Review

Israel Journal of Chemistry

DOI: 10.1002/ijch.202000014

Modelling Enzymatic Mechanisms with QM/MM Approaches: Current Status and Future Challenges

Rita P. Magalhães,[a] Henriques S. Fernandes,[a] and Sérgio F. Sousa*[b]

Abstract: Quantum mechanics/molecular mechanics (QM/MM) methods are presently a well-established alternative for the study of enzymatic reaction mechanisms. They enable the description of a small part of the enzyme, where reactions take place through QM, while the majority of the thousands of atoms that comprise these biomolecules are handled through MM. While different "flavors" and variations in the QM/MM field exist, this review will focus more on the application of the ONIOM methodology, presenting a fresh perspective on the application of this popular method in light of the growth in computational power and level of sophistication of the different methodologies that it can combine. In addition to a brief presentation of the basic principles behind these methods, this review will discuss different examples of applicability, common choices, practical considerations, and main problems involved, stemming from our experience in this field. Finally, a reflection on the future challenges for the next decade in the QM/MM modeling of enzymatic mechanisms is presented.

Keywords: Enzymatic Catalysis · Computational Enzymology · ONIOM · Polarized Embedding · Polarizable Force Fields · Linear Scaling DFT

1. Introduction

Enzymes play a central role in life, catalyzing many chemical and biological processes occurring in nature.[1–3] Understanding how enzymes catalyze their reactions is a problem of major importance, both from a fundamental and a practical perspective, with application in a variety of areas, from a more basic research that aims to understand how different events occur in the cell, to the development of new treatments for important diseases, and even in industrial biocatalytic applications.[4–7] Hence, it is not surprising that many studies are published each year trying to provide an atomic level understanding of the catalytic mechanism of different enzymes.

A plethora of experimental methods is routinely used to study enzymatic reactivity providing many important clues about the way how enzymes perform their reactions. Examples include different spectroscopic techniques, kinetic studies, studies with mutant enzymes, experiments in different pH conditions or temperatures, in the absence or presence of different metal atoms, among others. Such methods help to elucidate some of the structures and conformations adopted by enzymes along their reaction path and to identify some of the atoms and amino acid residues directly participating in their reactions. It also enables the determination of the kinetics of the reactions catalyzed. However, such methods often fail to present a full view of the enzymatic reaction, leaving many chemical questions unanswered.

Computational methods can be used to go beyond some of the limitations of the experimental methodologies, providing an alternative strategy to complement the information arising from these methods. When used with care, computational methods can provide a comparative analysis of different mechanistic proposals, enveloping an atomistic and even electronic perspective of the different proposals, helping researchers to discard mechanistic proposals and propose new ones.[8–12]

Many different computational methods have been used through the years to model enzymatic reactions.[9,13–24] Earlier examples adopted the popular cluster modelling approach,[25–28] which included atomistic models of the enzymes and reactions bearing only a small number of atoms around the active site (in some earlier cases as small as 15 to 20 atoms), treated with QM. With time, methods including the full enzyme gained popularity, particularly through the QM/MM framework, enabling the explicit inclusion of remaining atoms in the enzyme.

2. QM/MM Methods and the ONIOM Method

2.1 Basic Principles

The fundamental idea behind modern QM/MM methodologies has origins that trace back to the 1970s, following the works of Honig & Karplus,[29] and later of Warshel & Levitt.[30] These

[a] *R. P. Magalhães, H. S. Fernandes*
UCIBIO@REQUIMTE, BioSIM, Departamento de Biomedicina, Faculdade de Medicina da Universidade do Porto, Alameda Professor Hernâni Monteiro
4200-319 Porto, Portugal
[b] *S. F. Sousa*
UCIBIO@REQUIMTE, BioSIM, Departamento de Biomedicina, Faculdade de Medicina da Universidade do Porto, Alameda Professor Hernâni Monteiro
4200-319 Porto, Portugal
E-mail: sergiofsousa@med.up.pt

works paved the way for the methodological developments that emerged over the past 50 years and led to the attribution of the Nobel prize in Chemistry to Karplus, Warshel, and Levitt in 2013.

QM/MM methods assume that chemical systems of big dimensions, including enzymes and other biological systems, can be divided into two regions: an important electronic region that requires a quantum mechanical description (the QM region) and a surrounding region that can be treated by molecular mechanics (the MM region). The QM region includes the part of the chemical system directly involved in a chemical reaction under study, while the MM region comprehends the remaining of the system. The MM region is hence described as acting only indirectly on the electronically important region. Thus, in agreement with the underlying principle associated to QM/MM hybrid methods, when modeling an enzymatic reaction this is treated as a transformation involving only a relatively small number of atoms at the active site (the QM region), but that is influenced by the remaining of the enzyme and surrounding environment (the MM region).

Several different QM/MM methods have been made available through the years, differing in features like the type of scheme that is used to compute the QM/MM energies, the way how the boundary region is considered, the way how the interaction between the QM and MM regions is treated, and the inclusion or not of dynamics. Each method has its own advantages and disadvantages, which also depend on the type of enzyme and reaction under study. A comparison of different types of QM/MM methods falls outside the scope of the present work, but several excellent reviews can be found in the literature.[9,13,17–19,31–34]

Here, we will focus our attention on modeling QM/MM enzymatic reactions through the ONIOM method, presenting an updated and personal view on this methodology, in an attempt to link the historic evolution of the method, and its present application in light of the current challenges in the field, with some of the more anticipated developments for the next decade.

The ONIOM (our Own N-layer Integrated molecular Orbital molecular Mechanics) method is a subtractive multilayer multilevel method developed by Morokuma and coworkers in 1996,[35] following the publication of the IMOMM (Integrated Molecular Orbital Molecular Mechanics)[36] and IMOMO (Integrated Molecular Orbital Molecular Orbital)[37] methods. While it can be applied to n-layers, most commonly used implementations in computational enzymology adopt a 2-layers QM/MM approach.[31,38]

It has been described as "one of the most popular, successful, and easily-to-implement hybrid quantum mechanics/molecular mechanics (QM/MM) methods to treat complex molecular systems."[39] While opinions can greatly differ from researcher to researcher, it is certainly a robust alternative and one that we have been employing with success over the past 15 years.[40]

2.2 Applicability

QM/MM methods can have several different applications in the study of enzymes, in general, and enzymatic reactions in particular. In fact, the robustness of these methods offers a large variety of possibilities. Here, we highlight four typical applications.

Rita P. Magalhães received her BSc in Biochemistry from the Faculty of Sciences of the University of Porto. Her research project was developed at Stockholms Universitet on Spectroscopic Studies of Parallel Beta-Sheet Amiloyd Proteins. She recently completed her MSc in Chemistry, also at the University of Porto. Her dissertation was developed at BioSIM and focused on the Computational Identification of New Drugs Against Biofilm Formation and Inhibition, using Molecular Docking and Virtual Screening Techniques. She is currently a Research Fellow at BioSIM, and her work focuses on solving enzymatic catalytic mechanisms through QM/MM and Molecular Dynamics.

Henrique S. Fernandes is currently pursuing his Ph.D. under the supervision of Dr. Nuno M. F.S.A. Cerqueira and Dr. Sérgio F. Sousa at the BioSIM Research Group. The main topic of his research is understanding the catalytic mechanism of PLP-dependent enzymes using computational methods. Moreover, he has also been involved in the development of software addressed to (bio)chemical research, namely molUP and VMD Store.

Dr. Sergio F. Sousa received his PhD from the University of Porto in 2007, with a thesis focusing on computational enzymatic catalysis and biomolecular simulations. After a pos-doc in Biomembrane Simulations in 2008, he was appointed assistant researcher at REQUIMTE. His main areas of interest include docking, virtual screening, molecular dynamics, force field parameterization and QM/MM calculations. He currently leads the BioSIM research group at the Faculty of Medicine, University of Porto with research interests centred on the interface between chemistry and medicine, always with computers as the main partner.

2.2.1 Structural Refinement

QM/MM methods can be easily applied to refine specific interactions at particular locations in enzymes, starting from other computational methodologies or experimental structural data. The interactions of enzymes with their substrates or with possible or confirmed inhibitors are obvious examples. In fact, while X-ray crystallography and other experimental techniques can offer very interesting insight into the interaction of substrates (normally substrate analogs) with enzymes, QM/MM methods can be used to refine the specific hydrogen bonds involving the molecules at the active site and the surrounding amino acid residues.

In addition, QM/MM methods offer the possibility of quantifying the strength of the different interactions formed or of specific groups and to determine the change in electron density, spin, and other electronic properties on the substrate/inhibitor and interacting amino acid residues. In fact, in many enzymes, the sum of the number of atoms of the substrate or inhibitor bound at the active site, together with that of the atoms of the immediately surrounding amino acid residues, normally falls within an acceptable size for a QM region.

If these advantages are already evident in the refinement and analysis of experimental structures, the gain becomes even more evident when using QM/MM methods for refining interactions of substrates, ligands or inhibitors that result from the predictions of atomistic molecular dynamics, or protein-ligand docking. Hence, the combined use of these methodologies continues to increase.

2.2.2 Validating/Disproving Different Mechanistic Hypothesis

The most commonly used application for QM/MM methods in the study of enzymes is in computational enzymatic catalysis.[11,40] QM/MM methods are often applied to validate or disprove different mechanistic hypotheses regarding the catalytic pathway of a specific enzymatic reaction. In fact, for many enzymes, several different mechanistic proposals exist in the literature. These proposals arise from experimental data regarding the structure of the reactants or products, associated kinetics, data on specific mutants, known activity dependence on some cofactor, metal, pH condition, or other, or from analogy with other similar enzymes or reactions, etc.

In these cases, the chemical reactions associated with different mechanistic hypotheses are evaluated by QM/MM, with the atoms directly involved in the reaction being treated by QM. The corresponding potential energy surfaces are calculated, and the activation free energies are determined and compared with the experimental k_{cat} values. In addition, the intermediate and final structures are compared with the experimental structural and/or spectroscopic information available.

Normally, the QM/MM calculations provide relatively different values for the different mechanistic proposals, enabling a straightforward alternative to eliminate hypothesis, shown as computationally unrealistic or impossible. While this elimination of the "wrong" hypothesis can be done with relative safety, by demonstrating, for example, that a certain proposal previously made would result in an unrealistic activation free energy 10 kcal/mol higher than the experimental results, finding the "right" hypothesis is not so easy. First, different atomistic hypotheses can sometimes generate similar computed properties, or with values that are within 2–3 kcal/mol of each other, within what is considered the error of some of the most commonly used methods. Secondly, the very notion of "right hypothesis" is a false notion, as a hypothesis is only considered "right" until new experimental evidence appears that disproves it becomes available. Even so, QM/MM methods have been widely used to distinguish the "most likely" mechanistic hypothesis from other alternatives that were demonstrated as unrealistic or impossible. Recent examples involving our group include.[41–47]

In terms of computational enzymatic catalysis, and around the same general lines, QM/MM methods can also be used to reconcile and explain apparently contradicting experimental evidence, providing a distinct and independent alternative to analyze enzymatic reactions.[48]

2.2.3 Rational Protein Engineering

The study of the enzymatic mechanisms provided by QM/MM methods enables an atomic-level analysis of the main interactions formed in the reactants, transition state (TS), and products. In addition, the structure of the TS and of the distribution of charge at this important intermediate and its variation in relation to the reactants and products offers an opportunity for the rational selection of proposals for site-directed mutagenesis. A popular objective is that of lowering the activation free energy of the enzyme, thereby improving the rate of the resulting reaction, making the enzyme potentially more appealing for industrial applications as a biocatalyst.[49]

From the catalytic mechanism solved through the application of QM/MM methods, and by the analysis of the impact of the different amino acid residues around the active site in the wild-type enzyme, models are created, substituting key amino acid residues by others thought, for example, to stabilize better the transition state. The reaction is then simulated through QM/MM, and the resulting potential energy surface (PES) and activation free energies are compared with that of the wild-type enzyme. Several different possible mutations are normally evaluated by QM/MM, and the potential effects on the reaction-rate are estimated from the determination of the TS structures and associated energetics.[50] Only the most promising alternatives are then tested experimentally.

2.2.4 Role in Drug Discovery

Another potential application arising from the use of QM/MM methods in the study of enzymatic catalysis is the identification of new scaffolds for drug discovery efforts. It has long been thought that enzymes have evolved to stabilize the structures of the TS of the reactions that they catalyze.[51–61] Actually, when analyzing the structures of reactants, TS, and products, it is evident that active site residues stabilize much more the TSs.[62] Hence, it has been proposed that TS analogs could work as promising scaffolds for the development of novel enzyme inhibitors[62–65] with improved affinity and specificity for a certain enzyme known to behave as a therapeutic target.

One of the advantages of the application of QM/MM methods in computational enzymatic catalysis is the determination of the structures of the TS. Hence, some studies have tried to take advantage of these structures to generate pharmacophores to guide the search for new molecules that could be used as potential new inhibitors.[62,66]

With the increase of computational power and technological development, new and more effective applications of QM/MM methods are starting to emerge and to occupy a more important role in the study of enzymes.

3. Practical Considerations

This section reviews some practical considerations involving the application of QM/MM methods in the study of enzymes, stemming from our experience in the field. While most of the problems illustrated are general, a few are restricted or more relevant for the application of the ONIOM methodology. The molUP[67] plugin for QM/MM calculations is also presented.

3.1 Preparation of the Initial Structure

As with any other computational method for the study of biological systems, a critical stage of the application of QM/MM methods in the study of an enzymatic reaction mechanism is the preparation of the initial model.

While in theory models can be obtained from homology modeling or other techniques, the preferred alternative is an X-ray or NMR structure of the enzyme of relevance, ideally, with the substrate or with a close substrate analog. This is because the specific position of the sidechains around the active-site region is an important requirement for the accurate modeling of an enzymatic reaction. Logically, X-ray structures with a resolution better than 2 Å are preferred (ideally below 1.5 Å). Structures of mutants should be avoided whenever possible, especially those that contain mutations near the active site unless the aim of the study is studying the impact of certain mutations in the mechanism. However, when that is not possible, a careful structure selection should be made, ideally comparing multiple structures.

In the absence of ideal conditions, molecular modeling can be used to revert mutations, model substrates from substrate analogs or products. Other techniques, such as docking,[68,69] can be used to position specific substrates at the active site. Such efforts should, however, not be performed lightly, and the resulting models should be refined through small molecular dynamics simulations and initial QM/MM optimizations. Naturally, the maximum number of points of contact with the information arisen from the experimental studies should be sought.

Another important aspect concerns the exact protonation of the amino acid residues of the enzyme. Ionizable sidechains of amino acid residues should be considered with care. Software like PROPKA[70,71] or H++[72] are often used to analyze the most likely protonation states of aspartate, glutamate, lysine, arginine, and histidine residues in the specific environment of that particular enzyme, but other alternatives can also be employed.[73,74] Even so, visual inspection of the proposals is always advisable. In general, these alternatives provide reasonable proposals for the amino acid residues that are not located around the active site. For the latter, however, more care should be employed. For the amino acid residues directly participating in the reaction or very near the active site, in the absence of experimental evidence, it is normally advisable to repeat the study of the enzymatic reaction by QM/MM with the different protonation states and compare the results.

Taken together, modeling the substrate (or docking it), adding hydrogen atoms, and assigning protonation states introduces some potential strain into the model structure. Inherently, the magnitude of this effect depends on the level of modeling and changes introduced. Some authors prefer to perform a small energy minimization at the molecular mechanical level, prior to the QM/MM optimization, a common approach when small modeling is involved. Another alternative is that of performing a molecular dynamics simulation, with or without specific geometric constraints for 1 to 5 nanoseconds (sometimes less), depending on the extent of modelling. This can be done to remove "bad contacts" or unnatural strains in the structure arising from modeling, evaluate the effect of different protonation states in the structures, or in some cases, as a strategy to generate improved or different starting structures for the QM/MM reaction modelling process.

In some cases, large MD trajectories are generated and analyzed and the structure(s) that position the reactive atoms of the reaction to be evaluated in the closest or more favorable position is(are) selected as initial starting structures for reaction modeling. In other cases, this choice can be based on the average or most common structure. This approach can also be used to generate ensembles of starting structures for multi-PES QM/MM reaction modeling, starting from 20, 50, or even 100 structures. This results in an ability to sample the activation barriers.[40] Different conceptually possibilities exist also, such as selecting random structures taken at specific time intervals or clustering the structures of the trajectory and

selecting average or representative structures from each cluster for QM/MM reaction evaluation.

3.2 Choice of the QM/MM Boundary

Another important aspect is the choice of the boundary between the QM and the MM region. Ideally, the QM/MM boundary should be located as far away as possible from the reactive centre to obtain more accurate results. However, as the size of the QM region largely determines the computation cost of the calculation, choosing the smallest QM region possible that can yield reasonable results is a prerequisite for efficient QM/MM calculations. From the balance between these two main requirements, it results that the boundary should be chosen in a way that assures the most complete model that is still computationally feasible for the set of calculations required to solve the problem under study.

A few general guidelines are normally advisable to ensure representative results. The first logical requirement concerns bond formation and bond-breaking processes – all bonds being formed or broken have to be located in the QM region of the molecular system (Figure 1). This condition is also extensible to all atoms whose hybridization is altered during the chemical process under study. This requirement frequently leads to the inclusion of complete aromatic or conjugated systems in the QM region, with the resulting increase in computational expense. Conjugated and aromatic systems should not be divided between regions, but in case that is absolutely required, partition effects should be accounted for.

In addition, whenever possible, cases where second and third atoms effects are predicted to be relevant, those atoms should be included in the same region of the system. This involves, for example, positive/negative amino acid residues, the interaction between polar residues, and relevant hydrogen bonds. Relevant hydrophobic interactions, including π-π stacking between adjacent amino acid rings, should also be taken into consideration.

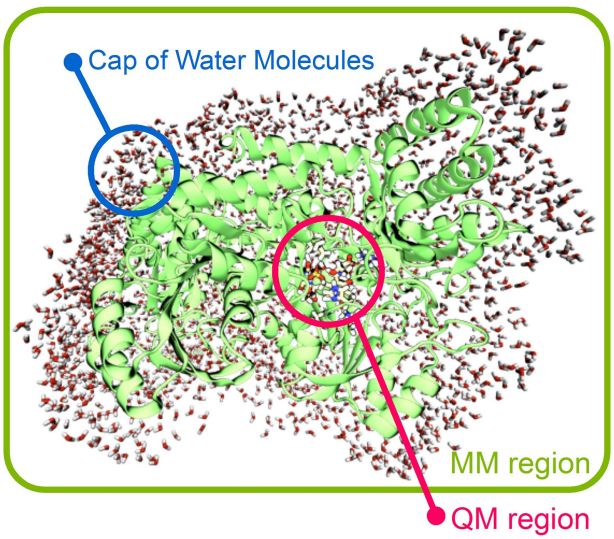

Figure 1. Representation of a 2-layers ONIOM QM/MM model of an enzyme. The QM and MM regions are identified, as well as, the cap of water molecules.

3.3 Choice of the QM Level

While many alternatives for treating the QM region are currently available, including semi-empirical, Hartree-Fock, and post-Hartree-Fock methods, most studies traditionally employ density functional theory (DFT) methods. B3LYP, despite its known limitations, has been historically the most widely used density functional to study enzymatic reactions.[75] Several more recent alternatives, as Minnesota[76–87] and Grimme's[88,89] density functionals are also frequently used.

Ideally, a rational choice of the density functional level should be employed. Some authors have taken particular care to evaluate the performance of different density functionals against values determined with a higher-level theoretical method or against experimental values of reference. These tests are normally conducted first on QM models bearing the same type of general chemistry of the reaction to be evaluated by QM/MM.

A common approximation in QM/MM enzymology involves performing the geometry optimizations and frequency calculations with one theoretical level (e.g. DFT/MM), and then using those structures to perform "single point" energy calculations with a more complete basis set (from a popular 6-31G(d,p) to 6-311++G(3df,3pd), for example) and a more accurate density functional or a higher QM level. Although, in some particular cases, this approach can lead to inaccurate results, it usually yields refined activation and reaction energies, as geometries are, in general, more insensitive to the computational method used than energies.

In recent years, the addition of dispersion correction to the density functional energies has gained popularity, especially with the D3[90] and D3-BJ[90,91] corrections. More recently, the developed domain-based local pair natural orbital-coupled cluster method with single, double, and perturbative triple excitations (DLPNO-CCSD(T))[92,93] has also become a popular choice. This method has been demonstrated to give results very close to those obtained with the canonical CCSD(T) method, with a cost close to that of a DFT functional,[94,95] and has been successfully applied in numerous chemistry/biochemistry studies.[42,96–104] In our studies, we apply "single point" energy calculations with DLPNO-CCSD(T) with the cc-pVDZ, and cc-pVTZ basis sets, and with the cc-pVDZ/C and the cc-pVTZ/C correlation fitting basis sets, keeping the default DLPNO cutoff parameters available in the ORCA software.[105] The DLPNO-CCSD(T) energies obtained with these basis sets are used to extrapolate to the complete basis set (CBS) limit following Truhlar's extrapolation method.[106]

3.4 Choice of the MM Level

In terms of the choice of the MM level, most efforts in QM/MM enzymology have traditionally relied on standard biomolecular force fields, like AMBER, CHARMM, or GROMACS. It has been generally assumed that the accuracy of a QM/MM results depends mostly on the choice of the QM level, at least if a good selection of a QM region has been made. This is quite true, at least up to a certain point. With the reactions taking place in the QM region, the MM region plays only a secondary role.

The magnitude of this secondary role depends mostly on the treatment of the QM/MM interactions, with different alternatives varying in terms of the embedding scheme used to handle the electrostatic interactions. Two alternatives have been traditionally used: (1) the mechanical embedding scheme and; (2) the electrostatic embedding scheme.

Mechanical embedding represents the simplest approach to incorporate the effect of the interactions between the QM and MM regions. This scheme neglects any electrostatic effect of the environment in the QM region, whose calculation is basically performed in the gas phase. The MM region plays only a structural role imposing limitations in the geometries adopted by the QM region. As for the MM region, a classical point charge model centred in the QM part is commonly used, which in practice means that the QM/MM interaction is calculated only at the MM level.

In contrast to the mechanical embedding approach, the electrostatic embedding approach already includes the polarization of the QM region by the MM charge distribution in the QM calculation. Electrostatic embedding is normally considered to be more accurate than mechanical embedding.[18] Some studies have shown that under some circumstances mechanical embedding can result in slow convergence with misleading energies, even for large QM sizes, whereas electrostatic embedding leads to faster and more reliable convergence of the QM region.[107,108] However, mechanical embedding remains a very useful alternative, particularly when treating larger QM regions, often providing results in excellent agreement with experimental data.[41] Electrostatic embedding methods still neglect the polarization of the MM system by the QM region – polarized embedding – an approach that requires the use of polarizable force fields in the treatment of the MM region.

3.5 Use of Link Atoms

The application of QM/MM methods in the study of enzymes frequently involves defining boundaries between the QM and MM region between atoms that are covalently bonded. Usually, the link atom method is employed. The QM regions are capped with "link" atoms in order to satisfy the free valence of the QM atoms that are covalently connected to atoms included in the MM part. In most cases, the boundary is made between the sidechains of the amino acid residues around the active-site and the backbone α-carbons. In these cases, the link atoms are typically hydrogen atoms. In other situations, other atoms and even functional groups can also be used to mimic the behavior of the MM atoms that are being replaced. The QM/MM boundary should be sufficiently far from the reactive center so that all the structural parameters involving the reactive center are contained in the QM region.

3.6 Inclusion of Solvent

The inclusion of the effect of the solvent is also a potentially important aspect of QM/MM enzymology. The active sites of enzymes can greatly differ in many properties. One of these properties is solvent exposure or solvent accessibility. Naturally, in more exposed active sites, this effect is expected to be more significant. Additionally, in many reactions or reaction hypotheses, specific water molecules are known or thought to play a role. Properly handling with solvent hence becomes important for the accuracy of the study. Depending on the specific problem, this can be done at different levels.

Concerning the reaction, when one or more water molecules are thought to participate directly in the reaction or to play a key important role in directly stabilizing a specific group or interaction, they should be included in the QM region and treated accordingly. When this role takes more indirect participation, it can be included in the MM region.

It is also important to consider that enzymes, as biomolecules, exist in an aqueous environment. Hence, normally, to properly simulate an enzymatic reaction, it is important to include also the effect of the surrounding aqueous environment. One way to handle this problem is through the use of a dielectric constant within a continuum model such as IEF-PCM[109–113] or C-PCM[114,115] in Gaussian software. These methods can be coupled with the QM/MM approach, providing an approximate non-atomistic "third" layer, enveloping the QM/MM model and accounting for the non-specific solvent effect. Another common alternative consists in adding to the enzyme a 5–10 Å cap of water molecules that are included in the MM region through the calculations (Figure 1). This second alternative is more commonly used. It is important to highlight that in the initial stages of model refinement, when MD simulations are employed, these are typically performed in rectangular water boxes subject to periodic boundary conditions. Hence, the caps of waters, used in the stages of QM/MM modelling of the reaction hypothesis are cut from the equilibrated water boxes.

3.7 Use of Constraints in the MM Region

The application of the QM/MM methods in the study of enzymatic catalysis assumes that the critically important part of the reaction under study takes place in the QM region and that the MM region plays a secondary role, partly structural,

partly environmental in the enzymatic reactions evaluated. Hence, the MM region influences the QM region.

Biologically, it is known that chemical changes in specific places of certain enzymes can result in changes (structural or other) that can sometimes be quite large in other regions of the enzymes through allosteric mechanisms and other processes. Therefore, it is not difficult to imagine cases in which the QM region could influence the MM region. However, in the enzymatic reaction hypotheses, typically studied through QM/MM, with the type of models, theoretical methods and approximations used, we are quite far from these scenarios, and our focus remains in the QM region.

While exploring a reaction coordinate through a linear transit scan, following an intrinsic reaction coordinate (IRC), optimizing a TS, or reoptimizing the reactants or products, it is not uncommon to witness a drastic conformational change in the MM region, often taking place in a single step. This change can normally be noticed by a disruption in the QM/MM energy profile that arises from the MM part, but not in the one from the QM region. Hence, it is normally not directly related to the reaction evaluated but results from technical issues with the optimization algorithms. Handling these artifacts is not always easy.

A common approach to handle this problem is that of constraining/freezing part of the MM region. Hence, the MM part is sometimes divided into an outer layer and an inner layer. Atoms in the outer layer are constrained to their initial positions, while atoms in the inner layer are free to optimize. Typically, the outer MM layer comprises just a small radius of the 3–5 Å outer-most amino acid residues or water molecules. In many cases, the frozen layer is made to include only part of the external cap of MM waters. This approach ensures a reasonable representation of the enzymatic reactions taking place in the QM region and prevents the technical artifacts arising from unnatural changes in the course of a QM/MM optimization.

3.8 MolUP – Making QM/MM Easier

To simplify the study of enzymatic reaction mechanisms through QM/MM calculations, we have been involved in the creation of molUP.[67] molUP is a VMD extension[116] developed at the University of Porto that provides a full-featured graphical user interface (GUI) to the computational chemistry software Gaussian, and was specially designed for preparing, and analyzing QM/MM calculations of enzymatic reaction mechanisms with ONIOM. molUP offers a series of options that enable users to easily create the QM and MM regions, define the boundaries, chose the QM and MM levels, assign charges, assign MM parameters for non-typical residues, assign link-atoms, handle the selection of solvent atoms and automatize the application of constrains by defining inner- and outer-MM regions. It also helps users to easily create inputs for frequency calculations, IRCs, TS optimizations, single point energy calculations with common choices, and creates inputs for use with the DLPNO/CCSD(T) approach in ORCA.[105] It also simplifies the process of analyzing and visualizing the output files using the advanced visualization facilities from VMD.

This extension includes a set of tools to set up any calculation supported by Gaussian; analyze energies through interactive plots; animate vibrational frequencies; draw the vectors associated with those frequencies; modify bonds, angles, and dihedrals; and collect bibliographic information on the employed methods. molUP is presently being maintained and developed at BioSIM (www.biosim.pt) and can be freely obtained through the VMD Store.[117]

4. Recent Examples

In this section, we highlight two recent applications of ONIOM in enzymatic catalysis using the molUP software, one employing a mechanical embedding scheme and one considering an electrostatic embedding approach.

4.1 Tryptophan Synthase

Tryptophan Synthase (TSase) is an emergent target in the treatment of tuberculosis (TB). This enzyme is present in *Mycobacterium tuberculosis*, where it plays an important role in the bacterial growth and replication, being involved in Trp biosynthesis. As this pathway is absent in mammals, TSase has become an attractive target for the development of new anti-TB antibiotics.

TSase is a multifunctional enzyme, with two dimers, each one with two independent active sites, interconnected by a 25 Å tunnel. We have recently studied the two steps of the catalytic mechanism of this enzyme with ONIOM QM/MM[41] employing a mechanical embedding scheme. Different models were considered, including one with 175 atoms in the QM region and close to 17,000 atoms in the MM region (ff99sb force field), which included a 2 Å cap of waters. The QM region was treated with B3LYP/6-31G(d) in geometry optimizations, while single-point energy calculations were performed with B3LYP/6-311++G(2df,pd).

The results allowed the determination of the catalytic mechanism of TSase with atomic detail, including the transition state structures and the free energy profile of the full catalytic process, in excellent agreement with the available experimental data.

4.2 Human Fatty Acid Synthase

Human Fatty Acid Synthase (hFAS) is a multidomain enzyme responsible for the biosynthesis of saturated fatty acids, a central process for most living organisms. During the biosynthetic reaction, hFAS incorporates acetyl and malonyl moieties through the action of its malony–acetyl transferase

(MAT) domain. In normal circumstances, the de novo synthesis of fatty acids (FA) is residual, as the FA demand is supplied by the normal diet. However, many studies have reported FAS gene overexpression in pathological conditions, including diabetes, obesity and cancer. Thus, significant efforts are being made to achieve a better understanding of the mFAS/hFAS structure, and of the catalytic machinery behind each individual domain.

We have recently been involved in the study of the MAT domain of hFAS, which is responsible for initiating the fatty acid synthesis pathway.[42] For that we employed ONIOM QM/MM with electrostatic embedding, with the QM region including close to 60 atoms, while the MM region included around 6300 atoms. B3LYP/6-31G(d):AMBER was employed to simulate the different reaction hypothesis (ff10 force field) through linear transit scans, minima and transition state optimization, and IRC and frequency calculations. B3LYP/6-311+G(2d,2p)-D3:AMBER level and DLPNO-CCSD(T)/CBS:AMBER were employed for the single point energy calculations, ensuring a more accurate determination of the energetics associated.

The results enabled the elucidation of the catalytic mechanism of MAT in hFAS, allowing also the clarification of the important role played by an oxyanion hole defined by the backbone amines of Met499 and Leu582 in stabilizing a critical intermediate in the reaction.

5. Future Challenges

During the last decades, there has been a continuous improvement in the application of QM/MM methods in modeling enzymatic reactions. Several QM/MM frameworks, like ONIOM, have already reached "majority", with two decades of widespread use. However, through the years, the increase in computational power and level of sophistication of the algorithms has enabled progressive improvements in the size of the systems and level of the computational methods that are combined. So, QM/MM methods continue to grow in strength, robustness, and ability to answer new questions. Here we discuss some of the challenges for the next years.

5.1 Bigger QM Regions

From a conceptual point of view, the application of QM/MM methods in the study of enzymatic reactions assumes that all the atoms involved in the enzymatic reaction are included in the QM region. This is normally done in most studies, at least for those atoms that are directly involved. However, many other atoms and amino acid residues around the active site also play an indirect but non-neglectable role in the enzymatic reactions. While part of these effects can be partially modeled by an MM representation (at least through an electrostatic embedding scheme, for example), others would greatly benefit from a QM representation. In fact, one of the dreams of every QM/MM modeler has always been to use very large QM regions. Historically, however, the choice of the QM region has been limited by the CPU power available, as the time required for a QM/MM calculation mostly depends on the time required for the QM calculation. This limitation also emerged from the fact that the parallelization of most QM codes is far from the level of that of MM implementations.

Many early QM/MM enzymology studies employed QM regions with up to 20 atoms treated with DFT. Presently, most studies already adopt QM regions with close to 200 atoms in DFT[41,118] and several examples with larger regions can be found.[118,119] The motivation to increase the size of the QM region continues well alive, particularly as more complex reactions involving larger conformational changes at the active site are studied.

Linear-scaling DFT methods[120–124] would be interesting alternatives to combine within QM/MM frameworks. Such methods enable the application of popular density functionals to QM models containing hundreds or thousands of atoms but are, in general, much adequate for periodic systems and for the application of plane wave schemes. Some implementations can be combined to obtained refined activation or reaction free energies through single point energy calculations with a much larger QM region. However, for proper QM/MM modeling, the determination of gradients, energy optimizations, and electronic polarization with this massive QM regions using linear-scaling DFT methods would be required.

5.2 Polarized Embedding and the Use of Polarizable Force Fields

Presently, one of the biggest limitations of most common QM/MM studies on enzymatic reactions is the difficulty in implementing a polarizable embedding scheme to account for the polarization on the MM region arising from the change in charge occurring along with a chemical reaction studied in the QM region. In fact, in many reactions, charge forms along the reaction path, with the TS structures, intermediates, or products displaying significant charge alternations in comparison with the reactants. This charge difference should naturally be mostly compensated by interactions formed with amino acid residues or other molecules that are included in the QM region. However, this condition can greatly depend on the magnitude of the effect, the size of the QM region and of course, the specific characteristics of the enzyme. In reality, the QM region reacts to the MM region, which in turn can be influenced by the chemical reactions taking place at the active site.

While this can, in principle, be tackled moving from an electrostatic embedding scheme to a polarized embedding scheme, in which both regions can mutually polarize each other, it is important to consider that such alternative implies a polarizable molecular mechanics force field. However, the vast majority of the QM/MM studies used today still apply the same type of biomolecular force fields that were in common

use 20 years ago. Improved variations in AMBER, CHARMM, GROMACS, and others have been made available, resulting in improved electrostatics and in several other structural improvements particularly noticeable in long MD simulations. However, all these alternatives continue to treat the MM part of the enzyme through pairwise additive potentials with fixed atom-centered charges.

Even when considering only the MM part, it is important to take into account that charged and polar groups in enzymes can polarize significantly when interacting with other charged, and polar groups, or with the solvent. Such polarization can affect the geometry of the MM region and its energetics. Commonly, force fields for biomolecular simulations treat polarization in an average way, implement in the parameterization, and that is fixed. However, charges in the amino acid residues depend on their surroundings, which vary among other features according to their position in the protein (more or less solvent exposure), conformation, interaction with other atoms or molecules. Hence, even when treating the MM region as a mere perturbation to the QM region, having a MM-charge dependent conformation could be very important, as this can influence differently the reactants, transition states and products, affecting the calculated activation, and reaction free energies.

Polarizable force fields[125,126] have been steadily improving over the last decades. Today, several excellent alternatives for the treatment of biomolecular systems are available. Notable examples include AMOEBA,[127–129] DRUDE,[130] CHARMM fluctuating charge,[131,132] SIBFA and GEM.[133]

Implementing polarizable force fields for use in QM/MM methods would enable a significant improvement in accounting for these effects. Some alternatives have been made available in recent years. Programs like LICHEM[134,135] (the layered interacting chemical models program) developed by Cisneros & co-workers, already enable the coupling of different unmodified quantum mechanics (QM) and molecular mechanics (MM) packages for QM/MM simulations involving polarizable force fields. Other alternatives have also been made available.[38,136,137] Programs like these are likely to play a very important role in helping to model enzymatic reactions for the next decade.

5.3 Improvements on the QM Level

During the past two decades, there have been important improvements in terms of density functionals.[75,138–146] Notable examples include the Minnesota density functionals and Grimme's density functionals and the introduction of dispersion corrections.[76–89] However, B3LYP continues to be a very popular choice. Today, the particular choice of density functional plays a secondary role when considering the CPU cost. Most studies employ several different alternatives and check for consistency in terms of the activation or reaction free energy barriers. The application of the DLPNO/CCSD(T) method in refining the single point energies has also become quite common.[42,44,46,96–101,103,104,147] The application of methods beyond DFT in QM/MM optimization of enzymatic systems is still relatively rare in comparison, but would be important particularly for treating enzymes, especially in cases in which the existence of excited states plays an important role.

6. Summary

QM/MM methods are among the most widely used computational methods for dealing with the reactivity of enzymes. With the first works published almost 50 years ago, the QM/MM field has grown and matured. In the last 20 years, it has blossomed with its implementation in popular computational chemistry software packages and is nowadays routinely used in the study of a large variety of chemical problems, and in particular, in the study of enzymatic reactivity. In this field, QM/MM methods offer a very appealing alternative to complement the information obtained from a large number of experimental techniques used, contributing the definition of an atomic level portrait of enzyme catalysis, incorporating electronic effects.

The computational development, the new possibilities brought by the use of GPUs and growth in parallelization efficiency, the availability of improved quantum mechanical methods, and of a new generation of polarizable biomolecular force fields, offer new possibilities to improve the QM/MM methods presently in use, enlarging the range of biological problems that can be tackled and our ability to answer some of the most fundamental questions regarding enzymatic reactivity.

Acknowledgements

This work was supported by the Applied Molecular Biosciences Unit – UCIBIO which is financed by national funds from FCT (UIDB/04378/2020). Henrique S. Fernandes acknowledges Fundação para a Ciência e a Tecnologia (FCT) for his PhD grant SFRH/BD/115396/2016.

References

[1] X. Zhang, K. N. Houk, *Acc. Chem. Res.* **2005**, *38*, 379–385.
[2] G. G. Hammes, S. J. Benkovic, S. Hammes-Schiffer, *Biochemistry* **2011**, *50*, 10422–10430.
[3] P. K. Agarwal, *Biochemistry* **2019**, *58*, 438–449.
[4] A. Guarneri, W. J. van Berkel, C. E. Paul, *Curr. Opin. Biotechnol.* **2019**, *60*, 63–71.
[5] J. Büchler, A. Papadopoulou, R. Buller, *Catalysts* **2019**, *9*, 1030.
[6] E. M. M. Abdelraheem, H. Busch, U. Hanefeld, F. Tonin, *React. Chem. Eng.* **2019**, *4*, 1878–1894.
[7] L. M. Schmitz, K. Rosenthal, S. Lütz, *Biotechnol. Bioeng.* **2019**, *116*, 3469–3475.
[8] M. J. Field, *J. Comput. Chem.* **2002**, *23*, 48–58.
[9] A. J. Mulholland, *Drug Discovery Today* **2005**, *10*, 1393–1402.

[10] M. W. van der Kamp, A. J. Mulholland, *Nat. Prod. Rep.* **2008**, *25*, 1001–1014.
[11] S. F. Sousa, P. A. Fernandes, M. J. Ramos, *Phys. Chem. Chem. Phys.* **2012**, *14*, DOI 10.1039/c2cp41180f.
[12] K. Świderek, I. Tuñón, V. Moliner, *WIREs Comput. Mol. Sci.* **2014**, *4*, 407–421.
[13] K. E. Ranaghan, A. J. Mulholland, *Int. Rev. Phys. Chem.* **2010**, *29*, 65–133.
[14] D. K. Remler, P. A. Madden, *Mol. Phys.* **1990**, *70*, 921–966.
[15] M. Dal Peraro, P. Ruggerone, S. Raugei, F. L. Gervasio, P. Carloni, *Curr. Opin. Struct. Biol.* **2007**, *17*, 149–156.
[16] K. Zinovjev, I. Tuñón, *WIREs Comput. Mol. Sci.* **2018**, *8*, e1329.
[17] H. M. Senn, W. Thiel, *Curr. Opin. Chem. Biol.* **2007**, *11*, 182–187.
[18] H. M. Senn, W. Thiel, *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 1198–1229.
[19] K. M. Merz, *Acc. Chem. Res.* **2014**, *47*, 2804–2811.
[20] A. Warshel, *Proc. Natl. Acad. Sci.* **1978**, *75*, 5250–5254.
[21] S. C. L. Kamerlin, A. Warshel, *Phys. Chem. Chem. Phys.* **2011**, *13*, 10401–10411.
[22] J. Åqvist, A. Warshel, *Chem. Rev.* **1993**, *93*, 2523–2544.
[23] R. Car, M. Parrinello, *Phys. Rev. Lett.* **1985**, *55*, 2471–2474.
[24] A. Laio, J. VandeVondele, U. Rothlisberger, *J. Chem. Phys.* **2002**, *116*, 6941–6947.
[25] P. E. M. Siegbahn, F. Himo, *WIREs Comput. Mol. Sci.* **2011**, *1*, 323–336.
[26] S. F. Sousa, P. A. Fernandes, M. J. Ramos, *Biophys. J.* **2005**, *88*, 483–494.
[27] S. F. Sousa, P. a Fernandes, M. J. Ramos, *J. Am. Chem. Soc.* **2007**, *129*, 1378–85.
[28] S. F. Sousa, P. A. Fernandes, M. J. Ramos, *Chemistry* **2009**, *15*, 4243–7.
[29] B. Honig, M. Karplus, *Nature* **1971**, *229*, 558–560.
[30] A. Warshel, M. Levitt, *J. Mol. Biol.* **1976**, *103*, 227–249.
[31] L. W. Chung, W. M. C. Sameera, R. Ramozzi, A. J. Page, M. Hatanaka, G. P. Petrova, T. V. Harris, X. Li, Z. Ke, F. Liu, H. B. Li, L. Ding, K. Morokuma, *Chem. Rev.* **2015**, *115*, 5678–5796.
[32] R. Lonsdale, J. N. Harvey, A. J. Mulholland, *Chem. Soc. Rev.* **2012**, *41*, 3025–3038.
[33] F. H. Wallrapp, V. Guallar, *WIREs Comput. Mol. Sci.* **2011**, *1*, 315–322.
[34] H. Lin, D. G. Truhlar, *Theor. Chem. Acc.* **2006**, *117*, 185.
[35] M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber, K. Morokuma, *J. Phys. Chem.* **1996**, *100*, 19357–19363.
[36] F. Maseras, K. Morokuma, *J. Comput. Chem.* **1995**, *16*, 1170–1179.
[37] S. Humbel, S. Sieber, K. Morokuma, *J. Chem. Phys.* **1996**, *105*, 1959–1967.
[38] J. R.-G. Pedregal, I. Funes-Ardoiz, G. Sciortino, J.-E. Sánchez-Aparicio, G. Ujaque, A. Lledós, J.-D. Maréchal, F. Maseras, *J. Comput. Chem.* **2019**, *40*, 381–386.
[39] L. W. Chung, H. Hirao, X. Li, K. Morokuma, *WIREs Comput. Mol. Sci.* **2012**, *2*, 327–350.
[40] S. F. Sousa, A. J. M. Ribeiro, R. P. P. Neves, N. F. Brás, N. M. F. S. A. Cerqueira, P. A. Fernandes, M. J. Ramos, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2017**, *7*, DOI 10.1002/wcms.1281.
[41] C. S. S. Teixeira, M. J. Ramos, S. F. Sousa, N. M. F. S. A. Cerqueira, *ChemCatChem* **2020**, *12*, 227–237.
[42] P. Paiva, S. F. Sousa, P. A. Fernandes, M. João Ramos, *ChemCatChem* **2019**, *11*, 3853–3864.
[43] A. C. C. Barbosa, R. P. P. Neves, S. F. Sousa, M. J. Ramos, P. A. Fernandes, *ACS Catal.* **2018**, *8*, DOI 10.1021/acscatal.8b01877.
[44] P. Paiva, S. F. Sousa, M. J. Ramos, P. A. Fernandes, *ACS Catal.* **2018**, *8*, 4860–4872.
[45] S. F. Sousa, J. F. M. Sousa, A. C. C. Barbosa, C. E. Ferreira, R. P. P. Neves, A. J. M. Ribeiro, P. A. Fernandes, M. J. Ramos, *J. Phys. Chem. A* **2016**, *120*, 5300–5306.
[46] H. S. Fernandes, M. J. Ramos, N. M. F. S. A. Cerqueira, *ACS Catal.* **2018**, *8*, 10096–10110.
[47] H. S. Fernandes, M. J. Ramos, N. M. F. S. A. Cerqueira, *Chem. Eur. J.* **2017**, *23*, 9162–9173.
[48] S. F. Sousa, N. M. F. S. A. Cerqueira, N. F. Brás, P. A. Fernandes, M. J. Ramos, *Int. J. Quantum Chem.* **2014**, *114*, DOI 10.1002/qua.24689.
[49] J. F. Rocha, A. F. Pina, S. F. Sousa, N. M. F. S. A. Cerqueira, *Catal. Sci. Technol.* **2019**, *9*, 4864–4876.
[50] P. Ferreira, S. F. Sousa, P. A. Fernandes, M. J. Ramos, *Chem. Eur. J.* **2017**, *23*, 17231–17241.
[51] L. Pauling, *Nature* **1948**, *161*, 707–709.
[52] S. F. Sousa, M. J. Ramos, C. Lim, P. A. Fernandes, *ACS Catal.* **2015**, *5*, 5877–5887.
[53] S. Marti, M. Roca, J. Andres, V. Moliner, E. Silla, I. Tunon, J. Bertran, *Chem. Soc. Rev.* **2004**, *33*, 98–107.
[54] T. C. Bruice, S. J. Benkovic, *Biochemistry* **2000**, *39*, 6267–6274.
[55] A. Warshel, P. K. Sharma, M. Kato, Y. Xiang, H. Liu, M. H. M. Olsson, *Chem. Rev.* **2006**, *106*, 3210–3235.
[56] M. H. M. Olsson, W. W. Parson, A. Warshel, *Chem. Rev.* **2006**, *106*, 1737–1756.
[57] J. Villà, A. Warshel, *J. Phys. Chem. B* **2001**, *105*, 7887–7907.
[58] M. Garcia-Viloca, J. Gao, M. Karplus, D. G. Truhlar, *Science* **2004**, *303*, 186–195.
[59] S. J. Benkovic, S. Hammes-Schiffer, *Science* **2003**, *301*, 1196–1202.
[60] A. Kohen, J. P. Klinman, *Acc. Chem. Res.* **1998**, *31*, 397–404.
[61] Z. D. Nagel, J. P. Klinman, *Nat. Chem. Biol.* **2009**, *5*, 543–550.
[62] V. L. Schramm, *Chem. Rev.* **2018**, *118*, 11194–11258.
[63] X. Wang, G. M. Bakanina Kissanga, E. Li, Q. Li, J. Yao, *Phys. Chem. Chem. Phys.* **2019**, *21*, 12163–12172.
[64] R. K. Harijan, O. Hoff, R. G. Ducati, R. S. Firestone, B. M. Hirsch, G. B. Evans, V. L. Schramm, P. C. Tyler, *J. Med. Chem.* **2019**, *62*, 3286–3296.
[65] S. Chen, K. Kapilashrami, C. Senevirathne, Z. Wang, J. Wang, J. A. Linscott, M. Luo, *J. Am. Chem. Soc.* **2019**, *141*, 8064–8067.
[66] I. Basu, G. Cordovano, I. Das, T. J. Belbin, C. Guha, V. L. Schramm, *J. Biol. Chem.* **2007**, *282*, 21477–21486.
[67] H. S. Fernandes, M. J. Ramos, N. M. F. S. A. Cerqueira, *J. Comput. Chem.* **2018**, *39*, 1344–1353.
[68] S. F. Sousa, A. J. M. Ribeiro, J. T. S. Coimbra, R. P. P. Neves, S. A. Martins, N. S. H. N. Moorthy, P. A. Fernandes, M. J. Ramos, *Curr. Med. Chem.* **2013**, *20*, 2296–2314.
[69] S. F. Sousa, P. A. Fernandes, M. J. Ramos, *Proteins Struct. Funct. Bioinf.* **2006**, *65*, 15–26.
[70] M. Rostkowski, M. H. M. Olsson, C. R. Søndergaard, J. H. Jensen, *BMC Struct. Biol.* **2011**, *11*, 6.
[71] C. R. Søndergaard, M. H. M. Olsson, M. Rostkowski, J. H. Jensen, *J. Chem. Theory Comput.* **2011**, *7*, 2284–2295.
[72] R. Anandakrishnan, B. Aguilar, A. V. Onufriev, *Nucleic Acids Res.* **2012**, *40*, W537–W541.
[73] S. Pahari, L. Sun, E. Alexov, *Database* **2019**, *2019*, DOI 10.1093/database/baz024.
[74] S. Witham, K. Talley, L. Wang, Z. Zhang, S. Sarkar, D. Gao, W. Yang, E. Alexov, *Proteins Struct. Funct. Bioinf.* **2011**, *79*, 3389–3399.

[75] S. F. Sousa, P. A. Fernandes, M. J. Ramos, *J. Phys. Chem. A* **2007**, *111*, 10439–10452.
[76] Y. Zhao, D. G. Truhlar, *J. Phys. Chem. A* **2005**, *109*, 5656–5667.
[77] Y. Zhao, D. G. Truhlar, *Theor. Chem. Acc.* **2008**, *120*, 215–241.
[78] T. Schwabe, S. Grimme, *J. Phys. Chem. Lett.* **2010**, *1*, 1201–1204.
[79] L. Goerigk, S. Grimme, *J. Chem. Theory Comput.* **2011**, *7*, 291–309.
[80] Y. Zhao, D. G. Truhlar, *Acc. Chem. Res.* **2008**, *41*, 157–167.
[81] J. Zheng, Y. Zhao, D. G. Truhlar, *J. Chem. Theory Comput.* **2009**, *5*, 808–821.
[82] R. Peverati, D. G. Truhlar, *J. Phys. Chem. Lett.* **2011**, *2*, 2810–2817.
[83] R. Peverati, D. G. Truhlar, *Phys. Chem. Chem. Phys.* **2012**, *14*, 13171–13174.
[84] H. S. Yu, X. He, S. L. Li, D. G. Truhlar, *Chem. Sci.* **2016**, *7*, 5032–5051.
[85] P. Verma, Y. Wang, S. Ghosh, X. He, D. G. Truhlar, *J. Phys. Chem. A* **2019**, *123*, 2966–2990.
[86] P. Verma, B. G. Janesko, Y. Wang, X. He, G. Scalmani, M. J. Frisch, D. G. Truhlar, *J. Chem. Theory Comput.* **2019**, *15*, 4804–4815.
[87] T. Schwabe, S. Grimme, *Phys. Chem. Chem. Phys.* **2007**, *9*, 3397–3406.
[88] S. Grimme, *J. Chem. Phys.* **2006**, *124*, 34108.
[89] F. Neese, T. Schwabe, S. Grimme, *J. Chem. Phys.* **2007**, *126*, 124115.
[90] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.
[91] S. Grimme, S. Ehrlich, L. Goerigk, *J. Comput. Chem.* **2011**, *32*, 1456–1465.
[92] C. Riplinger, F. Neese, *J. Chem. Phys.* **2013**, *138*, 34106.
[93] C. Riplinger, B. Sandhoefer, A. Hansen, F. Neese, *J. Chem. Phys.* **2013**, *139*, 134101.
[94] D. G. Liakos, F. Neese, *J. Chem. Theory Comput.* **2015**, *11*, 4054–4063.
[95] D. G. Liakos, M. Sparta, M. K. Kesharwani, J. M. L. Martin, F. Neese, *J. Chem. Theory Comput.* **2015**, *11*, 1525–1539.
[96] Y. Guo, C. Riplinger, D. G. Liakos, U. Becker, M. Saitow, F. Neese, *J. Chem. Phys.* **2020**, *152*, 24116.
[97] D. G. Liakos, Y. Guo, F. Neese, *J. Phys. Chem. A* **2020**, *124*, 90–100.
[98] M. F. Obst, A. Gevorgyan, A. Bayer, K. H. Hopmann, *Organometallics* **2020**, DOI 10.1021/acs.organomet.9b00710.
[99] B. Maity, Y. Minenkov, L. Cavallo, *J. Chem. Phys.* **2019**, *151*, 14301.
[100] D. C. Mielczarek, C. Nait Saidi, P. Paricaud, L. Catoire, *J. Comput. Chem.* **2019**, *40*, 768–793.
[101] Y. Minenkov, E. Chermak, L. Cavallo, *J. Chem. Theory Comput.* **2015**, *11*, 4664–4676.
[102] J. Chen, A. Draksharapu, D. Angelone, D. Unjaroen, S. K. Padamati, R. Hage, M. Swart, C. Duboc, W. R. Browne, *ACS Catal.* **2018**, *8*, 9665–9674.
[103] F. E. Medina, M. J. Ramos, P. A. Fernandes, *ACS Catal.* **2019**, *9*, 11404–11412.
[104] P. Janoš, T. Trnka, S. Kozmon, I. Tvaroška, J. Koča, *J. Chem. Theory Comput.* **2016**, *12*, 6062–6076.
[105] F. Neese, *WIREs Comput. Mol. Sci.* **2018**, *8*, e1327.
[106] D. G. Truhlar, *Chem. Phys. Lett.* **1998**, *294*, 45–48.
[107] S. Roßbach, C. Ochsenfeld, *J. Chem. Theory Comput.* **2017**, *13*, 1102–1107.
[108] J. Ho, Y. Shao, J. Kato, *Molecules* **2018**, *23*, DOI 10.3390/molecules23102466.
[109] S. Miertuš, E. Scrocco, J. Tomasi, *Chem. Phys.* **1981**, *55*, 117–129.
[110] S. Miertus˘, J. Tomasi, *Chem. Phys.* **1982**, *65*, 239–245.
[111] J. L. Pascual-ahuir, E. Silla, I. Tuñon, *J. Comput. Chem.* **1994**, *15*, 1127–1138.
[112] G. Scalmani, M. J. Frisch, *J. Chem. Phys.* **2010**, *132*, 114110.
[113] M. Cossi, N. Rega, G. Scalmani, V. Barone, *J. Comput. Chem.* **2003**, *24*, 669–681.
[114] A. Klamt, G. Schüürmann, *J. Chem. Soc. Perkin Trans. 2* **1993**, 799–805.
[115] V. Barone, M. Cossi, *J. Phys. Chem. A* **1998**, *102*, 1995–2001.
[116] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graphics* **1996**, *14*, 33–38.
[117] H. S. Fernandes, S. F. Sousa, N. M. F. S. A. Cerqueira, *J. Chem. Inf. Model.* **2019**, *59*, 4519–4523.
[118] H. J. Kulik, J. Zhang, J. P. Klinman, T. J. Martínez, *J. Phys. Chem. B* **2016**, *120*, 11381–11394.
[119] R. Mera-Adasme, M. Dominguez, O. Denis-Alpizar, *J. Mol. Model.* **2019**, *25*, 176.
[120] S. Mohr, L. E. Ratcliff, L. Genovese, D. Caliste, P. Boulanger, S. Goedecker, T. Deutsch, *Phys. Chem. Chem. Phys.* **2015**, *17*, 31360–31370.
[121] J. VandeVondele, U. Borštnik, J. Hutter, *J. Chem. Theory Comput.* **2012**, *8*, 3565–3573.
[122] J. Dziedzic, T. Head-Gordon, M. Head-Gordon, C.-K. Skylaris, *J. Chem. Phys.* **2019**, *150*, 74103.
[123] N. D. M. Hine, M. Robinson, P. D. Haynes, C.-K. Skylaris, M. C. Payne, A. A. Mostofi, *Phys. Rev. B* **2011**, *83*, 195102.
[124] C.-K. Skylaris, P. D. Haynes, A. A. Mostofi, M. C. Payne, *J. Chem. Phys.* **2005**, *122*, 84119.
[125] T. A. Halgren, W. Damm, *Curr. Opin. Struct. Biol.* **2001**, *11*, 236–242.
[126] A. Warshel, M. Kato, A. V. Pisliakov, *J. Chem. Theory Comput.* **2007**, *3*, 2034–2045.
[127] P. Ren, J. W. Ponder, *J. Phys. Chem. B* **2003**, *107*, 5933–5947.
[128] Y. Shi, Z. Xia, J. Zhang, R. Best, C. Wu, J. W. Ponder, P. Ren, *J. Chem. Theory Comput.* **2013**, *9*, 4046–4063.
[129] C. Zhang, C. Lu, Z. Jing, C. Wu, J.-P. Piquemal, J. W. Ponder, P. Ren, *J. Chem. Theory Comput.* **2018**, *14*, 2084–2108.
[130] J. A. Lemkul, J. Huang, B. Roux, A. D. MacKerell, *Chem. Rev.* **2016**, *116*, 4983–5013.
[131] J. E. Davis, S. Patel, *J. Phys. Chem. B* **2009**, *113*, 9183–9196.
[132] S. Patel, C. L. Brooks III, *J. Comput. Chem.* **2004**, *25*, 1–16.
[133] N. Gresh, G. A. Cisneros, T. A. Darden, J.-P. Piquemal, *J. Chem. Theory Comput.* **2007**, *3*, 1960–1986.
[134] E. G. Kratz, A. R. Walker, L. Lagardère, F. Lipparini, J.-P. Piquemal, G. Andrés Cisneros, *J. Comput. Chem.* **2016**, *37*, 1019–1029.
[135] H. Gökcan, E. A. Vázquez-Montelongo, G. A. Cisneros, *J. Chem. Theory Comput.* **2019**, *15*, 3056–3065.
[136] D. Loco, L. Lagardère, G. A. Cisneros, G. Scalmani, M. Frisch, F. Lipparini, B. Mennucci, J.-P. Piquemal, *Chem. Sci.* **2019**, *10*, 7200–7211.
[137] W. M. C. Sameera, F. Maseras, *J. Chem. Inf. Model.* **2018**, *58*, 1828–1835.
[138] N. Mardirossian, M. Head-Gordon, *Mol. Phys.* **2017**, *115*, 2315–2372.
[139] L. Goerigk, S. Grimme, *WIREs Comput. Mol. Sci.* **2014**, *4*, 576–600.
[140] S. F. Sousa, E. S. Carvalho, D. M. Ferreira, I. S. Tavares, P. A. Fernandes, M. J. Ramos, J. A. N. F. Gomes, *J. Comput. Chem.* **2009**, *30*, 2752–2763.
[141] J. M. L. Martin, G. Santra, *Isr. J. Chem.* **2019**, n/a, DOI 10.1002/ijch.201900114.

[142] A. Pribram-Jones, D. A. Gross, K. Burke, *Annu. Rev. Phys. Chem.* **2015**, *66*, 283–304.
[143] D. A. Wappett, L. Goerigk, *J. Phys. Chem. A* **2019**, *123*, 7057–7074.
[144] F. Yu, *J. Phys. Chem. A* **2014**, *118*, 3175–3182.
[145] M. Alipour, *J. Phys. Chem. A* **2016**, *120*, 3726–3730.
[146] G. Santra, N. Sylvetsky, J. M. L. Martin, *J. Phys. Chem. A* **2019**, *123*, 5129–5143.
[147] G. Bistoni, I. Polyak, M. Sparta, W. Thiel, F. Neese, *J. Chem. Theory Comput.* **2018**, *14*, 3524–3531.

Manuscript received: February 14, 2020
Revised manuscript received: March 9, 2020
Version of record online: April 30, 2020