EurJIC
European Journal of Inorganic Chemistry

Essays
doi.org/10.1002/ejic.202100330

Chemistry Europe
European Chemical Societies Publishing

Computational Organometallic Catalysis: Where We Are, Where We Are Going

Agustí Lledós*[a]

Dedicated to Professor Joan Bertran, who opened the doors to the quantum world for me, on the occasion of his 90th birthday

This essay gives my personal perspective of the current stage of computational methods applied to modeling organometallic catalysis, as well as the new directions the field is taking. The first part of the essay deals with what I consider the state-of-the-art to build up energy profiles, regarding both chemical and computational models. With a proper choice of the chemical model and computational methods, quantum mechanical calculations are nowadays able to provide accurate energy profiles of organometallic reactions in solution involving closed-shell species. However, in most cases they are still used to "predict the past", providing after-the-fact explanations and missing out the full potential of contemporary simulation techniques. Simulations are mature enough to be incorporated at the design stage and to guide the experimental exploration. The new directions the field is taking, incorporating automated exploration methods and combined with extensive data analysis and machine learning algorithms, approach the holy grail of catalyst discovering.

1. Introduction

Over almost forty years of theoretical research on organometallic reactivity, I have witnessed the transformation of quantum mechanical (QM) calculations from something uncommon, exclusively on specialist's hands and suitable just for qualitative analysis, to becoming a technique fully integrated in the chemist's toolbox and able to give quantitative measurements. In the last years, density functional (DFT) calculations have become the most common tool to study reaction mechanisms. A basic search in the Web of Science combining the words "DFT" and "European Journal of Inorganic Chemistry" evidences that, in the last years, more than a ten percent of the articles published in this journal include such calculations. A few recent reviews witness the importance that quantum chemical reaction modeling has gained in the field of homogeneous organometallic catalysis.[1–4] As an "old specialist" I am pleased to cope with the "democratization" of these techniques, nowadays within the reach of every chemist. But this fact also raises some philosophical questions, such as what the role of theoreticians is today or even if we are no longer needed. In this essay I will give my personal view about the current situation of theoretical calculations applied to the reactivity of molecular transition metal systems, particularly related to homogeneous catalysis, as well as the new developments I foresee for the near future.

The tremendous increase in computing power, together with the development of user-friendly software and quantum chemistry methods, specially DFT, makes it possible to accurately compute energy profiles of complex molecular systems. However, there is not always a clear understanding of the sources of error and limitations inherent to these techniques, nor about their quality control. Nowadays, calculations are usually employed in reactivity studies to measure a physical quantity (the potential energy of an ensemble of nuclei and electrons at fixed nuclei positions). While for experimental measurements such as elemental analysis, NMR chemical shifts and X-ray diffraction analysis there is a general consensus about the minimum standards required, the same does not apply to calculations. Even more demanding is the connection of these measurements with the proposal of a reaction mechanism. Calculations must be performed using state-of-the-art computational approaches. Of course, the standard to be chosen will depend on the nature of the system to be computed and evolves with time, very quickly in the case of computational techniques. The state-of-the-art can be defined as the best possible calculation at any given time. The first part of this essay will be devoted to what I consider the current state-of-the-art methods. Using them, reliable energy profiles can be constructed, but these do not answer all the mechanistic questions that chemists ask themselves: what is occurring in my flask and why, what will happen if…and even more, how can I improve one reaction or discover new reactions? These questions will be the subject of the second part of this essay. The scope and challenges,[5–10] as well as pitfalls[11] of computational methods for mechanistic studies in homogeneous catalysis have also been addressed in some recent articles collecting a number of examples, pointing out the current interest on the topic.

---

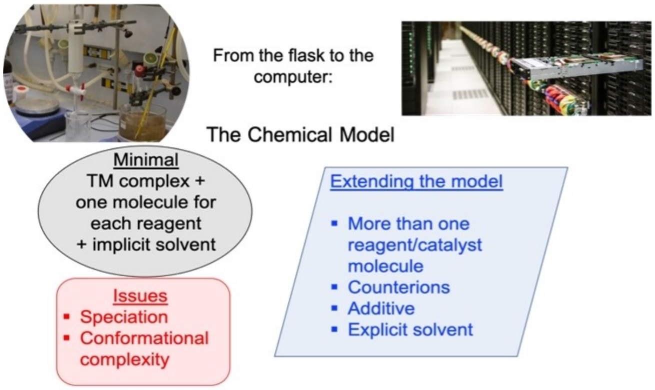
[a] *Prof. Dr. A. Lledós*
Departament de Química
Universitat Autònoma de Barcelona
Campus UAB, 08193 Cerdanyola del Vallès, Catalonia, Spain
E-mail: agusti@klingon.uab.es
https://compbionanocat.wordpress.com/

Part of the "RSEQ-GEQO Prize Winners" Special Collection.

2. Building Up Energy Profiles: The Choice of the Model

Until recently, computational tools were mainly used to explain chemical reactions after experimental results were obtained. The usual goal was to substantiate how the main product was formed and to understand the factors controlling the reaction. In order to do that, a sequence of intermediates and transition states (TS) along the pathway from reactants to products should be located and characterized. The connection of TSs with back and forward intermediates must be carefully checked. Ideally, this is done by two intrinsic reaction (IRC) calculations following both reaction path directions. In flat energy surfaces IRC often fails. In that case, displacement of the TS structure in both directions following the eigenvector corresponding to the imaginary frequency, and subsequent optimization of the resulting structures is recommended. QM calculations, combined with optimization techniques, provide absolute energies and geometries of chemically relevant structures in the potential energy surface of the system (PES). The relative energies of all these species, usually represented as an energy profile, describe the energy landscape for the reaction. Analysis of their geometries discloses the bond-forming and bond-breaking events taking place along the reaction pathway and allows the proposal of a mechanism for such reaction. In this approach the exploration of the PES to locate stationary structures (minima and TSs) is biased by the previous knowledge and the chemical intuition of the person performing the calculations.

One of the things that makes computational methods different from experimental techniques is that they do not work with samples, but with models. Two types of models operate in a QM calculation: on one side, the chemical complexity found in a reaction flask (millions of molecules of different types) should be reduced and adapted to something that a computer can handle. This is known as the **chemical model**. As calculations provide an atomistic description of the reaction, the most common way to do this is to include, in QM calculations, one molecule of the organometallic reagent and one molecule for each reactant. In many cases, this minimal model has to be extended in order to obtain reliable computational results (Scheme 1). On the other side, among the panoply of the available computational methods (for DFT calculations hundreds of different functionals and basis sets) one must be chosen to carry out the calculations. This is the **computational model** (formerly called the level of calculation). It should never be overlooked that all QM methods applied to molecular systems rely on approximations and it is safe to have in mind some estimation about the error that the chosen approximation can introduce in the results.



Scheme 1. Minimal and extended chemical models to perform Quantum-Mechanical based calculations on organometallic reactions.

The choice of both models is the first and most critical decision to be taken when starting a computational project, and it will have a major impact on the quality of results. Traditionally, particularly from the theoretical side, much more attention has been devoted to the computational model than to the chemical one, but a very accurate calculation in an incomplete chemical model can yield completely wrong results regarding reaction mechanisms.

3. The Chemical Model

This article focuses on organometallic chemistry in solution. That is, reactions performed in a solvent in a range of temperatures typically between 0–100 degrees Celsius. We want to simulate such a chemical reaction in a computer (i.e, *in silico*). Beforehand, it is necessary to have a clear perception of the experimental conditions under which the reaction takes place. As commented above, the most common chemical model includes one molecule of the organometallic reagent, nowadays with the actual ligands without further simplification, and one molecule for each reagent. Even when this model is reliable, the conformational flexibility of most of the bulky ligands employed in catalysis causes additional complications and can be a source of error.[12] An example of the importance of a wide exploration of the conformational space when computing energy profiles can be found in the recent computational analysis of the Buchwald-Hartwig amination of aryl halides using Pd-Josiphos complex.[13] The Gibbs energy difference (DFT calculations with M06 functional) between two conformations of the same intermediate (15.3 kcal·mol^{-1}) is much larger than the barrier of the step in which this intermediate is involved (12.3 kcal·mol^{-1}) (Figure 1). Recently

Agustí Lledós was born in Barcelona in 1955. He obtained a Ph.D. in Chemistry from the Universitat Autònoma de Barcelona (UAB) in 1984. After a postdoctoral stay at the Université de Paris-Sud, in 1987 he started computational research on organometallic catalysis at UAB. He has been Full Professor of Physical Chemistry at UAB since 1994. In 2018 he was awarded with the Gold Medal of the Organometallic Chemistry Group of the Spanish Royal Society of Chemistry (RSEQ).

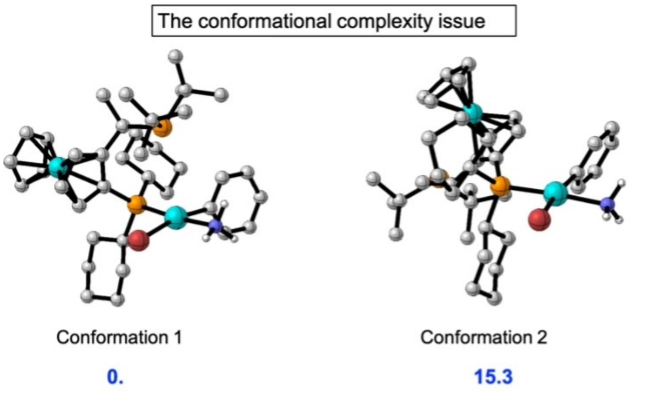


Figure 1. DFT-optimized geometries and relative Gibbs energies (kcal·mol^{-1}, computed with M06 functional) in 1,4-dioxane solvent of two conformations of the Pd(Josiphos)(NH_3)(Ph)Br intermediate in the Buchwald – Hartwig amination of arylhalides with one of the arms of the chelate phosphine open.[13] 3D-structures were generated using CYLview.[15]

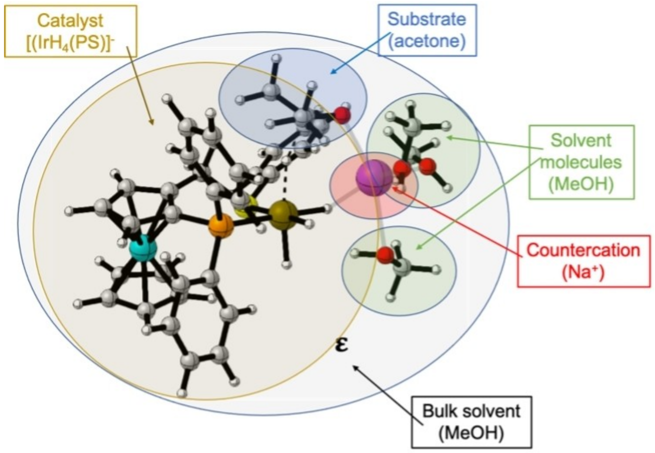


Figure 2. Extended chemical model for ketone hydrogenation catalyzed by a iridium-hydride complex, in alcohol solvent and with added base (NaOMe).[19] 3D-structures were generated using CYLview.[15]

developed methods combining low-level, semiempirical, DFT calculations with metadynamics simulations, enable fast QM exploration of the conformational space of large systems.[14]

Using this minimal model and without adding further molecules in the computational description, the speciation of reactive species in the reaction vessel arises as an important issue. The active form of the catalysts is often different from the species initially added, and in most cases it is not known. In such cases, several possibilities should be carefully scrutinized. Attention should also be paid to the speciation of simple reagents, which can be very solvent dependent. Reagents such as sodium *tert*-butoxide and copper diacetate keep aggregated structures in non-polar solvent environments, and their description as monomeric species in QM calculations could be misleading.[16,17] The same issue arises with iodosylbenzene (PhIO) a polymeric solid of low solubility.[18]

To build up a reliable energy profile, all the species that can influence the energy of the system in each step of a catalytic cycle should be taken into account. The correct mechanism cannot be represented using an incomplete model. Thus, when no convincing mechanism arises from the PES exploration with the minimal model, my first recommendation would be to extend the chemical model. The role in the reaction mechanism of these additional molecules that expand the minimal model should be scrutinized. Figure 2 displays an example of an extended chemical model for a ketone hydrogenation catalyzed by an iridium-hydride complex, in alcohol solvent and in the presence of NaOMe additive.[19]

A first extension of the initial model can be done by including additional molecules of one reagent, usually in excess in the experimental conditions. For instance, it is now well established in hydrofunctionalization reactions that a second molecule of the nucleophile can help in the proton migration step.[20] When ionic species are present there are always counterions in the reaction medium that might influence the reaction.[21] In some cases, there are experimental evidences of such effects, particularly regarding counteranions. Gold catalysis has been very prolific in showing the impact of counteranions in both kinetics and selectivity of transformations.[22] Counterions were generally considered as innocent partners and not included in the chemical model, but an increasing number of computational papers have started taking into account counteranions, particularly when dealing with proton transfer steps. Calculations have largely contributed to the understanding of the role of anions in gold-catalyzed reactions.[23] However, the inclusion of countercations in models is still a less common practice.[19] It must be highlighted that explicit treatment of counterions poses an added challenge for calculations: its solvation.

Additives are commonly used in homogeneous catalysis to improve the outcome of synthetically relevant reactions. However, in most cases their role is not well understood, and they are not commonly included in calculations. The synergistic use of calculations and experiments has allowed a deeper understanding of additive effects in a number of organometallic reactions.[24]

The reaction to be simulated in a computer proceeds in a solvent. Therefore, the solvent must be a consubstantial part of the chemical model and it needs to be taken into account in order to achieve a good computational description of the reaction. There are three main approaches for introducing the solvent in a QM calculation: i) implicit solvent model, in which the solvent molecules are replaced by a continuum polarizable medium characterized by its dielectric constant ε; ii) hybrid cluster-continuum model, in which a limited number of solvent molecules are introduced in the QM description of the system, and the rest is represented as a continuum, and iii) explicit solvent model, which includes a huge number of solvent molecules, usually suitable to reproduce the density of the chemical system. The most used and simplest solvent description is the continuum model. In most cases, particularly when dealing with low-polar, non-coordinating solvents, this model is enough to compute accurate energies in solution. However, it is striking, to my mind, that the most common way of using continuum models in recent literature is still performing single-point calculations at gas-phase optimized geometries. Today, most of the time there are no technical reasons that can justify this approach. Only in the cases in which optimization in solution of relevant species fails to properly converge despite repeated attempts, mainly due to numerical artifacts related with grids, gas-phase optimization, followed by single-point energy calculation in solution, is justified. Assuming the same geometries in the gas-phase and in the solvent can be specially misleading when there are charged species in the medium. Ionic mechanisms, implying steps in which ions are formed from neutral molecules, cannot be properly accounted for using this model. State-of-the-art calculations of reaction mechanisms demand the optimization and characterization of the stationary points *in* solution.

A continuum model is not able to describe specific interactions between solute and solvent. Usually, they can be important in protic solvents, such as alcohol solvents and water. Thus, the use of a purely continuum model is not recommended for such solvents. In the same way, if there is the possibility for solvent coordination or, even more, solvent molecules can participate in the reaction, the inclusion of explicit solvent molecules in the chemical model is mandatory. Hybrid cluster-continuum models can be a good alternative to tackle these issues. A major concern in these models is the number and relative position of explicit solvent molecules to be included. The influence of this number and their conformations on reaction barriers should be carefully checked until convergence is reached.[25] Automated methods to determine how many explicit solvent molecules need to be added in hybrid cluster-continuum schemes to capture most of the interaction between the solute and the environment have been recently proposed.[26]

The elucidation of reaction mechanisms when solvent molecules play a direct role in the mechanism and are involved in multiple ligand exchange processes has proven to be challenging. This occurs, for instance, in two textbook organometallic reactions: the Wacker process and the Grignard reaction. Calculations using explicit solvent models, placing the reagents in a box with a large number of solvent molecules, water and tetrahydrofuran (THF), respectively, and performing *ab initio* Molecular Dynamics (AIMD) simulations with them, have solved long standing controversies about the mechanism of both reactions.[27,28]

4. The Computational Model

Once the chemical model has been decided, the next step is the choice of the computational model. While the first election has consequences for the possible mechanisms to be explored (for instance, mechanisms with participation of dimers cannot be explored if only one molecule of the catalyst is included in the chemical system), the second one will have impact on the accuracy of the numbers provided by the calculations. QM calculations are devised to afford the internal energy (electronic energy, E^{elec}) of a molecular assembly in gas phase and at 0 K, very far from the actual experimental conditions.[29] Moreover, the key chemical magnitudes that govern the direction and rate of a reaction (equilibrium and rate constants) are not related to differences in internal energies, but in Gibbs energies. Energy profiles of reactions carried out in a solvent have to be built upon G values in solution. Transforming energies from E to G implies adding additional approximations to those inherent to the QM calculation and raises more choices to the modeler. Currently, exploration of the PES of real systems is habitually performed using a medium-size basis set (BS1), usually valence double-ξ+polarization. It is a common and highly recommended practice to refine the energy values with a very extended basis set (BS2) in order to remove basis-set superposition errors (BSSE). This is an approximation based on the assumption, usually true, that the optimized geometries with both basis set are very similar. In this way the Gibbs energies in solution are obtained from equation 1:

$$G_{sol} = E^{elec}(BS2) + [G_{sol}(BS1)\text{-}E^{elec}(BS1)] + \Delta G^{1atm \rightarrow 1M} \quad (1)$$

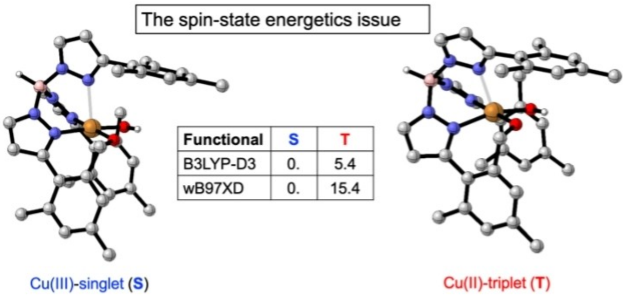
The first term (E^{elec}) is the QM energy in solution (usually DFT) and includes the Gibbs energy of solvation.[29] Its accuracy depends on the level of theory chosen (functional + basis set). The second term embraces the enthalpic and entropic contributions of the solute. It demands the partition functions for the solute, obtained from a frequency calculation performed at the optimized geometries (BS1). The last term accounts for the change of standard state from gas phase (1 atm) to solution (1 M) and at 298 K amounts 1.89 $kcal \cdot mol^{-1}$ for each species. It is only relevant when the number of moles changes in the reaction ($\Delta n \neq 0$).

The accuracy in computing E^{elec} is known as the electronic structure issue. The quality to price ratio have made DFT methods almost the only option for computational studies of organometallic reactivity. But which functional should be used? This is a question to ask ourselves before starting calculations, which does not have a general answer. Strictly speaking, it depends on the particular reaction to be studied. Moreover, catalytic reactions are multistep and the selected functional should be able to provide accurate results for all the steps. To further complicate matters, kinetic experimental studies that might serve as benchmarks are not very frequent. Despite that, some benchmarking studies have been carried out for one metal and one particular reaction[30] or several elementary steps of metal-catalyzed transformations.[31] In recent years local coupled-cluster methods, such as DLPNO-CCSD(T) have emerged as accurate and effective methods for benchmarking in big-sized systems.[32]

My personal impression is that the choice of the functional is not as important as it is generally thought to be. It is worthy to perform some calibration of the functional, but in general energy profiles with any non-local GGA or meta-GGA functional will differ by just a few $kcal \cdot mol^{-1}$ when computing closed-shell species. The most relevant aspect to take into account is probably the percentage of Hartree-Fock (HF) exchange in the functional, especially in reactions involving open-shell species. It is not optional, however, whether to take into account dispersion interactions or not. The use of dispersion-corrected functionals or the incorporation of empirical dispersion parameters[33] to non-dispersion-corrected functionals is mandatory. As already mentioned when describing continuum models, optimizations have to be performed taking into account dispersion. With the amount of evidence accumulated on the importance of dispersion effects when describing real organometallic systems,[34] and the low cost that its inclusion entails, it is hard to understand the reason why in many recent articles dispersion interactions are introduced only as single-point energy calculations at the geometries optimized without dispersion.

Unfortunately, the similarity of energy profiles obtained using different functionals of similar quality for closed-shell systems does not hold true when there are open-shell species (radicals) involved in a reaction mechanism. Comparison of DFT-computed relative energies of different spin-states of one molecule shows a great disparity, depending particularly on the HF exchange in the functional. An accurate calculation of the energy splitting between low-spin and high-spin states is difficult due to its strong functional dependence.[35] For instance, taking two functionals widely used in reactivity studies, such as wB97XD or B3LYP–D3, both incorporating dispersion effects, with a very large basis set, the high-spin [triplet, formally copper(II)] and low-spin [singlet, formally copper(III)] isomers of an intermediate formed in the Cu(I)-promoted aerobic C–H bond oxidation (Figure 3) differ by 5.4 (wB97XD) or 15.4 (B3LYP–D3) kcal·mol^{-1}.[36] When leaving the field of "comfortable" closed-shell PES, careful analysis of the functional dependence is needed.

As commented previously, electronic energies must be transformed into Gibbs energies by adding enthalpic and entropic contributions. This is usually done by means of frequency calculations in which the different contributions to the molecular entropy are estimated from the partition functions. In order to do that, the usual approach is the ideal gas/rigid rotor/harmonic oscillator approach (IGRRHO). It was thought that this approach overestimated entropic effects in solution, and several alternatives have been devised, most of them using empirical corrections, such as dividing by 1/2 or 2/3 the translational entropy. A number of recent works have pointed out the validity of the IGRRHO and the unnecessariness of further correction, provided that dispersion forces are taken into account in the calculation. The best currently available way to handle entropic effects for species in solution is the IGRRHO approach, combined with continuum solvent treatment of solvation Gibbs energy, and taking into account the change of reference standard state ($\Delta G^{1atm\rightarrow 1M}$ term in equation 1).[7–9]



Functional	S	T
B3LYP-D3	0.	5.4
wB97XD	0.	15.4

Figure 3. Functional dependence of the Gibbs energy difference between the low-spin and high-spin isomers of a copper-hydroxo complex.[36] Relative Gibbs energies in THF solvent in kcal·mol^{-1}. 3D-structures were generated using CYLview.[15]

5. Success… and Limitations

In the previous sections I have provided the ground for what I consider an accurate computing of an energy profile in the current times.[9] Nowadays, taking the necessary precautions regarding both the chemical and computational models, DFT calculations can afford in most cases a reliable quantitative description of the energy landscape for the formation of the major product of a transition metal promoted reaction in solution. In this way the determining states of a catalytic cycle can be found, and by using the energetic span model turnover frequencies (TOF) can be calculated from the energy profile.[37]

Commonly, the energetic information provided by the energy profile is used to infer and propose a reaction mechanism. It should never be overlooked that a reaction mechanism cannot be proven, but only disproved.[38,39] As commented before, the mechanism search often suffers from the "wish bias", defined as the tendency on the investigator to reach a desired result.[40] When dealing computationally with very well-known reactions, as it happens in most cases, no big surprises can be expected, the computed profile is just one more on this kind of reaction and the mechanistic proposal is sound….until proven otherwise. However, there is not a unique pathway for the most interesting reactions. As not all the possible routes have been explored, there could be a better mechanism, involving lower barriers than the initially proposed.

Another shortcoming is that often only the most productive part of the reaction is computationally explored. Rarely does one reaction give a 100% yield. In the reaction vessel, the catalyst can be deactivated and side reactions can happen, both facts reducing the reaction yield. Improving the reaction yield is a main goal in synthetic chemistry. Computational analysis of the off-cycle part of reaction mechanisms can help improving catalysts, preventing for instance their deactivation, as it has been shown for cross-coupling reactions.[41]

Performing calculations once the experimental part of a project has been finished, bringing after-the-fact explanations, is still the most common use of calculations in transition metal catalysis that can be found in the literature. This approach has proved to be very useful to understand experiments. However, proceeding solely in this way misses out the full potential of contemporary simulation techniques. Simulations are mature enough to be incorporated at the experimental design stage and to guide the experimental exploration throughout a project's development. In this way, and in synergistic collaboration with experiments, calculations can evolve from being a technique to "predict the past" to becoming a powerful tool to forecast the future. In chemistry that means to design new catalysts and to discover new reactions. In the next section I will outline how the combination of accurate QM calculations, described before, with emerging computational tools can make this possible.

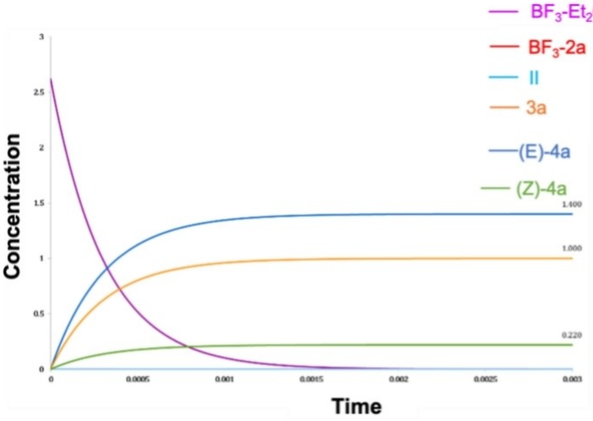


Figure 4. Microkinetic modeling of the isomerization of 2-phenyl-2-methyloxetane **2 a** to allylic alcohol catalyzed by $BF_3 \cdot Et_2O$.[45] **II**: zwitterionic intermediate; **3 a**: homoallylic alcohol product; **4 a**: E/Z-allylic alcohol products.

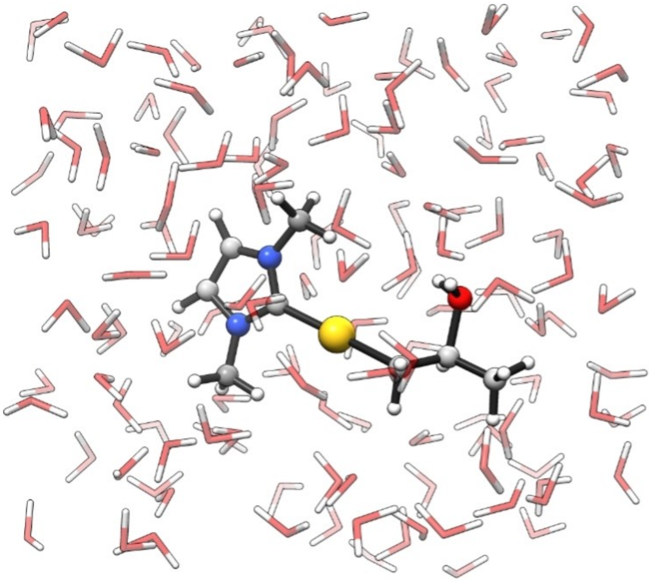


Figure 5. Snapshot from an explicit water AIMD simulation of the hydration of alkenes catalyzed by gold(I) complexes.

6. Designing Catalysts and Discovering New Reactions

Beyond calculation of energy profiles, computational organometallic chemistry is taking new directions to overcome shortcomings and to reach the goals pointed out in the previous section. On one side, calculations must be closer to the experiments, regarding both the outcome and the chemical models. On the other side, they have to explore the full PES of the system, overcoming the human bias. Finally, they can be combined with extensive data analysis and machine learning algorithms. All together, these advances approach the holy grail of catalyst design and can lead to the longed-for discovery of new reactions. The future is now: recent literature collects an explosion of articles devoted to these topics. For instance, in the last two years three review journals have published special issues devoted to machine learning in chemistry, focusing on the design of catalysts and materials.[42] Here, I will mention just a few examples.

Computational mechanistic studies are aimed to locate intermediates and transition states in a catalytic cycle and to determine their relative energies. In this way a Gibbs energy profile is obtained, and a mechanistic proposal is inferred from it. However, the direct outcome of experiments is the evolution of concentrations of reagents, intermediates, and products through time. These values depend on the rate constants of the reaction steps, related with Gibbs energy differences, but also on the concentrations of the reagents, not accounted for in the PES exploration. Microkinetics modeling is a simple and useful method for the introduction of concentration effects in the description provided by the Gibbs energy profiles, getting simulations closer to experimental conditions.[43,44]

The energy profile defines the steps of a complex reaction and their energy barriers. In the framework of the Transition State Theory, the rate constant of each reaction step can be computed from its Gibbs energy of activation. This allows the definition of a system of differential equations, which can be solved numerically by using the initial concentrations of all the compounds as starting conditions.[43,44] The outcome of the microkinetic modeling is the evolution of the concentration of each species with time (Figure 4),[45] which is precisely what most experiments measure. The computational cost of this procedure, which can also be used as additional testing of a proposed reaction mechanism, is very low.

Improving the solvent description is another steppingstone in approaching computer models to "real chemistry" in solution. This can be done employing extended explicit representations of the solvent, built to reproduce experimental densities (Figure 5). Dealing with large model systems, however, poses a challenge: how to explore their configurational space? Systematically investigating how the potential energy varies as a function of few nuclear coordinates, as done in typical PES exploration on reduced model systems, is not viable because of the large number of particles involved. First-principles molecular dynamics (also called *ab initio* Molecular Dynamics, AIMD), in which atomic forces are computed from electronic structure calculations (namely, density functional theory), is the technique of choice to investigate chemical events in solution when the solvent has a key role in the reaction and/or where solvent reorganization is important.[46]

These simulations make use of intensive computational resources, but the inclusion of explicit solvent allows the characterization of molecular events that would otherwise not be accessible in reduced model systems, as it has already been proved in Wacker[27] and Grignard[28] reactions. For some processes, such as transfer hydrogenation in alcohol solvent, the use of explicit solvent simulations may provide mechanistic insights that differ from those obtained in vacuo or in continuum models.[47]

As commented above, the usual way of proceeding in computational studies of organometallic reactivity has typically been to explore a limited number of pathways, inferred from previous knowledge and/or chemical intuition of the person performing the calculations, that lead to the major product of the reaction. That means to explore a very small fraction of the PES of the reacting system. The emergence, in the last years, of the automated reaction path methods, promises to revolutionize the way of exploring the PES. With these methods, which have been successfully applied to organometallic reactions,[48–50] an unbiased exploration of the full PES can be performed. They do not only allow exploration of multiple reaction paths accounting for product selectivity in connection with microkinetic models but also pave the way for the discovery of unexpected reactions in the full potential energy surface of the system.

Catalyst optimization and, furthermore, design and discovery, employing the usual approach of thoroughly PES exploration with accurate quantum mechanical calculations, becomes impractical when a large chemical space (thousands of catalysts and reactions) has to be explored. In the framework of the explosion of the use of data-driven, machine learning (ML) algorithms in all fields of knowledge, they have also burst onto catalyst design. Although these techniques are still at an early stage in this field, they have already achieved impressive successes, particularly in the area of enantioselective catalysis,[51,52] proving its potential. A number of very recent reviews prove the fact that they are called to change the way in which catalysts are discovered.[51–55]

The starting point of data-driven tools is establishing a relationship between a quantitative description of reactants, catalysts and reaction conditions with a property (for instance, activity: Quantitative Structure-Activity relationship, QSAR). The key ingredients of these models are the quantities (descriptors) that are correlated with the properties of interest. Multivariate linear regression models correlate data to molecular descriptors. For non-linear relationships in complex chemical data, machine learning offers a wide variety of algorithms that can be used to build accurate, transferable, and explainable predictive models. However, many of these algorithms require big data sets. Computational results from QM calculations can be used to generate these required data needed to feed the ML models (quantum-based ML).[56] The calculation of over 200,000 organic radical species[57] and over 80,000 organometallic compounds[58] are impressive examples of how DFT methods can be used to generate big data sets. These data sets can be exploited in data-driven predictive ML models for the discovery of new catalysts and reactions. However, the inclusion of many complexes which are not realistic/synthetically feasible to generate training data could lead to bias in a ML model's prediction.

DFT and ML methods can be combined to enable the prediction of reactivity within large chemical spaces containing thousands of complexes.[59] Around 2500 QM transition state structures in the chemical space of IrL_3X Vaska's complex were employed to train and test ML models that predict the H_2 activation barrier. Different ML algorithms were used for different purposes, including neural networks for accuracy, Gaussian processes for transferability, and Gradient boosting for explainability.[60] A reactivity descriptor database based on QM calculations of 130,000 organic molecules was used to predict regio-selectivity for three general types of substitution reactions.[61] Trends surrounding the thermodynamics of the hydroformylation reaction catalyzed by group 9 metals bearing phosphine ligands have been analyzed using a data-driven inspired workflow (data-powered volcano plots). The total data set used consisted of 1510 catalytic cycles derived from DFT computations and 491 catalytic cycles derived from machine learned profiles.[62] Electroanalytical techniques have been combined with parameterization tools, which include DFT calculations, to uncover reaction mechanism in redox catalysis.[63]

7. Conclusions

In this essay I wanted to give a personal insight on the current state of computational methods applied to homogenous catalysis processes and to point out some of the new directions they are taking. The basic assumption is that modern QM-based methods are able to give accurate energy profiles of an organometallic reaction in solution entailing closed-shell species, provided that a proper choice of the chemical and computational methods has been made. This proper choice, which describes what I consider the state-of-the-art of these techniques, has been analyzed in the first part of the article. In my opinion that means, basically, that all the species that can significantly affect the energy landscape of the reaction should be included in the QM description of the chemical model to be computed. In addition, the conformational space of big and flexible ligands, as well as the speciation issues of catalyst and reagents should be carefully analyzed. Regarding the computational model, optimization and characterization of stationary structures should be done in solution with functionals taking into account dispersion effects. Refinement of the energies with very extended basis sets is highly recommended, as well as some functional calibration for the particular reaction to be computationally studied. Entropic contributions can be computed using the IGRRHO approach, with no further corrections. In this way, accurate relative Gibbs energies of the species along the computed pathway for the formation of product(s) can be obtained. Success of the computational protocol that I have described cannot be assured when more than one spin-state can be involved in the reaction. Such energy profiles are still challenging and require careful functional calibration.

The translation from energy profile to "the" mechanism of the reaction is not as direct as it is usually thought. In the best cases, it can be considered a sound proposal…until a better one comes out. In most cases calculation of an energy profile is performed when the experimental research on the system has been concluded and is used only to explain the observed reactivity, missing out the actual potential of QM calculations. On the other side, an energy profile is the outcome of a biased exploration of a very small region of the PES of the system. Automated reaction path methods can explore the full reactivity of the system, accounting for the formation of side products and catalyst deactivation pathways and even discovering new reactions than can take place within the reaction mixture.

To fully benefit from the current potential of QM calculations in homogenous catalysis, these must be used to test and to guide the experimental research throughout all the project's development, from its initial design to the final optimization. However, the exploration of large chemical spaces required to optimize and discover new catalysts with the usual techniques of PES exploration is impractical. Increasingly, this is done using data-driven techniques and machine learning algorithms, able to explore very large chemical spaces, but these approaches often require big data sets. In this field, QM calculations have found their place to generate big data sets.

In the first lines of this essay, I asked myself if theoreticians are still needed. Throughout the text, I have sought to show that my answer is yes, but only for as long as we are able to do things that other chemists cannot. Routine, ex post calculation of energy profiles is nowadays within the reach of any chemist. However, exploring large chemical spaces and finding new, synthetically relevant reactions using state-of-the-art methods and exploiting them with machine learning methods, without losing the physical meaning of the quantities handled, can only be done by researchers combining a deep knowledge of computational methods with a molecular perspective. With this in mind, computational chemists can play an ever more central role in homogenous catalysis. There is still plenty of room for computational organometallic catalysis.

Acknowledgements

This essay summarizes forty years of human learning (HL), fed with discussions and talks with all the people who have been in my group and experimental and theoretical colleagues around the world. I am indebted to all of them. Special thanks to Gregori Ujaque (UAB), Giuseppe Sciortino (ICIQ), and David Balcells (Oslo) for a critical reading of the manuscript. Financial support from Spanish Ministry of Science and Innovation (project PID-2020-116861GB-I00) is acknowledged.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: Catalyst discovery · Chemical and computational models · Density functional calculations · Energy profile · Reaction mechanisms

[1] T. Sperger, I. A. Sanhueza, I. Kalvet, F. Schoenebeck, *Chem. Rev.* **2015**, *115*, 9532–9586.
[2] S. Santoro, M. Kalek, G. Huang, F. Himo, *Acc. Chem. Res.* **2016**, *49*, 1006–1018.
[3] K. D. Vogiatzis, M. V. Polynski, J. K. Kirkland, J. Townsend, A. Hashemi, C. Liu, E. A. Pidko, *Chem. Rev.* **2019**, *119*, 2453–2523.
[4] S. Ahn, M. Hong, M. Sundararajan, D. H. Ess, M.-H. Baik, *Chem. Rev.* **2019**, *119*, 6509–6560.
[5] L. Perrin, K. J. T. Carr, D. McKay, C. L. McMullin, S. A. Macgregor, O. Eisenstein, *Struct. Bond.* **2016**, *167*, 1–38.
[6] Z. Liu, C. Patel, J. N. Harvey, R. B. Sunoj, *Phys. Chem. Chem. Phys.* **2017**, *19*, 30647–30657.
[7] J. N. Harvey, F. Himo, F. Maseras, L. Perrin, *ACS Catal.* **2019**, *9*, 6803–6813.
[8] J. N. Harvey in: *Transition Metals in Coordination Environments. Challenges and Advances in Computational Chemistry and Physics, Vol. 29* (Eds: E. Broclawik, T. Borowski, M. Radoń), Springer, Cham, **2019**, pp 289–313.
[9] O. Eisenstein, G. Ujaque, A. Lledós, *Top. Organomet. Chem.* **2020**, *67*, 1–38.
[10] I. Funes-Ardoiz, F. Schoenebeck, *Chem* **2020**, *6*, 1904–1913.
[11] H. Ryu, J. Park, H. K. Kim, J. Y. Park, S.-T. Kim, M.-H. Baik, *Organometallics* **2018**, *37*, 3228–3239.
[12] M. Besora, A. A. C. Braga, G. Ujaque, F. Maseras, A. Lledós, *Theor. Chem. Acc.* **2011**, *128*, 639–646.
[13] P. Gómez-Orellana, A. Lledós, G. Ujaque, *J. Org. Chem.* **2021**, *86*, 4007–4017.
[14] S. Grimme, *J. Chem. Theory Comput.* **2019**, *15*, 2847–2862.
[15] CYLview20; C. Y. Legault, Université de Sherbrooke, 2020 (http://www.cylview.org).
[16] P. Villuendas, S. Ruiz, P. Vidossich, A. Lledós, E. P. Urriolabeitia, *Chem. Eur. J.* **2018**, *24*, 13124–13135.
[17] S.-T. Kim, S. Kim, M.-H. Baik, *Chem. Sci.* **2020**, *11*, 1017–1025.
[18] G. Barea, F. Maseras, A. Lledós, *New J. Chem.* **2003**, *27*, 811–817.
[19] J. M. Hayes, E. Deydier, G. Ujaque, A. Lledós, R. Malacea-Kabbara, E. Manoury, S. Vincendeau, R. Poli, *ACS Catal.* **2015**, *5*, 4368–4376.
[20] A. Couce-Rios, A. Lledós, I. Fernández, G. Ujaque, *ACS Catal.* **2019**, *9*, 848–858.
[21] A. Macchioni, *Chem. Rev.* **2005**, *105*, 2039–2074.
[22] M. Jia, M. Bandini, *ACS Catal.* **2015**, *5*, 1638–1652.
[23] G. Ciancaleoni, L. Belpassi, D. Zuccaccia, F. Tarantelli, P. Belanzoni, *ACS Catal.* **2015**, *5*, 803–814.
[24] T. Sperge, I. A. Sanhueza, F. Schoenebeck, *Acc. Chem. Res.* **2016**, *49*, 1311–1319.
[25] M. A. Ortuño, A. Lledós, F. Maseras, G. Ujaque, *ChemCatChem* **2014**, *6*, 3132–3138.
[26] G. N. Simm, P. L. Türtscher, M. Reiher, *J. Comput. Chem.* **2020**, *41*, 1144–1155.
[27] A. Stirling, N. N. Nair, A. Lledós, G. Ujaque, *Chem. Soc. Rev.* **2014**, *43*, 4940–4952.
[28] R. M. Peltzer, J. Gauss, O. Eisenstein, M. Cascella, *J. Am. Chem. Soc.* **2020**, *142*, 2984–2994.
[29] Calculations including a continuum solvent afford a "non-pure" thermodynamical magnitude, E_{solv}. This is an electronic energy that also includes the Gibbs energy effects of the solvent, but not those of the solute.
[30] K. Rohmann, M. Hölscher, W. Leitner, *J. Am. Chem. Soc.* **2016**, *138*, 433–443.
[31] K. H. Hopmann, *Organometallics* **2016**, *35*,3795–3807.
[32] D. G. Liakos, M. Sparta, M. K. Kesharwani, J. M. L. Martin, F. Neese, *J. Chem. Theory Comput.* **2015**, *11*, 1525–1539.
[33] S. Grimme, *WIREs Comput. Mol. Sci.* **2011**, *1*, 211–228.
[34] See, for instance: E. Lyngvi, I. A. Sanhueza, F. Schoenebeck, *Organometallics* **2015**, *34*, 805–812.
[35] M. Swart, *Top. Organomet. Chem.* **2020**, *67*, 191–226.
[36] M. Álvarez, F. Molina, M. R. Fructos, J. Urbano, E. Álvarez, M. Sodupe, A. Lledós, P. J. Pérez, *Dalton Trans.* **2020**, *49*, 14647–14655.
[37] S. Kozuch, S. Shaik, *Acc. Chem. Res.* **2011**, *44*, 101–110.
[38] O. Eisenstein, *Faraday Discuss.* **2019**, *220*, 489–495.

[39] S. L. Scott, *ACS Catal.* **2019**, *9*, 4706–4708.
[40] E. L. Wynder, I. T. Higgins, R. E. Harris, *J. Clin. Epidemiol.* **1990**, *43*, 619–621.
[41] D. Balcells, A. Nova, *ACS Catal.* **2018**, *8*, 3499–3515.
[42] a) D. Ess, L. Gagliardi, S. Hammes-Schiffer, *Chem. Rev.* **2019**, *119*, 6507–6508 (Special Issue: Computational Design of Catalysts from Molecules to Materials); b) H. J. Kulik, M. S. Sigman, *Acc. Chem. Res.* **2021**, *54*, 2335 – 2336 (Special issue: Data Science Meets Chemistry); c) R. Gómez-Bombarelli, A. Wiltschko, *Trends Chem.* **2021**, *3*, 75–76 (Special Issue: Machine Learning for Molecules and Materials).
[43] M. Besora, F. Maseras, *WIREs Comput. Mol. Sci.* **2018**, *8*, e1372.
[44] M. Jaraíz, *Top. Organomet. Chem.* **2020**, 67, 81–106.
[45] A. Cabré, S. Rafael, G. Sciortino, G. Ujaque, X. Verdaguer, A. Lledós, A. Riera, *Angew. Chem. Int. Ed.* **2020**, *59*, 7521–7527.
[46] P. Vidossich, A. Lledós, G. Ujaque, *Acc. Chem. Res.* **2016**, *49*, 1271–1278.
[47] J. W. Handgraaf, E. J. Meijer, *J. Am. Chem. Soc.* **2007**, *129*, 3099–3103.
[48] J. A. Varela, S. A. Vázquez, E. Martínez-Núñez, *Chem. Sci.* **2017**, *8*, 3843–3851.
[49] A. L. Dewyer, A. J. Argüelles, M. Zimmerman, *WIREs Comput. Mol. Sci.* **2018**, *8*, e1354.
[50] M. Hatanaka, Y. Takayoshi, S. Maeda, *Top. Organomet. Chem.* **2020**, *67*, 57–80.
[51] M. S. Sigman, K. C. Harper, E. N. Bess, A. Milo, *Acc. Chem. Res.* **2016**, *49*, 1292–1301.
[52] A. F. Zahrt, S. V. Athavale, S. E. Denmark, *Chem. Rev.* **2020**, *120*, 1620–1689.
[53] M. Foscato, V. R. Jensen, *ACS Catal.* **2020**, *10*, 2354–2377.
[54] L. C. Gallegos, G. Luchini, P. C. St. John, S. Kim, R. S. Paton, *Acc. Chem. Res.* **2021**, *54*, 827–836.
[55] D. J. Durand, N. Fey, *Acc. Chem. Res.* **2021**, *54*, 837–848.
[56] O. A. von Lilienfeld, K.-R. Müller, A. Tkatchenko, *Nat. Chem. Rev.* **2020**, *4*, 347–358.
[57] P. C. St. John, Y. G. Y. Kim, B. D. Etz, S. Kim, R. S. Paton, *Sci. Data* **2020**, 7:244.
[58] D. Balcells, B. B. Skjelstad, *J. Chem. Inf. Model.* **2020**, *60*, 6135–6146.
[59] G. d. P. Gomes, R. Pollice, A. Aspuru-Guzik, *Trends Chem.* **2021**, *3*, 96–110.
[60] P. Friederich, G. d. P. Gomes, R. De Bin, A. Aspuru-Guzik, D. Balcells, *Chem. Sci.* **2020**, *11*, 4584–4601.
[61] Y. Guan, C. W. Coley, H. Wu, D. Ranasinghe, E. Heid, T. J. Struble, L. Pattanaik, W. H. Green, K. F. Jensen, *Chem. Sci.* **2021**, *12*, 2198–2208.
[62] M. D. Wodrich, A. Fabrizio, B. Meyerac, C. Corminboeuf, *Chem. Sci.* **2020**, *11*, 12070–12080.
[63] T. Tang, C. Sandford, S. D. Minteer, M. S. Sigman, *Chem. Sci.* **2021**, *12*, 4771–4778.

Manuscript received: April 23, 2021
Revised manuscript received: May 24, 2021