Topical Perspectives

Challenges in computational studies of enzyme structure, function and dynamics

Alexandra T.P. Carvalho [a], Alexandre Barrozo [a], Dvir Doron [b], Alexandra Vardi Kilshtain [b], Dan Thomas Major [b,*], Shina Caroline Lynn Kamerlin [a,**]

[a] Science for Life Laboratory, Department of Cell and Molecular Biology, Uppsala University, BMC Box 596, S-751 24 Uppsala, Sweden
[b] Department of Chemistry and the Lise Meitner-Minerva Center of Computational Quantum Chemistry Bar-Ilan University, Ramat-Gan 52900, Israel

ARTICLE INFO

ABSTRACT

In this review we give an overview of the field of Computational enzymology. We start by describing the birth of the field, with emphasis on the work of the 2013 chemistry Nobel Laureates. We then present key features of the state-of-the-art in the field, showing what theory, accompanied by experiments, has taught us so far about enzymes. We also briefly describe computational methods, such as quantum mechanics-molecular mechanics approaches, reaction coordinate treatment, and free energy simulation approaches. We finalize by discussing open questions and challenges.

© 2014 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

1. Introduction

Enzymes are the guardians of life, reducing the timescales of the chemical reactions so crucial to biology from millions of years to fractions of seconds [1]. Therefore, to understand how enzymes work is to understand how life works, at the most fundamental, molecular level. In addition to their biological roles, there is also great interest in understanding how to use enzymes as artificial catalysts for a range of processes, from chemical synthesis to the generation of novel biofuels [2]. The history of contemporary enzymology is a long and distinguished one, which, despite its seeming experimental dominance, is nevertheless inseparable from theoretical contributions. The earliest work in enzymology dates back to the mid-to-late 1800s, with studies of fermentation (in fact, the word enzyme literally translates to "in yeast" [3]), followed by Emil Fischer's "lock-and-key" theory of enzyme–ligand binding in 1894 [4]. However, in the absence of any structural data, it was hard to infer what kind of molecules enzymes even *are*, let alone how they actually work. Here, the development of transition state theory (TST) by Eyring and Polanyi [5] proved crucial, as it paved the way for Linus Pauling's brave (for the time) assertion that the tremendous catalytic power of enzymes comes from the highly specific

tight binding of the transition states (TS) during the chemical reaction [6]. This observation is particularly impressive in light of the extremely limited information available about enzymes at the time. A major turning point in enzymology came in 1965, with the publication of the 2 Å resolution crystal structure of hen egg white lysozyme by Phillips and coworkers [7]. This small 14.3 kDa enzyme (mature form, Fig. 1) was the third protein and the first enzyme to have its structure solved [7], and as such, finally allowed an enzyme structure to be examined at atomic resolution, giving birth to structure-function studies of enzymes and allowing us to think of enzyme mechanism in structural terms. Increasing computational power and methodological developments at the time [8–13] meant that it was becoming possible to perform computer simulations on molecular structures. In subsequent years, *in silico* protein folding by minimization with normal mode jumps [14], and the earliest molecular dynamics (MD) simulation on a biological system [15], demonstrated that proteins are not in fact static structures, but rather dynamic entities. With time such protein dynamics have been accepted to be critical for their function. These major experimental and theoretical advances created a huge paradigm shift in the field, and marked the beginning of what is now contemporary enzymology.

Since these early days, computational enzymology has become an invaluable tool for studying enzyme activity. However, despite the many advances in this field, and the corresponding experimental progress [16–19], we still do not have a complete understanding of how enzymes really work. In the present perspective, we begin with a historical overview, showing the birth of computational

* Corresponding author.
** Corresponding author. Tel.: +46 18 471 4423.
  E-mail addresses: majort@biu.ac.il (D.T. Major),
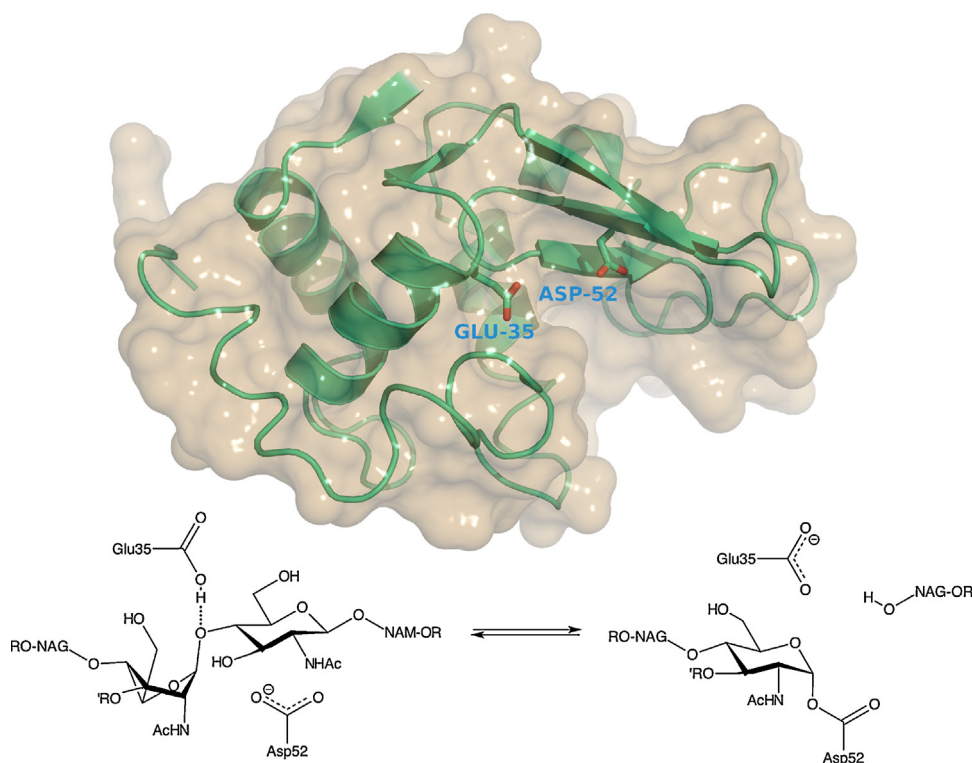kamerlin@icm.uu.se (S.C.L. Kamerlin).

Fig. 1. Overall structure [285] and proposed mechanism [235] of hen egg-white lysozyme. The figure shows two key residues in the cleavage of the ester bond holding two sugar rings. In the first step, the Michaelis complex adopts a skew-boat or distorted envelope conformation. The reaction proceeds through an undistorted covalent intermediate.

enzymology, with an emphasis on the work that provided the basis for awarding the 2013 Chemistry Nobel prize to Karplus, Levitt and Warshel [20–22]. Following this, we provide a concise summary of the current state-of-the-art in the field, showing what experiment and, in particular, theory, has taught us so far about enzymes. We also discuss some of the most popular theoretical techniques to study these biological machines. We finalize by highlighting and discussing open questions and challenges that we believe still need addressing.

2. The birth of a field

Choosing an exact event for the birth of a field is a difficult task. As discussed in the Introduction, a good "time-zero" for Computational Biochemistry could go back as far as the development of the first computer code for molecular mechanics (MM) methods by Allinger et al. [10,11], which was based on previous work on intermolecular potentials. Subsequently, Némethy and Scheraga developed simplified versions of these potentials for use in statistical mechanics simulations [9]. Concurrently, MD simulations (which date back to the early 1950s [23,24]), were gradually increasing in popularity at this point with the advent of the first supercomputers. However, it was still limited to small system sizes pertinent to problems in chemical and theoretical physics, such as the first simulations of liquids, a molten salt and a small organic molecule [25–29]. The first step enabling the leap from chemical physics to computational biology came in the late 1960s, when Lifson and Warshel developed the first consistent force field (CFF) [13]. This was combined with the fact that the Weizmann Institute had recently acquired what was then a powerful machine, aptly named the Golem (after the automaton from Jewish folklore) [30], which allowed them to perform comparatively computationally demanding calculations. While the first of these still, as would be expected, focused on small molecules [12], in 1969, Levitt and

Lifson used CFF to perform the first energy minimization of simple protein structures (myoglobin and lysozyme) [31]. At the time, Karplus, who was an established theoretical chemist in areas ranging from NMR to reaction dynamics, joined the Lifson group at the Weizmann Institute for a six-month leave and the first contact between the Nobel Prize trio was made [32]. In 1970, Warshel and Bromberg then published a QM(VB)+MM study of the oxidation of 4a,4b-dihydrophenanthrene oxidation [33] which is the first published QM+MM simulation of a chemical process and laid the foundation for subsequent QM/MM calculations of biological systems. Note that the key difference between a QM+MM and a QM/MM simulation is that the latter includes treatment of the coupling between the QM and MM regions, whereas the former does not, as also highlighted in [34].

At this stage, further progress was to some extent limited not just by computational power, but also by the lack of available structural data. The Protein Data Bank was only established in 1971 [35], at which point it only contained 7 protein structures. Another limitation was that traditional MM approaches, such as the ones used in the study of myoglobin and lysozyme, do not describe electrons explicitly, and cannot be used to describe bond making-and-breaking processes, making their usefulness for understanding chemical reactivity limited. However, numerous QM models were available at the time that *could* treat electronic structure properties. Thus, the ability to do separate QM and MM calculations existed in these early days, although a QM treatment was only possible for small molecules [33]. Still, there was a tremendous gap that needed to be bridged before it became possible to move to enzymes.

In 1971, Karplus, together with Barry Honig, were working on the retinal chromophore, and employed a hybrid Hamiltonian relying on a Hückel one-electron term for the π-electrons, and a pairwise non-bonded energy function for the sigma bond framework [36]. Subsequently, Warshel joined the Karplus group as a

post-doc, to continue the work on retinal, and in 1972, Warshel and Karplus published a hybrid method that combined more expensive QM methods with the cheaper MM methods [37]. The first iteration was restricted to planar systems, in which π-electrons were described by quantum mechanics, whereas the remaining system was treated classically, with the σ-electrons being incorporated into the nuclei. Further improvements were presented in the classic Warshel and Levitt 1976 paper [38], where these authors described what may be considered to be the first realization of current QM/MM approaches. Importantly, for the first time the crucial effect of the environment (treated at the MM level) on the QM region was included. During these very same years, the Nobel Prize trio engaged in groundbreaking work on protein simulations [14,15,39]. Gelin and Karplus used empirical potentials to study the side-chain dynamics in proteins using a precursor to the well-renowned CHARMM program [40], and, with McCammon, they studied the dynamics in a folded bovine pancreatic trypsin inhibitor protein [15]. Incidentally, this infant version of CHARMM relied heavily on the prior work of the Lifson and Scheraga groups [41], and on Warshel's presence in the Karplus group at the time [32]. Thus, the transition to biology was complete with the first computer simulations of actual biologically relevant problems.

Since these early days of computational biology, tremendous progress has been made over the years. In 1982, Berendsen and coworkers performed the first MD simulation of a membrane [42], followed by a protein in an aqueous crystal [43]. Subsequent work in the Karplus group saw the development of simulated annealing MD simulation approaches for X-ray protein structure refinement [44] and NMR structure determination [45]. At about the same time, the first MD simulation of DNA in solution was performed [46], followed by the first MD simulations of proteins in explicit water, showing the crucial role the solvent plays in stabilizing proteins [47–49]. Later, Warshel developed the empirical valence bond approach (EVB) [50], which is an empirical QM/MM approach based on valence bond theory. EVB has been demonstrated to be an extremely powerful tool for calculating the free energies of enzymatic reactions, as well as provides further insight into enzyme catalysis, as seen not just from several works by Warshel and collaborators [51–54], but also from some of the elegant work by Hammes-Schiffer and coworkers and many others [55–57]. Jorgensen brought further progress to the field with the development and application of the all atom OPLS-AA force field (FF) [58,59], TIPnP water molecules and free energy methods, and field with the implementation of electrostatic embedding and an interface between a FF and a semiempirical program [60]. Further QM/MM specific contributions were made by Mo and Gao [61], Morokuma and coworkers [62], Singh and Kollman [63], as well as Lin and Truhlar [64]. Additionally, these works laid the foundation for the development of the nowadays well-known computer simulation packages such as CHARMM [65,66], Amber [67], GROMACS [68], BOSS [69], NAMD [70], among others, which made these developments available for use by the general scientific community. In the interests of space, not all the excellent contributions can be outlined here, but for more detail we refer the reader to reviews by Senn and Thiel [71], and van der Kamp and Mulholland [72].

Together, these developments changed the overall shape of enzymology, demonstrating that not all experiments have to take place in the wet lab. We note that the Nobel trio were standing on the shoulders of giants – indeed, it is impossible to imagine the blossoming of multiscale modeling without the monumental contributions of Kohn and Pople to computational chemistry, for which they received the 1998 Nobel Prize in chemistry [19]. To help provide a chronological perspective for the reader, a crude timeline of some of these key events is shown in Fig. 2.

3. Current state-of-the-art and contemporary challenges

As outlined above, the past century has seen substantial advances, both in computational techniques and in our understanding of how enzymes really work. Indeed, the field of QM/MM has matured significantly, but novel approaches are continuously being developed to address new challenges to our understanding of enzyme reactivity and dynamics. Due to space limitations we cannot address all of these in this Perspective. Therefore, we are limiting ourselves to those closest to our own areas of research. The purpose of this section is to cover current state-of-the-art QM/MM approaches, including simulation techniques adopted in conjunction with QM/MM methods, and highlight some important challenges. Additionally, we would like to provide some technical warnings for newcomers to this field, as well as, hopefully, to highlight a number of less considered problems in order to inspire potential future research efforts.

3.1. An overview of QM/MM approaches for describing chemical and biological reactivity

Studying an enzyme at atomic resolution is a computationally demanding task. In this case, the complexity of the system is such that it is currently not possible to resort to pure QM methods to treat the entire system, and the phenomena under study cannot be accurately represented by MM methods (since MM methods are unable to describe the changes in electronic structure in the reacting part). Consequently, contemporary computational studies often use hybrid QM/MM approaches to elucidate the link between enzyme structure and function [71–73]. We note that recent work reporting minimization of small proteins at a pure QM level using code specially designed for graphical processing units, is a promising direction [74]. The QM/MM approach originally pioneered by Karplus, Levitt, and Warshel in the 1970s [38], has seen impressive advances over the past four decades in line with its increasing popularity. The overall concept behind this approach is deceptively simple: divide and conquer by treating the most important part of the system (the inner layer) using a high(er) level electronic structure method, and the rest of the system with a less expensive molecular mechanical method, which reduces computational cost and allows inclusion of the complete protein environment (the outer layer). Such methods may, in turn, be sub-divided into additive and subtractive schemes [75]. In the case of an additive scheme, the effective QM/MM Hamiltonian of the full system can be defined by:

$$\hat{H}_{Eff} = \hat{H}_{QM} + \hat{H}_{MM} + \hat{H}_{QM/MM} \tag{1}$$

Here, the energy of the full system is described by adding the energy obtained from the QM calculation in the inner layer with an MM calculation in the outer layer. Furthermore, an explicit coupling term is added that describes the interaction between both layers. The coupling term is crucial [20], and includes the bonded, electrostatic and van der Waals interactions between the atoms in both regions.

$$\hat{H}_{QM/MM} = -\sum_{i,n} \frac{q_n}{|R_i - r_n|} + \sum_{N,n} \frac{Z_N q_n}{|R_i - r_n|}$$
$$+ \sum_{N,n} \left(\frac{A_{Nn}}{(R_N - r_n)^{12}} - \frac{B_{Nn}}{(R_N - r_n)^6} \right) \tag{2}$$

Here the first term accounts for the interaction between the partial charges of the MM atoms (n), q_n, and the electrons of the QM region, i. The second term accounts for the interaction between
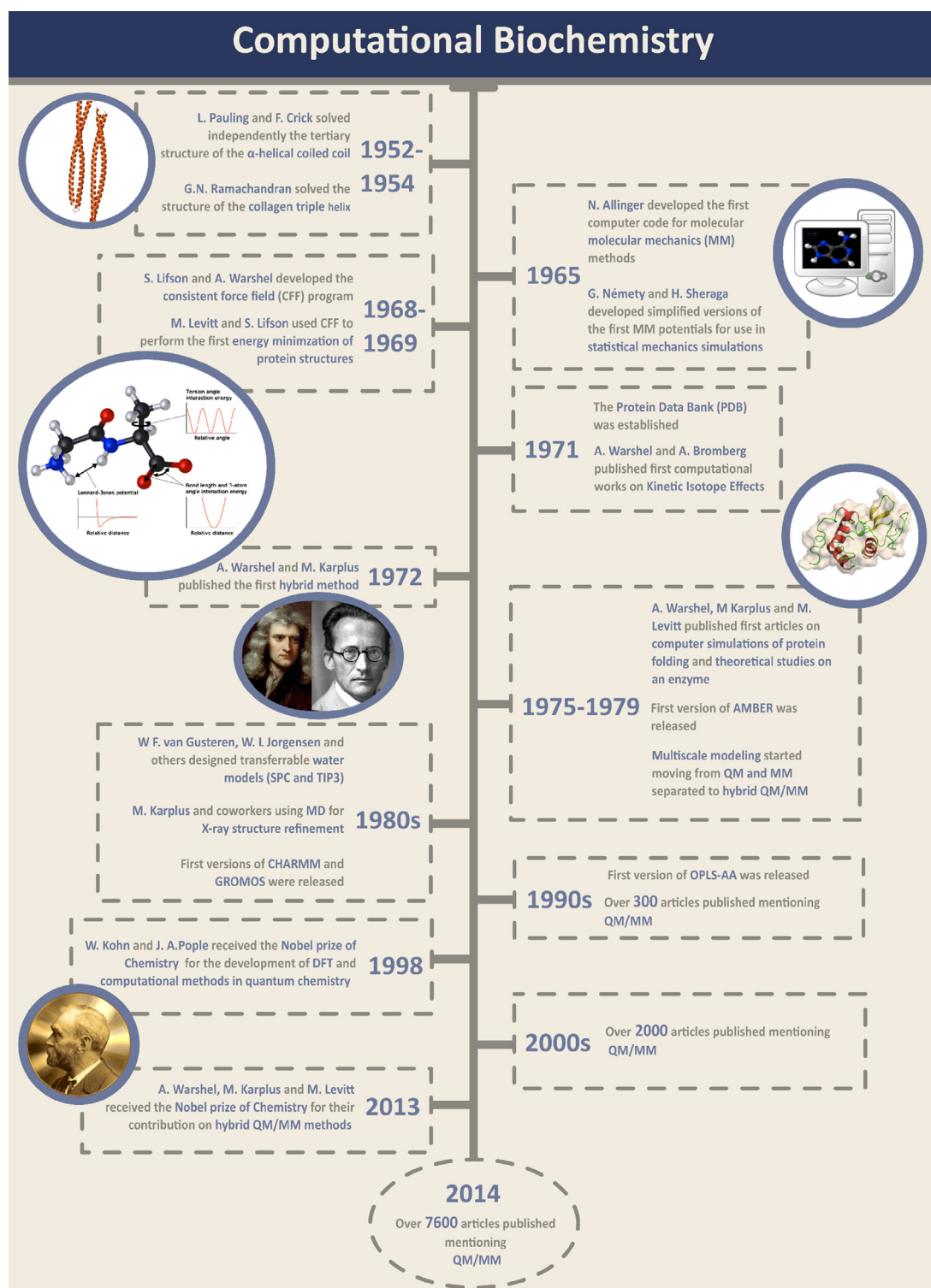
Fig. 2. A crude timeline of some key events in computational chemistry and biology. Statistics of published articles extracted from Thomson Reuters' Web of Science (https://webofknowledge.com/).

the MM partial charges and the nuclei, N, of the QM region, while the final term accounts for the van der Waals interactions between the QM and MM atoms. This latter term is needed to account for non-electrostatic interactions between QM and MM atoms. The terms of Eq. (2) are added to the Fock (*ab initio* or semiempirical methods) or Kohn-Sham (density functional theory (DFT) methods) matrix, and the wavefunction is polarized *via* the first term.

In the *ab initio* and semiempirical formalisms, the total energy may be obtained by solving the Schrödinger equation in the field of the MM partial charges

$$\hat{H}_{Eff}\,\psi = E_{Tot}\,\psi \tag{3}$$

and the total energy is then defined as

$$E_{Tot} = \left\langle \psi | \hat{H}_{QM} | \psi \right\rangle + \left\langle \psi \left| \hat{H}_{QM/MM} \right| \psi \right\rangle + E_{MM} \tag{4}$$
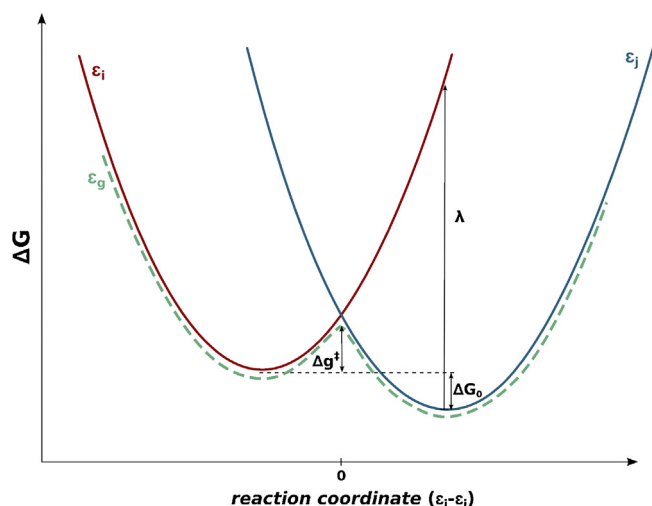
Fig. 3. Schematic overview of the empirical valence bond (EVB) approach, where ε_i and ε_j reflect the energies of the two diabatic states, ε_g is the energy of the ground state adiabatic free energy surface, λ is the reorganization energy, $\Delta g^{\ddagger}$ is the activation barrier for the reaction, ΔG_0 is the overall reaction free energy, and the reaction coordinate is taken as the energy difference between the two diabatic states ($\varepsilon_i - \varepsilon_j$). For further details, see main text.

Similar expressions arise for Kohn-Sham versions of DFT.

One of the major challenges inherent to *all* QM/MM approaches is the high computational cost associated with the QM part of the calculation. This is particularly important if one is interested in QM/MM free energies where extensive conformational sampling is required in order to obtain converged results. Ideally, one would employ high-level *ab initio* or DFT methods for the QM region. However, this is often not practical due to the high computational cost. Alternatives to circumvent this cost have been suggested, as is the case of the Car-Parrinello method, which has been used by several groups to study enzymes [76–78]. One may also employ a semiempirical Hamiltonian to describe the QM region, and there have been an increasing number of recent studies utilizing such approaches in studies of enzymatic reactivity [79,80]. However, standard semiempirical methods are often not accurate enough to give quantitative insight and it is necessary to perform reparameterization of the canonical parameter set to fit a specific target reaction [81]. Such specific reaction parameter approaches have been applied with great success for numerous enzymes [82–88]. Semiempirical versions of DFT present another possible accurate, yet fast, description of the QM region [89]. Additionally, some workers try to resolve this problem by performing the computationally expensive conformational sampling using the semiempirical approach, and then performing a perturbation to a higher level potential using single point calculations [90,91], although the success of this approach has been quite variable, depending on how different the two potentials actually are.

In spite of the usefulness of re-parameterized semiempirical methods, the QM calculation often remains the bottleneck for typical QM/MM calculations.

This may be overcome in the EVB approach [50,51,92–94], which has its history in the construction of potential energy surfaces (*e.g.* LEPS [95]). This method elegantly transforms a force field based description of individual electronic ground state (GS) species into a quantum chemical framework using valence bond theory, allowing for an empirically-based QM/MM description of chemical reactivity. Specifically, this approach describes the reaction through a combination of diabatic states that correspond to classical valence bond (VB) structures, [51,92–94], the energies of which can be described by ε_i and ε_j (Fig. 3). The potential energy

of each diabatic state is represented by classical MM force fields, which take the form:

$$\varepsilon_i = \alpha_{gas}^i + U_{intra}^i(\boldsymbol{R}, \boldsymbol{Q}) + U_{Ss}^i(\boldsymbol{R}, \boldsymbol{Q}, \boldsymbol{r}, \boldsymbol{q}) + U_{Ss}(\boldsymbol{r}, \boldsymbol{q}) \quad (5)$$

Here, the first term is the gas-phase energy of the reacting fragments in diabatic state i (ε_i) at infinite separation. $\boldsymbol{R}$ and $\boldsymbol{Q}$ represent the coordinates and charges of the reacting atoms in the relevant diabatic state, and $\boldsymbol{r}$ and $\boldsymbol{q}$ represent the coordinates and charges of the surrounding protein and solvent. U_{intra}^i and U_{Ss} represent the potentials of the inner (intramolecular) and outer (solute-solute) layers, and U_{Ss}^i represents the coupling between both layers (solute-solvent). The values obtained through Eq. (5) provide the diagonal elements of an $n \times n$ EVB Hamiltonian, which in the simple case of a two-state reaction takes the form shown in Eq. (6):

$$\hat{H}^{EVB} = \begin{pmatrix} H_{ii} & H_{ij} \\ H_{ij} & H_{jj} \end{pmatrix} = \begin{pmatrix} \varepsilon_i & H_{ij} \\ H_{ij} & \varepsilon_j \end{pmatrix} \quad (6)$$

The corresponding off-diagonal (H_{ij}) elements that describe the quantum mechanical coupling between the two states are related to the GS adiabatic energy, E_g, by:

$$H_{ij} = \sqrt{(\varepsilon_i - E_g)(\varepsilon_j - E_g)} \quad (7)$$

Typically, this coupling is obtained using a simple exponential form as a function of the distance between reacting atoms, although more complex functional forms have also been used [96–98]. It should also be noted that the H_{ij} elements have been rigorously verified to be phase-independent [51,92–94]. From Eq. (7), it can be seen that E_g can be obtained by simply diagonalizing the EVB Hamiltonian. Once the potential energy for each diabatic state has been obtained, the system can be adiabatically driven from one state to another to obtain the activation free energy, $\Delta g^{\ddagger}$, *vide infra*. Therefore, the different diabatic states provide an approximation for the basis functions used in MO-based QM calculations, and, like other semiempirical QM approaches, the EVB approach merely replaces the integrals related to every Hamiltonian matrix element with empirical functions. A core feature of the EVB philosophy is the rigorous calibration of a reference state (normally the background reaction in aqueous solution) to QM or experimental data, and then moving to a different environment such as an enzyme without the need for subsequent parameter adjustment, allowing one to quantify catalytic effects as well as pinpointing their exact origin. When this is combined with the fact that the EVB is an additive empirical QM/MM approach, it means that it is on the one hand fast enough to perform extensive conformational sampling of entire proteins, while at the same time describing both the bond-making and breaking process and the solute–solvent coupling in a consistent, physically meaningful fashion.

In passing, we note that non-empirical flavors of QM/MM VB approaches have been developed separately by Gao, Truhlar and Shurki [61,99–102] among others, as well as hybrid approaches such as the simple valence bond method (SVB) [103]. In this latter approach an additional term, which is a function of a pre-determined reaction coordinate, is added to the QM/MM potential energy in order to correct the energy of the high level layer (*e.g.* a semi-empirical approach). Another example of a valence bond approach is the MCMM of Truhlar et al. This method is similar to the EVB approach, mainly differing in the fact that the resonance integral (off-diagonal elements of the Hamiltonian matrix) are obtained by Shepard interpolation of quadratic expansions around a set of points where electronic structure data are available. A description of this method and the application to the study of two organic reactions: an OH radical reacting with propane at the primary carbon and an OH radical reacting with camphor can be found in [104].

Clearly, an ultimate long-term goal would be to be able to describe the entire system of interest with only QM, although

we are still far from a stage where the computational cost of doing so would be trivial. Current development of fast DFT codes combined with graphical processing units is bringing this goal closer [74]. Another approach is that used for instance by ONIOM, which allows a multi-layered description of the system. This entails using different levels of theory in the different layers, for example, using a correlated wavefunction method in the inner (high-level) layer, and a simpler DFT or wavefunction-based method for the remaining system. A less computationally expensive approach when dealing with biomolecular systems, such as enzymes, is to use a semiempirical ("cheap") QM approach or a density functional tight-binding (DFTB) [105] description of the outer layer. An example of a recent enzyme study using a combined DFT/semiempirical approach can be found in ref. [106]. Linear-scaling DFT approaches, such as the ONETEP method, have also been used in studying enzymes [107]. Yet another alternative is to move to constrained/frozen DFT approaches (CDFT/FDFT) [108–111]. In these methods the entire system is treated by DFT, but the electron densities of the atoms in the outer layer of the calculation are frozen (or constrained), thereby reducing the overall computational cost. The coupling between the inner and outer layers can then be evaluated using a non-additive kinetic energy functional (as discussed in *e.g.* [108–111]), allowing one to couple the two subsystems by means of orbital-free first-principles effective "embedding" potentials. CDFT/FDFT approaches have been successfully applied to a number of biochemical systems, for example studying ligand-to-metal charge transfer in metalloenzymes [112–114] and enzyme-catalyzed proton transfer [115]. For further details on recent developments in such orbital-free embedding approaches we refer the reader to *e.g.* refs. [116–118].

When discussing common current approaches to study enzyme reaction mechanisms, we would like to point the readers to the recent growing interest in QM cluster models of enzyme active sites (for reviews, see *e.g.* [119,120]). Briefly, in cluster models, such as that illustrated in Fig. 4, the full enzyme system is truncated to a smaller subset of key residues that are deemed important for the catalytic mechanism. Such approaches have gained in popularity in recent years due to increasing computational power, allowing routine DFT calculations of systems comprising several hundred atoms [119]. Excellent advances in our understanding of enzyme mechanism have been made by the use of QM cluster calculations [121], in particular the seminal work of Siegbahn on Photosystem II [122,123] and the work of Shaik on heme-related systems [124]. A major advantage of such approaches is their simplicity, substantially reducing the complexity of handling the full system computationally. Furthermore, it allows the treatment of complex electronic states, which might be too expensive with QM/MM approaches. However, this simplicity comes at the cost of a full-treatment of long-range interactions and conformational entropy, so depending on the nature of the problems one wants to address, one may want to move to more complete (albeit more challenging) models of the system [125,126]. Nevertheless, if treated with care, such models can provide very useful insight into the intrinsic reactivity of the relevant enzymes (see *e.g.* [127–129]). Nonetheless, there is no "correct" way of choosing the size of the cluster, as even distant charged residues can have significant electrostatic contribution to the total energy and can therefore also influence the energetics of reactions, in particular reactions involving charge migration [84,130–132].

Clearly, this Section is by no means exhaustive, and only touches on a small subset of popular current approaches. Ultimately, however, the reliability of all QM, QM/MM and QM/QM studies of enzyme mechanism will be limited by a wide range of issues. The first of these is the level of theory with which the QM region is described (usually DFT, semi-empirical, or EVB QM approach). The size of the high-level region is also crucial, as differently defined QM regions can give very different results (see for example [133,134]). The force field employed to describe the MM region is also very important, as this choice can greatly influence the accuracy of the results. Additionally, the polarization of the QM region *via* the MM partial charges must usually be fine-tuned on a per-system basis *via* the QM vdW parameters [135]. Furthermore, it might be necessary to employ a polarizable FF to account for the electronic response of the MM region to changes in the system [136–138]. Having reliably described both QM and MM regions, and their mutual polarization, there is also the question of how to best describe the covalent boundary between the two. This may be done by simple link-atom approaches [60], boundary atom approaches [139,140], localized orbitals [141] and generalized hybrid orbitals [142,143], all of which will affect the outcome of the calculations. Additionally, artifacts may arise if the boundary between the high and low level regions is too close to the reacting atoms.

From this section, it can be seen that there are ultimately a plethora of different approaches available for studying enzymatic activity using quantum chemical techniques, and, ultimately, the final decision should be based on the complexity of the system of interest.

Fig. 4. Illustrative cluster model of the promiscuous arylsulfatase from *Pseudomonas aeruginosa* actually used in the literature [286] showing the active site truncation with a *p*-nitrophenyl sulfate monoester bound to it. The nucleophile (FGly-51), initially in a hydrated form (as a geminal diol), is deprotonated by Asp-317, which acts as a general base. In its deprotonated form, the nucleophile attacks the sulfate. The product release occurs by His-115 deprotonating the other hydroxyl group from the FGly, followed by hemiacetal cleavage, yielding a free sulfate dianion and the aldehyde form of the nucleophile. The FGly is then hydrated and goes back to the form of the initial step.

3.2. Free energy simulations and the choice of an appropriate reaction coordinate

A crucial issue when embarking on an *in silico* enzyme study is the choice of free energy simulation approach and reaction coordinate. Many options exist, and the results may vary depending on the choices made and the system at hand. Typically, the goal is to obtain the free energy profile for the reaction as a function of some reaction, or progress, coordinate, x, *i.e.* $\Delta g(x)$. In order to obtain this, the simulation methods almost universally employ some form of biasing potential. This is needed in order to allow the sampling of rare events within a reasonable timeframe.
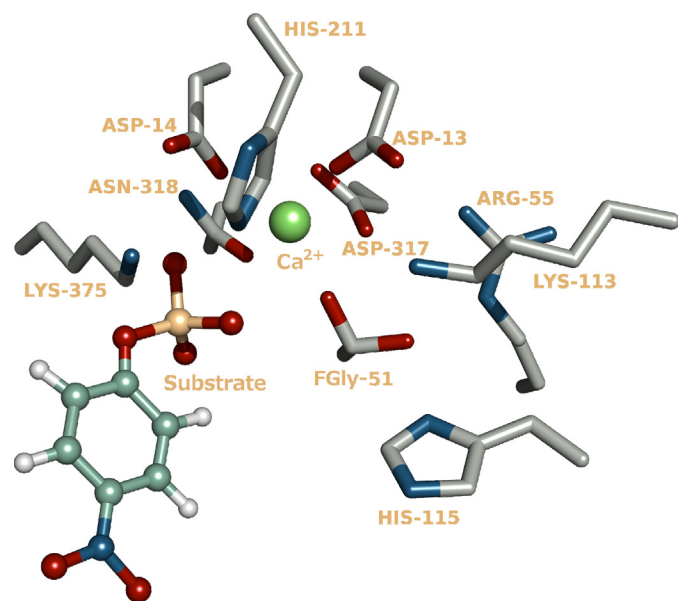
The reaction coordinate should allow the simulation to explore as much as possible of the configuration space relevant to the chemical change. When studying reactions in the gas-phase or using a simplified (implicit) solvent model, a geometric reaction coordinate is often adequate, *provided* that the reaction can be simplified as a function of one or two changing degrees of freedom in a meaningful way. Now, even in aqueous solution, this is not always the case and a lot of important mechanistic information can be lost by such a compression of the multidimensional landscape onto a limited description of the reaction coordinate [93,94]. However, the situation is aggravated when dealing with explicit solvent representations or complex biological systems such as enzymes, where it is important to not only sample over all possible solute conformations, but also to capture the response of the environment to changes in the conformation of the solute.

The simplest possible definition of a reaction coordinate is a geometric one. Such coordinates may be a simple bond-distance, angle, or dihedral angle, or the widely used anti-symmetric stretch coordinate [144]. Geometric coordinates are perhaps the most commonly used reaction coordinates in molecular simulations of chemical reactions [145], but require substantial prior insight into how the reaction is likely to proceed in the case of multidimensional systems, and arguably focus the search space too much as a result of mechanistic presumptions. An additional geometric coordinate may be defined as a bond order coordinate, which gained particular popularity in the experimental community due to seminal work by More O'Ferrall [146] and Jencks [147]. Using geometric metrics one may also define the reaction progress in terms of rehybridization, which provides additional insight [148].

A fundamentally different approach to the reaction coordinate definition is to use the energy gap between two diabatic states. Here, x is defined as the energy gap between the reactant and product diabatic states, $x = \varepsilon_1 - \varepsilon_2$, and this can then be used to describe the progress of the system from reactant to product. The energy gap reaction coordinate is a generalized reaction coordinate that does not require a pre-defined geometric path, but rather, it uses a superposition of the reactant and product potentials to drive the reaction [93,94]. The main advantages of the energy gap as a reaction coordinate are that it captures the main physics of the environmental response to the changes in solute geometry and charge distributions, while simultaneously allowing for accelerated convergence in free energy calculations in the condensed phase such as biological systems (see also discussion in [93,94]). The energy gap reaction coordinate has been successfully used in numerous studies of enzyme-catalyzed reactions, and it has been demonstrated to be a robust and effective reaction coordinate [52–54,149]. Note also that the energy gap coordinate is not limited to the EVB framework, but rather is a "universal" reaction coordinate that can be used for any system with a quantum chemical Hamiltonian [150]. Similarly to the energy gap coordinate, one may define an electrostatic potential (EP) difference coordinate, which describes the change in the EP difference between selected sites in the system [151]. For instance, one may define the difference between the EP at the donor–acceptor sites in hydrogen transfer reactions. Another category of reaction coordinates is collective variable coordinate, where one attempts to identify a set of coordinates, which describes the reaction progress. Usually, this is a set of geometric coordinates, which is found to change in concert during the reaction.

The choice of reaction coordinate is often a result of the QM treatment selected. With *ab initio*, semiempirical, or DFT methods, geometric reaction coordinates are often a natural choice, whereas in VB or EVB approaches the energy gap coordinate is a good choice. The EP difference potential has the advantage of being applicable to *ab initio*, semiempirical, or DFT methods, in conjunction with a geometric reaction coordinate, which drives the reaction between the reactant and product states. Choosing a poor reaction coordinate will result in increased recrossing at the dividing surface. That is, trajectories that reach the TS from the reactant state (RS) will recross and return to the reactant side. This may be corrected for quantitatively by computing a recrossing coefficient [152]. This may be done by rotation of the dividing surface or by running activated dynamics simulations. For further information on reaction coordinates, we refer the interested reader to an excellent recent review by Clementi and co-workers [144].

In addition to choosing an appropriate reaction coordinate, it is crucial to consider carefully the sampling method employed. It might initially seem tempting to simply identify a minimum potential energy path, which connects the reactant and product states *via* a TS. This is typically done in gas-phase reactions as well as in the solution phase with an implicit solvent description. However, this might be rather dangerous in explicit solvent simulations, as the energy landscape for such systems is quite complex, and one can easily get trapped in local minima [153]. The problem might become acute in enzyme simulations, and different starting structures can potentially give rather different results. We do note that in certain systems energy minimization approaches can be successful, presumably as a result of a good starting structure and a fairly rigid enzyme framework [71].

One of the most commonly used sampling methods is umbrella sampling (US) [154]. The free energy difference along the reaction coordinate is defined as:

$$\Delta g(x) = -RT \ \ln \rho(x) + C \tag{8}$$

where R the gas constant, T is the temperature, $\rho(x)$ is the probability density along the reaction coordinate, and C is an undetermined constant. In this approach, one applies an umbrella potential, which ideally equals the $U_{umb}(x) = -\Delta g(x)$. Since this function is not known *a priori* (*i.e.* the goal is to find $\Delta g(x)$), an iterative approach is usually adopted, wherein the umbrella potential is refined in subsequent free energy simulations (so-called "adaptive umbrella sampling" [155–158]). Additionally, the reaction coordinate is usually divided into windows, which allows focused sampling in designated regions along the progress coordinate. To keep the system in the simulation windows, a restraining potential is applied, usually in the form of a harmonic potential ($U_{res}(x) = k(x - x')^2$). The combined biasing potential is thus composed of the umbrella and restraining potentials. The effect of the bias is removed using statistical methods such as the weighted histogram analysis method [159].

Additional standard simulation techniques may be employed, such as free energy perturbation (FEP) and thermodynamic integration (TI). In the FEP approach one calculates free energy differences along the reaction coordinate using Zwanzig's perturbation formula [160]:

$$\Delta g^{x' \to x''} = -RT \ \ln \left\langle \exp\{-\beta \Delta E_{Tot}^{x' \to x''}\} \right\rangle_{x'} \tag{9}$$

In Eq. (9), the brackets $\langle \ldots \rangle$ represent an ensemble average, $\beta = (k_B T)^{-1}$, and k_B is Boltzmann's constant. Various schemes may be devised to do this efficiently, to reduce the number of QM calculations [161]. Analogously, one may define the free energy using TI, where the mean force acting on the reaction coordinate is integrated:

$$\Delta g^{x' \to x''} = \int_{x'}^{x''} dx \frac{dg(x)}{dx} = \int_{x'}^{x''} dx \left\langle \frac{\partial H(x)}{\partial x} \right\rangle_x \tag{10}$$

Practically, in TI simulations, constrained dynamics may be run at discrete values of x and one samples the average force at each location, and the average forces are then integrated numerically over the reaction coordinate. If one instead integrates over time,

one obtains slow-growth or steered MD, with the following expression:

$$\Delta g^{x'\to x''} = \int_{t'}^{t''} dt\, \frac{\partial H(x)}{\partial x}\, \frac{\partial x}{\partial t} \qquad (11)$$

If the dynamics is driven by a restraint potential, the method is dubbed slow-growth, while if a guiding potential is used, it is termed steered MD.

Warshel developed an approach, which combines FEP and US with the EVB Hamiltonian and the energy gap coordinate [162]. Initially, the potential energy for each diabatic state is obtained, and thereafter the system can be adiabatically driven from one state to another to obtain the activation free energy, $\Delta g^{\ddagger}$. This is done using a mapping potential that is written as a linear combination of reactant and product potentials:

$$\varepsilon_m = (1-\lambda_m)\varepsilon_i + \lambda_m \varepsilon_j \qquad (12)$$

where λ_m describes the weighting between the two states and is changed incrementally from 0 to 1 during the mapping procedure. The free energy, ΔG_m, for changing the system from state i to j can be obtained using the standard FEP/US methods described above [162], and the corresponding free energy surface along the reaction coordinate, x, can be obtained from the free energy functionals of the individual diabatic states:

$$\Delta g_i(x') = \Delta G_m - \beta^{-1} \ln \left\langle \delta(x-x') \exp[-\beta(\varepsilon_i(x)-\varepsilon_m(x))] \right\rangle_{\varepsilon_m} \qquad (13)$$

here, $\Delta g_i(x')$ corresponds to the free energy functional of the diabatic state i at position x' along the reaction coordinate x. Provided that the perturbations to $\Delta g_i(x')$ are sufficiently gradual, they can be patched together to give the full free energy surface. The GS free energy surface, $\Delta g_i(x')$, can be evaluated in a similar way, using the GS energy instead of the diabatic energy (see Ref. [93,94,162] for discussion and Fig. 3 for a schematic representation). Despite its apparent simplicity, when the EVB approach is combined with the energy-gap reaction coordinate and FEP/US, it is an extremely powerful tool for describing chemical reactivity in both chemical and biological systems (see e.g. Fig. 5) [162].

A powerful alternative to the standard simulation techniques described above, is the metadynamics approach of Laio and Parinello [163]. This method, which follows from the idea of using a reference potential approach, as well as the local elevation [164] and conformational flooding approaches [165], is an elegant way to focus conformational sampling by introducing the concept of memory to a simulations. Specifically, in this approach, repulsive markers are placed in a coarse timeline that is spanned by a small number of chemically relevant collective variables (which essentially define the "reaction coordinate" for the system). These markers are placed on top of the underlying free energy landscape for a system, thus discouraging the simulation from revisiting points that have already been visited in the conformational space. Therefore, the system can escape over the lowest energy TS as soon as the growing biasing potential and the underlying free energy well exactly counter-balance each other, allowing one to sample a much larger portion of the free energy landscape. In recent years, metadynamics has become a vastly popular tool, and has been successfully employed in the study of enzymatic reaction mechanisms [166–173], predominantly using DFT potentials. Additionally, recent advances in metadynamics allow estimation of the correct dynamic properties of barrier-crossings [174,175]. Despite its elegance, however, there are still two major challenges with the metadynamics approach. The first is the appropriate choice of collective variables (CVs): while there is a lot of flexibility in choosing these CVs, they nevertheless introduce bias into the simulation, and therefore the correct choice of CVs is both crucial and nontrivial. More of an issue, however, is the exceedingly high cost of repeated

QM calculations that are necessary to obtain converged results in *ab initio* metadynamics calculations. For example, in even the simple case of metadynamics simulations of the nucleophilic attack of Cl^- on MeCl in vacuum, 10^6 calls to the QM program were required in order to obtain convergent results [150].

An alternative to the above-mentioned methods, entails the family of "string" approaches, which have recently gained in popularity [176–178]. Once again, these approaches try to minimize bias in the system by assembling a set of isocommitor surfaces between reactant and product states, and then trying to obtain the minimum energy path between these two states by trying to find the curve with the maximal flux of transition current through these isosurfaces (see [144]). This is done while creating a string of conformations that connect state A to state B, and trying to move this string toward the minimum energy path connecting these two states. In principle, a limitation of this approach is that it is best suited for smooth energy surfaces, however, it has been successfully applied to a range of problems including, for example [148,179–182].

All of the above mentioned methods assume equilibrium behavior. This is in line with the basic presumptions of TST. However, if the chemical step is fast this assumption may break down, and non-equilibrium behavior becomes important. One way to introduce non-equilibrium behavior is by employing the Jarzynski equality [183,184], which connects between the irreversible work, W, for arbitrary switching times, t, and the equilibrium free energy:

$$\exp\left\{-\frac{\Delta g^{x\to x'}}{k_B T}\right\} = \left\langle \exp\left\{-\frac{W_t^{x\to x'}}{k_B T}\right\} \right\rangle_x \qquad (14)$$

where we move from a point x at equilibrium to a non-equilibrium point x' in time t. The average is performed over the equilibrium ensemble $\langle \ldots \rangle_x$ at point x, and one performs a series of non-equilibrium pulling simulations to point x'. The practical implementation of the Jarzynski equality in actual simulations is termed fast growth MD. Thus, it seems that careful comparison between such non-equilibrium fast-growth methods and standard equilibrium methods could be of great interest [161].

Additional concerns in enzyme simulations are the biases introduced in the simulations in order to cross barriers and the use of a pre-defined reaction coordinate. A method that alleviates both of these concerns is the transition path sampling (TPS) approach of Chandler and co-workers [185]. This method is a rare event sampling approach that focuses primarily on sampling the transition between two given states for a process, the reaction basins A and B. TPS entails a Monte Carlo (MC) search in reactive trajectory space, essentially moving between reactive trajectories. One of the major advantages of this approach is the fact that no information is required in advance about the actual reaction coordinate or TS structure, as long as the reaction basins are properly defined. An additional advantage is that the method does not apply any bias to the system, so in principle the dynamics of a TPS simulation is the true dynamics of the system. One of the disadvantages of TPS is the difficulty in reaching a converged transition path ensemble, resulting in a computationally costly approach [185]. Furthermore, it is not straightforward to obtain information regarding the free energy of a process from TPS since the search is performed in the reactive trajectory space, and little is known about the vast unreactive trajectory space. Several improvements to the TPS method have been suggested to deal with these issues [144,186–190]. Ideally, TPS can be used in conjunction with other methods, which provide free energy information, thus gaining both information regarding reaction rates and true dynamic behavior during reactions. The use of TPS for enzyme systems was pioneered by Schwartz and co-workers [191], and has since been adopted by others [192–194].
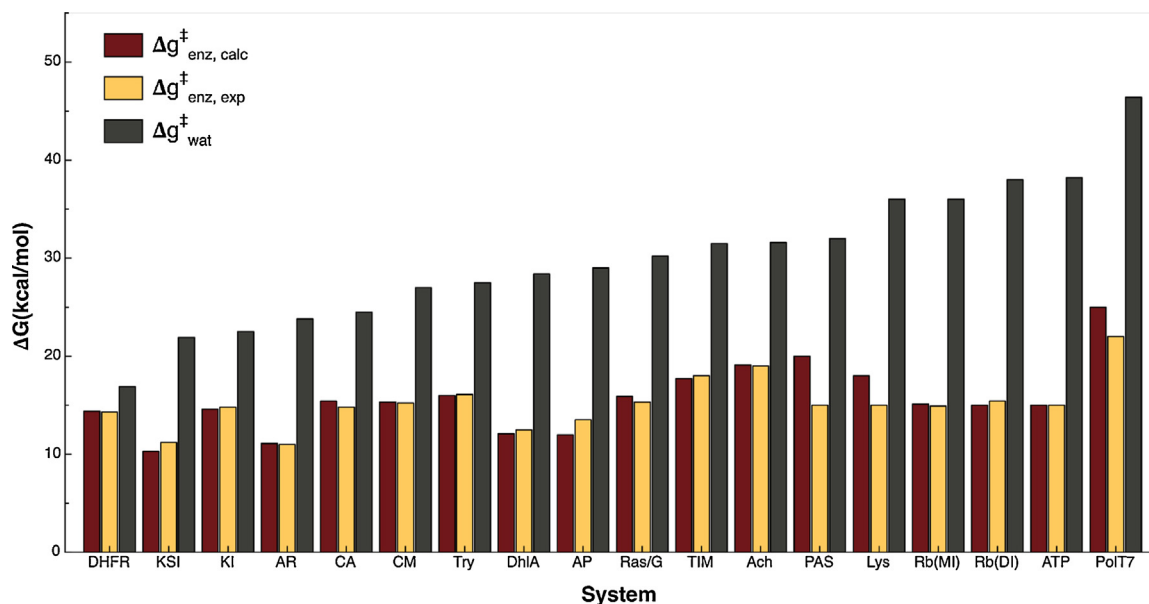
Fig. 5. Illustration of a range of systems for which the EVB approach has been able to reliably reproduce the (often very large) catalytic effect of different enzymes. Shown here are: dihydrofolate reductase (DHFR), ketosteroid isomerase (KI), aldose reductase (AR), carbonic anhydrase (CA), chorismate mutase (CM), haloalkane dehalogenase (DhlA), alkaline phosphatase (AP), Ras/GAP (Ras/G), triose phosphate isomerase (TIM), acetylcholine esterase (Ach), *Pseudomonas aeruginosa* arylsulfatase (PAS), lysozyme (Lys), F_1-ATPase (ATP), ribonuclease monoionic and diionic intermediate (Rb(MI) and Rb(DI)), DNA polymerase T7 (PolT7), orotidine 5′-monophosphate decarboxylase (ODC), staphylococcal nuclease (SN), Klenow fragment (Kf). $\Delta g^{\ddagger}_{wat}$, $\Delta g^{\ddagger}_{enz,calc}$ and $\Delta g^{\ddagger}_{enz,exp}$ denote the activation barriers to the reference reaction in water and to the (calculated and experimental) enzyme catalyzed reaction [53,149,233].

3.3. Where can things still go wrong?

Even after carefully choosing your preferred Hamiltonian, reaction coordinate and sampling method, there are still plenty of things that can go wrong in QM/MM simulations. In this section we will focus on selected issues that require careful attention.

The first point, and possibly the most important one, is the position of the substrate and possible co-factor in the active site. Typically, a QM/MM project starts with a crystal or NMR structure of the target protein. In cases where the enzyme is not fully closed in its pre-organized catalytic form or the active site is empty, great care must be taken in generating an active form of the enzyme. This may entail homology modeling or some form of *de-novo* design, and might have to be done in an iterative fashion with enzyme mechanistic simulations, until a properly closed and active form of the enzyme is obtained. If the enzyme contains a substrate analog in the active site it might still be necessary to do careful docking of the true substrate into the active site in an attempt to form an active enzyme form. Clearly, if this step is not carried out in the most careful manner, the entire project is in jeopardy.

An initial question when embarking on the project is the type of boundary to choose for the system. The most reliable options are periodic boundary conditions (PBC) or stochastic boundary conditions (SBC). In the former setup, the enzyme–substrate complex is embedded in a pre-equilibrated periodic water system, while in the latter it is embedded in a sphere of pre-equilibrated water molecules. The arguably preferred option is PBC, which allows full flexibility of the enzyme and in conjunction with the Ewald summation method [195–197] accounts for long-range electrostatics. SBC can in principle also include full enzyme flexibility, although this requires very large water spheres, and there will still be a boundary effect [198,199]. Other methods include the spherical solvent boundary potential model [200]. Similar but more general approaches include the generalized solvent boundary potential (GSBP), where not only solvent molecules, but also the macro-molecule atoms can be treated implicitly [201], and the QM/MM general boundary potential of Benighaus and Thiel [202–205].

The effect of long-range electrostatics may be mimicked *via* the extended Ewald method of Kuwajima and Warshel [206] or the analytic continuum electrostatic method of Schaefer and Karplus [207]. Still, the number of water molecules to include in the system, and hence the density, is a point of some concern. The ambiguity arises because of the need to delete water molecules overlapping with the enzyme–substrate complex, and it is not always clear which waters one should retain. If one includes too many waters, the system becomes too dense, which in turn will inhibit protein motion. If one includes too few water molecules, the protein might be too flexible and charges may not be sufficiently well screened. In particular, the precise position of water molecules in the active site can be crucial as was shown by Yang and co-workers [208]. Additionally, counter ions should be added to neutralize the system [209] and if possible match the experimental concentration [210].

An additional point of concern is the protonation states of protein residues as well as that of substrates and potential co-factors. Although automated tools exist for determining protein pK_a values, these often fail for active site residues. This is partly due to the lack of FF parameters for substrates/co-factors in the pK_a programs. Additionally, the enzyme is likely not in the active state in the initial structure. Thus, it is usually prudent to manually inspect the hydrogen-bond pattern of all titratable residues to arrive at reasonable protonation states. Following from this, occasional errors in crystal structures must be corrected prior to the *in-silico* setup (*e.g.* orientation of side-chains in His, Asn, and Gln residues). It may often be necessary to carry out the same enzyme mechanistic study using different protonation states of the enzyme, or possible using constant pH simulations [211].

Finally, although we appreciate that this may be a controversial view due to their prevalent usage, we believe that the enzyme simulations should ideally be free of any restraints (except, out of necessity, for those related to the reaction coordinate in biased simulations). That does not mean that the use of restraints is not a valid approach; only that it can be dangerous to include restraints or constraints, since their introduction can inhibit important local conformational changes occurring along the reaction coordinate.

The problem with introducing bias and its unknown impact on the system behavior was discussed in [52], where the study of adenylate kinase using the renormalization approach was conducted. It was also substantially discussed in Dryga and Warshel [212]. This is due to the fact that if one employs any form of restraints one might inhibit potentially crucial protein motions. Such restraint-free simulations are possible today, due to the reliability of modern force fields [213–218]. For instance, introducing even weak restraints on backbone atoms can introduce artifacts to the simulations, and possibly preclude conclusions relating to protein dynamics and their functional role. However, it may often be necessary to include restraints during the MD heating and equilibrium stages to allow the system to reach thermal equilibrium without undergoing large structural perturbations due to "hot pockets" in the enzyme.

This list is by no means exhaustive and numerous additional potential pitfalls exist. These include the choice of size of the QM region, the coupling between the QM and MM regions, the treatment of covalent boundaries at the QM/MM interface [219], as well as the choice of simulation thermostat for constant temperature simulations [220]. However, it aims to provide illustrations of some of the key concerns when preparing and assessing the outcome of simulations of enzyme function and catalysis.

3.4. What do barrier crossings really look like?

After discussing numerous aspects of enzyme modeling, ranging from methods to strategies and pitfalls, one should maybe ask the crucial question of whether our simulations really mimic the chemical events as they happen in nature. Thus we should address the fundamental question of how chemical systems cross barriers, and whether standard simulation methods faithfully reproduce the dynamic behavior of the true systems. While clearly not intending to be dismissive of previous excellent computational work, we believe this is an important step toward further improving our understanding of questions of great contemporary interest such as the role of protein dynamics in enzymes. Chemical dynamics has been studied extensively using high-level methods for simple gas-phase chemistry. One of the first such examples was the seminal work of Porter, Sharma, and Karplus on the simple $H_2 + H$ exchange reaction [221,222]. One of the startling conclusions of this work is that the chemical event (*i.e.* the reaction) occurs during the course of a few femtoseconds, in spite of the reaction having a considerable potential energy barrier. This illustrates that the rate constant observed in such simple reactions is not low because the individual reactive event is slow: indeed, the reactive event is fast. Rather, most attempts at reactions are unsuccessful because of the large activation energy, and it takes the system a long time to find reactive thermal collisions. This illustrates the nature of chemical reactions: a stochastic hunt in phase space (position and momentum space) for reactive states. Recent years have brought evidence of the accuracy possible with quantum dynamics simulations of simple gas-phase reactions [223,224].

Reactions in condensed phase environments are considerably more challenging to treat theoretically, although the nature of the chemical reaction is likely still the same: fast chemical events occurring on the femtosecond–picosecond time-scale [92,225–228]. Thus, crossing of the TS region is usually very fast in condensed phase reactions, and the reaction seems slow because there is a low probability of reaching the reactive phase-space state from which it is possible to rapidly climb to the TS. In the case of condensed phase environments, it is necessary to reorganize the surrounding medium into a phase-space configuration, which is conducive to such fast reactive events. Thus, the environment needs to reorganize itself around the different charge distributions in the reactant and TSs. However, considering the fast bond vibrational motion during the reaction, the environment needs to reach a state

capable of stabilizing the TS sufficiently to enable barrier crossing prior to the barrier crossing itself. Herein lies much of the catalytic effect – this reorganization is easier in a pre-organized enzyme matrix than in aqueous solution where the water molecules need to change the hydrogen-bonding environment prior to the chemistry [92,229]. This is not to say that no reorganization occurs during the chemical event, but rather that the majority of this reorganization is carried out before the reactive event. The question of whether enzyme residues have catalytic motions on the time-scale of the chemical event that are coupled to the chemistry is a topic of much current debate [52,230].

Recent work has shown that the choice of sampling method can influence the nature of the TS obtained in enzyme simulations [194]. In a comparison between US and the string method on the one hand, and TPS on the other hand, it was found that the biased methods resulted in slightly shorter donor–acceptor distances for the hydride transfer in the enzyme dihydrofolate reductase. This was suggested to possibly be due to the use of a biasing potential, which perturbs the true reaction coordinate dynamics, as well as the assumption of equilibrium solvation, which is intrinsic to the simulation approaches such as US and string methods. Thus, the perturbation of reaction coordinate dynamics and equilibrium solvation might lead to slight changes in the location of the TS, something which might be accounted for *via* recrossing estimations. Despite this potential limitation in the fine details of the location of the TS, it should be pointed out that both molecular-orbital and EVB based standard QM/MM simulations of this enzyme by many groups have provided excellent agreement with experiment in terms of other observables, such as the reaction rates, temperature dependence of KIEs and the effect of mutations [231,232]. Thus, effects such as reaction coordinate dynamics and equilibrium solvation are likely very small.

4. What computational enzymology has taught us about enzymes

Computational enzymology has contributed greatly to both our understanding of the phenomenal rate enhancement in enzymes as well as in elucidating enzyme mechanisms. Below we first bring a historical perspective on the first QM/MM study of an enzyme, namely lysozyme, and subsequently describe some of the main lessons we have learned from the computational work on enzymes, and bring some selected examples.

To concisely answer the question "how do enzymes really work?" is extremely challenging, and numerous hypotheses have been put forward to try to answer just that over the past few decades. These include preferential stabilization of the TS (*i.e.* TS stabilization), RS effects such as desolvation, dynamic effects, and nuclear QM tunneling [145,233].

Historically, the earliest structure–function hypotheses date back to the crystal structure of lysozyme, mentioned in the Introduction [7], which led to the idea that enzymes work by destabilizing the RS of the reaction. Specifically, lysozyme is a very small glycoside hydrolase (129 amino acids for the hen egg white variant, Fig. 1(A)), catalyzing the hydrolysis of the peptidoglycan layer found in the cell wall of Gram-positive bacteria [7]. The peptidoglycan layer is composed of alternating oligosaccharides (2-acetamido-2-deoxy-glucopyranoside (NAG) and 2-acetamido-2-deoxy-3-O-lactyl-glucopyranoside (NAM)) [234,235]. The enzyme has a cleft that binds six saccharide units of the peptidoglycan. According to the classical "Phillips" hypothesis, an intermediate binds in a distorted conformation; more precisely a NAM subunit adopts a half-chair conformation forming a kink in the position where the bond is being broken. This structure bears greater resemblance to the TS than to the RS of the reaction. Therefore, a logical consequence of this observation appeared at the time to be that

enzymes distort the RS of the reaction, and that strain plays a major role in enzyme catalysis.[236,237] In subsequent experimental and computational work on this enzyme, Vocadlo et al. [235] and Mulholland et al. [238] suggested a covalent mechanism in line with Koshland's proposal (Fig. 1) [239]. The reaction was suggested to proceed through an undistorted covalent intermediate; however, in the first step, the Michaelis complex presumably adopts a skew-boat or distorted envelope conformation.

The RS strain idea was extended to many different systems over the years, and has been the subject of substantial controversy. While the idea of GS destabilization is still disputed [240–242], there has been substantial progress in resolving the debate through simulations that can probe the details of the catalytic step at atomic resolution. Such simulations have allowed dissecting the different contributions to catalysis [233,243], obtaining information that can only be inferred indirectly through experiment. For example, in the specific case of lysozyme, in the Warshel and Levitt 1976 paper, the factors that affect the intermediate stabilization were analyzed [38]. At the time this intermediate was though to be an oxacarbenium ion, although later proved to be a covalent glycosyl-enzyme intermediate [235]. This work [38] showed that steric strain does not contribute significantly, since the protein/substrate complex relaxes rapidly, however the major difference in electrostatic stabilization of the substrate in the protein in relation to vacuum suggested that electrostatic stabilization was the main factor that accounted for acceleration of the reaction rate. In contrast, already in the 1940s, Pauling (correctly) deduced that a major contributor to the catalytic proficiency of enzymes is their ability to provide much tighter binding of the TS compared to the RS of the reaction [244]. However, in the absence of structural data, it was unclear what the origin of this tight binding of the TS actually *is*. The advent of not just structural data, but also sufficiently powerful supercomputers allowed this question to be addressed in a direct fashion, and it was suggested already in the 1970s that the tremendous catalytic power of enzymes is the result of the catalytic preorganization of the active site [229], which in turn provides optimal TS stabilization. Conceptually, this is a simple idea: enzymes accelerate the rate of the chemical reaction relative to the corresponding reaction in solution because protein active sites have their dipoles optimally oriented to require minimal reorganization during the course of the reaction. This does not exist in the background aqueous solution reaction, where there is a high energetic penalty for reorienting the solvent dipoles in response to changes in the solute charge distribution (Fig. 6). This groundbreaking contribution by Arieh Warshel has become one of the cornerstones of our contemporary understanding of enzyme catalysis.

Even still, it is a subject of debate, for instance in the case of the origin of the remarkable catalytic efficiency of orotidine decarboxylase [80,208,245–247], where desolvation of the RS appears to play an important role, based on experimental work [248], and theoretical studies have been contradictory [145,243]. Often the different conclusions drawn regarding the contributions of RS destabilization and TS stabilization to catalysis are a matter of reference reaction definition.

The hypotheses highlighted above (RS destabilization, TS stabilization) are only two of the many hypotheses that have been put forward over the years to try to explain how enzymes actually work. Several additional suggestions that are derivatives of the first two categories have been proposed. Examples of RS destabilization include desolvation [248], steric strain [249], "near-attack conformer" (NACs) [250–253] and entropic changes due to the freezing of key motions [254,255], all of which are hypotheses that have been computationally challenged as playing only a minor role in enzyme catalysis [243]. In contrast, electrostatic complementarity, geometric complementarity, and electrostatic pre-organization
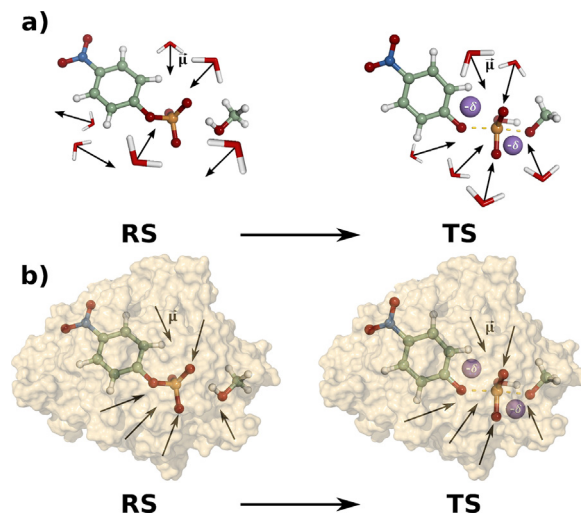

Fig. 6. Representation of electrostatic preorganization for an S_N2 reaction involving a charged substrate and a neutral nucleophile: (a) in water, where the dipoles of the environment must adjust themselves for the reaction; (b) in an enzyme, where the dipoles are already preorganized for the reaction from the beginning. The main contributor to the tremendous barrier reductions achieved by enzymes is the reduced cost of reorganizing preorganized dipoles in the active site of the enzyme (see [229]).

of the active site for a given reaction all provide crucial TS stabilization, which, despite Pauling's brave initial hypotheses would have been impossible to quantify without computational effort [229,233]. Moreover, recent years have seen an increased interest in "non-thermodynamic hypotheses" to explain enzyme catalysis, such as dynamical contributions/mode coupling (promoting vibrations) and quantum tunneling to name some examples [82,228,256–266]. There has been lively debate surrounding these issues [233,260,267], and while strong arguments are presented on both sides of this debate, it appears that such effects, while they can be crucial for function (*i.e.* not the catalytic step itself), appear to have only a minor effect on reducing the overall activation barrier (*i.e.* catalysis) compared to more conventional thermodynamic and electrostatic effects. Finally, nuclear quantum effects (NQE) such as tunneling have been suggested to play a role in catalysis, as well as being a reporter on enzyme dynamics [228,267]. Extensive work by Gao and Truhlar and co-workers [152], and Layfield and Hammes-Schiffer [268] have suggested that it is essential to include NQE to correctly reproduce absolute rates for hydrogen ($H^\bullet$, H^+, H^-) transfer reactions. However, NQE contributes only modestly to catalysis [82,269], although tunneling has been shown to have a small catalytic effect in the case of proton transfer in nitroalkane oxidase [82].

At this point, it is valuable to highlight an important aspect that is often lost in the discussion about the origins of enzyme catalysis and the possible contribution of dynamic effects. Per definition, the term "catalysis", in a chemical context, refers strictly to accelerating the rate of a chemical reaction, and therefore, in the most rigorous definition the catalytic effect of an enzyme can be taken to be the rate of the catalyzed reaction *vs.* the corresponding (intrinsic) uncatalyzed reaction [1,51,229]. Clearly, unlike the uncatalyzed counterpart, an enzymatic reaction is a complex, pre- and post-chemical, multistep process involving several conformational changes, substrate binding and formation of the Michaelis complex and subsequent product release, any of which can be rate limiting for a given system. However, under the strictest chemical definition, unless there is a clear memory effect between these steps (which has never been conclusively demonstrated), "catalysis" refers *only* to the rate of the chemical step in the enzyme compared to that of the corresponding uncatalyzed reaction, *i.e.*
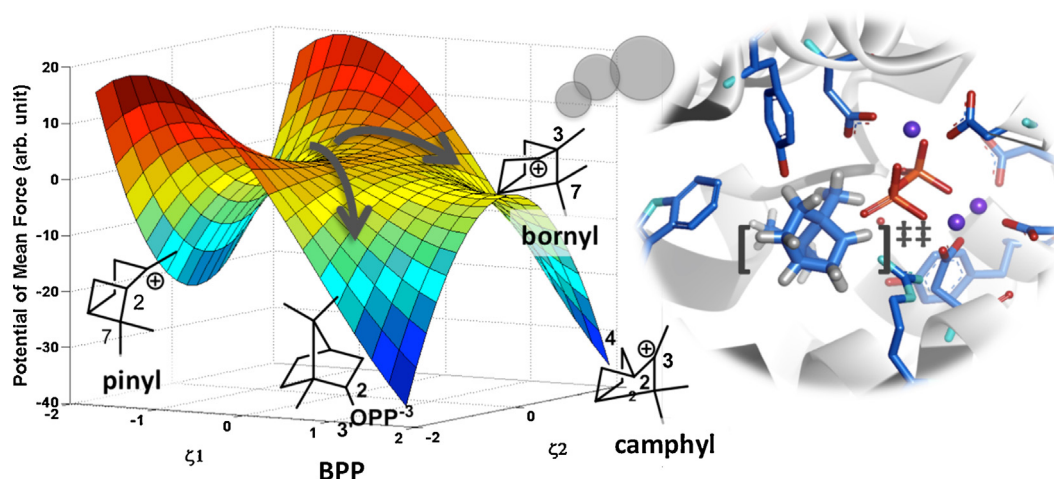
Fig. 7. Schematic depiction of the bifurcating free energy surface in the terpene cyclase bornyl diphosphate (BPP) synthase. The relevant carbocation species are marked, as well as the final product BPP. Copyright (2012) American Chemical Society.
Adapted with permission from ref. [277].

k_{cat}/k_{uncat}. Therefore, for a factor to be making a substantial contribution to the observed rate acceleration, it needs to be either absent in or substantially different from the background reference reaction (see discussion in *e.g.* [230]). This does not mean that protein flexibility is not crucial for enzyme activity; only that flexibility is usually also present in the background reaction. At the end of the day, enzymes use both pre-organization and flexibility to enhance the rate of chemical transformations. In the absence of proper pre-organization (*i.e.* folding), enzymes would not catalyze reactions, and without the inherent protein flexibility, most enzymes would likely be very poor catalysts, in spite of proper pre-organization. Indeed, enzymes ultimately only "care" about the fitness of an organism, and not the rate of the corresponding uncatalyzed reaction.

A main strength of computer simulations, in particular QM or QM/MM studies, is their power to predict and rationalize biological mechanism. For example, Siegbahn performed extensive DFT geometry optimizations on small models of the oxygen-evolving complex (OEC) in photosystem II [270]. These calculations allowed him to propose a position for the fourth Mn ion (in this complex three manganese metal ions and the calcium form a cubane-like complex, however the position of the fourth Mn ion is uncertain) and to propose a new mechanism for this reaction, where the oxyl radical bounded to the 4th Mn ion forms an O—O bond to a bridging oxo ligand in the Mn3Ca cube. A key feature of this mechanism is that the spins of the involved atoms are antiparallel aligned, which stabilizes the TS structure more than the reactants [270].

Another example is the work of Prasad et al. in determining the mechanism of phosphate hydrolysis for the activation of Ras GTPase activating protein (RasGAP) and the elongation factor-Tu (EF-Tu) [271]. These authors performed EVB calculations on these systems and found that the preferred mechanism for RasGAP is glutamine-61 assisted proton-transfer to the substrate (Gln-61 changes the P-loop interaction with the TS, stabilizing it) or substrate-assisted catalysis through a second mediating water molecule. For EF-Tu the preferred mechanism is proton transfer assisted by 2 water molecules accompanied by major transition state stabilization, presumably due to an allosteric electrostatic effect. In fact, all the best mechanisms involved electrostatic stabilization of the TS or central intermediate, suggesting that this is a fundamental aspect of GTPases activation [271].

Finally, not all enzymes are primarily intended to act as catalysts, as many enzymes control complex multistep reactions for which there simply exists no uncatalyzed counterpart (or the background reaction is so fast that no enzyme is needed). Terpene synthases provide such an example were chemical control is the principal "catalytic" effect [145,272–274]. In the absence of terpene cyclases, it is exceedingly difficult to obtain these molecules in the correct stereo- and regiochemical form [275]. This is due to the extreme reactivity of the intermediate carbocations, rather than insurmountable free energy barriers. Thus, in this family of enzymes chemical control, rather than rate enhancement is the key "catalytic" element [274]. Such control is achieved through an active site contour, which facilitates the correct folding of the initial substrate, sequestering and stabilization of intermediate reactive carbocations, as well as final reaction quenching [273]. QM/MM methods are ideally suited to treat such complex cases of chemical control, as one can readily map out various reaction pathways, including providing detailed information regarding highly reactive intermediates, which are not accessible from experiments. Interestingly, in certain terpene cyclases there is growing evidence for the existence of non-statistical dynamic effects, as has been demonstrated using multiscale simulations (Fig. 7) [276,277]. Such hitherto unknown dynamic effects in enzymes are due to underlying bifurcations on the free energy surfaces [278,279], with several open reaction channels. In order to control product distribution, these enzymes must be able to steer the reaction flux toward the desired reaction channel. This is achieved by using pre-organized electrostatic forces, which modulate the underlying free energy surface and selectively activate vibrations, which correctly channel the reaction flux. Again, in this case QM/MM simulations have been crucial in pinpointing these non-statistical dynamic effects.

In some of the above-mentioned cases, rather than accelerating a reaction, the enzyme is acting as a facilitator, allowing for the controlled formation and cleavage of chemical bonds on demand. Therefore, biology *does* conflate overall function with rate acceleration, and understanding how the features that control either are crucial to understanding how enzymes work. Recent years have shown how limited our knowledge of this area actually is: increasing awareness of enzyme multifunctionality for sometimes completely unexpected side reactions [280] show how little we know about the factors driving enzyme selectivity and specificity; the (sometimes quite drastic) unexpected effects of distant mutations [125,126] show how little we understand about long-range information transfer in biological systems; exhaustive analysis of enzyme kinetic parameters show that the popular enzymes of the 20th century, such as acetylcholine esterase and carbonic anhydrase, fall well away from the median of both k_{cat} and k_{cat}/K_M

distributions [281]. This shows that the "average" enzyme has not evolved to be as incredibly efficient as these "super enzymes", and we need to rethink how we approach our understanding of the evolution and the activity of average enzymes. This is not necessarily to say that the enzyme toolbox has to be very different for average and "super enzymes" – possibly, some enzymes just use the same toolbox more efficiently than others. Finally, enzymes work in networks with potentially discontinuous activity patterns [282], where preventing the enzyme from producing too much product can be as important as catalyzing the actual chemical reaction. As with the problems highlighted above, once again, theory is likely to play a major role in helping resolve these questions as we move further and further into the 21st Century.

5. Conclusions and future directions

One can only imagine the joy Lifson would have felt, had he lived to see the seeds he planted at the Weizmann Institute in the late 1960s culminate in the 2013 Nobel Prize in chemistry. In those early days when the first lines of computer codes were written, Karplus, Levitt, and Warshel laid the foundation for the entire field of biosimulations, which we today take for granted. Since the early work on myoglobin, lysozyme, and the retinal chromophore, the field has advanced immensely. There have been vast improvements in methodology accompanied by concomitant improvements in computer hardware. There is an abundance of software packages, which allows one to perform state-of-the-art simulations of chemical and biological processes at various levels of accuracy and time-scales. Today, one can model the direct folding of small proteins using brute force MD simulations as it happens in nature [283], and compute enzyme rates with near chemical accuracy [284]. Remarkably, with all the improvements in software and hardware and the huge number of researchers in the field, the early insights and views of the 2013 Chemistry Laureates on enzyme catalysis and protein dynamics remain dominant today [20–22]. The concept of enzyme pre-organization in biological catalysis introduced by Warshel in 1978 is a fundamental view on enzyme catalysis, and the idea of protein flexibility, which was born out of the first MD simulations of proteins conducted by Karplus, Levitt, and Warshel, is uniformly accepted today.

Fortunately for younger researchers (like the current authors, none of whom were born when the field of biosimulations was conceived), the field of computational enzymology is a dynamic one. At the methodological level, there are continuous improvements being made, which allow increasingly accurate simulations to be performed. At the biological level there are many unanswered questions regarding enzyme catalysis: How come enzymes are so big? Why are amino acid residues distal to the active site sometimes so catalytically important? Are protein dynamics important for chemistry? How come enzymes are often so promiscuous? In this Perspective, we have tried to highlight some of the multitude of ways in which we can now address such problems, however, these questions and many more are certain to keep the multi-scale modeler busy for some time to come.

Acknowledgements

References

[1] R. Wolfenden, M.J. Snider, The depth of chemical time and the power of enzymes as catalysts, Acc. Chem. Res. 34 (2001) 938–945.

[2] D. Hilvert, Design of protein catalysts, Annu. Rev. Biochem. 82 (2013) 447–470.

[3] W. Aehle, Enzymes in Industry, Wiley-VHC, 2007.

[4] E. Fischer, Einfluss der configuration auf der Wirkung der Enzyme, Ber. Dtsch. Chem. Ges 27 (1894) 2985–2993.

[5] H. Eyring, M. Polanyi, Über einfache Gasreaktionen, Z. Phys. Chem. B12 (1931) 279–311.

[6] L. Pauling, Nature of forces between large molecules of biological interest, Nature 161 (1948) 707–709.

[7] C.C. Blake, D.F. Koenig, G.A. Mair, A.C. North, D.C. Phillips, V.R. Sarma, Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Ångström resolution, Nature 206 (1965) 757–761.

[8] K.B. Wiberg, A scheme for strain energy minimization. Application to the cycloalkanes1, J. Am. Chem. Soc. 87 (1965) 1070–1078.

[9] G. Némethy, H.A. Scheraga, Theoretical determination of sterically allowed conformations of a polypeptide chain by a computer method, Biopolymers 3 (1965) 155–184.

[10] N.L. Allinger, M.A. Miller, L.W. Chow, R.A. Ford, J.C. Graham, The calculated electronic spectra and structures of some cyclic conjugated hydrocarbons[1,2], J. Am. Chem. Soc. 87 (1965) 3430–3435.

[11] N.L. Allinger, M.A. Miller, F.A. Van Catledge, J.A. Hirsch, Conformational analysis. LVII. The calculation of the conformational structures of hydrocarbons by the Westheimer-Hendrickson-Wiberg method, J. Am. Chem. Soc. 89 (1967) 4345–4357.

[12] M. Bixon, S. Lifson, Potential functions and conformations in cycloalkanes, Tetrahedron 23 (1967) 769–784.

[13] S. Lifson, A. Warshel, Consistent force field for calculations of conformations, vibrational spectra, and enthalpies of cycloalkane and n-alkane molecules, J. Chem. Phys. 49 (1968) 5116–5129.

[14] M. Levitt, A. Warshel, Computer simulation of protein folding, Nature 253 (1975) 694–698.

[15] J.A. McCammon, B.R. Gelin, M. Karplus, Dynamics of folded proteins, Nature 267 (1977) 585–590.

[16] H. Bisswanger, Enzyme Kinetics: Principles and Methods, second edition, Wiley, 2008.

[17] D. Kern, E.Z. Eisenmesser, M. Wolf-Watz, Enzyme dynamics during catalysis measured by NMR spectroscopy, in: L.J. Thomas (Ed.), Meth. Enzymol., vol. 394, Academic Press, 2005, pp. 507–524.

[18] M.M. Ling, B.H. Robinson, Approaches to DNA mutagenesis: an overview, Anal. Biochem. 254 (1997) 157–178.

[19] J. Alegre-Cebollada, R. Perez-Jimenez, P. Kosuri, J.M. Fernandez, Single-molecule force spectroscopy approach to enzyme catalysis, J. Biol. Chem. 285 (2010) 18961–18966.

[20] A. Warshel, Nobel lecture: Computer simulations of biological functions: from enzymes to molecular machines, Nobelprize.org. Nobel Media AB, 2013 http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2013/warshel-lecture.html

[21] M. Levitt, Nobel lecture: Birth & future of multi-scale modeling of biological macromolecules, Nobelprize.org. Nobel Media AB, 2013 http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2013/levitt-lecture.html

[22] M. Karplus, Nobel lecture: Development of multiscale models for complex chemical systems: From H+H₂ to biomolecules, Nobelprize.org. Nobel Media AB, 2013 http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2013/karplus-lecture.html

[23] B.J. Alder, T.E. Wainwright, Phase transition for a hard sphere system, J. Chem. Phys. 27 (1957) 1208–1209.

[24] B.J. Alder, T.E. Wainwright, Phase transition in elastic disks, Phys. Rev. (1962) 359–361.

[25] A. Rahman, Correlations in the motion of atoms in liquid argon, Phys. Rev. 136 (1964) A405-A11.

[26] A. Rahman, F.H. Stillinger, Molecular dynamics study of liquid water, J. Chem. Phys. 55 (1971) 3336–3359.

[27] F.H. Stillinger, A. Rahman, Improved simulation of liquid water by molecular dynamics, J. Chem. Phys. 60 (1974) 1545–1557.

[28] L.V. Woodcock, Isothermal molecular dynamics calculations for liquid salts, Chem. Phys. Lett. 10 (1971) 257–261.

[29] J.P. Ryckaert, A. Bellemans, Molecular dynamics of liquid n-butane near its boiling point, Chem. Phys. Lett. 30 (1975) 123–125.

[30] R.J. Solomon, The Computer Pioneers: Weizmann Institute Video Oral History., 1983.

[31] M. Levitt, S. Lifson, Refinement of protein conformations using a macromolecular energy minimization procedure, J. Mol. Biol. 46 (1969) 269–279.

[32] M. Karplus, Spinach on the ceiling: a theoretical chemist's return to biology, Annu. Rev. Biophys. Biomol. Struct. 35 (2006) 1–47.

[33] A. Warshel, A. Bromberg, Oxidation of 4a, 4b-dihydrophenanthrenes. III. A theoretical study of the large kinetic isotope effect of deuterium in the initiation step of the thermal reaction with oxygen, J. Chem. Phys. 52 (1970) 1262–1269.

[34] A. Warshel, Multiscale modeling of biological functions: from enzymes to molecular machines (nobel lecture), Angew. Chem. Int. Ed. (2014) http://onlinelibrary.wiley.com/doi/10.1002/anie.201403689/abstract

[35] H.M. Berman, G.J. Kleywegt, H. Nakamura, J.L. Markley, How community has shaped the Protein Data Bank, Structure (London, England: 1993). 21 (2013) 1485–1491.

[36] B. Honig, M. Karplus, Implications of torsional potential of retinal isomers for visual excitation, Nature 229 (1971) 558–560.

[37] A. Warshel, M. Karplus, Calculation of ground and excited state potential surfaces of conjugated molecules. I. Formulation and parametrization, J. Am. Chem. Soc. 94 (1972) 5612–5625.

[38] A. Warshel, M. Levitt, Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme, J. Mol. Biol. 103 (1976) 227–249.

[39] M. Levitt, A simplified representation of protein conformations for rapid simulation of protein folding, J. Mol. Biol. 104 (1976) 59–107.

[40] B.R. Gelin, M. Karplus, Sidechain torsional potentials and motion of amino acids in porteins: bovine pancreatic trypsin inhibitor, Proc. Natl. Acad. Sci. U. S. A. 72 (1975) 2002–2006.

[41] H.A. Scheraga, Calculations of conformations of polypeptides, Adv. Phys. Org. Chem. 6 (1968) 103–184.

[42] P. van der Ploeg, H.J.C. Berendsen, Molecular dynamics simulation of a bilayer membrane, J. Chem. Phys. 76 (1982) 3271–3276.

[43] W.F.V. Gunsteren, H.J.C. Berendsen, J. Hermans, W.G.J. Hol, J.P.M. Postma, Computer simulation of the dynamics of hydrated protein crystals and its comparison with X-ray data, Proc. Natl. Acad. Sci. U. S. A. 80 (1983) 4315–4319.

[44] A.T. Brünger, J. Kuriyan, M. Karplus, Crystallographic R factor refinement by molecular dynamics, Science 235 (1987) 458–460.

[45] L. Nilsson, G.M. Clore, A.M. Gronenborn, A.T. Brunger, M. Karplus, Structure refinement of oligonucleotides by molecular dynamics with nuclear Overhauser effect interproton distance restraints: application to 5′ d(C-G-T-A-C-G)₂, J. Mol. Biol. 188 (1986) 455–475.

[46] W.F. Van Gunsteren, H.J.C. Berendsen, R.G. Geurtsen, H.R.J. Zwinderman, A molecular dynamics computer simulation of an eight base-pair DNA fragment in aqueous solution: comparison with experimental two-dimensional NMR data, Ann. N.Y. Acad. Sci. 482 (1986) 287–303.

[47] H.J. Berendsen, W.F. Van Gunsteren, H.R. Zwinderman, R.G. Geurtsen, Simulations of proteins in water, Ann. N.Y. Acad. Sci. 482 (1986) 269–286.

[48] M. Levitt, R. Sharon, Accurate simulation of protein dynamics in solution, Proc. Natl. Acad. Sci. U. S. A. 85 (1988) 7557–7561.

[49] W.F. van Gunsteren, H.J.C. Berendsen, Computer simulation as a tool for tracing the conformational differences between proteins in solution and in the crystalline state, J. Mol. Biol. 176 (1984) 559–564.

[50] A. Warshel, R.M. Weiss, An empirical valence bond approach for comparing reactions in solutions and in enzymes, J. Am. Chem. Soc. 102 (1980) 6218–6226.

[51] A. Warshel, Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites, J. Biol. Chem. 273 (1998) 27035–27038.

[52] A.V. Pisliakov, J. Cao, S.C.L. Kamerlin, A. Warshel, Enzyme millisecond conformational dynamics do not catalyze the chemical step, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 17359–17364.

[53] S.C.L. Kamerlin, P.K. Sharma, Z.T. Chu, A. Warshel, Ketosteroid isomerase provides further support for the idea that enzymes work by electrostatic preorganization, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 4075–4080.

[54] A.J. Adamczyk, J. Cao, S.C.L. Kamerlin, A. Warshel, The catalytic power of dihydrofolate reductase and other enzymes arises from electrostatic preorganization, not conformational motions, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 14115–14120.

[55] K.F. Wong, T. Selzer, S.J. Benkovic, S. Hammes-Schiffer, Impact of distal mutations on the network of coupled motions correlated to hydride transfer in dihydrofolate reductase, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 6807–6812.

[56] D.K. Chakravorty, A.V. Soudackov, S. Hammes-Schiffer, Hybrid quantum/classical molecular dynamics simulations of the proton transfer reactions catalyzed by ketosteroid isomerase: Analysis of hydrogen bonding, conformational motions, and electrostatics, Biochemistry 48 (2009) 10608–10619.

[57] D.K. Chakravorty, S. Hammes-Schiffer, Impact of mutation on proton transfer reactions in ketosteroid isomerase: insights from molecular dynamics simulations, J. Am. Chem. Soc. 132 (2010) 7549–7555.

[58] W.L. Jorgensen, D.S. Maxwell, J. Tirado-Rives, Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids, J. Am. Chem. Soc. 118 (1996) 11225–11236.

[59] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926–935.

[60] M.J. Field, P.A. Bash, M. Karplus, A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations, J. Comput. Chem. 11 (1990) 700–733.

[61] Y. Mo, J. Gao, An ab initio molecular orbital–valence bond (MOVB) method for simulating chemical reactions in solution, J. Phys. Chem. A 104 (2000) 3012–3020.

[62] M. Svensson, S. Humbel, R.D.J. Froese, T. Matsubara, S. Sieber, K. Morokuma, 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-Bu)₃)₂ + H₂ oxidative addition, J. Phys. Chem. 100 (1996) 19357–19363.

[63] U.C. Singh, P.A. Kollman, 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. 7 (1986) 718–730.

[64] H. Lin, D.G. Truhlar, QM/MM: what have we learned, where are we, and where do we go from here? Theo. Chem. Acc. 117 (2007) 185–199.

[65] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, CHARMM: A program for macromolecular energy, minimization, and dynamics calculations, J. Comput. Chem. 4 (1983) 187–217.

[66] B.R. Brooks, C.L. Brooks, A.D. Mackerell, L. Nilsson, R.J. Petrella, B. Roux, et al., CHARMM: the biomolecular simulation program, J. Comput. Chem. 30 (2009) 1545–1614.

[67] D.A. Case, T.E. Cheatham, T. Darden, H. Gohlke, R. Luo, K.M. Merz, et al., The Amber biomolecular simulation programs, J. Comput. Chem. 26 (2005) 1668–1688.

[68] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation, J. Chem. Theory Comput. 4 (2008) 435–447.

[69] W.L. Jorgensen, J. Tirado-Rives, Molecular modeling of organic and biomolecular systems using BOSS and MCPRO, J. Comput. Chem. 26 (2005) 1689–1700.

[70] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, et al., Scalable molecular dynamics with NAMD, J. Comput. Chem. 26 (2005) 1781–1802.

[71] H.M. Senn, W. Thiel, QM/MM methods for biomolecular systems, Angew. Chem. Int. Ed. 48 (2009) 1198–1229.

[72] M.W. van der Kamp, A.J. Mulholland, Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology, Biochemistry 52 (2013) 2708–2728.

[73] R.A. Friesner, V. Guallar, Ab initio quantum chemical and mixed quantum mechanics/molecular mechanics (QM/MM) methods for studying enzymatic catalysis, Annu. Rev. Phys. Chem. 56 (2005) 389–427.

[74] H.J. Kulik, N. Luehr, I.S. Ufimtsev, T.J. Martinez, Ab initio quantum chemistry for protein structures, J. Phys. Chem. B 116 (2012) 12501–12509.

[75] H.M. Senn, W. Thiel, QM/MM methods for biological systems, in: M. Reiher (Ed.), Atomistic Approaches in Modern Biology, Springer, Berlin Heidelberg, 2007, pp. 173–290.

[76] A. Ardevol, C. Rovira, The molecular mechanism of enzymatic glycosyl transfer with retention of configuration: evidence for a short-lived oxocarbenium-like species, Angew. Chem. Int. Ed. 50 (2011) 10897–10901.

[77] S. Piana, D. Bucher, P. Carloni, U. Rothlisberger, Reaction mechanism of HIV-1 protease by hybrid carparrinello/classical MD simulations, J. Phys. Chem. B 108 (2004) 11139–11149.

[78] R. Car, M. Parrinello, Unified approach for molecular dynamics and density-functional theory, Phys. Rev. Lett. 55 (1985) 2471–2474.

[79] Y.-l. Lin, J. Gao, Kinetic isotope effects of L-dopa decarboxylase, J. Am. Chem. Soc. 133 (2011) 4398–4403.

[80] A. Vardi-Kilshtain, D. Doron, D.T. Major, Quantum and classical simulations of orotidine monophosphate decarboxylase: support for a direct decarboxylation mechanism, Biochemistry 52 (2013) 4382–4390.

[81] I. Rossi, D.G. Truhlar, Parameterization of NDDO wavefunctions using genetic algorithms. An evolutionary approach to parameterizing potential energy surfaces and direct dynamics calculations for organic reactions, Chem. Phys. Lett. 233 (1995) 231–236.

[82] D.T. Major, A. Heroux, A.M. Orville, M.P. Valley, P.F. Fitzpatrick, J. Gao, Differential quantum tunneling contributions in nitroalkane oxidase catalyzed and the uncatalyzed proton transfer reaction, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 20734–20739.

[83] D.T. Major, D.M. York, J. Gao, Solvent polarization and kinetic isotope effects in nitroethane deprotonation and implications to the nitroalkane oxidase reaction, J. Am. Chem. Soc. 127 (2005) 16374–16375.

[84] D.T. Major, K. Nam, J. Gao, Transition state stabilization and alpha-amino carbon acidity in alanine racemase, J. Am. Chem. Soc. 128 (2006) 8114–8115.

[85] D. Doron, D.T. Major, A. Kohen, W. Thiel, X. Wu, Hybrid quantum and classical simulations of the dihydrofolate reductase catalyzed hydride transfer reaction on an accurate semi-empirical potential energy surface, J. Chem. Theory Comput. 7 (2011) 3420–3437.

[86] A. Vardi-Kilshtain, D.T. Major, A. Kohen, H. Engel, D. Doron, Hybrid quantum and classical simulations of the formate dehydrogenase catalyzed hydride transfer reaction on an accurate semiempirical potential energy surface, J. Chem. Theory Comput. 8 (2012) 4786–4796.

[87] M.W. van der Kamp, R.F. Chaudret, A.J. Mulholland, QM/MM modelling of ketosteroid isomerase reactivity indicates that active site closure is integral to catalysis, FEBS J. 280 (2013) 3120–3131.

[88] L.Y.P. Luk, J.J. Ruiz-Pernía, W.M. Dawson, M. Roca, E.J. Loveridge, D.R. Glowacki, et al., Unraveling the role of protein dynamics in dihydrofolate reductase catalysis, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) 16344–16349.

[89] Q. Cui, H. Guo, M. Karplus, Combining ab initio and density functional theories with semiempirical methods, J. Chem. Phys. 117 (2002) 5617–5631.

[90] J.C. Corchado, J. Espinosa-Garcia, W.-P. Hu, I. Rossi, D.G. Truhlar, Dual-level reaction-path dynamics (the /// approach to VTST with semiclassical tunneling), Application to OH + NH₃ à H₂O + NH₂, J. Phys. Chem. 99 (1995) 687–694.

[91] K.A. Nguyen, I. Rossi, D.G. Truhlar, A dual level Shepard interpolation method for generating potential energy surfaces for dynamics calculations, J. Chem. Phys. 103 (1995) 5522–5530.

[92] A. Warshel, Dynamics of enzymatic reactions, Proc. Natl. Acad. Sci. U. S. A. 81 (1984) 444–448.

[93] S.C.L. Kamerlin, A. Warshel, The empirical valence bond model: theory and applications, WiRE: Comp. Mol. Sci. 1 (2011) 30–45.

[94] S.C.L. Kamerlin, A. Warshel, The EVB as a quantitative tool for formulating simulations and analyzing biological and chemical reactions, Faraday Discuss. 145 (2010) 71–106.

[95] S. Sato, On a new method of drawing the potential energy surface, J. Chem. Phys. 23 (1955) 592–593.

[96] H.B. Schlegel, J.L. Sonnenberg, Empirical valence-bond models for reactive potential energy surfaces using distributed gaussians, J. Chem. Theory Comput. 2 (2006) 905–911.

[97] Y.T. Chang, W.H. Miller, An empirical valence bond model for constructing global potential energy surfaces for chemical reactions of polyatomic molecular systems, J. Phys. Chem. 94 (1990) 5884–5888.

[98] Y.T. Chang, C. Minichino, W.H. Miller, Classical trajectory studies of the molecular dissociation dynamics of formaldehyde: $H_2CO \rightarrow H_2 + CO$, J. Chem. Phys. 96 (1991) 4341–4355.

[99] R. Valero, L. Song, J. Gao, D.G. Truhlar, Perspective on diabatic models of chemical reactivity as illustrated by the gas-phase S_N2 reaction of acetate ion with 1,2-dichloroethane, J. Chem. Theory Comput. 5 (2009) 1–22.

[100] A. Cembran, A. Payaka, Y.-l. Lin, W. Xie, Y. Mo, L. Song, et al., A non-orthogonal block-localized effective Hamiltonian approach for chemical and enzymatic reactions, J. Chem. Theory Comput. 6 (2010) 2242–2251.

[101] A. Shurki, H.A. Crown, Hybrid *ab initio* VB/MM method – a valence bond ride through classical landscapes, J. Phys. Chem. B 109 (2005) 23638–23644.

[102] A. Sharir-Ivry, T. Shnerb, M. Štrajbl, A. Shurki, VB/MM protein landscapes: a study of the S_N2 reaction in haloalkane dehalogenase, J. Phys. Chem. B 114 (2010) 2212–2218.

[103] L.S. Devi-Kesavan, M. Garcia-Viloca, J. Gao, Semiempirical QM/MM potential with simple valence bond (SVB) for enzyme reactions. Application to the nucleophilic addition reaction in haloalkane dehalogenase, Theor. Chem. Acc. 109 (2003) 133–139.

[104] H. Lin, Y. Zhao, O. Tishchenko, D.G. Truhlar, Multiconfiguration molecular mechanics based on combined quantum mechanical and molecular mechanical calculations, J. Chem. Theory Comput. 2 (2006) 1237–1254.

[105] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, et al., Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties, Phys. Rev. B 58 (1998) 7260–7268.

[106] D.F.A.R. Dourado, P.A. Fernandes, M.J. Ramos, B. Mannervik, Mechanism of glutathione transferase P1-1-catalyzed activation of the prodrug canfosfamide (TLK286, TELCYTA), Biochemistry 52 (2013) 8069–8078.

[107] M.M. El-Hendawy, N.J. English, D.A. Mooney, Comparative studies for evaluation of CO_2 fixation in the cavity of the Rubisco enzyme using QM, QM/MM and linear-scaling DFT methods, J. Mol. Model. 19 (2013) 2329–2334.

[108] T. Wesolowski, R.P. Muller, A. Warshel, *Ab initio* frozen density functional calculations of proton transfer reactions in solution, J. Phys. Chem. 100 (1996) 15444–15449.

[109] R.P. Muller, T. Wesolowski, A. Warshel, Calculations of chemical processes in solution by Density-functional and other Quantum-Mechanical Techniques, in: M. Springborg (Ed.), Density-Functional Methods: In Chemistry and Materials Science, John Wiley & Sons, Chichester, 1997.

[110] T.A. Wesolowski, A. Warshel, Frozen density functional approach for *ab initio* calculations of solvated molecules, J. Phys. Chem. 97 (1993) 8050–8053.

[111] T.A. Wesolowski, J. Weber, Kohn-Sham equations with constrained electron density: an iterative evaluation of the ground-state electron density of interacting molecules, Chem. Phys. Lett. 248 (1996) 71–76.

[112] M. Štrajbl, A. Shurki, M. Kato, A. Warshel, N.A.C. Apparent, effect in chorismate mutase reflects electrostatic transition state stabilization, J. Am. Chem. Soc. 125 (2003) 10228–10237.

[113] G. Hong, M. Štrajbl, T.A. Wesolowski, A. Warshel, Constraining the electron densities in DFT method as an effective way for *ab initio* studies of metal catalyzed reactions, J. Comput. Chem. 21 (2000) 1554–1561.

[114] M.H.M. Olsson, G. Hong, A. Warshel, Frozen density functional free energy simulations of redox proteins: computational studies of the reduction potential of plastocyanin and rusticyanin, J. Am. Chem. Soc. 125 (2003) 5025–5039.

[115] Y. Xiang, A. Warshel, Quantifying free energy profiles of proton transfer reactions in solution and proteins by using a diabatic FDFT mapping, J. Phys. Chem. B 112 (2008) 1007–1015.

[116] T.A. Wesolowski, One-electron equations for embedded electron density: challenge for theory and practical payoffs in multi-level modeling of complex polyatomic systems, in: J. Leszczynski (Ed.), Computational Chemistry: Reviews of Current Trends, vol. 10, World Scientific, 2006, p. 82.

[117] C.R. Jacob, J. Neugebauer, Subsystem density-functional theory, WiRE: Comp. Mol. Sci. 4 (2013) 325–362.

[118] F.R. Manby, M. Stella, J.D. Goodpaster, T.F. Miller, A simple, exact density-functional-theory embedding scheme, J. Chem. Theory Comput. 8 (2012) 2564–2568.

[119] P.M. Siegbahn, F. Himo, Recent developments of the quantum chemical cluster approach for modeling enzyme reactions, J. Biol. Inorg. Chem. 14 (2009) 643–651.

[120] L. Noodleman, T. Lovell, W.-G. Han, J. Li, F. Himo, Quantum chemical studies of intermediates and reaction pathways in selected enzymes and catalytic synthetic systems, Chem. Rev. 104 (2004) 459–508.

[121] P.E.M. Siegbahn, F. Himo, Recent developments of the quantum chemical cluster approach for modeling enzyme reactions, J. Biol. Inorg. Chem. 14 (2009) 643–651.

[122] P.E.M. Siegbahn, A structure-consistent mechanism for dioxygen formation in photosystem II, Chemistry 14 (2008) 8290–8302.

[123] P.E.M. Siegbahn, Structures and energetics for O_2 formation in photosystem II, Acc. Chem. Res. 42 (2009) 1871–1880.

[124] H. Chen, W.Z. Lai, S. Shaik, Multireference and multiconfiguration *ab initio* methods in heme-related systems: what have we learned so far? J. Phys. Chem. B 115 (2011) 1727–1742.

[125] M. Klvana, D.L. Murphy, P. Jerabek, M.F. Goodman, A. Warshel, J.B. Sweasy, et al., Catalytic effects of mutations of distant protein residues in human DNA polymerase β: theory and experiment, Biochemistry 51 (2012) 8829–8843.

[126] K. El Omari, S. Liekens, L.E. Bird, J. Balzarini, D.K. Stammers, Mutations distal to the substrate site can affect varicella zoster virus thymidine kinase activity: implications for drug design, Mol. Pharmacol. 69 (2006) 1891–1896.

[127] K.H. Hopmann, F. Himo, Quantum chemical modeling of the dehalogenation reaction of haloalcohol dehalogenase, J. Chem. Theory Comput. 4 (2008) 1129–1137.

[128] R. Sevastik, C.P. Whitman, F. Himo, Reaction mechanism of cis-3-chloroacrylic acid dehalogenase – a theoretical study, Biochemistry 48 (2009) 9641–9649.

[129] S.F. Sousa, P.A. Fernandes, M.J. Ramos, Computational enzymatic catalysis – clarifying enzymatic mechanisms with the help of computers, Phys. Chem. Chem. Phys. 14 (2012) 12431–12441.

[130] L. Wang, N.M. Goodey, S.J. Benkovic, A. Kohen, Coordinated effects of distal mutations on environmentally coupled tunneling in dihydrofolate reductase, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 15753–15758.

[131] Z. Wang, T. Abeysinghe, J.S. Finer-Moore, R.M. Stroud, A. Kohen, A remote mutation affects the hydride transfer by disrupting concerted protein motions in thymidylate synthase, J. Am. Chem. Soc. 134 (2012) 17722–17730.

[132] D.T. Major, J. Gao, A combined quantum mechanical and molecular mechanical study of the reaction mechanism and alpha-amino acidity in alanine racemase, J. Am. Chem. Soc. 128 (2006) 16345–16357.

[133] V. Lopez-Canut, M. Roca, J. Bertran, V. Moliner, I. Tunon, Promiscuity in alkaline phosphatase superfamily. Unraveling evolution through molecular simulations, J. Am. Chem. Soc. 133 (2011) 12050–12062.

[134] C.M. Isborn, A.W. Götz, M.A. Clark, R.C. Walker, T.J. Martínez, Electronic absorption spectra from MM and *ab initio* QM/MM molecular dynamics: environmental effects on the absorption spectrum of photoactive yellow protein, J. Chem. Theory Comput. 8 (2012) 5092–5106.

[135] J. Gao, X. Xia, A priori evaluation of aqueous polarization effects through Monte Carlo QM-MM simulations, Science 258 (1992) 631–635.

[136] T.A. Halgren, W. Damm, Polarizable force fields, Curr. Opin. Struct. Biol. 11 (2001) 236–242.

[137] A. Warshel, M. Kato, A.V. Pisliakov, Polarizable force fields: history, test cases, and prospects, J. Chem. Theory Comput. 3 (2007) 2034–2045.

[138] W. Xie, M. Orozco, D.G. Truhlar, J. Gao, X-Pol potential: an electronic structure-based force field for molecular dynamics simulation of a solvated protein in water, J. Chem. Theory Comput. 5 (2009) 459–467.

[139] I. Antes, W. Thiel, Adjusted connection atoms for combined quantum mechanical and molecular mechanical methods, J. Phys. Chem. A 103 (1999) 9290–9295.

[140] Y. Zhang, T.-S. Lee, W. Yang, A pseudobond approach to combining quantum mechanical and molecular mechanical methods, J. Chem. Phys. 110 (1999) 46–54.

[141] A. Monari, J.-L. Rivail, X. Assfeld, Theoretical modeling of large molecular systems. Advances in the Local Self Consistent Field method for mixed quantum mechanics/molecular mechanics calculations, Acc. Chem. Res. 46 (2013) 596–603.

[142] J. Gao, P. Amara, C. Alhambra, M.J. Field, A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations, J. Phys. Chem. A 102 (1998) 4714–4721.

[143] J. Pu, J. Gao, D.G. Truhlar, Generalized hybrid-orbital method for combining density functional theory with molecular mechanicals, Chem. Phys. Chem. 6 (2005) 1853–1865.

[144] M.A. Rohrdanz, W. Zheng, C. Clementi, Discovering mountain passes via torchlight: methods for the definition of reaction coordinates and pathways in complex macromolecular reactions, Annu. Rev. Phys. Chem. 64 (2013) 295–316.

[145] J. Gao, S. Ma, D.T. Major, K. Nam, J. Pu, D.G. Truhlar, Mechanisms and free energies of enzymatic reactions, Chem. Rev. 106 (2006) 3188–3209.

[146] R.A. More O'Ferrall, Relationships between E2 and E1$_c$B mechanisms of beta-elimination, J. Chem. Soc. B (1970) 274–277.

[147] W.P. Jencks, General acid-base catalysis of complex reactions in water, Chem. Rev. 72 (1972) 705–718.

[148] D. Doron, A. Kohen, D.T. Major, Collective reaction coordinate for hybrid quantum and molecular mechanics simulations: a case study of the hydride transfer in dihydrofolate reductase, J. Chem. Theory Comput. 8 (2012) 2484–2496.

[149] S. Hammes-Schiffer, Selected theoretical models and computational methods for enzymatic tunneling, in: R.K. Alleman, N. Scrutton (Eds.), Quantum Tunneling in Enzyme Catalyzed Reactions, Royal Society of Chemistry, Cambridge, UK, 2009.

[150] L. Mones, P. Kulhanek, I. Simon, A. Laio, M. Fuxreiter, The energy gap as a universal reaction coordinate for the simulation of chemical reactions, J. Phys. Chem. B 113 (2009) 7867–7873.

[151] R. Garcia-Meseguer, S. Marti, J.J. Ruiz-Pernia, V. Moliner, I. Tunon, Studying the role of protein dynamics in an S_N2 enzyme reaction using free-energy surfaces and solvent coordinates, Nat. Chem. 5 (2013) 566–571.

[152] J. Pu, J. Gao, D.G. Truhlar, Multidimensional tunneling, recrossing, and the transmission coefficient for enzymatic reactions, Chem. Rev. 106 (2006) 3140–3169.

[153] M. Klähn, S. Braun-Sand, E. Rosta, A. Warshel, On possible pitfalls in *ab initio* quantum mechanics/molecular mechanics minimization approaches for studies of enzymatic reactions, J. Phys. Chem. B 109 (2005) 15645–15650.

[154] G.M. Torrie, J.P. Valleau, Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling, J. Comput. Phys. 23 (1977) 187–199.

[155] M. Mezei, Adaptive umbrella sampling: self-consistent determination of the non-Boltzmann bias, J. Comp. Phys. 68 (1987) 237–248.

[156] R.W. Hooft, B.P. van Eijck, J. Kroon, An adaptive umbrella sampling procedure in conformational analysis using molecular dynamics and its application to glycol, J. Chem. Phys. 97 (1992) 6690.

[157] C. Bartels, M. Karplus, Multidimensional adaptive umbrella sampling: applications to main chain and side chain peptide conformations, J. Comp. Chem. 18 (1997) 1450–1462.

[158] C. Bartels, M. Karplus, Probability distributions for complex systems: adaptive umbrella sampling of the potential energy, J. Phys. Chem. B 102 (1998) 865–880.

[159] S. Kumar, J.M. Rosenberg, D. Bouzida, R.H. Swendsen, P.A. Kollman, The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method, J. Comput. Chem. 13 (1992) 1011–1021.

[160] R.W. Zwanzig, High-temperature equation of state by a perturbation method. I. Nonpolar gases, J. Chem. Phys. 22 (1954) 1420–1426.

[161] J. Kästner, H.M. Senn, S. Thiel, N. Otte, W. Thiel, QM/MM free-energy perturbation compared to thermodynamic integration and umbrella sampling: application to an enzymatic reaction, J. Chem. Theory Comput. 2 (2006) 452–461.

[162] A. Warshel, Computer Modeling of Chemical Reactions in Enzymes and Solutions, John Wiley & Sons, New York, 1991.

[163] A. Laio, M. Parrinello, Escaping free-energy minima, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 12562–12566.

[164] T. Huber, A.E. Torda, W.F. van Gunsteren, Local elevation: a method for improving the searching properties of molecular dynamics simulation, J. Comput-Aided Mol. Des. 8 (1994) 695–708.

[165] H. Grubmuller, Predicting slow structural transitions in macromolecular systems: Conformational flooding, Phys. Rev. E Stat. Nonlin. Soft Matter Phys. 52 (1995) 2893–2906.

[166] C.L. Stanton, I.F. Kuo, C.J. Mundy, T. Laino, K.N. Houk, QM/MM metadynamics study of the direct decarboxylation mechanism for orotidine-5'-monophosphate decarboxylase using two different QM regions: acceleration too small to explain rate of enzyme catalysis, J. Phys. Chem. B 111 (2007) 12573–12581.

[167] L. Petersen, A. Ardevol, C. Rovira, P.J. Reilly, Mechanism of cellulose hydrolysis by inverting GH8 endoglucanases: a QM/MM metadynamics study, J. Phys. Chem. B 113 (2009) 7331–7339.

[168] X. Biarnés, A. Ardevol, J. Iglesias-Fernández, A. Planas, C. Rovira, Catalytic itinerary in 1,3-1,4-beta-glucanase unraveled by QM/MM metadynamics. Charge is not yet fully developed at the oxocarbenium ion-like transition state, J. Am. Chem. Soc. 133 (2011) 20301–20309.

[169] X.L. Pan, W. Liu, J.Y. Liu, Mechanism of the glycosylation step catalyzed by human alpha-galactosidase: a QM/MM metadynamics study, J. Phys. Chem. B 117 (2013) 484–489.

[170] V. Rojas-Cervellera, A. Ardevol, M. Boero, A. Planas, C. Rovira, Formation of a covalent glycosyl-enzyme species in a retaining glycosyltransferase, Chemistry 19 (2013) 14018–14023.

[171] J. Yu, T. Ha, K. Schulten, Conformational model of the Holliday junction transition deduced from molecular dynamics simulations, Nucl. Acids Res. 32 (2004) 6683–6695.

[172] F.L. Gervasio, A. Laio, M. Parrinello, Flexible docking in solution using metadynamics, J. Am. Chem. Soc. 127 (2005) 2600–2607.

[173] B. Ensing, M.L. Klein, Perspective on the reactions between F⁻ and CH₃CH₂F: The free energy landscape of the E2 and S_N2 reaction channels, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 6755–6759.

[174] P. Tiwary, M. Parrinello, From metadynamics to dynamics, Phys. Rev. Lett. 111 (2013) 230602.

[175] M. Salvalaglio, P. Tiwary, M. Parrinello, Assessing the reliability of the dynamics reconstructed from metadynamics, J. Chem. Theory Comput. 10 (2014) 1420–1425.

[176] D. Sheppard, R. Terrell, G. Henkelman, Optimization methods for finding minimum energy paths, J. Chem. Phys. 128 (2008) 134106.

[177] E. Weinan, W. Ren, E. Vanden-Eijnden, String method for the study of rare events, Phys. Rev. B 66 (2002) 52301–52305.

[178] L. Maragliano, A. Fischer, E. Vanden-Eijnden, G. Ciccoti, String method in collective variables: minimum free energy paths and isocommittor surfaces, J. Chem. Phys. 125 (2006) 24106.

[179] Z. Chen, J.J. Concepcion, X. Hu, W. Yang, P.G. Hoertz, T.J. Meyer, Concerted O atom-proton transfer in the O-O bond forming step in water oxidation, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 7225–7229.

[180] A. Aleksandrov, M. Field, A hybrid elastic band string algorithm for studies of enzymatic reactions, Phys. Chem. Chem. Phys. 14 (2012) 12544–12553.

[181] K. Zinovjev, J.J. Ruiz-Pernía, I. Tuñón, Toward an automatic determination of enzymatic reaction mechanisms and their activation free energies, J. Chem. Theory Comput. 9 (2013) 3740–3749.

[182] L.Y.P. Luk, J.J. Ruiz-Pernía, W.M. Dawson, M. Roca, E.J. Loveridge, D.R. Glowacki, et al., Unraveling the role of protein dynamics in dihydrofolate reductase catalysis, Proc. Natl. Acad. Sci. U. S. A. (2013).

[183] C. Jarzynski, Nonequilibrium equality for free energy differences, Phys. Rev. Lett. 78 (1997) 2690–2693.

[184] C. Jarzynski, Equilibrium free-energy differences from nonequilibrium measurements: a master-equation approach, Phys. Rev. E 56 (1997) 5018–5035.

[185] P.G. Bolhuis, C. Dellago, D. Chandler, P. Geissler, Transition path sampling, throwing ropes over rough mountain passes, in the dark, Annu. Rev. Phys. Chem. 53 (2002) 291–318.

[186] D. Moroni, P.G. Bolhuis, T.S. van Erp, Rate constants for diffusive processes by partial path sampling, J. Chem. Phys. 120 (2004) 4055–4065.

[187] D. Moroni, T.S. van Erp, P.G. Bolhuis, Simultaneous computation of free energies and kinetics of rare events, Phys. Rev. E Stat. Nonlin. Soft Matter Phys. 71 (2005) 056709.

[188] T.S. van Erp, P.G. Bolhuis, Elaborating transition interface sampling methods, J. Comput. Phys. 205 (2005) 157–181.

[189] J. Rogal, P.G. Bolhuis, Multiple state transition path sampling, J. Chem. Phys. 129 (2008) 224107.

[190] N.B. Becker, R.J. Allen, P.R. ten Wolde, Non-stationary forward flux sampling, J. Chem. Phys. 136 (2012) 174118.

[191] D. Antonious, M.R. Abolfath, S.D. Schwartz, Transition path sampling study of classical rate-promoting vibrations, J. Chem. Phys. 121 (2004) 6442.

[192] R. Crehuet, M.J. Field, A transition path sampling study of the reaction catalyzed by the enzyme chorismate mutase, J. Phys. Chem. B 111 (2007) 5708–5718.

[193] B.C. Knott, M. Haddad Momeni, M.F. Crowley, L.F. Mackenzie, A.W. Götz, M. Sandgren, et al., The mechanism of cellulose hydrolysis by a two-step, retaining cellobiohydrolase elucidated by structural and transition path sampling studies, J. Am. Chem. Soc. 136 (2013) 321–329.

[194] D. Doron, A. Kohen, K. Nam, D.T. Major, How accurate are transition states from simulations of enzymatic reactions? J. Chem. Theory Comput. 10 (2014) 1863–1871.

[195] P.P. Ewald, Die berechnung optischer und elektrostatischer gitterpotentiale, Ann. Phys. 369 (1921) 253–287.

[196] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An N [center-dot] log(N) method for Ewald sums in large systems, J. Chem. Phys. 98 (1993) 10089–10092.

[197] K. Nam, J. Gao, D.M. York, An efficient linear-scaling Ewald method for long-range electrostatic interactions in combined QM/MM calculations, J. Chem. Theory Comput. 1 (2004) 2–13.

[198] A. Brünger, C.L. Brooks III, M. Karplus, Stochastic boundary conditions for molecular dynamics simulations of ST2 water, Chem. Phys. Lett. 105 (1984) 495–500.

[199] C.L. Brooks, A. Brünger, M. Karplus, Active site dynamics in protein molecules: a stochastic boundary molecular-dynamics approach, Biopolymers 24 (1985) 843–865.

[200] D. Beglov, B. Roux, Finite representation of an infinite bulk system: solvent boundary potential for computer simulations, J. Chem. Phys. (1994) 100.

[201] W. Im, S. Bernèche, B. Roux, Generalized solvent boundary potential for computer simulations, J. Chem. Phys. (2001) 114.

[202] T. Benighaus, W. Thiel, A general boundary potential for hybrid QM/MM simulations of solvated biomolecular systems, J. Chem. Theory Comput. 5 (2009) 3114–3128.

[203] T. Benighaus, W. Thiel, Long-range electrostatic effects in QM/MM studies of enzymatic reactions: application of the solvated macromolecule boundary potential, J. Chem. Theory Comput. 7 (2010) 238–249.

[204] E. Boulanger, W. Thiel, Solvent boundary potentials for hybrid QM/MM computations using classical drude oscillators: a fully polarizable model, J. Chem. Theory Comput. 8 (2012) 4527–4538.

[205] J. Zienau, Q. Cui, Implementation of the solvent macromolecule boundary potential and application to model and realistic enzyme systems, J. Phys. Chem. B 116 (2012) 12522–12534.

[206] S. Kuwajima, A. Warshel, The extended Ewald method: a general treatment of long-range electrostatic interactions in microscopic simulations, J. Chem. Phys. 89 (1988) 3751–3759.

[207] M. Schaefer, M. Karplus, A comprehensive analytical treatment of continuum electrostatics, J. Phys. Chem. 100 (1996) 1578–1599.

[208] H. Hu, A. Boone, W. Yang, Mechanism of OMP decarboxylation in orotidine 5′-monophosphate decarboxylase, J. Am. Chem. Soc. 130 (2008) 14493–14503.

[209] J.S. Hub, B.L. de Groot, H. Grubmüller, G. Groenhof, Quantifying artifacts in Ewald simulations of inhomogeneous systems with a net charge, J. Chem. Theory Comput. 10 (2013) 381–390.

[210] M.A. Kastenholz, P.H. Hünenberger, Influence of artificial periodicity and ionic strength in molecular dynamics simulations of charged biomolecules employing lattice-sum methods, J. Phys. Chem. B 108 (2003) 774–788.

[211] J. Mongan, D.A. Case, Biomolecular simulations at constant pH, Curr. Opin. Struct. Biol. 15 (2005) 157–163.

[212] A. Dryga, A. Warshel, Renormalizing SMD: the renormalization approach and its use in long time simulations and accelerated PAU calculations of macromolecules, J. Phys. Chem. B 114 (2010) 12720–12728.

[213] P.L. Freddolino, C.B. Harrison, Y. Liu, K. Schulten, Challenges in protein-folding simulations, Nat. Phys. 6 (2010) 751–758.

[214] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J.L. Klepeis, R.O. Dror, et al., Improved side-chain torsion potentials for the Amber ff99SB protein force field, Proteins 78 (2010) 1950–1958.

[215] R.B. Best, X. Zhu, J. Shim, P.E.M. Lopes, J. Mittal, M. Feig, et al., Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ1 and χ2 dihedral angles, J. Chem. Theory Comput. 8 (2012) 3257–3273.

[216] S. Piana, K. Lindorff-Larsen, D.E. Shaw, Atomic-level description of ubiquitin folding, Proc. Natl. Acad. Sci. U. S. A. (2013).

[217] T.J. Lane, D. Shukla, K.A. Beauchamp, V.S. Pande, To milliseconds and beyond: challenges in the simulation of protein folding, Curr. Opin. Struct. Biol. 23 (2013) 58–65.

[218] S. Piana, J.L. Klepeis, D.E. Shaw, Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations, Curr. Opin. Struct. Biol. 24 (2014) 98–105.

[219] A.J. Mulholland, Modelling enzyme reaction mechanisms, specificity and catalysis, Drug Discov. Today 10 (2005) 1393–1402.

[220] H. Hu, H. Liu, Pitfall in quantum mechanical/molecular mechanical molecular dynamics simulation of small solutes in solution, J. Phys. Chem. B 117 (2013) 6505–6511.

[221] R.N. Porter, M. Karplus, Potential energy surface for H3, J. Chem. Phys. 40 (1964) 1105–1115.

[222] M. Karplus, R.N. Porter, R.D. Sharma, Exchange reactions with activation energy. I. Simple barrier potential for (H, H2), J. Chem. Phys. 43 (1965) 3259–3287.

[223] D.C. Clary, Quantum dynamics of chemical reactions, Science 321 (2008) 789–791.

[224] R. Otto, J. Ma, A.W. Ray, J.S. Daluz, J. Li, H. Guo, et al., Imaging dynamics on the F + H₂O → HF + OH potential energy surfaces from wells to barriers, Science 343 (2014) 396–399.

[225] B.J. Gertner, R.M. Whitnell, K.R. Wilson, J.T. Hynes, Activation to the transition state: reactant and solvent energy flow for a model S_N2 reaction in water, J. Am. Chem. Soc. 113 (1991) 74–87.

[226] E. Neria, M. Karplus, Molecular dynamics of an enzyme reaction: proton transfer in TIM, Chem. Phys. Lett. 267 (1997) 23–30.

[227] S.D. Schwartz, V.L. Schramm, Enzymatic transition states and dynamic motion in barrier crossing, Nat. Chem. Biol. 5 (2009) 551–558.

[228] J.P. Klinman, A. Kohen, Hydrogen tunneling links protein dynamics to enzyme catalysis, Annu. Rev. Biochem. 82 (2013) 471–496.

[229] A. Warshel, Energetics of enzyme catalysis, Proc. Natl. Acad. Sci. U. S. A. 75 (1978) 5250–5254.

[230] S.C.L. Kamerlin, A. Warshel, At the dawn of the 21st century: is dynamics the missing link for understanding enzyme catalysis? Proteins: Struct. Func. Bioinformat. 78 (2010) 1339–1375.

[231] H.B. Liu, A. Warshel, Origin of the temperature dependence of isotope effects in enzymatic reactions: the case of dihydrofolate reductase, J. Phys. Chem. B 111 (2007) 7852–7861.

[232] N. Kanaan, S. Ferrer, S. Marti, M. Garcia-Viloca, A. Kohen, V. Moliner, Temperature dependence of the kinetic isotope effects in thymidylate synthase. A theoretical study, J. Am. Chem. Soc. 133 (2011) 6692–6702.

[233] M.H.M. Olsson, W.W. Parson, A. Warshel, Dynamical contributions to enzyme catalysis: critical tests of a popular hypothesis, Chem. Rev. 106 (2006) 1737–1756.

[234] A.J. Kirby, The lysozyme mechanism sorted - after 50 years, Nat. Struc. Biol. 8 (2001) 737–739.

[235] D.J. Vocadlo, G.J. Davies, R. Laine, S.G. Withers, Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate, Nature 412 (2001) 835–838.

[236] C.C.F. Blake, L.N. Johnson, G.A. Mair, A.C.T. North, D.C. Phillips, V.R. Sarma, Crystallographic studies of activity of hen egg-white lysozyme, Proc. R. Soc. Ser. B-Biol. 167 (1967) 378–388.

[237] L.O. Ford, P.A. Machin, D.C. Phillips, R. Tjian, L.N. Johnson, Crystal structure of a lysozyme-tetrasaccharide lactone complex, J. Mol. Biol. 88 (1974) 349–371.

[238] A.L. Bowman, I.M. Grant, A.J. Mulholland, QM/MM simulations predict a covalent intermediate in the hen egg white lysozyme reaction with its natural substrate, Chem. Commun. (2008) 4425–4427.

[239] D.E. Koshland, Stereochemistry and the mechanism for enzymatic reactions, Biol. Rev. 28 (1953) 416–436.

[240] F.E. Romesberg, B.D. Santarsiero, B. Spiller, J. Yin, D. Barnes, P.G. Schultz, et al., Structural and kinetic evidence for strain in biological catalysis, Biochemistry 37 (1998) 14404–14409.

[241] J. Yin, A.E. Beuscher, S.E. Andryski, R.C. Stevens, P.G. Schultz, Structural plasticity and the evolution of antibody affinity and specificity, J. Mol. Biol. 330 (2003) 651–656.

[242] P.K. Sharma, Z.T. Chu, M.H. Olsson, A. Warshel, A new paradigm for electrostatic catalysis of radical reactions in vitamin B12 enzymes, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 9661–9666.

[243] A. Warshel, Computer simulations of enzyme catalysis: methods, progress, and insights, Annu. Rev. Biophys. Biomol. Struct. 32 (2003) 425–443.

[244] L. Pauling, Molecular architecture and biological reactions, Chem. Eng. News 24 (1946) 1375–1377.

[245] L.P. Kotra, E.F. Pai, Inhibition of orotidine-5′-monophosphate decarboxylase – discoveries and lessons, Nucl. Acids Symp. Ser. 52 (2008) 85–86.

[246] M. Fujihashi, T. Ishida, S. Kuroda, L.P. Kotra, E.F. Pai, K. Miki, Substrate distortion contributes to the catalysis of orotidine 5′-monophosphate decarboxylase, J. Am. Chem. Soc. 135 (2013) 17432–17443.

[247] N. Wu, Y. Mo, J. Gao, E.F. Pai, Electrostatic stress in catalysis: structure and mechanism of the enzyme orotidine monophosphate decarboxylase, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 2017–2022.

[248] J.K. Lee, K.N. Houk, A proficient enzyme revisited: the predicted mechanism for orotidine monophosphate decarboxylase, Science 276 (1997) 942–945.

[249] O. Tapia, J. Andres, V.S. Safont, Enzyme catalysis and transition structures in vacuo. Transition structures for the enolization, carboxylation and oxygenation reactions in ribulose-1,5-bisphosphate carboxylase/oxygenase enzyme (Rubisco), J. Chem. Soc., Faraday Trans. 90 (1994) 2365–2374.

[250] T.C. Bruice, Some pertinent aspects of mechanism as determined with small molecules, Annu. Rev. Biochem. 45 (1976) 331–373.

[251] F.C. Lightstone, T.C. Bruice, Ground state conformations and entropic and enthalpic factors in the efficiency of intramolecular and enzymatic reactions. 1. Cyclic anhydride ormation by substituted glutarates, succinate, and 3,6-endoxo-(4-tetrahydrophthalate monophenyl esters, J. Am. Chem. Soc. 118 (1996) 2595–2605.

[252] T.C. Bruice, S.J. Benkovic, Chemical basis for enzyme catalysis, Biochemistry 39 (2000) 6267–6274.

[253] J.L. Griffin, M.W. Bowler, N.J. Baxter, K.N. Leigh, H.R. Dannatt, A.M. Hounslow, et al., Near attack conformers dominate beta-phosphoglucomutase complexes where geometry and charge distribution reflect those of substrate, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 6910–6915.

[254] M.I. Page, W.P. Jencks, Entropic contributions to rate accelerations in enzymic and intramolecular reactions and the chelate effect, Proc. Natl. Acad. Sci. U. S. A. 68 (1971) 1678–1683.

[255] W.P. Jencks, Catalysis in Chemistry and Enzymology, Dover Publications, 1987.

[256] S. Hammes-Schiffer, Catalytic efficiency of enzymes: a theoretical analysis, Biochemistry 52 (2013) 2012–2020.

[257] K. Henzler-Wildman, D. Kern, Dynamic personalities of proteins, Nature 450 (2007) 964–972.

[258] K.A. Henzler-Wildman, M. Lei, V. Thai, S.J. Kerns, M. Karplus, D. Kern, A hierarchy of timescales in protein dynamics is linked to enzyme catalysis, Nature 450 (2007) 913–916.

[259] K.A. Henzler-Wildman, V. Thai, M. Lei, M. Ott, M. Wolf-Watz, T. Fenn, et al., Intrinsic motions along an enzymatic reaction trajectory, Nature 450 (2007) 838–844.

[260] S.J. Benkovic, S. Hammes-Schiffer, A perspective on enzyme catalysis, Science 301 (2003) 1196–1202.

[261] T.H. Rod, J.L. Radkiewicz, C.L. Brooks, Correlated motion and the effect of distal mutations in dihydrofolate reductase, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 6980–6985.

[262] D.R. Glowacki, J.N. Harvey, A.J. Mulholland, Taking Ockham's razor to enzyme dynamics and catalysis, Nat. Chem. 4 (2012) 169–176.

[263] D.R. Glowacki, J.N. Harvey, A.J. Mulholland, Protein dynamics and enzyme catalysis: the ghost in the machine, Biochem. Soc. Trans. 1 (2012) 515–521.

[264] S. Hay, L.O. Johannissen, M.J. Sutcliffe, N.S. Scrutton, Barrier compression and its contribution to both classical and quantum mechanical aspects of enzyme catalysis, Biophys. J. 98 (2010) 121–128.

[265] S. Hay, N.S. Scrutton, Good vibrations in enzyme-catalysed reactions, Nat. Chem. 4 (2012) 161–168.

[266] K. Francis, A. Kohen, Protein motions and the activation of the C–H bond catalyzed by dihydrofolate reductase, Curr. Opin. Chem. Biol. 21 (2014) 19–24.

[267] Z.D. Nagel, J.P. Klinman, A 21st century revisionist's view at a turning point in enzymology, Nat. Chem. Biol. 5 (2009) 543–550.

[268] J.P. Layfield, S. Hammes-Schiffer, Hydrogen tunneling in enzymes and biomimetic models, Chem. Rev. 114 (2013) 3466–3494.

[269] S.C.L. Kamerlin, J. Mavri, A. Warshel, Examining the case for the effect of barrier compression on tunneling, vibrationally enhanced catalysis, catalytic entropy and related issues, FEBS Lett. 584 (2010) 2759–2766.

[270] P.E.M. Siegbahn, O-O bond formation in the S-4 state of the oxygen-evolving complex in photosystem II, Chemistry 12 (2006) 9217–9227.

[271] B.R. Prasad, N.V. Plotnikov, J. Lameira, A. Warshel, Quantitative exploration of the molecular origin of the activation of GTPase, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) 20509–20514.

[272] J. Clardy, C. Walsh, Lessons from natural molecules, Nature 432 (2004) 829–837.

[273] D.W. Christianson, Unearthing the roots of the terpenome, Curr. Opin. Chem. Biol. 12 (2008) 141–150.

[274] D.T. Major, Y. Freud, M. Weitman, Catalytic control in terpenoid cyclases: multiscale modeling of thermodynamic, kinetic, and dynamic effects, Curr. Opin. Chem. Biol. 21 (2014) 25–33.

[275] K.C. Nicolaou, Z. Yang, J.J. Liu, H. Ueno, P.G. Nantermet, R.K. Guy, et al., Total synthesis of taxol, Nature 367 (1994) 630–634.

[276] M. Weitman, D.T. Major, Challenges posed to bornyl diphosphate synthase: diverging reaction mechanisms in monoterpenes, J. Am. Chem. Soc. 132 (2010) 6349–6360.

[277] D.T. Major, M. Weitman, Electrostatically guided dynamics—the root of fidelity in a promiscuous terpene synthase? J. Am. Chem. Soc. 134 (2012) 19454–19462.

[278] Y.J. Hong, D.J. Tantillo, A potential energy surface bifurcation in terpene biosynthesis, Nat. Chem. 1 (2009) 384–389.

[279] Y.J. Hong, D.J. Tantillo, Biosynthetic consequences of multiple sequential post-transition-state bifurcations, Nat. Chem. 6 (2014) 104–111.

[280] M. Harel, A. Aharoni, L. Gaidukov, B. Brumshtein, O. Khersonsky, R. Meged, et al., Structure and evolution of the serum paraoxonase family of detoxifying and anti-atherosclerotic enzymes, Nat. Struc. Mol. Biol. 11 (2004) 412–419.

[281] O. Khersonsky, D.S. Tawfik, Enzyme promiscuity: a mechanistic and evolutionary perspective, Annu. Rev. Biochem. 79 (2010) 471–505.

[282] H. Engelkamp, N.S. Hatzakis, J. Hofkens, F.C. De Schryver, R.J.M. Nolte, A.E. Rowan, Do enzymes sleep and work? Chem. Commun. (2006) 935–940.

[283] R.O. Dror, R.M. Dirks, J.P. Grossman, H. Xu, D.E. Shaw, Biomolecular simulation: a computational microscope for molecular biology, Annu. Rev. Biophys. 41 (2012) 429–452.

[284] F. Claeyssens, J.N. Harvey, F.R. Manby, R.A. Mata, A.J. Mulholland, K.E. Ranaghan, et al., High-accuracy computation of reaction barriers in enzymes, Angew. Chem. Int. Ed. 45 (2006) 6856–6859.

[285] R. Diamond, Real-space refinement of the structure of hen egg-white lysozyme, J. Mol. Biol. 82 (1974) 371–391.

[286] T. Marino, N. Russo, M. Toscano, Catalytic mechanism of the arylsulfatase promiscuous enzyme from *Pseudomonas aeruginosa*, Chem-Eur J. 19 (2013) 2185–2192.