

Authors: María Silvia Polzella and Penélope Lodeyro

Title: Re-evaluating Semi-Empirical Computer Simulations in Quantum Chemistry

Affiliations and addresses of the authors:

María Silvia Polzella College of Philosophy and Humanities Research Center (CIFYH), National University of Córdoba (UNC) - Secretary of Science and Technology (SECyT), Haya De La Torre, Ciudad Universitaria S/n, 5016, Córdoba, Argentina)

Penélope Lodeyro Humanities Institute (IDH), National University of Córdoba (UNC) – National Scientific and Technical Research Council (CONICET), Haya De La Torre, Ciudad Universitaria S/n, 5016, Córdoba, Argentina)

E-mail address and telephone number of the authors: María Silvia Polzella (marsipol@ffyh.unc.edu.ar; +5493512462438) and Penélope Lodeyro (loopenelope@gmail.com; +5493513081410)

Abstract

Usually within the context of computer simulation in quantum chemistry practices (and solid-state physics), there is a distinction between ab initio and semi-empirical methods. Related to this, a controversy within the scientific and philosophical communities came about regarding the superiority of the ab initio methods due to their theoretical rigor. In this article we re-evaluate the condition of the semi-empirical simulations in this area of research. We examine some of the aspects of this debate that have been considered in philosophy and provide additional elements to the analysis.

Keywords: computer simulations; quantum chemistry; semi-empirical; ab initio; scientific practices

This is a pre-print version of an article published in *Foundation of Chemistry*. The final authenticated version is available online at: <http://dx.doi.org/10.1007/s10698-018-09329-w>

1. Introduction

Usually within the context of computer simulation in chemistry (and solid-state physics), there is a distinction between *ab initio* and semi-empirical methods. The *ab initio* methods are characterized as those calculations that are derived from first principles, or are based on fundamental equations of physics or chemistry. In this sense, they are considered to operate on a purely theoretical foundation. On the other hand, the semi-empirical ones owe their name to the fact that they incorporate empirical parameters in their calculations. That is, the values of some of the equation's parameters that, in principle, can be theoretically determined are fixed by values that have been established experimentally. Related to these characterizations, a controversy within the scientific community came about regarding the superiority of the *ab initio* methods due to their theoretical rigor. The dispute stimulated the philosophical reflection, becoming a debate on methods based on "principles and parameters" (Ramsey 2000).

Although both the semi-empirical and *ab initio* methods originated before the use of computers in scientific practices, the advancement of both has been propelled with the introduction of computers, giving way to the emergence of computer simulations in quantum chemistry.

In this article we re-evaluate the condition of the semi-empirical simulations in this area of research. To achieve this, we examine some of the aspects of the debate between the *ab initio* and semi-empirical methods that have been considered in philosophy and provide additional elements to the analysis. With the purpose of reinforcing the argument that the distinction is a question of degree and not of type (Ramsey 1997, 2000; Scerri 2004a, 2004b), we stress that the *ab initio* simulations do not have the theoretical rigor that they claim to possess. We begin by highlighting that the *ab initio* methods generally incorporate drastic approximations such as the Born-Oppenheimer approach, which gives them semi-classical features. Then, we point out that the basis sets, besides being "put in 'by hand'" (Scerri 2004a), although they must be complete, in practice are restricted due to the limited capacity of the calculation.¹ Next, we review the difficulties that the consideration of the density functional theory (DFT) introduces in trying to distinguish between the *ab initio* and semi-empirical methods. Once it has been shown that the *ab initio* simulations do not have the theoretical rigor that is often claimed, we will emphasize the fact that their reliability mainly comes from the good agreement with the experimental data, just as for those simulations that are semi-empirical. Therefore, it begs the question, what makes the *ab initio* methods preferable? With the purpose of delineating an answer, we consider the main objections that have been brought out against the semi-empirical simulations. For this, we analyze some of the aspects that make a semi-empirical simulation a good one. Regarding the theoretical foundation, we show the relevance of taking great care in selecting the approximation as well as the parametrization. Furthermore, we propose the degree of generality, the range of application, and precision as the key features that depend on the computational development. Finally, we stress that scientific practices show the prominent role played by

¹ In quantum chemistry, the "basis set" usually refers to the set of (nonorthogonal) one-particle functions used to build molecular orbital. The basis set, in theory, should be complete, but a complete basis set means that an infinite number of functions must be used, which is impossible in actual calculations.

semi-empirical simulations in the investigative frontier of quantum chemistry, by enabling to study larger and larger systems.

2. Some traits of the *ab initio* computer simulations that question their theoretical purity

Schrödinger's equation in wave mechanics (1926) proved to be useful for solving problems in chemistry, beginning with the Heitler and London published work in 1927. It showed on the one hand that wave mechanics works, and on the other, provided a theoretical explanation as to why two hydrogen atoms combine to form one molecule (Park 2009). They found a new "*pure quantum effect*, the exchange energy" (Kostas Gavroglu and Simões 2012, p. 16), of attraction or repulsion between like particles and developed a schematic theory for homopolar valence (Jammer 1966) which, together with the electron spin and the Pauli exclusion principle, completed the new quantum theory. According to Kostas Gavroglu (1997), the Heitler and London work began a new chapter in the field of study of chemical problems. It created an expectation that one could gather information on molecular systems using wave mechanics. This optimism was captured in Dirac's famous opinion (1929), according to which only a few adjustments were needed to get an agreement to the ideas of special relativity, even though the latter was not of major importance to molecular structure and ordinary chemical reactions. The necessary laws of physics for giving a mathematical description of a grand portion of physics as well as the whole of chemistry were already known, but the equations were "much too complicated to be solvable". From classical mechanics, the impossibility of finding exact solutions for interactions in systems with more than two bodies was already known; whereas, for producing insight on the main features of complex atomic or molecular systems, quantum physics confronts the n-body problem. Thus, it was necessary to develop approximation methods that allow for the treatment of chemical phenomena and also have sufficient predictive capability.

Regarding the status of the approximation methods developed between 1927 and 1935² van Vleck and Sherman (1935, p.169) questioned the following:

"How, then, can it be said that we have a quantum theory of valence? *The answer is that to be satisfied one must adopt the mental attitude and procedure of an optimist rather than a pessimist about it.* The latter demands a rigorous postulational theory, and calculations devoid of any questionable approximations or of empirical appeals to known facts. The optimist, on the other hand, is satisfied with approximate solutions of the wave equation. If they favor, say, tetrahedral and plane hexagonal models of methane and benzene, respectively, or a certain order of sequence among activation energies, or a paramagnetic oxygen molecule, he is content that these same properties will be possessed by more accurate solutions. He appeals freely to experiment to determine constants, the direct calculation of which would be too difficult. The pessimist, on the other hand, is eternally worried because the omitted terms in the approximations are usually rather large, so that any pretense of rigor should be lacking. The optimist replies that the approximate calculations do nevertheless give one an excellent "steer" and a very good idea of

² The text refers to the method of molecular orbital developed by Lennard-Jones, Hund, Herzberg and Mulliken; and the strictly homopolar method of Heitler, London, Slater and Pauling.

“how things go,” permitting the systematization and understanding of what would otherwise be a maze of experimental data codified by purely empirical valence rules. In particular, he finds that a mechanism is really provided by quantum mechanics for the Lewis electron pair bond, and for the stereochemistry of complicated organic compounds.”

As one can observe, this quote shows that in the initial stages of quantum chemistry there was a controversy among scientists regarding how these approximation methods should be elaborated. Some of them adhered to a demand for a theory of rigorous postulations, one that does neither appeal to questionable approximations nor to empirical facts. Meanwhile, others not only allowed for approximations, but also did not have any objections to using empirical values when the calculation was too difficult. Kostas Gavroglu and Ana Simões (1994) state that this attitude of the chemists, of implicitly and explicitly taking on questions about the role of theory and the methodological status of empirical observations for theory building in chemistry, was the cornerstone of the genesis and development of quantum chemistry as an autonomous sub-discipline. This brief review is intended to provide a framework for the following discussion about the distinction between *ab initio* and semi-empirical methods.

The *ab initio* methods have been characterized as those calculations that derive from first principles or are based on fundamental equations of physics or chemistry. In this sense, these methods are considered to operate upon a base, which is purely theoretical. Within the field of non-relativistic quantum mechanics, they seek to solve Schrödinger’s equation by using well-defined approximations that can be systematically improved, converging into the exact solution and only using fundamental physical constants for their calculations (Thiel 2005). Meanwhile, the methods denominated as semi-empirical are characterized as those that involve assumptions, approximations, or generalizations designed to simplify the calculation or to produce a result that has a sufficient empirical basis (Thiel 2005; Dewar 1992). They are formulated within the same theoretical framework of quantum mechanics as those using the *ab initio* methods, but, in order to speed up the calculation, the values of some parameters, instead of being calculated, are fulfilled by empirically-obtained values. With this distinction came a disagreement in the scientific community regarding whether *ab initio* methods were inherently better than semi-empirical ones, in virtue of their theoretical rigor. Philosophers took notice of the controversy, which was originally analyzed by Ramsey (1997, 2000) and Scerri (2004b) in terms of “principle and parameter” approaches, as a debate that “has been cast as a ‘right’ vs a ‘wrong’ way to produce theory” (Ramsey, 2000, p. 549). As we mentioned, the dispute between the *ab initio* and semi-empirical methods came from the origins of quantum chemistry and had been dealt with in the time prior to the incorporation of electronic digital computers (Park 2009; Ramsey 1997). However, the incorporation of electronic digital computers boosted the development of quantum chemistry. In terms of approximation methods, it produced a radical change and gave rise to computer simulations enabling the study of the behavior of previously inaccessible atomic and molecular systems. Currently, there is a massive deployment of computer simulations that have their own characteristics, which differentiate them greatly from mere number crunching. On the contrary, computer simulations constitute a rich inferential process that ranges from the choice of mathematical and theoretical model and its implementation to the running of the algorithm,

visualization and analysis of the produced data (Winsberg 2003, 2010). In addition, in the construction of computer simulations, due to the limited computational resources available, the theoretical model undergoes drastic transformations by the introduction of idealizations, approximations and even fictions to enable the calculation. In this way, the simulation acquires its own identity and is semi-autonomous with respect to the theory (Winsberg 2010). This contributes some particularities to the debate. Therefore, our work is restricted to analyzing the debate in the context of computer simulations.

A point that deserves to be treated first is that the methods called *ab initio* have some features that put into question the supposed theoretical rigor and pose a difficulty in evaluating their merits. Regarding this, Ramsey (1997, 2000) and Scerri (2004b) state that the *ab initio* methods are not purely theoretical and the discrimination by the semi-empirical methods is a matter of degree more than a clear-cut distinction. From a conceptual point of view, there is a sense in which no calculation can ever be said to be fully *ab initio*. Even the equations of the most fundamental of theories of particle physics, for example, do not generally compute the masses of elementary particles from first principles, as things stand at present (Scerri 2004b). Furthermore, in practice, the use of approximations and idealizations is inevitable.

Molecular quantum chemistry attempts to solve the time-independent Schrödinger equation. In principle, solving this equation would yield the wave function of a molecular system from which its molecular properties could be derived. However, in a strict sense, by using the wave equation, one can only analytically solve two-body systems, like that of the hydrogen atom which has only one electron and one proton (with the valence bond theory the hydrogen molecule could be studied, however this is already an approximation). Therefore, all quantum chemical methods are necessarily approximate. This leads us to the topic concerning how the selection of approximation impacts the results of these methods (Grant 2016).

Indeed, there is a series of coordinated idealizations and approximations at the base of *ab initio* simulations for the study of molecular systems. They assume a Schrödinger equation that is independent of time, which is an idealization, since physical and chemical natural systems depend on time. At the same time, the relativistic mass effects are usually ignored. Likewise, the majority of the computational simulations assume the Born-Oppenheimer approximation as valid, which consists basically in separating the molecule into electronic structure and nuclear movement.³ Due to the fact that the nuclear mass is much greater than the mass of the electron, it provides a good simplification to suppose that the movement of the nucleus is much less than that of its electrons (Tully 2000). Therefore, for a small movement of the nucleus, the electrons will react instantaneously, readjusting the optimal configuration as if the nucleus was in a fixed position. Thus, a stationary state is assumed for the nucleus in any given position and the nuclear coordinates are separated from the electronic coordinates. In a strict sense, this

³ In general, the Born Oppenheimer approximation assumes the adiabatic approximation. In this approach, the nuclear motion is confined to the lowest electronic potential surface of the system. There are procedures for which the Born Oppenheimer approximation is not valid: non radiative transitions in molecules and solids electron transfer, quenching of excited electronic states, collisional electron excitation, and inelastic electron scattering. “To describe such processes, “non-adiabatic transitions” among different potential-energy surfaces must be accounted for ...” (Tully 2000, p. 175).

approximation does not correspond to a quantum-mechanical image; rather it gives a semi-classical trait to the theoretical model underlying the simulation. Thus, the separation of electron and nuclear motions depends on assuming the nuclei as classical-like particles at rest in a definite position and the assumption of nuclei at rest and in a fixed position is in direct contradiction with the Heisenberg principle of quantum mechanics which prevents quantum systems from having simultaneously defined values for position and velocity (Lombardi and Castagnino 2010; Fortin and Martinez Gonzalez 2013). So, our argument is that ab initio simulations have as their foundation a semi-classical model, not quantum-mechanical, and in this sense, they stray from the assumed theoretical rigor. They do not provide an analytic solution to the “full Schrödinger equation” as that would imply that “*every* particle contributes to the kinetics and potential energy operators on an equal footing” (Wolley 1978, p. 1078 note 16).

Another aspect that shows ab initio simulations are not so ab initio as they claim is the basis set. As Scerri (2004a, p.105) affirmed, “Quantum mechanical calculations are not capable of actually generating their own basis sets that must instead be put in ‘by hand’”. Furthermore, they are often selected by reference to some empirical data (Scerri 2004b). An orbital, whether it be atomic or molecular, in principle, is a mathematical function in tridimensional space and is a characteristic solution to a mono-electronic Schrödinger wave equation. The set of basis functions, called basis set, is used to represent the electronic wave function in Hartree-Fock y Post Hartree-Fock Methods and Density functional theory. The computer simulations achieve an efficient implementation of the mathematical model, transforming the partial differential equations into algebraic equations. The orbital is then expressed as linear combinations of the basis functions.

In principle the basis set should be complete, however, in practice one is restricted to a basis set of finite size due to the limited capacity in calculating. For this, the basis set should be very carefully selected. The choice of the basis set strongly conditions the computational demands and the accuracy of the computational simulations ab initio.⁴ It is difficult to find a single set of functions that produces good results, that is, with an acceptable accuracy, in a broad spectrum of molecules and that which is computationally economical.⁵ In order to evaluate the accuracy, in practice, the majority of the existing numerous basis sets are tested with experimental data, as can be appreciated in work such as that of Gill et. al. (1992). Usually, a minimal basis set is sought out for the construction of a molecular orbital to be applied to the study of moderately large molecular systems, although an extended basis-set would produce much more accurate results. Achieving a good representation of the molecular orbital requires

⁴ A paradigmatic example of the tension in the choice of the basis set between a reasonable accuracy and the speed of calculation is the technique developed by Pople and colleagues (Hehre, Steward and Pople 1969). It consists of replacing each Slater-type orbitals (STO) by a linear combination of small number of Gaussian-type orbitals (GTO's) to accelerate the calculation, although it implied some loss of accuracy regarding the STO's. It is worth noting that, currently, the basis set of gaussian orbitals to model the wave functions are the ones that are most frequently used for ab initio simulations that use Hartree-Fock and Post-Hartree-Fock methods.

⁵ It's worth noting that a single basis set is applied to a broad range of molecules, but it does not have the same accuracy for all of them. Also, a chosen basis set does not have the same accuracy for all of the properties for any given molecule.

“wisdom” upon the election of the basis set. As Lowe (1993) states, “the choice of basis is one of the places where human decision can affect the outcome. Different choices of basis set will produce different wave functions and energies” (van Brakel 2000). In this sense, we have argued that the basis sets are not provided by the theory, they are chosen based on the speed of the calculation and the accuracy of the results. The selection of an extended basis set is in tension with the computational resources demand. Furthermore, the results of the methods, using different basis sets, are compared with experimental data. The reasonable agreement with these values enables the use of a determined basis set. The success of a basis set among others, is justified by an acceptable empirical adequacy to a minor computational cost. Thus, the validation of a basis set does not arise from quantum theory, but from empirical adequacy. In this sense, the choosing of a basis set undermines the pretension of theoretical purity of ab initio computational simulations.

Finally, we should take into consideration an approach of the computer simulation practices that further complicates any attempt to demarcate between ab initio and parametric methods, as has been stressed by Scerri (2004b). We are referring to the development of the density functional theory (DFT) computer simulations. DFT attempts to overcome the many-body problem by computing the energy written in terms of the electron density (Hohenberg and Kohn 1964). This captures the electron probability density or charge density⁶, since energy is a functional of density and the density of the system minimizes this functional. The total energy of a system is a unique functional of the electron density and the density uniquely determines the potential acting on the electrons, and vice-versa. Additionally, DFT assures the existence of an ‘exact’ functional that by construction handles exchange and correlation in any situation. By treating the electrons around the nucleus as a homogenous gas, it is possible to write an equation for orbitals of a particle, from which density is obtained (Kohn and Sham 1965). On the other hand, the function of the electron density can be determined experimentally by x-ray diffraction or electron diffraction, which allows for an experimental comparison of the results.

This seems to be a huge step, throwing out the fearsome complexity of a multidimensional wave function, but the problem that arises is that the theory does not tell us what that functional is, or how to find it (Segall et al. 2002). The greatest strength of the density functional theory lies in the possibility of reducing the problem of N-bodies, which implies going from considering N electrons with 3N spatial dimensions, plus the spin of the electron (4N variables), to only three spatial dimensions (x, y, z) by using the density functional theory. However, the complexity of many electrons interacting persists and it is necessary to find the exchange and correlation energy as a functional of the density, which requires approximations. The local-density approximation (LDA) is the simplest. This was already presented in the original formulation of the Kohn-Sham approximation method, which considers the electrons around the nucleus as a homogeneous electron gas. This idealization did not give an accurate result; thus, later a gradient in the electron density distribution was introduced, and, as Scerri (2004b) states, this is made in a semi-empirical way, incorporating experimental results. Becke (1998) has developed the generalized

⁶ Designated as $\rho(x, y, z)$ is one probability per unit of volume. If the charge of an electron is considered to be a unit of charge, then it has electron charge units.

gradient approximations (GGA) as a way of considering a gradient in the electronic density distribution. These approximations have improved the calculated results greatly, but they have introduced some numerical parameters of adjustment, determined through the optimization of the atomization energies accuracy from a standard set of molecules. Later, Becke (1993a, 1993b, 1996) improved the approach introducing a parametrized hybrid method, the Hybrid Hartree-Fock/GGA method.

In summary, even though as theory DFT is formally exact, in practice, it needs approximations for its application. The accuracy depends on the quality of the approximated density functional employed. Additionally, these approximations are parametrized. This shows a clear distancing from the ideal of theoretical rigor assumed for ab initio methods. As Segall et al. (2002, p. 2721-22) affirms: “Practical E_{xc} functionals (exchange and correlation energy functionals) are the major approximation made to DFT: they are not derived from first principles, but are postulated from physically reasonable assumptions, and their use is justified a posteriori by their success. (...) Almost all workers use the LDA or GGA in one of the several parametrizations available”. In effect, its success derives from an increase in the accuracy at an equal or lower computational cost, which is achieved by introducing empirical parameters. The question that then ensues is to what extent is DFT an ab initio method. Especially, given that the improvement in the accuracy that boosts the broad diffusion of the DFT method in the last few decades has been associated, in particular, with the development of hybrids between traditional wave function and DFT methods.⁷

3. Assessing the value of semi-empirical computer simulations

In the previous section we discussed the assertions made by Ramsey and Scerri, according to whom the ab initio methods are not purely theoretical and the discrimination against semi-empirical ones is more a matter of degree than a clear-cut distinction. Related to this, Scerri (2004b) stressed that even approaches considered to be ab initio, such as the density functional theory, have some semi-empirical traits, which makes even more difficult to distinguish between the methods.

It would seem as though theoretical rigor constitutes a *desideratum* linked to the idea that greater dependence on a theoretical framework gives it a higher degree of reliability. However, as we previously noted, various resources are introduced which give a degree of independence to the simulation with respect to the theory (Winsberg 2010). The process of construction of the computer simulations is very complex and involves an amount of elements that do not always derive from theory. Theory does not determine which is the best approximation for the phenomenon, at best it serves as a guide (Cartwright 1999). In this sense, we consider that the good agreement with the experimental data is for both, ab initio and semi-empirical, which grants them a good deal of reliability. For the purpose of validation, the

⁷ It must be noted that DFT forged the development of numerous hybrid approaches. Special attention deserves to be given to the application of DFT in the field of molecular dynamics that gave way to the formulation of ab initio molecular dynamics (Car and Parrinello 1985), making it possible to investigate large systems. For its importance, this approach deserves a thorough treatment that exceeds the aims of this work.

simulations are tested by comparing their results with known chemical facts.⁸ Accuracy is one of the main goals, thus, an acceptable range of accuracy is established for distinct properties. This is done in such a way that the quantitative results of the simulations should not have a greater discrepancy with the experimental data than that of the predetermined accuracy. For example, a global accuracy of 1 K/mole is appropriate for energies, such as heats of formation or ionizations potentials (Pople 1998).

Therefore, if the so-called *ab initio* methods do not have the assumed theoretical rigor and the reliability comes in greatly measure from empirical testing, it begs the question as to what makes them preferable against semi-empirical methods. We will analyze some of the considerations relating to semi-empirical simulations with the aim of delineating an answer. The first matter is to consider the objections that have been given. We deem that there are two lines of criticism that are the most attended. First, due to the drastic approximations and parametrizations, the *ab initio* theorists sustain that the semi-empirical methods are “bereft of any true theoretical foundations” (Freed 1995, p. 26). Second, it is believed that the semi-empirical methods are lacking of inferential content, in the sense that the values are adjusted so frequently that the model is a model for just one case (Ramsey 1997).

Regarding the first objection, it is worth reviewing some of the constructive aspects of semi-empirical simulations. They are based on the Hartree-Fock method within of the wave function approach. This method is also commonly used in *ab initio* simulations. However, in constructing the diverse semi-empirical simulations, distinct approximations are introduced in order to simplify the calculation.⁹ Many integrals are discarded and the remaining ones are modeled by empirical functions with adjustable parameters that are compared against a great number of precise experimental or theoretical data (benchmark) (Wu et al. 2016). The theoretical approach underlying and the approximations to the chosen integrals, determine the interactions that will be included within the simulation. If the simulation “retains the essential physics to describe the properties of interest, the parametrization may account for all other effects in an average sense, and it is then a matter of validation and benchmarking to establish the numerical accuracy of such methods.” (Thiel 2013, p.1).

It seems that there are not enough elements for maintaining that semi-empirical simulations bereft of theoretical foundation. Thus, as we have highlighted, they are based on the same theoretical framework

⁸ Even though here we are only referring to external validation, the notion of validation has been widely discussed in the context of computational simulations. In general, these treatments are based on the distinction between internal and external validation. They are said to be externally valid when the computational model represents, to an acceptable degree, the studied realm (Guala 1999; Morgan 2002, 2003; Morrison 2009). On the other hand, a simulation is internally valid if the solutions of the numerical model approach, to the level of desired accuracy, are relevant solutions to the equations of the original mathematical model (Winsberg 2003). The process of internal validation gives legitimacy to the system, consistency between systems, and does not have any direct implications on the reliability of the model in representing phenomena.

⁹ The zero-differential-overlap (ZDO) approximation was the approximation introduced in the initial stages of quantum chemistry to overcome problems in evaluating numerous integrals of three and four centers. This approximation could be addressed at different levels (CNDO, complete neglect of differential overlap; INDO, intermediate neglect of differential overlap; and NDDO, neglect of diatomic differential overlap).

commonly used in *ab initio* simulations, the wave function approach. Likewise, in the choice of the approximation, they strive to retain the fundamental physical traits that are necessary for describing the properties of interest. Furthermore, parametrization seeks to determine the optimal values of the parameters, comparing them with reference data, preferably experimental.¹⁰ For the empirical testing, a great number of molecular systems are selected, the greatest amount possible, all kinds of molecules with different bonding types; and, if possible, some reactions. This is done with a thorough evaluation, numerous hours of calculating and a great deal of patience (Dewar 1985).

As can be seen, the main elements that intervene for the realization of a good semi-empirical simulation are, on the one hand, the approximation that is chosen, and on the other, the efforts placed in the parametrization. In this sense, by the choice of the suitable parameters, these simulations can get even better results than *ab initio* Hartree-Fock or self-consistent field (SCF) calculations, due to the fact that an adequate parametrization can compensate for the partial neglect of electron correlation in *ab initio* SCF theory (Levine 2014). Additionally, the ability to reproduce experimental results that are distinct from those used for determining the parameters is what gives these simulations a greater reliability.

Regarding the second objection we consider that, even though there are some semi-empirical simulations with more specific purposes that can have a lesser inferential content, this does not preclude the fact that there are also many with a high degree of generality. We use generality to refer to the ability to describe distinct types of molecular systems and equally well many of their properties, in the sense of accuracy of results. In fact, an objective is to achieve simulations of general purpose: “One of the major assets of MINDO/3 and MNDO was their demonstrated ability to reproduce *all* ground-state properties of molecules of *all* kinds, including properties and types of molecules not used in parametrizing them” (Dewar 1985, p. 3902-3).¹¹ As can be noted, by 1985 general-purpose semi-empirical simulations had been achieved.¹² Likewise, another issue that must be taken into consideration about the computer simulations is the “range of applicability,” that is, the size of the systems to address (Pople 1998). This is directly related to the computational cost. A lower computational cost broadens the range of applicability. Semi-empirical simulations have shown their merits by achieving results that are close or comparable to those of *ab initio* simulations, with the time requirement for calculations being on a scale of magnitude three times lesser. Furthermore, *ab initio* simulations also have to make some concessions to achieve a broad range of applicability. For example, often, “small empirical corrections” are introduced to correct certain errors produced in using limited basis sets in order to speed up the calculation. The method becomes “slightly empirical” given that the origin of the parameters is partially understood (Pople 1998).

¹⁰ In absence of reliable experimental reference data, accurate theoretical data (e.g., from high-level *ab initio* calculations) are accepted as substitutes.

¹¹ MINDO/3: modified intermediate neglect of differential overlap; MNDO: modified neglect of diatomic overlap.

¹² Properties reproduced by MNDO include heats of formation, molecular geometries, dipole moments, ionization energies, electron affinities, polarizabilities, molecular vibration frequencies, thermodynamic properties, kinetic isotope effects and properties of polymers.

The quest of generality, of a broad range of applicability and of accuracy in the results is intimately related to the computational resources available. For this, the advancements in hardware are one of the key factors that enable the study of complex molecular systems. Among the most significant innovations, the work of the processors in parallel and the combination of the central processing units (CPUs) and the graphics processing units (GPUs) were developments that made huge strides in accelerating the calculation.¹³ Of course, to fully take advantage of these advances, developing effective programs is vital. Running the calculations in parallel in multiple computer centers can greatly reduce the required computational time for the study complex molecular systems; since the problems can be divided in parts so every part can be solved in parallel fashion, simultaneously. Other significant advances in the development of hardware are the sustained growth in the number of transistors in a microprocessor and the increase in their speed, which has resulted in a doubling in the performance of chips approximately every two years.¹⁴

These transformations in hardware revitalized the use of *ab initio* simulations based on wave function, as it made possible the study of larger systems. However, for systems of very many electrons the wave function methods “encounter and are stopped by the ‘exponential wall’” (Kohn 1999, p. 1258). That is, with an increase in the number of electrons comes an exponential increase in the number of configurations and this restricts the use of these methods to molecules with small total number of chemically active electrons.¹⁵ In the density functional theory, the Kohn-Sham equation is a *single-particle* wave function. Thus, the simulations that employ this approach can handle systems which are much more complex with lesser computational demand than those founded on wave function approach. This, in addition to the advances in hardware and in the development of approximations (Becke 1998, 1993a), provided the groundwork for the simulations based on the density functional theory to gain ground against those based on wave function approach. They allow systems to be treated closer to reality, as excitatory states, with a reasonable demand of time and computational resources. However, thinking of molecular systems like DNA molecules, enzyme-substrate interactions, or proteins in their aqueous solution, computational resources are still a limiting factor to treat them with these methods. It is in this

¹³ There are two kinds of parallel electronic computers. One is shared memory, in which the same machine has multiple processing elements, and the other is distributed memory, where multiple machines are linked to form a cluster of processing elements. Both types enable for the execution of calculations concurrently across multiple processing elements (Needham et al. 2016).

¹⁴ According to Gordon Moore’s projection (1965, 1975), approximately every two years the number of transistors in a microprocessor will double, until reaching the limit of what is physically possible. Likewise, David House predicted that it would take eighteen months for the same increase, taking into consideration their speed as well. “... the landscape of computer hardware is constantly changing, but not all changes are equally disruptive to scientific software programmers. After approximately three decades of continuous performance improvement from steadily increasing clock frequencies and concomitant decrease in semiconductor feature size, microprocessor design underwent a fundamental shift toward parallelism at the beginning of the last decade. Having reached the limits of frequency-scaling due to power constraints, node-level parallelism is now the basis for keeping pace with Moore’s law.” (DePrince III et al. 2016, p. 279-80).

¹⁵ Furthermore, Kohn highlight that, even if there were no computational limits, other limiting factors emerge when the number of electrons increases and correcting for these physical factors, such as relativistic or radioactive, would be limiting (Kohn 1999).

area where semi-empirical simulations continue to have a broad range of application for researching these systems that are so interesting to the chemistry, going to the limits of what is possible to investigate. The value of the semi-empirical computer simulations is more evident in very many atom systems, to calculate effects such as those of transference and long-range charge fluctuation which can only be captured by computational quantum chemistry approach which cover the entire system; in the study of reaction mechanisms in large biomolecules, non-covalent interactions, electronically excited states, and hydrogen bonds; in medical chemistry and drug design for determining structure-property and structure-activity relationships (Thiel 2005, 2013).¹⁶ Likewise, for taking on very large systems, there are also multi-scale computer simulations (Winsberg 2010). In many cases, they are built using a semi-empirical method to treat the active part of the reaction from a quantum approach, and the rest of the molecular system with a classical approach of molecular dynamic (MD), which has a lesser computational demand.

As we have shown, the semi-empirical simulations overcame their main objections. Improving the approximations together with the mentioned computational advances, achieved a high degree of generality with good level of accuracy, and with a broader range of applicability, since regarding to speed they are better by three orders of magnitude compared to that of DFT.

4. Final considerations

We believe that the analysis of computer simulations demands a dynamic perspective, given that the elements which are combined for its development are constantly changing. We refer to the fact that the accuracy of the instruments is increased and experimental techniques are perfected providing improved values for parametrization and empirical testing. Advancements are constantly occurring in both the capacity and speed of calculations in computation, thus things that were once untreatable are now able to be studied. Likewise, new mathematical and algorithmic developments are being incorporated, resulting in greater reliability and accuracy of the results compared with experimentation. Thus, ab initio and DFT methods can accurately address larger systems, however, there are many interesting molecular systems that exceed their limits and can only be studied using semi-empirical methods.

The broad spectrum of simulation packages available in the market shows that the semi-empirical methods are well established. Due to improvements in the approximations and parametrizations, they achieve a high degree of generality together with high accuracy, and with the lowest computational cost within the methodologies of quantum chemistry. This makes these simulations highly useful, expanding the frontiers of investigation on complex systems at the molecular level.

As a reflection on the distinction among the methods, we see that the ab initio and DFT attempt to respond to the theoretical rigor, mainly employing top-down strategies; but they cannot avoid the use of some bottom-up strategies to enable the calculation. Whereas, semi-empirical methods in trying to

¹⁶ With semi-empirical simulations, one can deal with systems containing thousands of atoms. An improvement of the conventional semi-empirical method allows for executing energy calculations with high accuracy on molecules that have thousands of atoms: polyglycine chains of 20,000 atoms; a group of water molecules with more than 1,800 atoms, and nucleic acids of more than 6,300 atoms (Daniels et al. 1997; Wu et al. 2016).

address larger systems resort more to the use of bottom-up strategies to speed-up the calculation. Can it be said that any of these methods are misguided in the strategy of research that they employ? We don't believe so; they are only different strategies associated to our ways to knowing. However, in the context of computer simulations, both strategies are combined in various proportions to enable new knowledge, which confers a singular character upon the computer simulations as a research methodology.

References

Becke, A. D.: Density-functional thermochemistry. IV. A new dynamical correlation functional and implications for exact-exchange mixing. *Journal of Chemical Physics* 104 (3), 1040 - 1046 (1996)

Becke, A. D.: A new mixing of Hartree-Fock and local density-functional theories. *Journal of Chemical Physics* 98 (2), 1372 – 1377 (1993a)

Becke, A. D.: Density-functional thermochemistry. III. The role of exact Exchange. *Journal of Chemical Physics* 98 (7), 5648 - 5652 (1993b)

Becke, A. D.: Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A* 38 (6), 3098-2100 (1988).

Car, R., Parrinello, M.: Unified Approach for Molecular Dynamics and Density-Functional Theory. *Physical Review Letters* 55 (22), 2471-2474 (1985)

Cartwright N.: Quantum Halmiltonians and the BCS model of superconductivity. In: Morgan, M. S., Morrison, M. (eds.) *Models as mediators: perspectives on natural and social science*, pp. 241-281. Cambridge University Press, Cambridge (1999).

Daniels, A. D., Millam, J. M., Scuseria, G. E.: Semiempirical methods with conjugate gradient density matrix search to replace diagonalization for molecular systems containing thousands of atoms. *The Journal of Chemical Physics* 107 (2), 425–431 (1997)

DePrince III, A. E., Hammond, J. R., Sherrill, C. D. Iterative Coupled-Cluster Methods on Graphics Processing Units. In: Walker, R. C., Götz, A. W. (eds.) *Electronic structure calculations on graphics processing units: from quantum chemistry to condensed matter physics*, pp. 279-280. John Wiley & Sons, Ltd., Chichester (2016)

Dewar, M. J. S.: *A Semiempirical Life*. DC: American Chemical Society, Washington (1992)

Dewar, M. J. S., Zoebisch, E. G., Healy, E. F., Stewart, J. J. P.: AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* 107, 3902-3909 (1985)

Dirac, P.: Quantum Mechanics of Many-Electron Systems. *Proceedings of the Royal Society of London A*. 123, 714-733 (1929)

Fortin, S., Martinez Gonzalez J. C.: La relación entre química y física: isomerismo óptico y la paradoja de Hund. *Revista colombiana de filosofía* 13, 199-224 (2013)

- Freed, K. F.: Building A Bridge Between Ab Initio and Semiempirical Theories of Molecular Electronic Structure. In: Calais, J. L., Kryachko, E. S. (eds.) *Structure and Dynamics of Atoms and Molecules: Conceptual Trends*, pp. 25–67. Springer Netherlands, Dordrecht (1995)
- Gavroglu, K., Simões, A.: *Neither Physics nor chemistry. A history of quantum chemistry*. MIT Press, Cambridge (2012)
- Gavroglu, K.: Philosophical issues in the history of chemistry. *Synthese*, 111 (3), 283–304 (1997)
- Gavroglu, K., Simões, A. The Americans, the Germans, and the beginnings of quantum chemistry: the confluence of diverging traditions. *Historical Studies in the Physical Sciences* 25 (1), 47 - 110 (1994)
- Gill, P. M. W.: Density Functional Theory (DFT), Hartree-Fock (HF), and the Self-consistent Field. In: von Ragué Schleyer, P. (ed.) *Encyclopedia of Computational Chemistry Vol. 1*, pp. 678–689. John Wiley & Sons, Chichester (1998)
- Grant, F.: Divergence, diagnostic, and a dichotomy of methods. In: Scerri, E., Fisher, G. (eds.) *Essays in the Philosophy of Chemistry*, pp. 306-331. Oxford University Press., New York (2016)
- Guala, F.: The problem of external validity (or “parallelism”) in experimental economics. *Social science information* 38 (4), 555-573 (1999)
- Hehre, W. J., Stewart, R. F., Pople J. A.: Self-Consistent Molecular Orbital Methods. I. Use of Gaussian Expansions of Slater Type Atomic Orbitals. *Chem. Phys.* 51, 2657-2664 (1969)
- Heitler, W., London, F.: Wechselwirkung neutraler Atome und homopolare Bindung nach der Quantenmechanik. *Zeitschrift für Physik* 44, 455-472 (1927)
- Hohenberg, P., Kohn, W.: Inhomogeneous Electron Gas. *Physical Review* 136 (3B), 864–871 (1964)
- Jammer, M.: *The Conceptual Development of Quantum Mechanics*. McGraw-Hill, New York (1966)
- Kohn, W. (1999): Electronic structure of matter—wave functions and density functionals. *Nobel Lectures Chemistry 1996 – 2000*. World Scientific Publishing Co., Singapur (2003)
- Kohn, W., Sham, L. J.: Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review* 140 (4A), 1133–1138 (1965)
- Levine, I.: *Quantum Chemistry*. Pearson Education, Inc., New York (2014)
- Lombardi, O., Castagnino, M.: Matters are not so clear on the physical side. *Foundations of Chemistry*, 12, 159-166 (2010)
- Lowe, J. P.: *Quantum Chemistry*. Academic Press., New York (1993)
- Moore, G.: Progress in Digital Integrated Electronics. IEEE, IEDM (International Electron Devices Meeting) *Tech Digest* 3, 11-13 (1975)

- Moore, G.: Cramming More Components onto Integrated Circuits. *Electronics Magazine* 38 (8), 114-117 (1965)
- Morgan, M. S.: Experiments without material intervention: model experiments, virtual experiments and virtually experiments. In: Radder, H. (ed.) *The Philosophy of Scientific Experimentation*, pp. 216-235. University of Pittsburgh Press., Pittsburgh, PA (2003)
- Morgan, M. S.: Model experiments and models in experiments. In: Magnani, L., Nersessian, N. J. (eds.) *Model-Based Reasoning: Science, Technology, Values*. International Conference on Model-Based Reasoning: Scientific Discovery, Technological Innovation, Values, pp. 41-58. Kluwer Academic/Plenum, London (2002)
- Morrison, M.: Fictions, representations, and reality. In: Suárez, M. (ed.), *Fictions in Science: Philosophical Essays on Modeling and Idealization*, pp. 4-110. Routledge, London (2009)
- Needham, P., Götz, A. W., Walker, R. C.: Why graphics processing units. In: Walker, R. C., Götz, A. W. (eds.) *Electronic structure calculations on graphics processing units: from quantum chemistry to condensed matter physics*, pp.1-19. John Wiley & Sons, Ltd., Chichester (2016)
- Park, B. S.: Between Accuracy and Manageability: Computational Imperatives in Quantum Chemistry. *Historical Studies in the Natural Sciences* 39 (1), 32–62 (2009)
- Pople, J. A. (1998): Quantum Chemical Models. *Nobel Lectures Chemistry 1996 – 2000*. World Scientific Publishing Co., Singapur (2003)
- Ramsey, J.: Of Parameters and Principles: Producing Theory in Twentieth Century Physics and Chemistry. *Studies in History and Philosophy of Science Part B* 31 (4), 549–567 (2000)
- Ramsey, J. L.: Between the Fundamental and the Phenomenological: The Challenge of “Semi-Empirical” Methods. *Philosophy of Science* 64 (4), 627–653 (1997)
- Scerri, E.: Just How Ab Initio is Ab Initio Quantum Chemistry? *Foundations of Chemistry*. 6, 93–116 (2004a)
- Scerri, E.: Principles and Parameters in Physics and Chemistry. *Philosophy of Science* 71 (5), 1082–1094 (2004b)
- Schrödinger, E.: On the Relation between the Quantum Mechanics of Heisenberg, Born, and Jordan, and that of Schrödinger. In: Schrödinger, E. (ed.) *Collected Papers on Wave Mechanics*, pp. 45-61. Chelsea Publishing Company, New York (1982). Translated from the original: Über das Verhältnis der Heisenberg-Born_Jordanschen Quantenmechanik zu der Meinen. *Annalen der Physik* 4 (79): 734-756 (1926)

- Segall, M. D., Lindan, P. J. D., Probert, M. J., Pickard, C. J., Hasnip, P. J., Clark, S. J., Payne, M. C.: First-principles simulation: ideas, illustrations and the CASTEP code. *Journal of Physics: Condensed Matter* 14 (11), 2717-2744 (2002)
- Thiel, W.: Semiempirical quantum-chemical methods. *WIREs Comput Mol Sci* (2013). doi: 10.1002/wcms.1161
- Thiel, W.: Semiempirical quantum-chemical methods in computational chemistry. In: Frenking, G., Kim, K. S. & Scuseria, G. E. (eds.) *Theory and Applications of Computational Chemistry*, pp. 559-580. Elsevier, Amsterdam (2005)
- Tully, J. C.: Perspective on "Zur Quantentheorie der Molekeln" Born M, Oppenheimer R (1927) *Ann Phys.* 84: 457. *Theor Chem Acc* 103, 173-176 (2000)
- Van Brakel, J.: *Philosophy of Chemistry. Between the Manifest and the Scientific Image*. Leuven University Press, Leuven (2000)
- Van Vleck, J. H., Sherman, A.: The Quantum Theory of Valence. *Rev. Mod. Phys.* 7, 167-228 (1935)
- Winsberg, E.: *Science in the age of computer simulation*. University of Chicago Press, Chicago (2010)
- Winsberg, E.: Simulated experiments: Methodology for a virtual world. *Philosophy of science* 70 (1), 105-125 (2003)
- Woolley, R. G.: Must a Molecule Have a Shape? *Journal of the American Chemical Society* 100, 1073-8, (1978)
- Wu, X., Koslowski, A., Thiel, W.: Semiempirical Quantum Chemistry. In: Walker, R. C., Götz, A. W. (eds.) *Electronic structure calculations on graphics processing units: from quantum chemistry to condensed matter physics*, pp. 239-253. John Wiley & Sons, Ltd., Chichester (2016)