Computational chemistry as an analytical tool: thermochemical examples in atmospheric chemistry

Oscar N. Ventura* and Martina Kieninger

MTC-Lab, DEQUIFIM, Facultad de Química, Universidad de la República, CC 1157, Montevideo 11800, Uruguay

Abstract: Computational chemistry is one area of research which has evolved to the point in which it can be used profitably in cooperation with experimental, traditional chemistry. The ever increasing power of modern computers coupled with the development of new theoretical approaches can be used for accurate and precise prediction of molecular properties. One of the most useful applications of computational chemistry is the prediction of properties difficult or impossible to measure experimentally. In some cases, computational chemistry can be used to predict certain properties with accuracy that rivals or even excels that of the best experiments available. Thus, nowadays, the computer can be looked at as another analytical instrument, which can be used like an spectrometer, for instance, to search for solutions to difficult chemical problems. In this paper we introduce in a simplified manner the way in which computational chemistry can be used for that purpose. Some examples are given, taken from our own work on the stratospheric chemistry of substances that may destroy the ozone layer.

INTRODUCTION

Chemistry is an experimental science. This is a truism which nobody currently challenges. However, there are different kind of experiments, from traditional wet chemistry to laser spectroscopy in crossed molecular beams. And, as of recently, computational chemistry.

Computational chemistry (ref. 1) is a branch of physical chemistry (ref. 2). It aims to predict the same properties that can be measured in the laboratory (different types of energies, molecular structures, spectroscopic characteristics, reaction mechanisms, etc) through the use of models of the real world. To that end, physical theories (classical and quantum mechanics) are applied to the specific realm of chemistry, providing mathematical recipes which link observed properties (like, for instance, the geometry of a molecule as determined by means of X-ray spectroscopy) with specific features of the models (for instance, a minimum on a potential energy surface). These models (refs. 3,4) can be coded into computer programs which were developed in earlier times only for mainframe computers, but are now even available in personal computers (ref. 5).

Although nobody seriously discusses that theories are what transform common data gathering into a well-organized scientific body, the acceptance of computational chemistry methods on equal foot to other laboratory experimentation methods has needed time. Different reasons exist for that, not the least important of them being the over-enthusiastic expectations about the possibilities of theoretical methods as applied to the study of molecules. On the other side, traditional chemists have been slow in accepting that one could obtain accurate models of the real world from physics and mathematics, without a hands-on commitment to the laboratory bench. Finally, the inherent mathematical complexity of this subject, contained into the already complex field of physical chemistry, has contributed to the lack of serious consideration given to the field until recently. The last 15 years, however, witnessed the advent of powerful computers, accessible even to modest laboratories, the popularization of personal computers, its transformation from word processors to powerful calculating and all-purpose machines, and the development of computer programs powerful enough on the one hand, and user-friendly on the other. All these facts contributed to the fact that computational methods are accepted and employed widely nowadays (between 10 and 20% of the papers published in 1997 by the Journal of the American Chemical Society, the best journal in all fields of chemistry in the world, have contained some theoretical or computational approach).

In the following, we will first give a very summary description of what is computational chemistry and how it does work. Secondly, we will describe, a little more in depth, the new methods developed in the last ten years. Finally, we will give examples of how computational chemistry can be used as an analytical tool for determining some properties of systems, including especial cases in which its use helps to solve experimental problems.

**COMPUTATIONAL CHEMISTRY**

Computational chemistry is mainly based on the physical theory known as Quantum Mechanics (ref. 6). This theory, well tested and fully accepted at present, allows to write an equation for each physical system, known as the Schrödinger equation (ES). The ES says basically that, given the number and type of particles in the system and knowing their way of interacting, we can obtain the total energy of the system solving a quite complicated partial differential equation. Quantum mechanics says also that, within the frame of one of the possible realizations of the theory, any physical system (including, for instance, atoms and molecules) can be described by a function, known as the wave function, which contains all the information one can obtain on the given system. From this function, through convenient mathematical manipulations, one can obtain all the properties of the system, including those that can be measured in the laboratory. Moreover, quantum mechanics also says that the systems exist in certain states (called stationary) which energy can be determined from the ES. Systems (including atoms and molecules) do not change continuously their energy, as they do in classical mechanics, but in jumps, absorbing or emitting quanta of energy, thus the name of the theory.

The writing of the ES for a system of nuclei and electrons (which we identify as a molecular system) does not make any distinction among the particles. However, a clever simplification was introduced by Born and Oppenheimer, who, based on the fact that nuclei are much more massive than electrons, postulated that one can solve the corresponding equation obtained for the electrons after fixing the nuclei, independently of the movement of the nuclei. This approximation, known nowadays with the name of Born-Oppenheimer approximation (BOA), gives rise to the electronic energy, a function of the position of the nuclei within a molecule, which varies when the position of the nuclei vary, and which is know in Quantum Chemistry with the name of potential energy surface (PES).

Quantum Chemistry is basically a collection of methods for solving the ES under the BOA, the calculation of the PES or only some portion of it, and the computing of the properties of the molecules for the different position of the nuclei as dictated by the point on the PES. In fact, one of the most important operations within Quantum Chemistry is the optimization of nuclear geometry. This amounts to determine the points on the PES where the energy of the system is minimal (most negative) implying that these especial arrangements of nuclei have extra stability. In fact, the nuclear arrangement at the points of the PES where the energy is a minimum correspond to stable species (for instance, different isomers of a molecule) or different points in a reaction mechanism (reactants, products, transition states). The study of the wave function at the minimum-energy structures allow to determine the molecular orbitals and to build up theories of reactivity (ref. 4). Moreover, special integrals over the wave function allow to determine the value of properties like electrostatic potentials, charges, etc, which are used in different ways to explain chemical structure and reactivity.

The solution of the ES under the BOA is a very complicated affair and we can not describe it here. Suffice it to say that certain methods, known with the name of Hartree-Fock and post-Hartree-Fock, have become standard in the field, and their use made relatively easy by the existence of programs that started to be developed around 1970. Only from about 1985 the computers have been powerful enough to tackle real chemical problems, but still until few years ago the applications were somehow confined to small systems, because the difficulty of applying these methods runs parallel to the size of the system. What is more, the difficulty increases as a power (usually 4 but it can be up to 8) of the size of the system, so that sometimes, even adding one more atom to the system turns a solvable problem into an unsolvable one. Fortunately, in the last 10 years some new approaches have appeared, known as density functional methods, which can be applied to much larger systems with good precision and accuracy. I will describe them in the following section.
Most of the general theory of DFT can be found in the well-known books of Parr and Yang (ref. 7) and Dreizler and Gross (ref. 8). Most practical electronic structure DFT calculations involve solving the Kohn-Sham equations (ref. 9). In these equations, the unknown quantity is the exchange-correlation energy:

$$E_{xc} = T - T_s + V_{ee} - J$$

where $T$ is the exact kinetic energy, $T_s$ is the Kohn-Sham approximate kinetic energy (i.e. non-interacting), $V_{ee}$ is the exact Coulomb repulsion between electrons, and $J$ is the classical Coulomb energy associated with the given density. Modern density functional calculations are all based on obtaining a suitable approximate expression for the exchange-correlation energy $E_{xc}$ (and/or its functional derivative, i.e. the exchange-correlation potential $v_{xc}$).

The oldest version of DFT is the so-called Local Spin Density (LSD) approximation, where the expression for $E_{xc}$ is derived from the exchange-correlation energy per particle for a uniform electron gas (otherwise called jellium). In this case, the energy depends only on the density itself. But modern density functional theory starts really developing with the advent of Generalized Gradient Approximations (GGA) (refs. 10-17), mostly associated with the names of Perdew, Becke, Yang and Parr. In this case, the exchange-correlation potential depends not only on the density itself but also on its gradient.

Perdew 1991 form of the exchange-correlation functional is considered to be the best on formal grounds (see the papers by Ernzerhof (refs. 18-19) and references therein Perdew-Wang 1991 (PW91) GGA derives from a model for the exchange-correlation hole. For instance, it has been shown that PW91 models accurately the change in the exchange hole of the H2 molecule upon atomization. However, the most popular exchange potential nowadays (for reasons to be explained later) is Becke 1988 (ref. 13). In this case, the LSD exchange was complemented with a term depending on the gradient of the density designed in such a way that the exchange energy density has the correct -1/r asymptotic behavior. Becke, together with Roussel (ref. 20), developed another potential which depends also on the Laplacian (i.e. second derivatives) of the density, but this one has been much less used (ref. 21). An interesting paper on the comparison of different GGAs was published recently by Perdew and Burke (ref. 22).

The reason why Becke exchange functional became the most used one in computational studies for the moment being, has to be found in the successful attempts made by Becke to mix exact Hartree-Fock exchange with DFT in the expression for the exchange-correlation potential. In 1992 (published 1993) Becke (ref. 23) called the attention to the so-called adiabatic connection formula (see references [18,19] of Becke's paper (ref. 23)). This formula leads from the non-interacting Kohn-Sham reference system to the fully interacting real system. Becke proposed in this paper a linear interpolation which ended up in an admixture of exact exchange and LSD exchange in equal proportions. This method is called nowadays half and half method. Also in this paper appeared for the first time the use of the molecules included in the G1 set of Pople (a set of molecules used to derive the G1 and, later, G2 model chemistries) (ref. 24)) for testing purposes. After that, the G2 set has been much used (and abused) by proponents of new potentials.

The analysis that Becke did of the adiabatic connection formula and the introduction of the exact exchange potential as one component of the exchange-correlation energy, opened the field of the Adiabatically Connected Functionals (ACM in popular nomenclature). Becke again, in a paper sent only two months later, introduced the formula (ref 25):

$$E_{xc} = E_{xc}^{LSDA} + a_x(E_{HF}^{xc} - E_{LSDA}^{xc}) + a_z\Delta E_{x}^{B88} + a_c\Delta E_{c}^{PW91}$$

where $X$ stands for exchange, $C$ for correlation and the rest of the symbols are self-explaining. The $a_x$, $a_z$ and $a_c$ parameters were adjusted for reproducing the data in the G1 database. This method was immediately known as Becke three-parameters method, or B3. Notice that Becke used the PW91 method for the correlation part of the exchange-correlation formula. Almost immediately this method was coded in the immensely popular suite of programs Gaussian92, but with the Lee-Yang-Parr (ref. 14) formula for correlation, instead of the Perdew-Wang. Thus, the B3LYP method was born with somehow inappropriate parameters. Later on, also the B3PW91 method was coded, for which the parameters determined by Becke were appropriate. Becke's 1993 paper (ref. 25) showed that his latest functional performed significantly better than GGAs, an observation that has been the base of most of the studies performed in the last years.
SOME APPLICATIONS

There is some interest in atmospheric chemistry in molecules containing fluorine-oxygen bonds. Such molecules represent a difficult subject for normally reliable ab initio methods (refs. 26-32). These problems have been frequently addressed (refs. 32, 33) by the use of singles and doubles plus perturbational triples coupled-cluster theory, CCSD(T) (ref. 34). Regrettably, the computational cost of such theory is so high, that it can be applied routinely only to very small molecules. Therefore, this is fertile soil for applying DFT methods. Different authors reported different degree of success concerning DFT calculations on several F-O containing compounds. Some authors concluded that DFT gives wrong results when applied to compounds containing these bonds (refs. 30, 35-38). Other authors, however, report surprisingly good results of DFT methods in the study of thermodynamics of reactions containing such compounds (refs. 39, 40). It has been suggested, based on some of these calculations (ref. 40), that the accepted enthalpy of formation of FOOF is wrong and should be revised. We will describe in the following how we did arrive to that conclusion and what are the best results produced. More detailed accounts of the present facts can be found in the original works (ref. 39-45).

Our main example in this section concerns to the determination of the use of the computer as a calorimeter to determine the enthalpies of several reactions involving fluorine oxides and peroxides (FO, FOH, FOO, FOF, FOOF) important in the stratosphere, in processes related with rocket propulsion and nuclear waste. In our first paper in 1995 (ref. 39), we did a thorough study of several reactions, from which we deduced the enthalpies of formation at 0K of some of these species, as shown in Table 1. Shortly afterwards, there appeared some experimentally recommended data in the J. Phys. Chem. Ref. Data, which is the best journal in the field of standards and recommended magnitudes for thermochemistry. Since these values showed a marked discrepancy with ours in the case of FOOF, and only in this case, we performed even more accurate calculations (ref. 40) which showed our initial estimations to be right. In all cases except FOOF the agreement between theory and experiment is excellent, and thus the discrepancy was harder to explain. It turned out that all the information existent on FOOF is coming from only one experiment done in 1959. All later papers were either quoting this original one, or refining some of its conclusions. Because of the accuracy of the calculations performed, we suggested that this experiment was presumably wrong. As a consequence, the experimentalists are now reviewing the experimental evidence, in order to (probably) correct the estimation.

<table>
<thead>
<tr>
<th>Species</th>
<th>Calculated in 1995</th>
<th>Recommended experimental</th>
<th>Calculated in 1997</th>
</tr>
</thead>
<tbody>
<tr>
<td>FO(g)</td>
<td>+25.4 ± 1.0</td>
<td>+25.8 ± 2.4</td>
<td>+25.2 ± 1.0</td>
</tr>
<tr>
<td>FOH(g)</td>
<td>-19.0 ± 1.0</td>
<td>-19.1 ± 1.0</td>
<td></td>
</tr>
<tr>
<td>FOO(g)</td>
<td>+7.2 ± 2.0</td>
<td>+6.5 ± 0.5</td>
<td>+7.0 ± 1.0</td>
</tr>
<tr>
<td>FOF(g)</td>
<td>+6.4 ± 0.5</td>
<td>+6.2 ± 1.0</td>
<td></td>
</tr>
<tr>
<td>FOOF(g)</td>
<td>+8.2 ± 1.0</td>
<td>+5.5 ± 0.5</td>
<td>+9.0 ± 1.0</td>
</tr>
</tbody>
</table>

The second example is a determination of enthalpies of formation using different theoretical methods (refs. 44, 45). In Table 2 it is shown how one can determine very accurate values, employing the DFT methods pointed out above, and how these values can be even better than those obtained with the much more effort-demanding methods employed before, CCSD(T). Standard statistical thermodynamics procedures were employed to calculate enthalpies of reaction, \( \Delta H^\circ(298.15) \), and formation \( \Delta H^\circ(298.15) \). The enthalpy of formation of the atoms H, N, O and F were obtained as half the enthalpy of reaction for the atomization of the corresponding homonuclear diatomic molecule. Enthalpies of formation of the three diatomic radical oxides HO, NO and FO, as well as of the closed-shell triatomics HOH, FOH and FOF were obtained by three different procedures. In the first place, the isodesmic reactions

\[
\begin{align*}
    \text{FO + HOH} & \rightarrow \text{FOH} + \text{HO} + \Delta H^\circ[1] \\
    \text{HO + FOF} & \rightarrow \text{FOH} + \text{FO} + \Delta H^\circ[2]
\end{align*}
\]

© 1998 IUPAC, Pure & Applied Chemistry 70, 2301–2307
TABLE 2  Enthalpies of formation at 298.15 K of some first-row atoms and molecules calculated with the procedures explained in the text, using the large, noncontracted basis set a

<table>
<thead>
<tr>
<th></th>
<th>Exp. d</th>
<th>B3PW91</th>
<th>B3LYP</th>
<th>CCSD(T)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>H</td>
<td>ISO²</td>
<td>DIA³</td>
<td>At⁴</td>
</tr>
<tr>
<td>H</td>
<td>52.10</td>
<td>70.99</td>
<td>51.24</td>
<td>50.63</td>
</tr>
<tr>
<td>N</td>
<td>112.97</td>
<td>110.45</td>
<td>112.43</td>
<td>108.32</td>
</tr>
<tr>
<td>O</td>
<td>59.56</td>
<td>60.68</td>
<td>59.99</td>
<td>56.22</td>
</tr>
<tr>
<td>F</td>
<td>18.97</td>
<td>16.94</td>
<td>17.01</td>
<td>16.71</td>
</tr>
<tr>
<td>rmse b</td>
<td>4.80</td>
<td>0.56</td>
<td>1.58</td>
<td></td>
</tr>
<tr>
<td>Max c</td>
<td>18.89</td>
<td>1.96</td>
<td>4.65</td>
<td></td>
</tr>
</tbody>
</table>

rmse b in kcal/mol, c root mean square error, d maximum deviation from the experimental values, e taken from the JANAF Thermochemical Data Tables, see text, f using isodesmic reactions 1 (first entry) or 2 (second entry), g using the reaction of formation from homonuclear diatomics, h using the reactions of atomization were employed together with the experimental enthalpies of formation for three of the molecules involved, to obtain the enthalpy of formation of the fourth. In the second procedure, the enthalpy of formation of the molecules was obtained calculating the $\Delta H_f(298.15)$ of reactions involving the homonuclear diatomic molecules. For instance, the enthalpy of formation of NO was obtained as one-half the $\Delta H_f(298.15)$ of the reaction

$$\text{NO} + \text{NO} \rightarrow \text{N}_2 + \text{O}_2$$

No experimental value is involved in this way of calculating the enthalpies of formation, except for the definition of the enthalpies of formation of the elements in their natural state (i.e., the homonuclear diatomic molecules in the gaseous state at room temperature involved in these calculations) as zero. The third procedure employed for the calculation of the enthalpies of formation of the molecules resorts to the atomization reactions supplemented with the experimental enthalpies of formation of the atoms involved.

Provided that the isodesmic reactions are used, B3PW91 (errors: 0.30 and 0.19 kcal/mol respectively) is marginally better than B3LYP (errors: 0.45 and 0.24 kcal/mol) and DFT is better than CCSD(T), which exhibits errors of 1.96 and 1.12 kcal/mol. Nonetheless, the error is within so-called “chemical accuracy” (understood usually of an error smaller than 2 kcal/mol). The second procedure, i.e. using the reactions of formation from the homonuclear diatomic molecules, gives worse results than with the isodesmic reactions for DFT, with maximum errors of up to 3.4 and 3.7 kcal/mol for B3PW91 and B3LYP, respectively. Contrary to accepted belief, CCSD(T) thermochemical results are better in this case than when using the isodesmic reactions. In fact, the maximum error here (for the HO molecule) is only 1.4 kcal/mol (half that of DFT). Furthermore, the enthalpies of formation of the other four molecules are within 0.4 kcal/mol of the experimental values, and the agreement with experiment is better for the larger molecules than for the smaller ones.

© 1998 IUPAC, Pure & Applied Chemistry 70, 2301-2307
CONCLUSIONS
We have briefly described computational chemistry and how it can be used to transform the computer into an analytical tool, showing as an example its performance as a calorimeter for determining the enthalpies of reaction and formation of certain species important in the atmosphere. We have shown that the precision and accuracy of the methods rivals that of the experimental setup and can conveniently be employed, on the one hand, to pinpoint errors in the experimental determinations or, in the other, as one more weapon available to the chemist for exploring the real world. We hope thus to encourage the collaboration between experimentalists and computational chemists in the explorations of problems where the computer can be used as routinely as a spectrometer for measuring/calculating properties of molecules.

ACKNOWLEDGMENTS
We gratefully acknowledge the support to our work by the following agencies: Commission of the European Communities (European Union), Inter American Bank for Development (IABD/BID), National Council for Scientific and Technological Research (CONICYT), CSIC and Alexander von Humboldt-Stiftung.

REFERENCES
24 See refs. [7-9] in the previous paper.
29 P. N. Noble, P. N. J. Phys. Chem. 95, 4695 (1991)
Computational chemistry as an analytical tool

42 O. N. Ventura. TMMeC 1, 58-65 (1997).