Schulten, K. Photochemical Reaction Dynamics of the Primary Event of Vision Studied by Means of a Hybrid Molecular Simulation. *Biophys. J.* **2009**, *96*, 403−416.
(336) (a) Polli, D.; Altoè, P.; Weingart, O.; Spillane, K. M.; Manzoni, C.; Brida, D.; Tomasello, G.; Orlandi, G.; Kukura, P.; Mathies, R. A.; Garavelli, M.; Cerullo, G. Conical Intersection Dynamics of the Primary Photoisomerization Event in Vision. *Nature* **2010**, *467*, 440−443. (b) Chung, W. C.; Nanbu, S.; Ishida, T. QM/MM Trajectory Surface Hopping Approach to Photoisomerization of Rhodopsin and Isorhodopsin: The Origin of Faster and More Efficient Isomerization for Rhodopsin. *J. Phys. Chem. B* **2012**, *116*, 8009−8023.
(337) González-Luque, R.; Olaso-González, G.; Merchán, M.; Coto, P. B.; Serrano-Andrés, L.; Garavelli, M. On the Role of the Triplet State in the Cis/Trans Isomerization of Rhodopsin: A CASPT2//CASSCF Study of a Model Chromophore. *Int. J. Quantum Chem.* **2011**, *111*, 3431−3437.
(338) (a) Pecourt, J.-M. L.; Peon, J.; Kohler, B. Ultrafast Internal Conversion of Electronically Excited Rna and DNA Nucleosides in Water. *J. Am. Chem. Soc.* **2000**, *122*, 9348−9349. (b) Schreier, W. J.; Schrader, T. E.; Koller, F. O.; Gilch, P.; Crespo-Hernández, C. E.; Swaminathan, V. N.; Carell, T.; Zinth, W.; Kohler, B. Thymine Dimerization in DNA Is an Ultrafast Photoreaction. *Science* **2007**, *315*, 625−629.
(339) (a) Rupert, C. S.; Goodgal, S. H.; Herriott, R. M. Photoreactivation in Vitro of Ultraviolet-Inactivated Hemophilus Influenzae Transforming Factor. *J. Gen. Physiol.* **1958**, *41*, 451−471. (b) Todo, T.; Takemori, H.; Ryo, H.; Ihara, M.; Matsunaga, T.; Nikaido, O.; Sato, K.; Nomura, T. A New Photoreactivating Enzyme That Specifically Repairs Ultraviolet Light-Induced (6−4)-Photoproducts. *Nature* **1993**, *361*, 371−374. (c) Yarosh, D. B.;

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

Kripke, M. L. DNA Repair and Cytokines in Antimutagenesis and Anticarcinogenesis. *Mutat. Res.* **1996**, *350*, 255−260.
(340) Sobolewski, A. L.; Domcke, W. Computational Studies of the Photophysics of Neutral and Zwitterionic Glycine in an Aqueous Environment: The Glycine−$(H_2O)_2$ Cluster. *Chem. Phys. Lett.* **2008**, *457*, 404−407.
(341) Siemiarczuk, A.; Petersen, C. E.; Ha, C.-E.; Yang, J.; Bhagavan, N. V. Analysis of Tryptophan Fluorescence Lifetimes in a Series of Human Serum Albumin Mutants with Substitutions in Subdomain 2a. *Cell Biochem. Biophys.* **2004**, *40*, 115−122.
(342) Cascella, M.; Cuendet, M. A.; Tavernelli, I.; Rothlisberger, U. Optical Spectra of Cu(II)-Azurin by Hybrid Tddft-Molecular Dynamics Simulations. *J. Phys. Chem. B* **2007**, *111*, 10248−10252.
(343) Cascella, M.; Magistrato, A.; Tavernelli, I.; Carloni, P.; Rothlisberger, U. Role of Protein Frame and Solvent for the Redox Properties of Azurin from Pseudomonas Aeruginosa. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 19641−19646.
(344) du Penhoat, M. A. H.; López-Tarifa, P.; Ghose, K. K.; Jeanvoine, Y.; Gaigeot, M. P.; Vuilleumier, R.; Politis, M. F.; Bacchus-Montabonel, M. C. Modeling Proton-Induced Damage on 2-Deoxy-D-Ribose. Conformational Analysis. *J. Mol. Model.* **2014**, *20*, 2221.
(345) López-Tarifa, P.; Du Penhoat, M.-A. H.; Vuilleumier, R.; Gaigeot, M.-P.; Rothlisberger, U.; Tavernelli, I.; Le Padellec, A.; Champeaux, J.-P.; Alcamí, M.; Moretto-Capelle, P. Time-Dependent Density Functional Theory Molecular Dynamics Simulation of Doubly Charged Uracil in Gas Phase. *Cent. Eur. J. Phys.* **2014**, *12*, 97−102.
(346) Curchod, B. F. E.; Rothlisberger, U.; Tavernelli, I. Trajectory-Based Nonadiabatic Dynamics with Time-Dependent Density Functional Theory. *ChemPhysChem* **2013**, *14*, 1314−1340.
(347) Pearl, J. *Causality: Models, Reasoning and Inference*; Cambridge University Press: New York, 2000; Vol. *29*.
(348) Saeys, Y.; Inza, I.; Larrañaga, P. A Review of Feature Selection Techniques in Bioinformatics. *Bioinformatics* **2007**, *23*, 2507−2517.
(349) Sanders, M. P. A.; McGuire, R.; Roumen, L.; de Esch, I. J. P.; de Vlieg, J.; Klomp, J. P. G.; de Graaf, C. From the Protein's Perspective: The Benefits and Challenges of Protein Structure-Based Pharmacophore Modeling. *Med. Chem. Res.* **2011**, *3*, 28−38.
(350) Kampf, G.; Kapinos, L. E.; Griesser, R.; Lippert, B.; Sigel, H. Comparison of the Acid−Base Properties of Purine Derivatives in Aqueous Solution. Determination of Intrinsic Proton Affinities of Various Basic Sites. *J. Chem. Soc., Perkin Trans. 2* **2002**, 1320−1327.
(351) Francis, A. W.; Helquist, S. A.; Kool, E. T.; David, S. S. Probing the Requirements for Recognition and Catalysis in Fpg and Muty with Nonpolar Adenine Isosteres. *J. Am. Chem. Soc.* **2003**, *125*, 16235−16242.
(352) (a) Guan, Y.; Manuel, R. C.; Arvai, A. S.; Parikh, S. S.; Mol, C. D.; Miller, J. H.; Lloyd, S.; Tainer, J. A. Muty Catalytic Core, Mutant and Bound Adenine Structures Define Specificity for DNA Repair Enzyme Superfamily. *Nat. Struct. Biol.* **1998**, *5*, 1058−1064. (b) Manuel, R. C.; Hitomi, K.; Arvai, A. S.; House, P. G.; Kurtz, A. J.; Dodson, M. L.; McCullough, A. K.; Tainer, J. A.; Lloyd, R. S. Reaction Intermediates in the Catalytic Mechanism of Escherichia Coli Muty DNA Glycosylase. *J. Biol. Chem.* **2004**, *279*, 46930−46939.
(353) Hall, M. A. Correlation-Based Feature Selection for Machine Learning. Ph.D. Thesis, The University of Waikato, 1999.
(354) Hoyer, P. O.; Hyvarinen, A.; Scheines, R.; Spirtes, P. L.; Ramsey, J.; Lacerda, G.; Shimizu, S. Causal Discovery of Linear Acyclic Models with Arbitrary Distributions. *arXiv [stat.ML]*, 2012.
(355) Alderson, R. G.; De Ferrari, L.; Mavridis, L.; McDonagh, J. L.; Mitchell, J. B. O.; Nath, N. Enzyme Informatics. *Curr. Top. Med. Chem.* **2012**, *12*, 1911−1923.
(356) (a) Brown, D. R.; Qin, K.; Herms, J. W.; Madlung, A.; Manson, J.; Strome, R.; Fraser, P. E.; Kruck, T.; von Bohlen, A.; Schulz-Schaeffer, W.; Giese, A.; Westaway, D.; Kretzschmar, H. The Cellular Prion Protein Binds Copper in Vivo. *Nature* **1997**, *390*, 684−687. (b) Prusiner, S. B. Prions. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 13363−13383.
(357) (a) Brown, D. R.; Wong, B. S.; Hafiz, F.; Clive, C.; Haswell, S. J.; Jones, I. M. Normal Prion Protein Has an Activity Like That of Superoxide Dismutase. *Biochem. J.* **1999**, *344*, 1−5. (b) Viles, J. H.; Cohen, F. E.; Prusiner, S. B.; Goodin, D. B.; Wright, P. E.; Dyson, H. J. Copper Binding to the Prion Protein: Structural Implications of Four Identical Cooperative Binding Sites. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 2042−2047. (c) Aronoff-Spencer, E.; Burns, C. S.; Avdievich, N. I.; Gerfen, G. J.; Peisach, J.; Antholine, W. E.; Ball, H. L.; Cohen, F. E.; Prusiner, S. B.; Millhauser, G. L. Identification of the Cu2+ Binding Sites in the N-Terminal Domain of the Prion Protein by Epr and Cd Spectroscopy. *Biochemistry* **2000**, *39*, 13760−13771. (d) Burns, C. S.; Aronoff-Spencer, E.; Dunham, C. M.; Lario, P.; Avdievich, N. I.; Antholine, W. E.; Olmstead, M. M.; Vrielink, A.; Gerfen, G. J.; Peisach, J.; Scott, W. G.; Millhauser, G. L. Molecular Features of the Copper Binding Sites in the Octarepeat Domain of the Prion Protein. *Biochemistry* **2002**, *41*, 3991−4001. (e) Miura, T.; Hori-i, A.; Mototani, H.; Takeuchi, H. Raman Spectroscopic Study on the Copper(II) Binding Mode of Prion Octapeptide and Its Ph Dependence. *Biochemistry* **1999**, *38*, 11560−11569. (f) Garnett, A. P.; Viles, J. H. Copper Binding to the Octarepeats of the Prion Protein. Affinity, Specificity, Folding, and Cooperativity: Insights from Circular Dichroism. *J. Biol. Chem.* **2003**, *278*, 6795−6802. (g) Kramer, M. L.; Kratzin, H. D.; Schmidt, B.; Römer, A.; Windl, O.; Liemann, S.; Hornemann, S.; Kretzschmar, H. Prion Protein Binds Copper within the Physiological Concentration Range. *J. Biol. Chem.* **2001**, *276*, 16711−16719. (h) Morante, S.; González-Iglesias, R.; Potrich, C.; Meneghini, C.; Meyer-Klaucke, W.; Menestrina, G.; Gasset, M. Inter- and Intra-Octarepeat Cu(II) Site Geometries in the Prion Protein: Implications in Cu(II) Binding Cooperativity and Cu(II)-Mediated Assemblies. *J. Biol. Chem.* **2004**, *279*, 11753−11759. (i) Rachidi, W.; Mangé, A.; Senator, A.; Guiraud, P.; Riondel, J.; Benboubetra, M.; Favier, A.; Lehmann, S. Prion Infection Impairs Copper Binding of Cultured Cells. *J. Biol. Chem.* **2003**, *278*, 14595−14598. (j) Rachidi, W.; Vilette, D.; Guiraud, P.; Arlotto, M.; Riondel, J.; Laude, H.; Lehmann, S.; Favier, A. Expression of Prion Protein Increases Cellular Copper Binding and Antioxidant Enzyme Activities but Not Copper Delivery. *J. Biol. Chem.* **2003**, *278*, 9064−9072. (k) Shiraishi, N.; Ohta, Y.; Nishikimi, M. The Octapeptide Repeat Region of Prion Protein Binds Cu(II) in the Redox-Inactive State. *Biochem. Biophys. Res. Commun.* **2000**, *267*, 398−402. (l) Wong, B. S.; Pan, T.; Liu, T.; Li, R.; Gambetti, P.; Sy, M. S. Differential Contribution of Superoxide Dismutase Activity by Prion Protein in Vivo. *Biochem. Biophys. Res. Commun.* **2000**, *273*, 136−139. (m) Cui, T.; Daniels, M.; Wong, B. S.; Li, R.; Sy, M.-S.; Sassoon, J.; Brown, D. R. Mapping the Functional Domain of the Prion Protein. *Eur. J. Biochem.* **2003**, *270*, 3368−3376. (n) Sakudo, A.; Lee, D.-C.; Nishimura, T.; Li, S.; Tsuji, S.; Nakamura, T.; Matsumoto, Y.; Saeki, K.; Itohara, S.; Ikuta, K.; Onodera, T. Octapeptide Repeat Region and N-Terminal Half of Hydrophobic Region of Prion Protein (Prp) Mediate Prp-Dependent Activation of Superoxide Dismutase. *Biochem. Biophys. Res. Commun.* **2005**, *326*, 600−606.
(358) (a) Wong, B. S.; Chen, S. G.; Colucci, M.; Xie, Z.; Pan, T.; Liu, T.; Li, R.; Gambetti, P.; Sy, M. S.; Brown, D. R. Aberrant Metal Binding by Prion Protein in Human Prion Disease. *J. Neurochem.* **2001**, *78*, 1400−1408. (b) Sigurdsson, E. M.; Brown, D. R.; Alim, M. A.; Scholtzova, H.; Carp, R.; Meeker, H. C.; Prelli, F.; Frangione, B.; Wisniewski, T. Copper Chelation Delays the Onset of Prion Disease. *J. Biol. Chem.* **2003**, *278*, 46199−46202.
(359) (a) Requena, J. R.; Groth, D.; Legname, G.; Stadtman, E. R.; Prusiner, S. B.; Levine, R. L. Copper-Catalyzed Oxidation of the Recombinant Sha (29−231) Prion Protein. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 7170−7175. (b) Giese, A.; Levin, J.; Bertsch, U.; Kretzschmar, H. Effect of Metal Ions on De Novo Aggregation of Full-Length Prion Protein. *Biochem. Biophys. Res. Commun.* **2004**, *320*, 1240−1246.
(360) Van Doorslaer, S.; Cereghetti, G. M.; Glockshuber, R.; Schweiger, A. Unraveling the Cu2+ Binding Sites in the C-Terminal Domain of the Murine Prion Protein: A Pulse Epr and Endor Study. *J. Phys. Chem. B* **2001**, *105*, 1631−1639.
(361) (a) Whittal, R. M.; Ball, H. L.; Cohen, F. E.; Burlingame, A. L.; Prusiner, S. B.; Baldwin, M. A. Copper Binding to Octarepeat Peptides

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263

of the Prion Protein Monitored by Mass Spectrometry. *Protein Sci.* **2000**, *9*, 332−343. (b) Burns, C. S.; Aronoff-Spencer, E.; Legname, G.; Prusiner, S. B.; Antholine, W. E.; Gerfen, G. J.; Peisach, J.; Millhauser, G. L. Copper Coordination in the Full-Length, Recombinant Prion Protein. *Biochemistry* **2003**, *42*, 6794−6803. (c) Hasnain, S. S.; Murphy, L. M.; Strange, R. W.; Grossmann, J. G.; Clarke, A. R.; Jackson, G. S.; Collinge, J. Xafs Study of the High-Affinity Copper-Binding Site of Human Prp(91−231) and Its Low-Resolution Structure in Solution. *J. Mol. Biol.* **2001**, *311*, 467−473. (d) Stöckel, J.; Safar, J.; Wallace, A. C.; Cohen, F. E.; Prusiner, S. B. Prion Protein Selectively Binds Copper(II) Ions. *Biochemistry* **1998**, *37*, 7185−7193.

(362) Schellenberger, J.; Park, J. O.; Conrad, T. M.; Palsson, B. Ø. Bigg: A Biochemical Genetic and Genomic Knowledgebase of Large Scale Metabolic Reconstructions. *BMC Bioinf.* **2010**, *11*, 213.

(363) (a) Reed, J. L.; Famili, I.; Thiele, I.; Palsson, B. O. Towards Multidimensional Genome Annotation. *Nat. Rev. Genet.* **2006**, *7*, 130−141. (b) Thiele, I.; Palsson, B. Ø. A Protocol for Generating a High-Quality Genome-Scale Metabolic Reconstruction. *Nat. Protoc.* **2010**, *5*, 93−121.

(364) Orth, J. D.; Thiele, I.; Palsson, B. Ø. What Is Flux Balance Analysis? *Nat. Biotechnol.* **2010**, *28*, 245−248.

(365) Hatzimanikatis, V.; Li, C.; Ionita, J. A.; Henry, C. S.; Jankowski, M. D.; Broadbelt, L. J. Exploring the Diversity of Complex Metabolic Networks. *Bioinformatics* **2005**, *21*, 1603−1609.

(366) (a) Beltrao, P.; Kiel, C.; Serrano, L. Structures in Systems Biology. *Curr. Opin. Struct. Biol.* **2007**, *17*, 378−384. (b) Chang, R. L.; Andrews, K.; Kim, D.; Li, Z.; Godzik, A.; Palsson, B. O. Structural Systems Biology Evaluation of Metabolic Thermotolerance in Escherichia Coli. *Science* **2013**, *340*, 1220−1223. (c) Chang, R. L.; Xie, L.; Bourne, P. E.; Palsson, B. O. Antibacterial Mechanisms Identified through Structural Systems Pharmacology. *BMC Syst. Biol.* **2013**, *7*, 102.

(367) (a) Lesley, S. A.; Kuhn, P.; Godzik, A.; Deacon, A. M.; Mathews, I.; Kreusch, A.; Spraggon, G.; Klock, H. E.; McMullan, D.; Shin, T.; Vincent, J.; Robb, A.; Brinen, L. S.; Miller, M. D.; McPhillips, T. M.; Miller, M. A.; Scheibe, D.; Canaves, J. M.; Guda, C.; Jaroszewski, L.; Selby, T. L.; Elsliger, M.-A.; Wooley, J.; Taylor, S. S.; Hodgson, K. O.; Wilson, I. A.; Schultz, P. G.; Stevens, R. C. Structural Genomics of the Thermotoga Maritima Proteome Implemented in a High-Throughput Structure Determination Pipeline. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11664−11669. (b) Zhang, Y.; Thiele, I.; Weekes, D.; Li, Z.; Jaroszewski, L.; Ginalski, K.; Deacon, A. M.; Wooley, J.; Lesley, S. A.; Wilson, I. A.; Palsson, B.; Osterman, A.; Godzik, A. Three-Dimensional Structural View of the Central Metabolic Network of Thermotoga Maritima. *Science* **2009**, *325*, 1544−1549.

(368) Chang, R. L.; Xie, L.; Xie, L.; Bourne, P. E.; Palsson, B. Ø. Drug Off-Target Effects Predicted Using Structural Analysis in the Context of a Metabolic Network Model. *PLoS Comput. Biol.* **2010**, *6*, e1000938.

(369) (a) Zhang, Q. C.; Petrey, D.; Deng, L.; Qiang, L.; Shi, Y.; Thu, C. A.; Bisikirska, B.; Lefebvre, C.; Accili, D.; Hunter, T.; Maniatis, T.; Califano, A.; Honig, B. Structure-Based Prediction of Protein-Protein Interactions on a Genome-Wide Scale. *Nature* **2012**, *490*, 556−560. (b) Wang, X.; Wei, X.; Thijssen, B.; Das, J.; Lipkin, S. M.; Yu, H. Three-Dimensional Reconstruction of Protein Networks Provides Insight into Human Genetic Disease. *Nat. Biotechnol.* **2012**, *30*, 159−164.

(370) Cheng, T. M. K.; Goehring, L.; Jeffery, L.; Lu, Y.-E.; Hayles, J.; Novák, B.; Bates, P. A. A Structural Systems Biology Approach for Quantifying the Systemic Consequences of Missense Mutations in Proteins. *PLoS Comput. Biol.* **2012**, *8*, e1002738.

(371) (a) Nam, H.; Lewis, N. E.; Lerman, J. A.; Lee, D.-H.; Chang, R. L.; Kim, D.; Palsson, B. O. Network Context and Selection in the Evolution to Enzyme Specificity. *Science* **2012**, *337*, 1101−1104. (b) Notebaart, R. A.; Szappanos, B.; Kintses, B.; Pál, F.; Györkei, Á.; Bogos, B.; Lázár, V.; Spohn, R.; Csörgő, B.; Wagner, A.; Ruppin, E.; Pál, C.; Papp, B. Network-Level Architecture and the Evolutionary Potential of Underground Metabolism. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 11762−11767. (c) Guzmán, G. I.; Utrilla, J.; Nurk, S.; Brunk, E.; Monk, J. M.; Ebrahim, A.; Palsson, B. O.; Feist, A. M. Model-Driven Discovery of Underground Metabolic Functions in Escherichia Coli. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 929−934.

(372) Gerhart, J.; Kirschner, M.; Moderbacher, E. S. *Cells, Embryos, and Evolution: Toward a Cellular and Developmental Understanding of Phenotypic Variation and Evolutionary Adaptability*; Blackwell Science: Malden, MA, 1997.

(373) Jensen, R. A. Enzyme Recruitment in Evolution of New Function. *Annu. Rev. Microbiol.* **1976**, *30*, 409−425.

(374) (a) Jacob, F. Evolution and Tinkering. *Science* **1977**, *196*, 1161−1166. (b) O'Brien, P. J.; Herschlag, D. Catalytic Promiscuity and the Evolution of New Enzymatic Activities. *Chem. Biol.* **1999**, *6*, R91−R105.

(375) Meier, K.; Choutko, A.; Dolenc, J.; Eichenberger, A. P.; Riniker, S.; van Gunsteren, W. F. Multi-Resolution Simulation of Biomolecular Systems: A Review of Methodological Issues. *Angew. Chem., Int. Ed.* **2013**, *52*, 2820−2834.

(376) Ren, Q.; Ranaghan, K. E.; Mulholland, A. J.; Harvey, J. N.; Manby, F. R.; Balint-Kurti, G. G. Optimal Control Design of Laser Pulses for Mode Specific Vibrational Excitation in an Enzyme−Substrate Complex. *Chem. Phys. Lett.* **2010**, *491*, 230−236.

(377) Markram, H. The Blue Brain Project. *Nat. Rev. Neurosci.* **2006**, *7*, 153−160.

DOI: 10.1021/cr500628b
Chem. Rev. 2015, 115, 6217−6263