Published in final edited form as:

Methods Enzymol. 2016 ; 577: 213–250. doi:10.1016/bs.mie.2016.05.016.

QM/MM Analysis of Transition States and Transition State Analogues in Metalloenzymes

Daniel Roston and Qiang Cui

Department of Chemistry and Theoretical Chemistry Institute, University of Wisconsin, Madison, 1101 Univ. Ave., Madison, WI 53706

Abstract

Enzymology is approaching an era where many problems can benefit from computational studies. While ample challenges remain in quantitatively predicting behavior for many enzyme systems, the insights that often come from computations are an important asset for the enzymology community. Here we provide a primer for enzymologists on the types of calculations that are most useful for mechanistic problems in enzymology. In particular, we emphasize the integration of models that range from small active site motifs to fully solvated enzyme systems for cross-validation and dissection of specific contributions from the enzyme environment. We then use a case study of the enzyme alkaline phosphatase to illustrate specific application of the methods. The case study involves examination of the binding modes of putative transition state analogues (tungstate and vanadate) to the enzyme. The computations predict covalent binding of these ions to the enzymatic nucleophile and that they adopt the trigonal bipyramidal geometry of the expected transition state. By comparing these structures with transition states found through free energy simulations, we assess the degree to which the transition state analogues mimic the true transition states. Technical issues worth treating with care as well as several remaining challenges to quantitative analysis of metalloenzymes are also highlighted during the discussion.

Keywords

Metalloenzymes; QM/MM; phosphoryl transfer; transition state; transition state analog

1. Introduction

Enzymologists can now use computations to understand a wide variety of problems. Phenomena such as ligand binding,[55] chemical mechanisms,[43] transition states,[108, 142] and enzyme dynamics[11, 75, 121, 124] are all accessible through computations of one form or another. In recent years, computational design of novel enzymes[13, 115, 125] has emerged as one of the most promising avenues of biotechnology research. The ability to readily design novel enzymes from basic principles would herald in an era where enzymologists could truly claim to understand enzyme catalysis. Not long ago, the field of computational enzymology was still in its infancy, with relatively limited claims of successful predictive power.[77] The computational design of novel activity, therefore, is

quite promising, but as of yet, computationally designed enzymes are sluggish relative to naturally evolved ones.[78, 82] Thus, despite much progress in recent years, it remains challenging to predict behavior for an arbitrary enzyme, suggesting that there are missing pieces in general theories of enzyme catalysis[11, 45] or in specific computational models.

From a technical point of view, metalloenzymes are particularly challenging because the metal co-factors are generally more difficult to treat reliably with computational approaches. Therefore, balancing computational efficiency and accuracy, which is the key to computational studies of complex systems, is most essential to the analysis of metalloenzymes. To this end, it is important to choose the computational model and methods carefully based on the problem in hand. Moreover, we find it is instructive to integrate models of different complexity and methods at different levels of theory; comparing the results from different calculations helps to cross-validate robustness of mechanistic models and to evaluate the importance of specific factors (e.g., protein motion) to the properties of interest.

Another challenge to the analysis of complex enzymes is a gap between practitioners of experimental techniques and computational methods for studying enzymes. Experimentalists may assert that there is a dearth of computational models making “blind” predictions, while the opposite side of that coin is that rather rarely are experimentalists willing to test predictions that emerge from computational studies. Much of the problem likely lies in the seeming lack of communication and understanding between experimentalists and theoreticians. Computational chemists often have little sense of what experimental tests are feasible or how to interpret experimental measurements properly. Similarly, experimental chemists may not grasp what or how something has been calculated in a computational study, or indeed how calculated properties ought to correspond to “reality” as measured in a lab. Enhancing the communication between experimental and computational enzymologists is essential to pushing computational enzymology out of its adolescence.

Bearing those thoughts in mind, in the present work, we aim to provide an understanding of the types of calculations that can illuminate various problems in enzymology that should be accessible to those with limited background on computational methods, but still useful for more experienced computational chemists. While some types of calculations are fairly specialized, others are readily accomplished with user-friendly software and can therefore be adopted and used effectively by experimental chemists themselves to obtain the desired information. We begin by providing a brief survey of the types of calculations commonly used for studying (metallo)enzymes; more details can be found in several recent review articles by us[48, 90] and others[21, 122], as well as in the other chapters in this volume. We will then use a case study of the metalloenzyme alkaline phosphatase (AP) to illustrate details of how one may use calculations for understanding enzyme chemistry and make clear connections to experimental measurements.

2. Background on Computational Methods

Many considerations go into choosing an appropriate approach for studying a biomolecular system. First, there is the question of what property one wants to calculate. The goal of a

calculation should be to obtain an experimentally testable property that arises out of a model for biological behavior. In principle, calculations can predict kinetic and thermodynamic properties, geometrical properties, spectral signatures, etc. But even given a particular experimentally observable property, there are many considerations that go into how to calculate it. To begin, one must choose a level of theory that is appropriate for the specific system of interest. This usually entails balancing reliability of calculations with computational cost in order to achieve the highest accuracy possible required for the model size and timescale one needs. Biological systems of interest can range in size from small peptides in solution to something as large as a ribosome translating a gene that is kilobases long. Furthermore, enzymological processes occur on timescales ranging from femtoseconds to many seconds,[119] and for many processes, one cannot easily separate these timescales. For example, a chemical reaction's passage through a transition state (TS) could occur in a matter of tens of femtoseconds, but the vibrations that are necessary to align active site dipoles in order to allow for that barrier crossing may occur in picoseconds. In turn, this alignment might be dependent upon other conformational transitions that may range from a sidechain isomerization to domain reorientations, which may occur in nanoseconds or microseconds and beyond. Thus the fact that a TS only lasts a matter of tens of femtoseconds has little bearing on how long one must simulate a system in order to understand properties of a TS because motions at various time scales facilitate TS crossing. The size of a system and the timescale of the processes, therefore, are important considerations in choosing a level of theory to use for a computational study.

2.1. Levels of Theory

Due to practical limits on computational power, the most reliable computational methods (i.e., *ab initio* quantum mechanics) cannot be used for large systems or for long simulations. Thus, one generally has to consider how best to balance computational cost and accuracy. When chemistry (i.e., bond breaking/formation) is not involved, oftentimes one can use a molecular mechanics (MM) force field to model an enzymological system.[1, 107, 116] For many purposes, a well-parameterized force field can provide very reliable results for a question of interest, such as binding affinity of ligands to an enzyme active site.[71] Force fields have their limits, though. For one, a force field must be parameterized for a specific system, which is no simple task despite some recent advances[69, 136]. If one wishes to know how a promising new drug binds to a particular protein, for example, one needs to find a set of force field parameters (vibrational force constants, electrostatic charges, etc.) specific to that drug. Second, most force fields for biomolecules involve rather simple functional forms and physical terms (e.g., even polarizable force fields[30] often assume atom-centered isotropic polarizabilities), which limit the transferability and accuracy, especially when metal ions are involved[68].

For a process that involves bond reorganization, one generally needs to include electrons in the calculation using some QM method. In addition to being capable of modeling bond reorganization events, QM calculations have the advantage that they are more general and need not be parameterized for a specific purpose (although some QM methods can be specifically parameterized to achieve greater accuracy[29, 35, 112]). Different QM methods have different domains where they are either more or less reliable; for example, Density

Functional Theory methods[9, 101] are useful in most metalloenzyme studies[12], although their limitations are also well-documented[26] and highly-correlated *ab initio* QM methods[22, 24] are needed even for qualitative insights in some cases[80]. A drawback is that QM calculations can be very costly in terms of computational time, and even the cheapest QM methods are 100–1000 fold slower than MM calculations, depending on the size of the system. Thus, even using a semi-empirical QM method (e.g. AM1,[33] PM3, [128] DFTB,[28, 37] etc.) for an entire enzyme system is not feasible, although recent work[79, 86] suggests that an era may be approaching when computers are capable of treating a wide range of biological problems strictly quantum mechanically. Nonetheless, it is not obvious that approximate QM methods will necessarily achieve more accurate results than those of well-parameterized MM methods in describing, for example, protein conformations. At any rate, one solution to the cost of QM calculations is to use a small model that includes just the active site atoms.[12] Oftentimes one can extract very useful information from small models, especially if there are good experimental constraints to guide the model. This requires restraints on certain atoms, though, and otherwise ignores the role of the enzyme environment outside of the active site.

A more sophisticated approach is to use hybrid QM/MM calculations. QM/MM calculations have been developing for a few decades now[21, 39, 41, 67, 72, 88, 92, 97, 109, 122, 141] and the value of these calculations was recently recognized by the 2013 Nobel Prize in chemistry. QM/MM calculations take advantage of the best aspects of the two methods within the same calculation. In these calculations, the parts of the system that must be treated quantum mechanically (e.g. the active site of a chemical reaction or a ligand for which no force field exists) are treated with QM, but the remainder of the system is treated with MM. This allows one to use a reliable and versatile method for the active site atoms, while still accounting for the effects of the enzyme environment. There have been many recent and excellent review articles on QM/MM methods for biological applications[21, 43, 67, 92, 122], thus we will not repeat them here. We only hope to emphasize several technical points often not emphasized in the literature.

First, it is important to balance QM-MM and MM-MM interactions. Since QM-MM interaction terms[39, 88] are an inexpensive component of QM/MM calculations, they are often computed without any cutoff distance; MM-MM interactions, by contrast, in many non-periodic setups are usually treated with a cutoff, meaning that interactions beyond a certain distance are ignored to limit computational cost. Such a combination can lead to over-polarization of the MM region and thus structural artifacts in QM/MM simulations[117]. One solution is to employ a periodic boundary condition with Particle-Mesh-Ewald type of treatment[31] for MM-MM interactions, which divides the infinite summation of interactions into a short range component and a long range component and the latter is computed in Fourier space to achieve fast convergence. Another solution is to use a finite-spherical setup with extended electrostatics[129] for MM-MM interactions and implicit solvation at the edge of the sphere[70, 117, 132, 149]; the size of the spherical region needs to be carefully chosen to avoid too much structural constraint in the reactive site[89]. Second, QM-MM interactions need to be carefully calibrated[40, 51, 62], even when a rather large QM region is used. As illustrated in Fig. 1, relatively small errors in QM-MM interactions favor an altered hydrogen-bonding pattern at the QM-MM boundary;

although being $>10 \text{ \AA}$ away from the center of interest, the altered hydrogen bonding pattern ultimately propagates to the active site. Therefore, QM-MM interactions need to be calibrated such that artifactual structural changes can be distinguished from realistic structural transitions. Finally, as emphasized by several recent studies[50, 52, 65], it is essential to carefully determine the water occupancy of active sites and the titration states of residues; in some cases, cumulative effects of titration state of residues that are distant from the active site may still have a notable impact on the structural and energetic properties of the reactive center.

2.2. Types of Calculations

For mechanistic studies, the simplest type of calculation is a local geometry optimization that leads the system to a local minimum on the potential energy surface (PES). Starting from a local energy minimum, one can further explore nearby saddle point(s) and minimum energy pathway(s) that lead to, for example, intermediate and product for the catalyzed reaction. This type of analysis is most effective for active-site models of enzymes[12], for which the number of important stationary points on the PES is quite limited. For full enzyme models, minimization based calculations are generally less useful due to the enormously large number of stationary points and therefore minimum energy pathways (MEPs) on the high-dimensional PES, and it is challenging (if not impossible) to estimate the proper statistical weights from these MEPs[109, 138]. Another important limitation of minimization type of calculations is the lack of thermal fluctuations and therefore entropic contribution to the reaction energetics; along this line, it is important to recall that protein and solvent fluctuations (not just the reactive motifs) can't be ignored when considering the entropic contribution[74, 140]. Therefore, it is generally difficult to compare the results from a limited number of MEP calculations to experimental free energy data, although qualitative insights may still be obtained from the analysis of MEP calculations.

To obtain free energies, it is important to conduct adequate sampling of the PES, which may involve simulations lasting nanoseconds to microseconds and beyond, depending on the characteristics of the system. There is a vast literature on free energy calculations for biomolecules, some of which is also discussed in chapters in this volume. A difficulty with free energy simulations is often the problem of sampling rare configurations. For example, in a typical protein system, even the most advanced computers can only run simulations lasting on the order of microseconds (using classical potential functions),[123] yet enzymes typically have turnover numbers in the ms to seconds timescales.[38] These slow events correspond to high barriers on the free energy surface and to observe them requires enhanced sampling techniques, such as umbrella sampling[135] or metadynamics.[7] Simulations using enhanced sampling add biasing potentials of one form or another to the PES with the goal of attenuating the surface so that all relevant regions are sampled evenly and efficiently. In these cases, one has to carefully choose an "order parameter", often corresponding to the reactive degrees of freedom of interest, along which to bias the PES and sample the system. Some problems benefit from using collective order parameters such as energy gap[140] or local level of hydration[90]; identifying the proper degrees of freedom to thoroughly sample is indeed the major challenge to free energy simulations, especially when the bottleneck is entropic in nature. Other types of simulations use completely non-

physical order parameters; for example an “alchemical” mutation[126, 130] that changes one atom to another (or even one residue or ligand to another) can assess properties like proton affinities[85] or relative binding affinities of ligands.[32, 126, 130]

Alchemical mutations highlight the fact that often times simulations are limited only by a chemist’s creativity, but difficulty can arise if the question of interest does not contain an obvious choice for an order parameter. In such cases, a viable alternative is transition path sampling (TPS),[16] which finds dynamic trajectories connecting two states without the biasing of umbrella sampling or metadynamics; closely related techniques are the finite-temperature string methods[36]. If one is interested in dynamic aspects of an enzyme, and not just properties of reactants and transition states, TPS may be vital, and some have suggested that understanding enzyme reactivity requires understanding dynamic trajectories. [121] Likewise, others have found that the ensemble of transition states differs in important ways when examined by umbrella sampling and TPS.[34] TPS is ideal for transition events (such as barrier crossing) that are intrinsically fast; for processes that involve slower or diffusive events, other techniques such as string[36] or Markov State Modeling[17] are likely more effective. The controversial question of the role of dynamics in enzyme catalysis has been argued extensively,[11, 15, 57, 58, 73, 98] and the practical concern of which computational methods work best for specific step(s) in the catalytic cycle may play a role in guiding that debate.

In addition to structures and energetics for the catalytic cycle, computational studies are also able to calculate various experimental observables such as NMR chemical shifts, vibrational spectra and isotope effects. These are important quantities that allow one to make quantitative connection to experiments and provide validation/test of computational results; in many cases, comparison between computation and experiments is essential to the assignment and molecular interpretation of the experimental data[44, 96, 104, 143, 144]. Depending on the system and observable(s) of interest, the degree of required sampling varies; generally speaking, however, averaging over a set of structures is likely important even to observables (e.g., nuclear quadrupole coupling constant) that are often regarded as reporting “static” structural properties.[110]

The type of calculation dictates what level of theory should be used. For minimization type of calculations, it is feasible to use *ab initio* QM or QM/MM potentials. For simulations that require much more extensive sampling (e.g., free energy and TPS), it is usually unrealistic to use *ab initio* QM/MM potentials in the framework of standard molecular dynamics. The solution is to either decouple the sampling of QM and MM regions[67, 111], so that QM is effectively treated with minimization while the MM region is sampled using molecular dynamics with frozen QM structure, or to effectively integrate results from low-level QM/MM and *ab initio* QM/MM calculations[24, 90, 95, 105, 106]. For spectral observables or isotope effects, results are usually most robust with *ab initio* QM(/MM) potentials, although carefully calibrated semi-empirical QM(/MM) methods can be useful for specific properties, such as vibrational spectra[104, 143, 144] and UV adsorption spectra[59, 139].

2.3. Summary

Just as there is a wide variety of experimental methods one can use to study an enzyme, there is a wide variety of computational methods. Here we have provided an outline of some of the most useful strategies for calculating properties and behavior of enzymes. Some techniques, such as active site modeling, can be used successfully by (primarily) experimental enzymologists to add a layer of quantitative understanding to the interpretation of experimental results.[76, 114] Other methods, such as free energy simulations, require more specialized training and until programs like CHARMM and AMBER become even more transparent to general users, computational chemists will have to communicate the results and implications of simulations to the experimental community as clearly as possible. We now turn our attention to a case study that demonstrates the usefulness of some of the above techniques.

3. Case Study: The Transition State of Alkaline Phosphatase

Here we describe some recent studies on transition state (TS) structure in the enzyme alkaline phosphatase (AP) to illustrate the use of a few of the methods above. The studies involve QM calculations using active site models, as well as QM/MM calculations and simulations of the entire enzyme. AP catalyzes the hydrolysis of phosphate monoesters, its native substrates, in addition to exhibiting promiscuous activity toward phosphate diesters, [100, 147] phosphorylated pyridines,[81] phosphorothioates,[60] and even sulfonic esters. [99] The chemical mechanism of hydrolysis by AP (Fig. 2) is similar to phosphoryl transfer in many enzymes:[25] the substrate phosphorylates an enzymatic nucleophile (S102), and this phosphorylated-enzyme intermediate is subsequently hydrolyzed to release phosphate. Important questions remain about how AP and other phosphoryl transferases catalyze the first phosphorylation step.

We have examined the binding mode(s) of putative TS analogue (TSA) inhibitors in AP using geometry optimizations and PES scans, and compared those geometries with the geometries of TS structures for phosphate monoester hydrolysis recently found by free energy simulations.[113] In addition to illustrating some of the computational methods available to enzymologists, we provide insights into the ability of TSAs to mimic the actual TS of the enzyme and the scope and limits of what can be gleaned from experimental observations of TSAs in this enzyme.

3.1. TS Analogue Binding

TSAs serve as important inhibitors for enzymes and are currently used clinically as treatments for a variety of diseases.[118, 119] In addition to their clinical relevance, TSAs are useful in many experimental probes of enzyme structure and dynamics, including x-ray crystallography,[61] NMR,[15] IR,[6] and others. Furthermore, TSAs have been useful in the development of catalytic antibodies[10, 94, 120] and other biomimetic catalysts. One can imagine a variety of experimental criteria that would define a “good” TSA ligand: the ligand binds tightly to the enzyme and competitively inhibits catalysis;[118] mutation affects the inhibition constant of the TSA in the same way that it affects the rate of the native enzymatic reaction;[103] and even the absence of certain kinds of dynamics have been proposed to

indicate that an enzyme-ligand complex resembles the TS.[6] Perhaps the most obvious criterion is that the TSA and TS have similar structural properties and charge distributions. While crystallographic studies of TSA complexes illuminate the structures of TSAs bound to active sites, crystallographic studies cannot probe actual TSs for proper comparison. Despite sensationalistic claims of “trapping” TSs,[23, 102] actual TSs are unstable points on a free energy surface and thus cannot be trapped. Their structures can, however, be obtained through computational methods and our recent study of TS structure in AP[113] allows us to compare TSAs with actual TSs.

Here we use a variety of computational methods to understand the binding modes of the tungstate (WO_4^{2-}) and vanadate (VO_4^{3-}) ions in the active site of AP, with an eye to answering the question of how well TSAs truly mimic TSs. In previous work, crystallographers used vanadate as a TSA for AP,[61] as well as other phosphatases.[87, 145, 148] VO_4^{3-} binds covalently to the enzymatic nucleophile (S102 , O_{nuc}), but it has no leaving group (or at least a very poor one), so the ligand adopts the trigonal bipyramidal geometry one expects[83] in the TS of a concerted transphosphorylation reaction. Vanadate therefore fulfills many roles of a good TSA, but it has limitations. The magnitude of the charge of VO_4^{3-} is larger than native substrates (phosphate monoesters, charge=-2) as well as promiscuous substrates (diesters and sulfate esters, charge=-1) and this difference in charge could result in misleading experimental results. The magnitude of this charge depends on the protonation state of the ion,[27] and uncertainty as to the protonation state when bound to the enzyme clouds the interpretation of experimental results. WO_4^{2-} , on the other hand, has the same charge as the native substrate and may have a low enough pK_a to avoid ambiguity of charge.[27, 54] Until recently,[103] however, WO_4^{2-} had not been used as a TSA for AP or any other enzyme with a serine nucleophile. We note that since initiating this project, Herschlag and coworkers have solved a crystal structure of AP with bound tungstate.[103] The results of their crystallographic study were not made available to us, though, until we deposited our results with a third-party. Thus, the results we present here on tungstate should be viewed as a “blind” prediction except where noted.

In this work, we use a variety of computational methods to test whether WO_4^{2-} could serve as a TSA in wt E. coli AP. We will move back and forth between methods and results/discussion in order to highlight the use of specific computational methods, the results that they yield, and how that information might inspire additional calculations (or experiments). We describe *ab initio* QM calculations on a small model of the enzyme active site, as well as *ab initio* QM/MM calculations on the full enzyme. We also test the computational methods using vanadate, which we were able to compare to a published crystal structure. Overall, our calculations support the use of WO_4^{2-} as a TSA that binds covalently to the serine nucleophile. After comparing our prediction with the recently solved crystal structure, we find that the geometry of the predicted covalent complex aligns reasonably well with the crystallographic model, but we consider whether protonation state is, in fact, an issue with tungstate as well as vanadate. Our original assumption of a deprotonated tungstate has similar bonding to the nucleophile and leaving group oxygen (O_{nuc} and O_{lg} , respectively) as the TS for phosphate monoesters with good leaving groups (LGs). The modeled ligands also

interact with nearby enzymatic residues (e.g. R166) in a similar manner to the TS of those monoesters, though there are still questions about the protonation states of the ligands – especially vanadate – which may be manifested in the effects of mutations on binding constants.

3.2. Computational Benchmarks

Prior to employing any computational method to predict unknown behavior it is necessary to examine the accuracy of the method for the property of interest. In this case, we wished to find a QM method that could accurately model geometric properties of tetra- and penta-coordinate tungstate and vanadate species; the method also needs to be efficient for QM/MM calculations of full enzyme models. To this end, we tested the ability of density functional theory[101] at the B3LYP/6-31+G* level[8, 84] with an effective core potential[56] on the central metal atom to reproduce the geometries of small molecule crystals. The species chosen were WO_4^{2-} , VO_4^{3-} , and two entries in the Cambridge Structural Database (CSD), CINMES and YIQLEP, which are WO_5 and VO_5 derivatives, respectively. We performed geometry optimizations of these species using Gaussian[42] and the optimized geometries are listed in Table 1. We find that the computations reproduce the structural properties quite well and that the QM method is well suited for modeling the ligands in the AP active site.

3.3. Active Site Optimizations

In an initial assessment of using WO_4^{2-} as a TSA in wt AP we sought to compute the binding mode of the ligand in a small model of the enzyme active site. The starting structure of the active site model was the crystal structure of wt AP with bound PO_4^{3-} at 1.75 Å resolution (PDB: 1ED8)[127] and consisted of the groups in Fig. 3. The enzymatic residues were truncated at their α carbons, which were converted to methyl groups and frozen during optimizations, so the overall model contained 124 atoms and had a net charge of -1 . The structure was optimized using the same *ab initio* QM method as above. In addition to the vacuum model, the optimization was performed using implicit solvent models[134] corresponding to a range of solvents in order to understand the magnitude and nature of any environmental effects.

Our optimizations of tungstate bound to the active site model (Fig. 4) suggest that the ligand could, indeed, serve as a TSA. Optimized W- O_{nuc} distances range from 2.13–2.16 Å, depending on solvent model, indicating a clear covalent nature to the bonding between the two atoms. For reference, the sum of the two atoms' van der Waals radii is ca. 4.1 Å.[3] Additionally, the geometry around the W is trigonal bipyramidal, which is expected[83] for the TS of transphosphorylation reactions. For reference, the other W-O bonds (i.e., bonds to oxygens besides O_{nuc}) range from 1.75–1.90 Å, so the bond to the nucleophile is somewhat longer than a typical W-O bond. This is a desirable characteristic if tungstate is to serve as a TSA, because the bond to the central atom is presumably not fully formed at the TS of the reaction.[83]

3.4. QM/MM Optimizations

To test how the ligand would behave in a more realistic model, we conducted optimizations of WO_4^{2-} in the full enzymatic environment in stages using CHARMM.[19] Since initial configuration could affect the optimized structure, we obtained a variety of starting structures for the optimizations, both from crystal structures with PO_4^{3-} (PDB: 1ED8, ref. [127]) and VO_4^{3-} (PDB: 1B8, ref. [61]), and from snapshots taken from QM/MM molecular dynamics simulations of the enzyme with PO_4^{3-} or with p-nitrophenyl phosphate (pNPP, a monoester substrate) bound to the active site. The simulations were conducted in accordance with the general methods described in refs. [63, 64, 113]. The starting structure was the crystal structure with PO_4^{3-} used above, and the ligand was changed to pNPP *in silico* for the simulations with that ligand. The active site was overlaid with a spherical droplet of water with 25 Å radius centered at one of the Zn^{2+} ions and water molecules within 2.8 Å of any crystallographic atoms were removed. The 25 Å radius sphere (“inner region”) was treated using fully exible molecular dynamics, and outside of that sphere (“outer region”), the protein was frozen and the outer region was treated with the generalized solvent boundary potential (GSBP).[70, 117] A four Å shell that was treated by Langevin dynamics served as a buffer region connecting the outer region with the inner region[20]. The simulations used DFTB3/3OB as the QM method[46, 47, 49, 91] with the CHARMM force field[1, 93] for the MM region. The QM region consisted of the atoms shown in Fig. 3 with link atoms between the α and β carbons of QM sidechains. The system underwent a short optimization and then the system was heated from 48 K to 298 K over the course of 100 ps using 1 fs timesteps, after which it was equilibrated at 298 K for at least 50 ps before snapshots were taken to use as starting structures for optimizations with WO_4^{2-} . During the simulations, mild restraints helped to tune the QM-MM boundary, which must be managed carefully as discussed above[53]. This included restraints on the H-bond between the sidechain of Asp330 and the backbone of Ser347 that is apparent in the crystal structure and a restraint on the C-O bond of Asp51 (QM region) that interacts with the nearby Mg^{2+} (MM region). The snapshots used for subsequent optimizations were all taken at intervals of at least 100 ps.

The snapshots obtained from the simulations, as well as the original crystal structures (1ED8 and 1B8J) were first optimized at the DFTB3/MM level since that method is 100× faster than B3LYP/MM. Unfortunately, a complete parameterization of DFTB3 for tungsten is not yet available,[137] so 16 these initial optimizations used PO_4^{3-} as the ligand with the goal of optimizing the enzyme environment. Snapshots with pNPP as the ligand were converted to PO_4^{3-} and S102 was protonated for the optimizations with PO_4^{3-} . Thus, the overall charge remained constant. These structures underwent 10,000 steps of adapted-basis Newton-Raphson (ABNR) optimization[18]. At that point, the PO_4^{3-} was converted to WO_4^{2-} , S102 was deprotonated, and optimization using B3LYP/MM calculations began using the Gaussian09 interface with CHARMM[149]. All waters were treated as MM during this stage of the optimization. The QM region was treated with the same QM method as in the active site model. ABNR optimization continued for 200–500 steps using the solvent macromolecular boundary potential (SMBP)[132, 149] instead of GSBP. Gradients for the

QM atoms were used as the primary criterion for determining convergence of the optimizations.

The optimized structures (Fig. 5) are consistent with the active site models, suggesting that within the enzyme environment, tungstate can form a (long) covalent bond with the serine nucleophile and adopt the trigonal bipyramidal geometry of the TS. Given the range of starting structures for these calculations, the resulting structures are significantly more heterogeneous than the active site models (Table 2). The W-O_{nuc} distances, for example, range from 2.12–4.42 Å. Some of these structures, therefore, cannot be thought of as exhibiting covalent bonds to the ligand; they are clearly in the van der Waals region.[3] The structures that do not form covalent complexes add ambiguity to the ability of tungstate to serve as a TSA, but we were able to clarify this ambiguity with a PES scan.

3.5. QM/MM PES Scan

To clarify the question of why some structures converged to geometries without a covalent bond, we conducted a PES scan along the W-O_{nuc} coordinate using a structure that optimized to one of the longest W-O_{nuc} distances (4.34 Å). At each point in the scan, the system was optimized with a harmonic restraining potential (force constant=1000 kcal/mol·Å²) maintaining the W-O_{nuc} distance. Each point along this coordinate was generated following 10 steps of ABNR optimization at the neighboring point, and all points were subsequently optimized for an additional 100 steps; a modest number of optimization steps was used because the enzyme structure was thoroughly optimized prior to the PES scan. In the initial scan, points were calculated every 0.25 Å, but additional points were added in the region of the barrier separating the van der Waals complex and the covalently bonded complex and those points were treated the same way as the other points.

The PES scan (Fig. 6) demonstrates that – at least in the case of the structure chosen to scan – the non-covalently bound complex was simply trapped in a local minimum, and that in an experimental setting with thermal energies, it would be able to form the covalent complex. The true minimum along the W-O_{nuc} distance is at around 2.25 Å, which is in accordance with the structures that originally optimized to covalent complexes. This covalently bound state is more stable than the van der Waals complex by around 15 kcal/mol. Furthermore, the barrier to forming the covalently bound state from the van der Waals complex is only around 8 kcal/mol, and it is therefore accessible under crystallographic or other experimental conditions. As mentioned previously, a limitation of PES scans like this is that they calculate only the relative energies of the covalent and van der Waals complexes, and the results can be sensitive to subtle structural changes (e.g., a water ip) during geometry optimization[133]. Therefore, free energies are desirable for making more quantitative predictions of the relative stability of the covalent and van der Waals complexes. Unfortunately, conducting free energy simulations of the tungstate complex is not straightforward because the B3LYP method with about 120 QM atoms is very costly and there is no semi-empirical alternative that is reliable for tungsten; we are currently pursuing B3LYP/MM calculations using the QM/MM minimum free energy path approach[66] (D. Fang and Q. Cui, work in progress). We note, however, testing the results of such free energy calculations experimentally would be difficult. Measurement of binding constants, for

example, do not distinguish between the covalent and van der Waals complex, while crystallographic measurements do not contain meaningful information on relative free energies (for example, there are crystal structures of AP available with PO₄ either covalently[127] or non-covalently[14] bound, but information on the relative stabilities of the two binding modes is not available from the crystals. Given that both complexes can be measured, all that can be said is that the relative stabilities are dependent on crystallographic conditions). Further spectroscopic measurements might provide additional insights into the potential diversity in the binding mode of tungstate.

3.6. Tests of Protonation States

An important question to ask at this point is how reliable are B3LYP/MM calculations of the interactions between tungstate and AP? We note that previous calculations of the interaction between vanadate and AP[63] found that B3LYP/MM calculations predicted a significantly longer V-O_{nuc} bond than that observed in the crystal structure (2.42 Å calculated vs. 1.72 Å observed). A V-O_{nuc} length of 2.42 Å would indicate an interaction with very little covalent character,[3] and further concern comes from the fact that the geometry of the optimized complex appeared to be more tetrahedral than trigonal bipyramidal. Those calculations, therefore, did not precisely reproduce the experimentally resolved crystal structure,[61] which demonstrated a much shorter V-O_{nuc} bond and trigonal bipyramidal geometry. One could question, then, whether B3LYP/MM calculations underestimate the strength of M-O_{nuc} bonds. Considering the overall satisfying performance of B3LYP for several model compounds shown in Table 1, another possibility is that the protonation state of the vanadate played a role in that discrepancy. The previous calculations were with VO₄³⁻, but it is very possible that this species is protonated under crystallographic conditions since HVO₄²⁻ has a pK_a of 13.3 in solution.[27] The positively charged groups (two Zn²⁺, an arginine, and a Mg²⁺-bound water) in the immediate vicinity of this ligand in the active site almost certainly lower its pK_a substantially, but it is difficult to know by precisely how much. The serine nucleophile, for example, is believed to be deprotonated in the apo enzyme, and have a pK_a <5,[4] suggesting that the active site lowers its pK_a relative to that in solution by at least 10 pH units. Given that vanadate binds covalently, however, the additional negative charge from O_{nuc} may raise the pK_a. Thus, it is not obvious which species will predominate under experimental conditions. To test this question, we calculated structures of multiple protonation states for vanadate in the active site model and in the full enzyme model.

In both the active site model and the full enzyme model, optimizations with vanadate (either VO₄³⁻, HVO₄²⁻, or H₂VO₄⁻) were initiated from an optimized structure with tungstate covalently bound to O_{nuc}. The W was replaced with V and where applicable, the O opposite the S102 nucleophile (O_{lg}) was protonated. In the doubly protonated species, an additional phosphoryl oxygen was protonated. The active site model was optimized as above. In the full enzyme model, water molecules within 15 Å of the vanadate were first optimized for 1000 steps of ABNR with all other atoms frozen and treated at the MM level, in order to allow the hydrogen bonding network to adapt to the different protonation state. The system then underwent 100–200 steps of ABNR optimization using the B3LYP/MM method used for tungstate. Again, the gradients on the ligand and the nucleophile were the primary

criteria for convergence. Both the active site model and the full enzyme model (Fig. 7) find that the singly protonated species converges to a structure most consistent with that found experimentally, suggesting that vanadate is singly protonated in the active site of AP. In contrast to the deprotonated form calculated previously, the protonated form converges to a structure where the V-O_{nuc} bond is shorter than the V-O_{lg} bond (1.97 Å vs. 2.14 Å), which is the case in the crystal structure (1.72 Å vs. 1.92 Å). Additionally, as is the case in the crystal structure, the protonated structure exhibits trigonal bipyramidal geometry. We note that the doubly protonated species in the active site model had large distortions of the active site geometry (e.g., Zn-Zn distance of 5.3 Å vs. 4.1 Å in the crystal). Together, these support the hypothesis that HVO_4^{2-} is the form of vanadate that predominates in the crystal structure and that if one chooses the correct protonation state, B3LYP/MM calculations can adequately reproduce the binding modes of MO_4^{n-} ligands. Thus, knowing nothing about the recent x-ray studies of tungstate in AP, we were able to predict that tungstate is able to bind covalently in a trigonal bipyramidal geometry reminiscent of the TS for phosphoryl transfer.

A potential criticism of the vanadate analysis could be that it is, in a sense, circular: we have sampled different protonation states in order to find one that fits the experimental results, and then claimed that the computational method is accurate if we choose the correct protonation state. Another possibility is that the method is inaccurate, but if we choose an incorrect protonation state, it gives results in agreement with the crystal structure. Experimental measurements of the pH dependence of the crystal structure, therefore, are necessary to test the model. Another issue with this analysis is that we are fitting our computational model to an experimental model (i.e., atomic coordinates from refinement of the X-ray diffraction data); a better comparison would have been with the raw experimental data (i.e., the electron density). When judging the veracity of a computational model, one should ask whether it correctly predicts/reproduces experimental results, not whether it correctly predicts/reproduces an experimentalist's interpretation of those results. Unfortunately, electron density is not available in the PDB for the vanadate structure.

3.7. Comparison with Experiment

At this point, we can examine whether the recently solved structure with tungstate[103] supports or refutes our findings. We have overlaid our calculated structures and the experimental structures in Fig. 8 and we find relatively good agreement: the crystallographic structure contains a covalent W-O_{nuc} bond and trigonal bipyramidal geometry. We do note, however, that the crystallographic structure has a substantially longer W-O_{lg} bond than we obtain (Table 3). Initially we assumed that tungstate would be deprotonated in the active site because the pK_a of HWO_4^- is 3.6 in solution[27, 54] and the positively charged moieties in the active site would lower that even further. The large deviation with the crystal structure made us question this assumption and so we calculated the structure of the protonated species in the active site model. As Table 3 and Fig. 9 indicate, the results for active site models do not differ in substantial ways from the full enzyme models of the same species. The structure of HWO_4^- in the active site model optimized to a structure whose W-O_{lg} bond was more similar to that in the crystal structure, but that model has a somewhat shorter W-O_{nuc} bond than the crystal structure. Thus, the question of the protonation state of $\text{WO}_4\text{-AP}$ is not as clear as in $\text{VO}_4\text{-AP}$.

Further adding to this uncertainty, the electron density and the anomalous diffraction in the crystal structure with tungstate[103] has peculiarities and ambiguities that warrant discussion. In contrast to most crystal structures of wt E. coli AP, which have two Zn^{2+} and one Mg^{2+} in the active site, the Mg^{2+} site had too much electron density to model a Mg^{2+} there, so the model contains an octahedrally-coordinated Zn^{2+} at that site. While octahedral Zn^{2+} is not unheard of, it is relatively rare in proteins[2] and is not typically observed in alkaline phosphatase, although one structure from the same lab also reported a Zn in place of the Mg[5]. Furthermore, the tungsten atom in the crystal had weaker electron density than expected, despite normal density for 3 oxygens around the tungsten (of the 5 oxygens expected around tungsten). The authors interpreted this as indicating a half-occupied tungstate covalently bound to S102. The unoccupied active site contains waters in similar positions as two of the tungstate oxygens and the position of S102 is independent of tungstate binding. This configuration of waters differs from a structure of the *apo* enzyme[127] and results in a linear $\text{Zn-OH}_2\text{-Zn}$, which is a surprising configuration based on structures with similar di- Zn^{2+} moieties (e.g., metallo β -lactamase, PDB: 4HL2). In the end, the way the authors chose to fit the electron density may be the most reasonable given the ambiguities. Nonetheless, we discuss these ambiguities, as well as those with the protonation state of the computational structure, to urge caution in the way computationalists interpret experimental results and the way experimentalists judge the validity of computational results. Often times the situation is not as clear-cut as crystallographic or computational structures may imply and human choices like protonation states or the best way to fit electron density may lower one's confidence in the particular model's applicability.

Despite the overall similarity between the calculated and experimental TSA structures, Fig. 9 shows that there are potentially important differences in bond lengths in the calculated TSA structures and their crystallographic counterparts. For example, the W-O_{lg} bond is nearly 0.5 Å longer in the crystallographic model than in the calculated WO_4^{2-} model. The protonated structure, HWO_4^- , achieves closer agreement, but the ambiguities mentioned above with the crystal structure make it difficult to use the crystal structure to rule out the deprotonated structure as a possibility. The long bond length in the crystal structure, for example, may result from an artifact of deconvoluting electron density contributions from both O_{lg} and a water in similar positions. We could find no other structures in the PDB containing a Ser-bound tungstate to determine what typical bond lengths might be. A few structures are available with Asp-bound tungstate (e.g., 3ET5, 2I34, 1J9K) and these have W-O_{nuc} bonds in the range of 2.4–2.8 Å indicating that WO_{nuc} bonds in enzymes might generally be longer than our model predicts. Clearly, additional work will be necessary to determine where the discrepancy in bond lengths between the calculated and crystallographic models arises.

The differences in bond lengths with vanadate are also intriguing. In this case, the proper protonation state appears to be HVO_4^{2-} , given that VO_4^{3-} does not bind covalently and H_2VO_4^- results in large deformations of the di- Zn^{2+} moiety (Table 3). The difference between the calculated structures and the experimental structure is ca. 0.2–0.3 Å in both the V-O_{nuc} and V-O_{lg} lengths. As mentioned, the electron density for the vanadate structure is

not available to offer any hints from where this discrepancy may originate. A high resolution structure of a closely related enzyme, nucleotide pyrophosphatase (PDB: 2GSO),[145] has $V-O_{nuc}$ and $V-O_{lg}$ lengths of 1.9 and 2.0 Å, respectively, which is closer to what we obtain computationally, but at this time we cannot suggest that there are clear problems with the vanadate-AP structure.

3.8. TSAs vs. TSs

Now we wish to ask whether either tungstate or vanadate is a good mimic of the actual TS. To answer this computationally, we must examine models for the actual TS, which we have recently done using DFTB3/MM free energy simulations;[113] the models of the TS are qualitatively consistent with diverse experimental data that include free energy relations, mutation effects and kinetic isotope effects. We note that DFTB3 was required because the active site of AP is very solvent accessible and therefore adequate sampling is crucial to the proper modeling of the TS structure (cumulatively more than 10 ns is used for each substrate in the DFTB3/MM umbrella sampling simulations); the DFTB3 method was also calibrated by a set of relevant phosphoryl transfer reactions and pK_a calculations against higher level calculations as well as relevant experimental data (see SI of ref. [113]). What we found in that study is that the structure of the TS depends strongly on the LG for both phosphate mono- and diesters. The reaction pathways can all be characterized as tight, where nucleophilic attack precedes cleavage of the bond to the LG, but worse LGs have TSs that are later in the reaction coordinate and involve significantly more $P-O_{lg}$ cleavage. Fig. 8 shows the TSAs overlaid with representative snapshots from the TS region for phosphate monoesters with a good LG and a poor LG. The overall active site architectures are quite similar for the TSs and TSAs, but in terms of ligation to the Zn^{2+} ions, the TSAs seem to better resemble the TS for poor LGs. In Fig. 9, which contains a summary of the relationships between enzymatic TSs and the structures of the TSAs in terms of bonding to the central atom (P, V, or W), one can see that in terms of the primary components of the reaction coordinate, the calculated structure of WO_4^{2-} appears to be most similar to the TS of phosphate monoesters with good LGs, despite differences in Zn^{2+} ligation apparent in Fig. 8 (and Table 3). If, however, HWO_4^- is the relevant species, it may be more similar to TSs with poor LGs, similar to the calculated structures for HVO_4^{2-} .

While the ligation pattern to the Zn^{2+} ions is somewhat different at the TS for phenyl phosphate than for tungstate (or vanadate), the interactions with the H-bonding network in the active site,[131] and most notably R166, are quite similar. Thus, the model predicts that mutations of this network will affect tungstate and vanadate binding in a similar manner as they will affect catalysis of phosphate ester hydrolysis. Measurements[103] showed that mutations (including double- and triple-mutants, etc.) cause similar changes in tungstate binding and catalytic rate for hydrolysis of p-nitrophenyl phosphate. While many mutations also show such a relationship with vanadate binding, most of the mutants that involve the E322Y mutation, which disrupts Mg^{2+} -binding, do not fit the trend. The E322Y mutation appears to have little effect on vanadate binding, despite the fact that it hinders both tungstate binding and catalytic rate. It is not obvious why this should be the case from our calculated geometries or the geometries in the crystal structures. The catalytic role of Mg^{2+} appears to be to position a water to donate a hydrogen bond to the phosphoryl group at the TS

of the reaction.[146] The fact that ablation of this Mg^{2+} does not disrupt vanadate binding may indicate that the H-bond donated by the Mg^{2+} -bound water is itself disrupted by protonation of the vanadate at that position.

4. Summary/Conclusions

Despite the many challenges in using computational methods to understand enzymes, when they are applied carefully they can make reliable predictions of behavior in a variety of contexts. We have shown that active site modeling using *ab initio* QM calculations can be used to understand ligand binding, and this can be supplemented by QM/MM calculations using the full enzyme. Our results indicate the importance of sampling many configurations, as the starting configuration can bias the outcome of optimizations. In many cases the most reliable and predictive method for understanding behavior will be free energy simulations, but for systems that contain heavy metals, computational cost generally prohibits straightforward *ab initio* QM/MM free energy simulations and force fields are generally unavailable. Thus combining insights from *ab initio* QM/MM based optimization and PES scans with semi-empirical QM/MM (e.g., DFTB3/MM) based molecular dynamics simulations can provide valuable insights on questions such as binding mode, geometry and similarity between TSA and TS. Several important directions of development for metalloenzyme studies include improving the efficiency and robustness of *ab initio* QM/MM based free energy simulations, developing semi-quantitative semi-empirical QM methods for metal ions, and better integrating multi-level QM/MM methods for quantitative computations. These developments will make it possible to make more quantitative and testable predictions from computational studies of metalloenzymes. Along this line, we echo the suggestion from experimental community that computational studies ought to aim at not only analyzing published experimental results but also making clearly spelled out predictions that can be tested by future experiments. One format might be to summarize such predictions in the supplementary materials of a publication[113].

Our comparison of the binding geometry of TSAs with the structures of actual TS models provides some confidence in using those TSAs for experimental studies of TS structure. Still, the variation of TS structure as a function of LG ability[113] urges caution in which TSA should be used to model which TS. Tungstate is apparently a good mimic of the TS of monoesters with good LGs, but vanadate may better model those with poor LGs. Measurements of inhibition by the TSAs vs. rate for different LGs can test this prediction. [103] Additionally, the fact that AP binds to tungstate with the geometry of the TS speaks to the great catalytic power of AP. Many structures in the PDB, for example, contain tungstate bound to the active site of a phosphatase but not covalently. The fact that AP stabilizes this ligand in the conformation of the TS is likely related to APs extraordinary ability to stabilize the TS. Continued study of how AP is so successful at stabilizing its TS will provide important insights into the general mechanisms of biological catalysis.

Acknowledgments

We thank Dr. Dan Herschlag and co-workers (A. Peck and F. Sunden) for providing the coordinates to the crystal structure with tungstate prior to publication. We thank James Fraser for serving as an intermediary to verify that the computational work was accomplished prior to sharing of x-ray crystallography results. DR is supported by an NIH

NRSA Post-Doctoral Fellowship (1F32GM112371-01A1). This research was also supported by an NIH grant (R01GM106443) and an XSEDE allocation (TG-MCB110014) to QC. Computational resources from the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by NSF grant number OCI-1053575, are greatly appreciated; computations are also partly supported by the National Science Foundation through a major instrument grant (CHE-0840494).

References

1. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evensen JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE III, Roux B, Schlenker M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B*. 1998; 102:3586–3616. [PubMed: 24889800]
2. Alberts IL, Nadassy K, Wodak SJ. Analysis of zinc binding sites in protein crystal structures. *Prot Sci*. 1998; 7:1700–1716.
3. Alvarez S. A cartography of the van der waals territories. *Dalton Trans*. 2013; 42:8617–8636. [PubMed: 23632803]
4. Andrews LD, Deng H, Herschlag D. Isotope-edited FTIR of alkaline phosphatase resolves paradoxical ligand binding properties and suggests a role for ground-state destabilization. *J Am Chem Soc*. 2011; 133:11621–11631. [PubMed: 21692505]
5. Andrews LD, Fenn TD, Herschlag D. Ground state destabilization by anionic nucleophiles contributes to the activity of phosphoryl transfer enzymes. *PLoS Biol*. 2013; 11:e1001599. [PubMed: 23843744]
6. Bandaria JN, Dutta S, Nydegger MW, Rock W, Kohen A, Cheatum CM. Characterizing the dynamics of functionally relevant complexes of formate dehydrogenase. *Proc Natl Acad Sci USA*. 2010; 107:17974–17979. [PubMed: 20876138]
7. Barducci A, Bonomi M, Parrinello M. *Metadynamics*. Wiley Inter Rev Comp Mol Sci. 2011; 1:826–843.
8. Becke A. Density-functional thermochemistry. iii. the role of exact exchange. *J Chem Phys*. 1993; 98:5648–5652.
9. Becke AD. Perspective: Fifty years of density-functional theory in chemical physics. *J Chem Phys*. 2014; 140:18A301.
10. Benkovic SJ. Catalytic antibodies. *Annu Rev Biochem*. 1992; 61:29–54. [PubMed: 1497313]
11. Benkovic SJ, Hammes-Schiffer S. A perspective on enzyme catalysis. *Science*. 2003; 301:1196–1202. [PubMed: 12947189]
12. Blomberg MRA, Borowski T, Himo F, Liao RZ, Siegbahn PEM. Quantum chemical studies of mechanisms for metalloenzymes. *Chem Rev*. 2014; 114:3601–3658. [PubMed: 24410477]
13. Blomberg R, Kries H, Pinkas DM, Mittl PRE, Grutter MG, Privett HK, Mayo SL, Hilvert D. Precision is essential for efficient catalysis in an evolved kemp eliminase. *Nature*. 2013; 503:418. [PubMed: 24132235]
14. Bobyr E, Lassila JK, Wiersma-Koch HI, Fenn TD, Lee JJ, Nikolic-Hughes I, Hodgson KO, Rees DC, Hedman B, Herschlag D. High-resolution analysis of zn²⁺ coordination in the alkaline phosphatase superfamily by exafs and x-ray crystallography. *J Mol Biol*. 2012; 415:102–117. [PubMed: 22056344]
15. Boehr DD, Dyson HJ, Wright PE. An nmr perspective on enzyme dynamics. *Chem Rev*. 2006; 106:3055–3079. [PubMed: 16895318]
16. Bolhuis PG, Chandler D, Dellago C, Geissler PL. Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annu Rev Phys Chem*. 2002; 53:291–318. [PubMed: 11972010]
17. Bowman GR, Beauchamp KA, Boxer G, Pande VS. Progress and challenges in the automated construction of markov state models for full protein systems. *J Chem Phys*. 2009; 131:124101. [PubMed: 19791846]

18. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. Charmm: A program for macromolecular energy, minimization and dynamics calculations. *J Comput Chem.* 1983; 4:187–217.
19. Brooks BR, CLB, Mackerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoseck M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. Charmm: The biomolecular simulation program. *J Comput Chem.* 2009; 30:1545–1614. [PubMed: 19444816]
20. Brooks CL III, Karplus M. Solvent effects on protein motion and protein effects on solvent motion: Dynamics of the active site region of lysozyme. *J Mol Biol.* 1989; 208:159–181. [PubMed: 2769750]
21. Brunk E, Rothlisberger U. Mixed quantum mechanical/molecular mechanical molecular dynamics simulations of biological systems in ground and electronically excited states. *Chem Rev.* 2015; 115:6217–6263. [PubMed: 25880693]
22. Chan GKL, Sharma S. The density matrix renormalization group in quantum chemistry. *Annu Rev Phys Chem.* 2011; 62:465–481. [PubMed: 21219144]
23. Chen J, Sharma S, Quirocho FA, Davidson AL. Trapping the transition state of an atp-binding cassette transporter: Evidence for a concerted mechanism of maltose transport. *Proc Natl Acad Sci USA.* 2001; 98:1525–1530. [PubMed: 11171984]
24. Claeysens F, Harvey JN, Manby FR, Mata RA, Mulholland AJ, Ranaghan KE, Schutz M, Thiel S, Thiel W, Werner HJ. High-accuracy computation of reaction barriers in enzymes. *Angew Chem Int Ed.* 2006; 45:6856–6859.
25. Cleland WW, Hengge AC. Enzymatic mechanisms of phosphate and sulfate transfer. *Chem Rev.* 2006; 106:3252–3278. [PubMed: 16895327]
26. Cohen AJ, Mori-Sanchez P, Yang WT. Challenges for density functional theory. *Chem Rev.* 2012; 112:289–320. [PubMed: 22191548]
27. Cruywagen JJ. Protonation, oligomerization, and condensation reactions of vanadate(v), molybdate(vi), and tungstate(vi). *Adv Inorgan Chem.* 2000; 49:127–182.
28. Cui Q, Elstner M. Density functional tight binding: values of semi-empirical methods in an *ab initio* era. *Phys Chem Chem Phys.* 2014; 16:14368–14377. [PubMed: 24850383]
29. Cui Q, Karplus M. Qm/mm studies of the triosephosphate isomerase (tim) catalyzed reactions: The effect of geometry and tunneling on proton transfer rate constants. *J Am Chem Soc.* 2002; 124:3093–3124. [PubMed: 11902900]
30. Cui, Q. Meuwly, M., Ren, PY., editors. *Many-Body Effects and Electrostatics in Biomolecules*. Pan Stanford Publishing; 2016.
31. Darden T, York D, Pedersen L. Particle mesh ewald - an $n \cdot \log(n)$ method for ewald sums in large systems. *J Chem Phys.* 1993; 98:10089–10092.
32. Deng YQ, Roux B. Computations of standard binding free energies with molecular dynamics simulations. *J Phys Chem B.* 2009; 113:2234–2246. [PubMed: 19146384]
33. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP. Development and use of quantum mechanical molecular models. 76. am1: a new general purpose quantum mechanical molecular model. *Journal of the American Chemical Society.* 1985; 107:3902–3909. DOI: 10.1021/ja00299a024
34. Doron D, Kohen A, Nam K, Major DT. How accurate are transition states from simulations of enzymatic reactions? *J Chem Theory Comput.* 2014; 10:1863–1871. [PubMed: 24860275]
35. Doron D, Major DT, Kohen A, Thiel W, Wu X. 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.* 2011; 7:3420–3437. [PubMed: 26598171]
36. EW, Vanden-Eijnden E. Transition-path theory and path-finding algorithms for the study of rare events. *Annu Rev Phys Chem.* 2010; 61:391–420. [PubMed: 18999998]
37. Elstner M, Porezag D, Jungnickel G, Elsner J, Haugk M, Frauenheim T, Suhai S, Seifert G. Self-consistent-charge densityfunctional tight-binding method for simulations of complex materials properties. *Phys Rev B.* 1998; 58:7260–7268.
38. Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*. W. H. Freeman and Company; 1999.

39. Field MJ, Bash PA, Karplus M. A combined quantum-mechanical and molecular mechanical potential for molecular-dynamics simulations. *J Comput Chem.* 1990; 11:700–733.
40. Freindorf M, Gao JL. Optimization of the Lennard-Jones parameters for a combined ab initio quantum mechanical and molecular mechanical potential using the 3–21G basis set. *J Comput Chem.* 1996; 17:386–395.
41. Friesner RA, Guallar V. Ab initio qm and qm/mm methods for studying enzyme catalysis. *Annu Rev Phys Chem.* 2005; 56:389–427. [PubMed: 15796706]
42. Frisch, MJ., Trucks, GW., Schlegel, HB., Scuseria, GE., Robb, MA., Cheeseman, JR., Montgomery, JA., Jr, Vreven, T., Kudin, KN., Burant, JC., Millam, JM., Iyengar, SES., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Petersson, GA., Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Knox, JE., Hratchian, HP., Cross, JB., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, RE., Yazyev, O., Austin, AJ., Cammi, R., Pomelli, C., Ochterski, JW., Ayala, PY., Morokuma, K., Voth, GA., Salvador, P., Dannenberg, JJ., Zakrzewski, VG., Dapprich, S., Daniels, AD., Strain, MC., Farkas, O., Malick, DK., Rabuck, AD., Raghavachari, K., Resman, JBF, Ortiz, JV., Cui, Q., Baboul, AG., Clifford, S., Cioslowski, J., Stefanov, BB., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Martin, RL., Fox, DJ., Keith, T., Laham, MAA., Peng, CY., Nanayakkara, A., Challacombe, M., Gill, PMW., Johnson, B., Chen, W., Wong, MW., Gonzalez, C., Pople, JA. Gaussian 03, Revision C.01. Gaussian, Inc; Wallingford CT, U.S.A: 2004.
43. Gao JL, Ma SH, Major DT, Nam K, Pu JZ, Truhlar DG. Mechanisms and free energies of enzymatic reactions. *Chem Rev.* 2006; 106:3188–3209. [PubMed: 16895324]
44. Gao JL, Truhlar DG. Quantum mechanical methods for enzyme kinetics. *Annu Rev Phys Chem.* 2002; 53:467–505. [PubMed: 11972016]
45. Garcia-Viloca M, Gao J, Karplus M, Truhlar DG. How enzymes work: Analysis by modern rate theory and computer simulations. *Science.* 2004; 303:186–195. [PubMed: 14716003]
46. Gaus M, Cui Q, Elstner M. Dftb-3rd: Extension of the self-consistent-charge density-functional tight-binding method scc-dftb. *J Chem Theory Comput.* 2011; 7:931–948.
47. Gaus M, Goez A, Elstner M. Parametrization and benchmark of dftb3 for organic molecules. *J Chem Theory Comput.* 2012; 9:338–354. [PubMed: 26589037]
48. Gaus, M., Goyal, P., Hou, G., Lu, X., Pang, X., Zienau, J., Xu, X., Elstner, M., Cui, Q. Molecular modeling at the atomic scale: Methods and applications in quantitative biology. CRC Press; 2014. p. 33-82.
49. Gaus M, Lu X, Elstner M, Cui Q. Parameterization of dftb3/3ob for sulfur and phosphorus for chemical and biological applications. *J Chem Theory Comput.* 2014; 10:1518–1537. [PubMed: 24803865]
50. Ghosh N, Prat-Resina X, Cui Q. Towards a reliable molecular model of cytochrome c oxidase: insights from microscopic pK_a calculations. *Biochem.* 2009; 48:2468–2485. [PubMed: 19243111]
51. Giese TJ, York DM. Charge-dependent model for many-body polarization, exchange, and dispersion interactions in hybrid quantum mechanical/molecular mechanical calculations. *J Chem Phys.* 2007; 127:194101. [PubMed: 18035873]
52. Goyal P, Lu J, Yang S, Gunner MR, Cui Q. Changing hydration level in an internal cavity modulates the proton affinity of a key glutamate in cytochrome c oxidase. *Proc Natl Acad Sci USA.* 2013; 110:18886–18891. [PubMed: 24198332]
53. Goyal P, Qian HJ, Irle S, Lu X, Roston D, Mori T, Elstner M, Cui Q. *Feature Article*: Molecular simulation of water and hydration effects in different environments: Challenges and developments for DFTB based models. *J Phys Chem B.* 2014; 118:11007–11027. [PubMed: 25166899]
54. Griffith WP, Lesniak PJB. Raman studies on species in aqueous solutions.3. vanadates molybdates and tungstates. *J Chem Soc A - Inorg Phys Theor.* 1969
55. Halperin I, Ma BY, Wolfson H, Nussinov R. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins: Struct Funct & Gene.* 2002; 47:409–443.
56. Hay PJ, Wadt WR. Ab initio effective core potentials for molecular calculations - potentials for the transition-metal atoms sc to hg. *J Chem Phys.* 1985; 82:270–283.
57. Hay S, Scrutton NS. Good vibrations in enzyme-catalysed reactions. *Nat Chem.* 2012; 4:161–168. [PubMed: 22354429]

58. Henzler-Wildman KA, Lei M, Thai V, Kerns SJ, Karplus M, Kern D. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature*. 2007; 450:913–916. [PubMed: 18026087]
59. Hoffman M, Wanko M, Strodel P, Konig PH, Frauenheim T, Schulten K, Thiel W, Tajkhorshid E, Elstner M. Color tuning in rhodopsins: The mechanism for the spectral shift between bacteriorhodopsin and sensory rhodopsin ii. *J Am Chem Soc*. 2006; 128:10808–10818. [PubMed: 16910676]
60. Hollfelder F, Herschlag D. The nature of the transitionstate for enzyme-catalyzed phosphoryl transfer - hydrolysis of o-aryl phosphorothioates by alkaline-phosphatase. *Biochem*. 1995; 38:12255–12264.
61. Holtz KM, Stec B, Kantrowitz ER. A model of the transition state in the alkaline phosphatase reaction. *J Biol Chem*. 1999; 274:8351–8354. [PubMed: 10085061]
62. Hou G, Zhu X, Elstner M, Cui Q. A modified qm/mm hamiltonian with the self-consistent-charge density-functional-tight-binding theory for highly charged qm regions. *J Chem Theory Comput*. 2012; 8:4293–4304. [PubMed: 23275762]
63. Hou GH, Cui Q. QM/MM analysis suggests that Alkaline Phosphatase (AP) and Nucleotide pyrophosphatase/phosphodiesterase (NPP) slightly tighten the transition state for phosphate diester hydrolysis relative to solution: implication for catalytic promiscuity in the AP superfamily. *J Am Chem Soc*. 2012; 134:229–246. [PubMed: 22097879]
64. Hou GH, Cui Q. Stabilization of different types of transition states in a single enzyme active site: QM/MM analysis of enzymes in the alkaline phosphatase superfamily. *J Am Chem Soc*. 2013; 135:10457–10469. [PubMed: 23786365]
65. Hu H, Boone A, Yang WT. Mechanism of omp decarboxylation in orotidine 5'-monophosphate decarboxylase. *J Am Chem Soc*. 2008; 130:14493–14503. [PubMed: 18839943]
66. Hu H, Lu ZY, Parks JM, Burger SK, Yang WT. Quantum mechanics/molecular mechanics minimum free-energy path for accurate reaction energetics in solution and enzymes: Sequential sampling and optimization on the potential of mean force surface. *J Chem Phys*. 2008; 128:034105. [PubMed: 18205486]
67. Hu H, Yang WT. Free energies of chemical reactions in solution and in enzymes with ab initio quantum mechanics/molecular mechanics methods. *Annu Rev Phys Chem*. 2008; 59:573–601. [PubMed: 18393679]
68. Hu LH, Ryde U. Comparison of methods to obtain force-field parameters for metal sites. *J Chem Theory Comput*. 2011; 7:2425–2463.
69. Huang L, Roux B. Automated force field parameterization for nonpolarizable and polarizable atomic models based on ab initio target data. *J Chem Theory Comput*. 2013; 9:3543–3556.
70. Im W, Berneche S, Roux B. Generalized solvent boundary potential for computer simulations. *J Chem Phys*. 2001; 114:2924–2937.
71. Jiao D, Golubkov PA, Darden TA, Ren P. Calculation of protein-ligand binding free energy by using a polarizable potential. *Proc Natl Acad Sci USA*. 2009; 105:6290–6295.
72. Kamerlin SCL, Haranczyk M, Warshel A. Progress in ab initio QM/MM free-energy simulations of electrostatic energies in proteins: Accelerated qm/mm studies of pk(a), redox reactions and solvation free energies. *J Phys Chem B*. 2009; 113:1253–1272. [PubMed: 19055405]
73. Kamerlin SCL, Warshel A. At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis? *Proteins: Struct, Funct, & Bioinf*. 2010; 78:1339–1375.
74. Kazemi M, Himo F, Aqvist J. Enzyme catalysis by entropy without circe effect. *Proc Natl Acad Sci USA*. 2016 In press.
75. Kerns SJ, Agafonov RV, Cho YJ, Pontiggia F, Otten R, Pachov DV, Kutter S, Phung LA, Murphy PN, Thai V, Alber T, Hagan MF, Kern D. The energy landscape of adenylate kinase during catalysis. *Nat Struct Mol Biol*. 2015; 22:124–131. [PubMed: 25580578]
76. Kipp DR, Hirschi JS, Wakata A, Goldstein H, Schramm VL. Transition states of native and drug-resistant HIV-1 protease are the same. *Proc Natl Acad Sci USA*. 2012; 109:6543–6548. [PubMed: 22493227]
77. Kraut DA, Carroll KS, Herschlag D. Challenges in enzyme mechanism and energetics. *Annu Rev Biochem*. 2003; 72:517–571. [PubMed: 12704087]

78. Kries H, Blomberg R, Hilvert D. De novo enzymes by computational design. *Curr Opin Chem Biol.* 2013; 17:221–228. [PubMed: 23498973]
79. Kulik HJ, Leuhr N, Ufimtsev IS, Martinez TJ. Ab initio quantum chemistry for protein structures. *J Phys Chem B.* 2012; 116:12501–12509. [PubMed: 22974088]
80. Kurashige Y, Chan GKL, Yanai T. Entangled quantum electronic wavefunctions of the mn4cao5 cluster in photosystem ii. *Nat Chem.* 2013; 5:660–666. [PubMed: 23881496]
81. Labow BI, Herschlag D, Jencks WP. Catalysis of the hydrolysis of phosphorylated pyridines by alkaline-phosphatase has little or no dependence on the pk(a) of the leaving group. *Biochem.* 1993; 32:8737–8741. [PubMed: 8395879]
82. Lassila JK, Baker D, Herschlag D. Origins of catalysis by computationally designed retroaldolase enzymes. *Proc Natl Acad Sci USA.* 2010; 107:4937–4942. [PubMed: 20194782]
83. Lassila JK, Zalatan JG, Herschlag D. Biological phosphoryl transfer reactions: Understanding mechanism and catalysis. *Annu Rev Biochem.* 2011; 80:669–702. [PubMed: 21513457]
84. Lee C, Yang W, Parr RG. Development of the collesalvetti correlation-energy formula into a functional of the electron density. *Phys Rev B.* 1988; 37:785–789. DOI: 10.1103/PhysRevB.37.785
85. Li G, Cui Q. pk_a calculations with qm/mm free energy perturbations. *J Phys Chem B.* 2003; 107:14521–14528.
86. Liao RZ, Thiel W. Convergence in the qm-only and qm/mm modeling of enzymatic reactions: A case study for acetylene hydratase. *J Comp Chem.* 2013; 34:2389–2397. [PubMed: 23913757]
87. Lindqvist Y, Schneider G, Vihko P. Crystal-structures of rat acid-phosphatase complexed with the transition-state analogs vanadate and molybdate - implications for the reaction-mechanism. *Eur J Biochem.* 1994; 221:139–142. [PubMed: 8168503]
88. Gao, J. In *Reviews in Computational Chemistry VII*. Lipkowitz, KB., Boyd, DB., editors; VCH; New York: 1995.
89. Lu X, Cui Q. Charging free energy calculations using the generalized solvent boundary potential (gsbp) and periodic boundary condition: a comparative analysis using ion solvation and reduction potential in proteins. *J Phys Chem B.* 2013; 117:2005–2018. [PubMed: 23347181]
90. Lu X, Fang D, Ito S, Okamoto Y, Ovchinnikov V, Cui Q. Qm/mm free energy simulations: recent progress and challenges. *Mol Simul.* 2016 In press.
91. Lu X, Gaus M, Elstner M, Cui Q. Parameterization of DFTB3/3OB for Magnesium and Zinc for chemical and biological applications. *J Phys Chem B (Jorgensen Festschrift).* 2015; 119:1062–1082.
92. van der Kamp MW, Mulholland AJ. Combined quantum mechanics/molecular mechanics (qm/mm) methods in computational enzymology. *Biochem.* 2013; 52:2708–2728. [PubMed: 23557014]
93. MacKerell AD Jr, Feig M, Brooks CL III. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comp Chem.* 2004; 25:1400–1415. [PubMed: 15185334]
94. Mader MM, Bartlett PA. Binding energy and catalysis: The implications for transition-state analogs and catalytic antibodies. *Chem Rev.* 1997; 97:1281–1301. [PubMed: 11851452]
95. Marti S, Moliner V, Tuñón I. Improving the qm/mm description of chemical processes: A dual level strategy to explore the potential energy surface in very large systems. *J Chem Theory Comput.* 2005; 1:1008–1016. [PubMed: 26641916]
96. Molina PA, Jensen JH. A predictive model of strong hydrogen bonding in proteins: The $n\delta$ -h-o δ hydrogen bond in low-ph alpha-chymotrypsin and alpha-lytic protease. *J Phys Chem B.* 2003; 107:6226–6233.
97. Monard G, KM, Merz J. Combined quantum mechanical/molecular mechanical methodologies applied to biomolecular systems. *Acc Chem Res.* 1999; 32:904–911.
98. Nagel ZD, Klinman JP. A 21(st) century revisionist's view at a turning point in enzymology. *Nat Chem Biol.* 2009; 5:543–550. [PubMed: 19620995]
99. O'Brien PJ, Herschlag D. Sulfatase activity of e-coli alkaline phosphatase demonstrates a functional link to arylsulfatases, an evolutionarily related enzyme family. *J Am Chem Soc.* 1998; 120:12369–12370.

100. O'Brien PJ, Herschlag D. Functional interrelationships in the alkaline phosphatase superfamily: phosphodiesterase activity of escherichia coli alkaline phosphatase. *Biochem.* 2001; 40:5691–5699. [PubMed: 11341834]
101. Parr, RG., Yang, WT. *Density-Functional Theory of Atoms and Molecules.* New York: Oxford University Press; 1989.
102. Pearson AD, Mills JH, Song Y, Nasertorabi F, Han GW, Baker D, Stevens RC, Schultz PG. Trapping a transition state in a computationally designed protein bottle. *Science.* 2015; 347:863–867. [PubMed: 25700516]
103. Peck A, Sunden F, Andrews LD, Pande VS, Herschlag D. Tungstate as a transition state analog for catalysis by alkaline phosphatase. *Biochem.* 2015 Submitted.
104. Phatak P, Ghosh N, Yu H, Cui Q, Elstner M. Amino acids with an intermolecular proton bond as the proton storage site in bacteriorhodopsin. *Proc Acad Natl Sci USA.* 2008; 105:19672–19677.
105. Plotnikov NV, Kamerlin SCL, Warshel A. Para-dynamics: An effective and reliable model for ab initio qm/mm free-energy calculations and related tasks. *J Phys Chem B.* 2011; 115:7950–7962. [PubMed: 21618985]
106. Polyak I, Benighaus Boulanger TE, Thiel W. Quantum mechanics/molecular mechanics dual hamiltonian free energy perturbation. *J Chem Phys.* 2013; 139:064105. [PubMed: 23947841]
107. Ponder JW, Wu CJ, Ren PY, Pande VS, Chodera JD, Schnieders MJ, Haque I, Mobley DL, Lambrecht DS, DiStasio RA, Head-Gordon M, Clark GNI, Johnson ME, Head-Gordon T. Current status of the amoeba polarizable force field. *J Phys Chem B.* 2010; 114:2549–2564. [PubMed: 20136072]
108. Pu JZ, Gao JL, Truhlar DG. Multidimensional tunneling, recrossing, and the transmission coefficient for enzymatic reactions. *Chem Rev.* 2006; 106:3140–3169. [PubMed: 16895322]
109. Riccardi D, Schaefer P, Yang Y, Yu H, Ghosh N, Prat-Resina X, Konig P, Li G, Xu D, Guo H, Elstner M, Cui Q. *Feature Article:* Development of effective quantum mechanical/molecular mechanical (QM/MM) methods for complex biological processes. *J Phys Chem B.* 2006; 110:6458–6469. [PubMed: 16570942]
110. Riccardi D, Yang S, Cui Q. Proton transfer function of carbonic anhydrase: Insights from qm/mm simulations. *Biochim Biophys Acta.* 2010; 1804:342–351. [PubMed: 19679196]
111. Rod TH, Ryde U. Accurate qm/mm free energy calculations of enzyme reactions: Methylation by catechol o-methyltransferase. *J Chem Theory Comput.* 2005; 1:1240–1251. [PubMed: 26631668]
112. Rossi I, Truhlar DG. Parameterization of nddo wave-functions using genetic algorithms - an evolutionary approach to parameterizing potential-energy surfaces and direct dynamics calculations for organic-reactions. *Chem Phys Lett.* 1995; 233:231–236.
113. Roston D, Demapan D, Cui Q. Leaving group ability affects transition state structure for phosphoryl transfer in a single enzyme active site. 2016 Submitted.
114. Roston D, Kohen A. Elusive transition state of alcohol dehydrogenase unveiled. *Proc Natl Acad Sci USA.* 2010; 107:9572–9577. [PubMed: 20457944]
115. Rothlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, Gallaher JL, Althoff EA, Zanghellini A, Dym O, Albeck S, Houk KN, Tawfik DS, Baker D. Kemp elimination catalysts by computational enzyme design. *Nature.* 2008; 453:190–194. [PubMed: 18354394]
116. Salomon-Ferrer R, Case D, Walker R. An overview of the amber biomolecular simulation package. *WIREs Comput Mol Sci.* 2013; 3:198–210. DOI: 10.1002/wcms.1121
117. Schaefer P, Riccardi D, Cui Q. Reliable treatment of electrostatics in combined QM/MM simulation of macromolecules. *J Chem Phys.* 2005; 123 Art No. 014905.
118. Schramm VL. Enzymatic transition states, transition-state analogs, dynamics, thermodynamics, and lifetimes. *Annu Rev Biochem.* 2011; 80:703–732. [PubMed: 21675920]
119. Schramm VL. Transition states and transition state analogue interactions with enzymes. *Acc Chem Res.* 2015; 48:1032–1039. [PubMed: 25848811]
120. Schultz PG. Catalytic antibodies. *Angew Chem Int Ed.* 1989; 28:1283–1295.
121. Schwartz SD, Schramm VL. Enzymatic transition states and dynamic motion in barrier crossing. *Nat Chem Biol.* 2009; 5:552–559.

122. Senn HM, Thiel W. QM/MM methods for biomolecular systems. *Angew Chem Int Ed*. 2009; 48:1198–1229.
123. Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP, Bank JA, Jumper JM, Salmon JK, Shan YB, Wriggers W. Atomic-level characterization of the structural dynamics of proteins. *Science*. 2010; 330:341–346. [PubMed: 20947758]
124. Shukla D, Meng YL, Roux B, Pande VS. Activation pathway of src kinase reveals intermediate states as targets for drug design. *Nat Comm*. 2014; 5:3397.
125. Siegel JB, Zanghellini A, Lovick HM, Kiss G, Lambert AR, Clair JLS, Gallaher JL, Hilvert D, Gelb MH, Stoddard BL, Houk KN, Michael FE, Baker D. Computational design of an enzyme catalyst for a stereoselective bimolecular dielsalder reaction. *Science*. 2010; 329:309–313. [PubMed: 20647463]
126. Simonson T, Archontis G, Karplus M. Free energy simulations come of age: Protein-ligand recognition. *Acc Chem Res*. 2002; 35:430–437. [PubMed: 12069628]
127. Stec B, Holtz KM, Kantrowitz ER. A revised mechanism for the alkaline phosphatase reaction involving three metal ions. *J Mol Biol*. 2000; 299:1303–1311. [PubMed: 10873454]
128. Stewart JJP. Optimization of parameters for semiempirical methods ii. applications. *Journal of Computational Chemistry*. 1989; 10:221–264. URL: <http://dx.doi.org/10.1002/jcc.540100209>. DOI: 10.1002/jcc.540100209
129. Stote RH, States DJ, Karplus M. On the treatment of electrostatic interactions in biomolecular simulation. *Journal of Chimie Physique*. 1991; 88:2419–2433.
130. Straatsma TP, McCammon JA. Computational alchemy. *Annu Rev Phys Chem*. 1992; 43:407–435.
131. Sunden F, Peck A, Salzman J, Ressler S, Herschlag D. Extensive site-directed mutagenesis reveals interconnected functional units in the alkaline phosphatase active site. *eLife*. 2015; 4:e06181.
132. Benighaus T, Thiel W. Long-Range Electrostatic Effects in QM/MM Studies of Enzymatic Reactions: Application of the Solvated Macromolecule Boundary Potential. *J Chem Theory Comput*. 2011; 7:238–249. [PubMed: 26606237]
133. Tao P, Hodoscek M, Larkin JD, Shao YH, Brooks BR. Comparison of three chain-of-states methods: Nudged elastic band and replica path with restraints or constraints. *J Chem Theory Comput*. 2012; 8:5035–5051. [PubMed: 23526888]
134. Tomasi J, Mennucci B, Cammi R. Quantum mechanical continuum solvation models. *Chem Rev*. 2005; 105:2999–3093. [PubMed: 16092826]
135. Torrie GM, Valleau JP. Non-physical sampling distributions in Monte-Carlo free-energy estimation: Umbrella sampling. *J Comput Phys*. 1977; 23:187–199.
136. Vanommeslaeghe K, Raman EP, MacKerell AD Jr. Automation of the charmm general force field (cgenff) ii: assignment of bonded parameters and partial atomic charges. *J Chem Inf Model*. 2012; 52:3155–3168. [PubMed: 23145473]
137. Wahiduzzaman M, Oliveira AF, Philipsen P, Zhechkov L, Evan Lenthe Witek HA, Heine T. Dftb parameters for the periodic table: Part 1, electronic structure. *J Chem Theory Comput*. 2013; 9:4006–4017. [PubMed: 26592396]
138. Wales, DJ. *Energy Landscapes*. Cambridge University Press; 2003.
139. Wanko M, Hoffman M, Strodel P, Koslowski A, Thiel W, Neese F, Frauenheim T, Elstner M. Calculating absorption shifts for retinal proteins: Computational challenges. *J Phys Chem B*. 2005; 109:3606–3615. [PubMed: 16851399]
140. Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solution*. Wiley; New York: 1991.
141. Warshel A, Levitt M. Theoretical studies of enzymic reactions - dielectric, electrostatic and steric stabilization of carbonium-ion in reaction of lysozyme. *J Mol Biol*. 1976; 103:227–249. [PubMed: 985660]
142. Warshel A, Sharma PK, Kato M, Xiang Y, Liu HB, Olsson MHM. Electrostatic basis for enzyme catalysis. *Chem Rev*. 2006; 106:3210–3235. [PubMed: 16895325]
143. Wolf S, Freier E, Cui Q, Gerwert K. Infrared spectral marker bands characterizing a transient water wire inside a hydrophobic membrane protein. *J Chem Phys*. 2014; 141:22D524.

144. Wolf S, Freier E, Gerwert K. A delocalized proton-binding site within a membrane protein. *Biophys J*. 2014; 107:174–184. [PubMed: 24988352]
145. Zalatan JG, Fenn TD, Brunger AT, Herschlag D. Structural and functional comparisons of nucleotide pyrophosphatase/phosphodiesterase and alkaline phosphatase: Implications for mechanism and evolution. *Biochem*. 2006; 45:9788–9803. [PubMed: 16893180]
146. Zalatan JG, Fenn TD, Herschlag D. Comparative enzymology in the alkaline phosphatase superfamily to determine the catalytic role of an active-site metal ion. *J Mol Biol*. 2008; 384:1174–1189. [PubMed: 18851975]
147. Zalatan JG, Herschlag D. Alkaline phosphatase monoand diesterase reactions: Comparative transition state analysis. *J Am Chem Soc*. 2006; 128:1293–1303. [PubMed: 16433548]
148. Zhang M, Zhou M, VanEtten RL, Stauffacher CV. Crystal structure of bovine low molecular weight phosphotyrosyl phosphatase complexed with the transition state analog vanadate. *Biochem*. 1997; 36:15–23. [PubMed: 8993313]
149. Zienau J, Cui Q. Implementation of the solvent macromolecule boundary potential and application to model and realistic enzyme systems. *J Phys Chem B*. 2012; 116:12522–12534. [PubMed: 22985044]

Figure 1.

Artifacts may arise from uncalibrated QM-MM interactions in enzyme simulations. A) The crystal structure of Alkaline Phosphatase (AP) with bound PO_4^{3-} (PDB: 1ED8) shows a H-bond between the sidechain of His331 and the backbone of Gln410. B) A representative snapshot from a DFTB3/MM simulation of the system shows that the His331-Gln410 H-bond is broken and a new one has formed between the sidechains of His331 and Asp330. This lengthens the Zn-Zn distance and perturbs the binding mode of the PO_4^{3-} . In panel B, the yellow ball in the His331 sidechain indicates the location of the QM link atom separating the QM and MM subsystems.

Figure 2.

The hydrolysis of phosphate esters in AP occurs by phosphorylation of the S102 nucleophile followed by hydrolysis of that phosphorylated enzyme intermediate. Phosphoryl transfer pathways vary as a function of the lengths of the breaking and forming $P-O_{lg}$ and $P-O_{nuc}$, respectively) and can occur through tight or loose pathways, or somewhere in between (synchronous). Here we explore how different TSs and different TS analogue inhibitors vary in their geometries and interactions with active site motifs.

Figure 3.

The atoms in the active site model of AP, as well as the QM region used for the QM/MM calculations of the full enzyme. Note that in the B3LYP/MM optimizations of the full enzyme model, all waters were treated as MM.

Figure 4.

Structures of the active site model of AP with tungstate using different solvent models (indicated by color). Structures were optimized at the B3LYP/6-31+G* level, using an effective core potential for the metal atoms. The bond from the W to the nucleophilic serine alkoxide (O_{nuc}) is indicated by the arrow. Alpha carbons were frozen to their crystal structure[127] positions during the optimizations.

Figure 5.

Optimized structures from B3LYP/MM calculations of tungstate in the active site of AP. Starting structures were obtained from crystal structures and simulations with other ligands as indicated by color. In cases where the W-O_{nuc} distance is less than 2.5 Å, a bond is drawn between the two atoms, indicated by the arrow.

Figure 6.

PES scan at the B3LYP/MM level of the W-O_{nuc} distance, starting from a structure that converged to one of the longest W-O_{nuc} distances (4.34 Å). The covalently bound complex is approximately 15 kcal/mol more stable than the van der Waals complex, and the barrier between the two is around 8 kcal/mol. Note that these should be regarded as semi-quantitative estimates since only potential energy is included in the computations.

Figure 7.

Optimized structures with VO_4^{3-} (A) and HVO_4^{2-} (B) in AP. Bond lengths between V-O_{nuc} and V-O_{lg} are indicated in the figure. For reference, those bond lengths are 1.72 Å and 1.92 Å, respectively, in the crystal structure (See Fig. 9).[61] The bond lengths in (A) differ slightly from those calculated previously for that form79 because the starting structures for the optimizations differed. Also visible in these structures is the notable difference in geometry around the V, which is tetrahedral in (A), but trigonal bipyramidal in (B). These optimizations were both initiated from the same optimized structure with tungstate covalently bound to the serine nucleophile.

Figure 8.

Overlays of representative structures of TSAs (calculated, red; crystallographic, blue) and TS structures (computed with DFTB3/MM free energy simulations[113], in green) for two different phosphate monoesters. A) Phenyl phosphate and vanadate. B) Ethyl phosphate and vanadate. C) Phenyl phosphate and tungstate. D) Ethyl phosphate and tungstate. These structures suggest that in terms of overall geometry especially ligation by the Zn^{2+} ion the TSAs are more representative of the TS for ethyl phosphate, but in terms of M-O_{nuc} and M-O_{lg} bond lengths, tungstate more resembles phenyl phosphate (cf. Fig. 9). The structures were aligned based on the alpha carbons of the residues composing the QM region.

Figure 9.

The TS region of the phosphate ester hydrolysis reaction, showing the various TS structures for phosphate monoesters found in free energy simulations,[113] as well as the crystallographic TSA structures (including one vanadate structure bound to a closely related enzyme, NPP[145]) and the most relevant calculated TSA structures. The guides as to the tight, synchronous, and loose regions are determined from the sum of bond order of the P- O_{lg} and P- O_{nuc} bonds as described in ref. [113]. The calculated structure of tungstate here is the average of the structures that converged to a covalent bond in the initial optimizations. The points labeled QM/MM refer to the full enzyme model and those labeled QM refer to the active site model.

Table 1Optimized geometries of small molecule crystals^a

	M-O1		M-O2		M-O3		M-O4		M-O5	
	Calc	Exp	Calc	Exp	Calc	Exp	Calc	Exp	Calc	Exp
WO ₄ ²⁻	1.79	1.80	1.77	1.80	1.78	1.80	1.81	1.80	NA	NA
VO ₄ ³⁻	1.76	1.72	1.76	1.72	1.72	1.72	1.72	1.72	NA	NA
CINMES ^b	1.91	1.87	1.91	1.87	1.91	1.87	1.91	1.87	1.71	1.70
YIOLEP ^c	1.83	1.87	1.88	1.86	1.91	1.89	2.15	2.09	2.11	2.10

^a Each entry is the distance from the central metal to the oxygen (in Å), calculated at the B3LYP/6-31+G* level using an effective core potential for the metal ion.^b A pentacoordinate square pyramidal WO₅-derivative.^c A pentacoordinate trigonal bipyramidal VO₅-derivative. Experimental geometries are from the Inorganic Crystal Structure Database (ICSD) and the Cambridge Structural Database (CSD).

Table 2Initial and final W-O_{nuc} distance (in Å) for each optimized structure

Source ^a	Initial value ^b	Final value
VO ₄ -xtal	1.72 ^c	2.12
PO ₄ -xtal	3.11	2.16
pNPP MD	4.08	4.26
pNPP MD	4.06	4.34
pNPP MD	2.86	2.24
pNPP MD	2.94	2.25
pNPP MD	3.10	2.22
pNPP MD	3.42	3.26
pNPP MD	4.05	4.42
pNPP MD	3.22	3.08

^aSource of initial coordinates for optimization.^bFollowing the initial optimization with PO₄³⁻ in the active site.^cSince a covalent bond is visible in the crystal structure but is unlikely to remain when the V is replaced with P, the initial optimization with PO₄³⁻ had a restraint on this distance to maintain the interactions in the crystal structure.

Table 3

Geometries^a of calculated and experimental TSAs and calculated TSs

	M(P)-O _{lg}	M(P)-O _{nuc}	Zn-Zn	Zn1-O _{lg}	Zn2-O _{nuc}	Zn1-O _{nb}	Zn2-O _{nb}
WO ₄ -xtal (SC66)	2.2	2.1	4.0	2.0	2.1	2.1	2.0
WO ₄ ²⁻ - active site	1.9	2.2	4.1	2.2	2.1	2.1	2.2
HWO ₄ ⁻ - active site	2.1	2.0	4.2	2.2	2.2	2.2	2.2
WO ₄ ²⁻ - QM/MM ^b	1.8	2.2	4.1	2.4	2.0	2.1	2.1
VO ₄ -xtal (1B8D)	1.9	1.7	4.1	2.4	1.9	2.1	2.2
VO ₄ ³⁻ - active site	1.8	3.6	4.3	2.1	2.0	2.4	2.2
HVO ₄ ²⁻ - active site	2.2	2.0	4.1	2.1	2.1	2.1	2.2
H ₂ VO ₄ ⁻ - active site	2.2	1.9	5.2	2.1	2.1	2.2	3.3
VO ₄ ³⁻ - QM/MM	1.8	2.5	4.0	2.2	2.0	2.1	2.1
HVO ₄ ²⁻ - QM/MM	2.1	2.0	4.1	2.1	2.0	2.1	2.2
PhOP-QM/MM ^c	1.9	2.1	4.2	3.5	2.0	2.0	3.1
EtOP-QM/MM ^c	2.1	1.8	4.2	2.1	2.3	2.3	2.1

^aDistances are in Å. “active site” indicates active site model optimized at the pure QM level; “QM/MM” indicates QM/MM optimized result.

^bAverage of the structures that converged to covalent W-O_{nuc}-bonds.

^cFrom ref.[113]; PhOP: Phenyl Phosphate; EtOP: Ethyl Phosphate.