UNIVERSIDAD DE OVIEDO

FACULTAD DE QUÍMICA

BASIS FUNCTIONS IN COMPUTATIONAL CHEMISTRY: ANALYSIS OF THE SECOND PERIOD DIATOMIC HYDRIDES

(Physical Chemistry)

BACHELOR THESIS IN CHEMISTRY

Rosa Mary López Álvarez

Oviedo, July, 2021

THESIS CONTENTS

0. OBJECTIVE

The development of this thesis is carried out from a theoretical chemistry point of view, applying the field of Physical Chemistry through its branches of Computational Chemistry and Quantum Chemistry. The use of computers as a tool is a key aspect in all the experimental work to obtain the results to solve the proposed chemical systems based on the principles and theoretical foundations studied, such as the Schrödinger equation, which is one of the most important equations of Quantum Chemistry to obtain information about the behavior of an electron bound to a nucleus.

The objective of this work is the study and comparison of some diatomic hydride molecules and several polyatomic molecules corresponding to the second period. For this purpose, several mathematical expressions are used and the Hartree-Fock and Density Functional theories are applied. This computational thesis, use some free software programs. And the three main programs are: ORCA, for the electronic structure part, Avogadro, to visualize the molecules and AIMALL as a more sophisticated program to analyze the results obtained for each of the molecules.

1. INTRODUCTION

Chemistry is the science that deals with the properties, construction and transformation of molecules and, within this huge field, there is a more theoretical part, the Computational Chemistry, as part of it. This focuses on obtaining relevant results by creating efficient mathematical approximations and computer programs that can be relied upon depending on the desired accuracy and the nature of the system [1].

One of the purposes of using the Computational and Theoretical branches of chemistry is to solve the Schrödinger equation for the wave function which is not an exact value and which contains all the data that give information about a system of molecules. For this purpose, ab initio methods are used to generate the solutions of the Schrödinger equation. These solutions do not include any empirical or semi-empirical parameters and are obtained directly from theoretical principles, without any reference to experimental data.

One of the most widely used ab initio methods is the Hartree-Fock (HF) theory. This is an independent particle model that replaces instantaneous interactions by average interactions. Thus, each electron is described as an orbital and when all the orbitals are obtained, a determinant is constructed with all of them, the total wave function. The wave function is obtained by a simple Slater determinant, which the Hartree-Fock method creates and is composed of orbitals of the occupied space that are approximated as a linear combination of basis functions. Mathematical functions (basis functions) are used to expand the molecular orbitals and the appropriate choice of these will lead to more or less accurate results. In addition, basis sets can be optimized by applying the variational principle for the total energy.

Density functional theory (DFT) can also be used, which takes into account the correlation between electrons and is therefore a more sophisticated method than HF theory.

2. THEORETICAL FOUNDATIONS

One of the main objectives of Computational Chemistry is to obtain solutions for the Schrödinger equation and more specifically for the time-independent molecular Schrödinger equation, which is written in operator form as

$$H\Psi = E\Psi \, , \tag{1}$$

where H is the Hamiltonian operator that contains the potential and kinetic energies of electrons and atomic nuclei in a system. And E is the total energy of the system. Finally, the wave function Ψ depends on the coordinates of the nuclei and electrons and gives all the information about the system.

The treatment of the Schrödinger equation is difficult and exact solutions cannot be obtained, except for single-electron systems. If more complicated systems are analyzed, it will be necessary to introduce approximations. They are useful to obtain a reduced and adequate version of the Schrödinger equation that can be solved, but they decrease the accuracy of the method. Thus, there are two different types of methods: semi-empirical methods, which can be used when empirical data are available, and ab initio methods, which do not require the use of experimental data.

In practice, for all the molecules studied, only the Hartree-Fock (HF) theory is applied, where the approximate energies obtained from a variational calculation must be equal to or greater than the exact energy and a Hartree-Fock limit value appears. Some of the approximations that the HF method employs to arrive at the solution of the equation are: Born-Oppenheimer approximation, the use of Slater's determinant to write the wave function, approximation methods, the assumption of the variational solution as a linear combination of a finite number of basis functions or the avoidance of relativistic effects.

2.1. The Born – Oppenheimer approximation

The Born-Oppenheimer approximation is of great importance when solving the Schrödinger equation since it allows to determine the electronic part for the nuclear positions considering them as parameters and disregards the coupling between the electronic motion and the nuclei. Then, the motion of the electrons is separated from that of the nuclei and the latter is omitted due to the assumption of a much larger mass for an atomic nucleus than for an electron. Thus, the electrons move much faster and this makes it possible to say that the nuclei are almost fixed with respect to the electrons [2].

In practice, the approximation is incorporated into the Hamiltonian operator which can be written in terms of the potential and kinetic energies of electrons (*e*) and nuclei (*n*), the particles in a system. Therefore, the non-relativistic Hamiltonian is expressed as

$$H = T_e(r) + T_n(R) + V_{ne}(r,R) + V_{nn}(R) + V_{ee}(r) \ ,$$

(2)

which includes the particles collected as a set of electronic coordinates (r) and as a set of nuclear coordinates (R). This operator has five contributions to the total energy of the system: the kinetic energies of the electron (T_e) and the nucleus (T_n), the attraction of the electrons to the nucleus (V_{en}) and the internuclear (V_{nn}) and interelectronic (V_{ee}) repulsions.

The Hamiltonian operator is simplified by following the Born-Oppenheimer approximation which does not consider the motion of the nuclei and allows the T_n term to be neglected and the V_{nn} term to be considered constant. In the end, since the operator contains only the electronic part, it will be called the electronic Hamiltonian operator

$$H_{el} = T_e(r) + V_{ne}(R,r) + V_{ee}(r) + V_{nn}(R) \ ,$$

(3)

and involves writing the electronic Schrödinger equation as

$$H_{el}\Psi(r;R) = E_{el}\Psi(r;R) \ .$$

(4)

2.2. Approximate methods

The manipulation of the Schrödinger equation often leads to dead ends and to avoid this, approximation methods such as the Variational method or the Perturbation theory are employed. Although both methods are widely used, here we will focus on the Variational method, which is used for the Hartree-Fock theory.

The variational principle is the starting point of variational methods and finds the most accurate solution from a test function. The principle states that the energy value will be equal to or greater than the actual value when a function is well defined and normalized.

$$\langle \Psi | A | \Psi \rangle \geq E_0 \ ,$$

(5)

E_0 being the lowest possible energy value.

Now, this principle introduces the so-called "variational parameters", which are adjustable parameters used to construct a wave function for testing purposes. The ultimate goal of these parameters is to minimize the energy of the constructed wave functions. Thus, the obtained wave function and its energy are approximations of the variational method that are considered to be accurate.

Thus, the wave function (Φ) can be expressed as a linear combination of exact eigenfunctions (Ψ_i)

$$\Phi = \sum_i c_i \Psi_i \; , \tag{6}$$

and the energy approximation corresponding to the wavefunction will be:

$$E(\Phi) = \frac{\langle \Phi | H | \Phi \rangle}{\langle \Phi | \Phi \rangle} \; , \tag{7}$$

The expansion substituted in the exact wavefunction

$$E(\Phi) = \frac{\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} c_i c_j \; \langle \Psi | H | \Psi \rangle}{\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} c_i c_j \; \langle \Psi | \Psi \rangle} \; . \tag{8}$$

Now, taking into account that $H\psi_i = E\psi_i$ and the orthonormality of the ψ_i

$$E(\Phi) = \frac{\sum_{i=0}^{\infty} c_i^2 E_i}{\sum_{i=0}^{\infty} c_i^2} \; . \tag{9}$$

Subtracting the exact ground state energy on both sides of the expression

$$E(\Phi) - E_0 = \frac{\sum_{i=0}^{\infty} c_i^2 E_i - E_0}{\sum_{i=0}^{\infty} c_i^2} \; . \tag{10}$$

In the end it is concluded that $E(\Phi) \geq E_0$ since in the right hand side of the Eq.10, each term has to be greater than or equal to zero and this implies that the left hand side must also be greater than or equal to zero. Therefore, the energy of any approximate wave function is always greater than or equal to the energy of the ground state (E_0).

To summarize the development of the variational method, these two conclusions are drawn: for any approximate test wave function its energy is always greater than the actual energy and the decrease of the energy in the test functions by any variation makes the approximate energy close to the exact value.

The variational method is applied as the expansion of molecular orbitals as the linear combination of molecular orbitals (LCMO) in the Hartree-Fock theory.

2.3. Hartree-Fock theory

The study of different properties of the systems chosen to analyze throughout this thesis is not an easy matter and involves the use of elaborate computational methods. To simplify the work, independent particle models are introduced, since they consider the motion of electrons independent of the dynamics of all other electrons. An independent particle model means that the interaction between the particles is approximated by taking the average of all interactions occurring in the particle. In the HF theory, each electron is described by an orbital, and the total wave function is given as a product of orbitals, called the Hartree Product. However, this product does not satisfy the antisymmetry principle which states that a wavefunction describing fermions must be antisymmetric with respect to the exchange of any set of spin-space coordinates. The following description is largely based on chapter 3 of Jensen's book [1].

To solve this problem, Slater determinants (SDs) can be constructed to make the wave function antisymmetric. The structure of SDs always follows the same pattern: the columns are single-electron wave functions, called orbitals, and the rows represent the electron coordinates, Eq. 11. As a general example, N electrons and N spin orbitals (ϕ) are taken to represent the following Slater determinant

$$\Phi_{SD} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(1) & \phi_2(1) \cdots & \phi_N(1) \\ \vdots & \ddots & \vdots \\ \phi_1(N) & \phi_2(N) \cdots & \phi_N(N) \end{vmatrix}. \tag{11}$$

Furthermore, by taken the trial wave function as a single Slater determinant, the approximation is improved since the electron correlation is neglected and the electron-electron repulsion is only included as an average effect. Next to the selection of the proper single-determinant trial wave function, the variational principle is used to derive the Hartree-Fock (HF) equations, by minimizing the energy.

Let state the Hamiltonian again

$$H_{el} = T_e + V_{ne} + V_{ee} + V_{nn} \ . \tag{12}$$

It should be noted that terms such as T_e and V_{ne} depend on only one electron coordinate, while V_{ee} depends on two electron coordinates. This implies the classification of operators according to the number of electron indices:

- The one-electron operator is denoted as h_i and describes the motion of the electron in the all-nuclei field.

$$h_i = -\frac{1}{2} \nabla_i^2 - \sum_a^{N\,nuclei} \frac{Z_a}{|R_a - r_i|} \ . \tag{13}$$

- The two-electron operator is called g_{ij} and gives the electron-electron repulsion.

$$g_{ij} = \frac{1}{|r_i - r_j|} \ . \tag{14}$$

With these two types of operators, Eq.13 and Eq.14, a simpler electronic Hamiltonian can be written

$$H_{el} = \sum_i^{N\,elec} h_i + \sum_{j>i}^{N\,elec} g_{ij}. \tag{15}$$

Once the Slater determinant and the Hamiltonian have been written in simplified form, the energy of the system can be calculated starting from $E_e = \langle \Phi | H | \Phi \rangle$ and with the normalized wave function. Then as

the variational principle is applied to the energy expression, the parameters are said to be varied to achieve the minimum possible energy value.

If some algebraic relations are applied, the electronic energy can be written in terms of the two types of operators explained above, Eq.13 and Eq.14

$$E = \sum_{i=1}^{Nelec} h_i + \frac{1}{2} \sum_{i=1}^{Nelec} \sum_{j>i}^{Nelec} (J_{ij} - K_{ij}) + V_{nn} \ , \tag{16}$$

where each of the operators correspond to the following integrals

$$h_i = \langle \phi_i(1) | h_i | \phi_i(1) \rangle \ , \tag{17}$$

$$J_{ij} = \langle \phi_i(1)\phi_j(2) | g_{12} | \phi_i(1)\phi_j(2) \rangle \ , \tag{18}$$

$$K_{ij} = \langle \phi_i(1)\phi_j(2) | g_{12} | \phi_j(1)\phi_i(2) \rangle \ , \tag{19}$$

and ϕ_i denotes the molecular orbitals (MOs).

To derive the energy variation, it is expressed in terms of Coulomb (J) and exchange (K) operators.

$$E = \sum_{i=1}^{Nelec} \langle \phi_1(19 | h_i | \phi_i(1) \rangle + \frac{1}{2} \sum_{ij}^{Nelec} (\langle \phi_j | J_i | \phi_i \rangle - \langle \phi_j | K_i | \phi_j \rangle) + V_{nn} \ , \tag{20}$$

$$J_i | \phi_j(2) \rangle = \left\langle \langle \phi_i(1) | g_{12} | \phi_i(1) \rangle \big| \phi_j(2) \right\rangle \ , \tag{21}$$

$$K_i | \phi_j(2) \rangle = \left\langle \langle \phi_i(1) | g_{12} | \phi_j(1) \rangle \big| \phi_i(2) \right\rangle \ . \tag{22}$$

Starting from the energy expression which is symmetric and follows the variational principle, it is necessary to determine the set of MOs that makes the minimum energy. To obtain this energy, the "best" Slater determinant is sought without altering the MOs which must remain orthogonal and normalized. This optimization is carried out using Lagrange multipliers for which a small infinitesimal change in the coordinates of the MOs should not change the Lagrange function.

$$L = E - \sum_{i}^{Nelec} \lambda_{ij}(\langle \phi_i | \phi_j \rangle - \delta_{ij}) \ , \tag{23}$$

$$\delta L = \delta E - \sum_{ij}^{Nelec} \lambda_{ij}(\langle \phi_i | \phi_j \rangle - \langle \phi_i | \phi_j \rangle) = 0 \ , \tag{24}$$

and the variation of the energy is given by

$$\delta E = \sum_{i}^{Nelec} (\langle \delta\phi_i | h_i | \phi_i \rangle + \langle \phi_i | h_i | \delta\phi_i \rangle) + \frac{1}{2} \sum_{ij}^{Nelec} \begin{pmatrix} \langle \delta\phi_i | J_j - K_j | \phi_i \rangle + \langle \phi_i | J_j - K_j | \delta\phi_i \rangle + \\ \langle \delta\phi_j | J_i - K_i | \phi_j \rangle + \langle \phi_j | J_i - K_i | \delta\phi_j \rangle \end{pmatrix} . \tag{25}$$

The third and fifth terms are identical and the fourth and sixth terms are identical, in Eq.25. They are lumped to cancel the ½ factor and the energy variation is then written in terms of the Fock operator, F_i.

$$\delta E = \sum_i^{Nelec} (\langle \delta\phi_i|h_i|\phi_i\rangle + \langle \phi_i|h_i|\delta\phi_i\rangle) + \sum_{ij}^{Nelec} (\langle \delta\phi_i|J_j - K_j|\phi_i\rangle + \langle \phi_i|J_j - K_j|\delta\phi_i\rangle) , \tag{26}$$

$$\delta E = \sum_i^{Nelec} (\langle \delta\phi_i|F_i|\phi_i\rangle + \langle \phi_i|F_i|\delta\phi_i\rangle) . \tag{27}$$

Thus, the Fock operator will be equivalent to

$$F_i = h_i + \sum_j^{Nelec} (J_j - K_j) . \tag{28}$$

The Fock operator is an effective energy operator of an electron and describes the kinetic energy of that electron and the attraction to all nuclei and repulsion to the rest of the electrons. Also, it is important to note that the Hamiltonian operator is not the sum of the Fock operators.

Returning to the variation of the Lagrange function mentioned before, it can be rewritten as

$$\delta L = \sum_i^{Nelec} (\langle \delta\phi_i|F_i|\phi_i\rangle + \langle \phi_i|F_i|\delta\phi_i\rangle) - \sum_{ij}^{Nelec} \lambda_{ij}(\langle \delta\phi_i|\phi_j\rangle - \langle \phi_i|\delta\phi_j\rangle) . \tag{29}$$

According to the variational principle, the desired orbitals are those that make $\delta L=0$. And with the help of the complex conjugate properties of the Lagrange function, the final set of Hartree-Fock equations is obtained.

$$F_i\phi_i = \sum_j^{Nelec} \lambda_{ij}\phi_j \tag{30}$$

Further simplifications can be made by choosing a unitary transformation that makes the Lagrange multiplier matrix diagonal, $\lambda_{ij} = 0$ and $\lambda_{ij} = \varepsilon_i$.

$$F_i\phi_i = \varepsilon_i\phi_i . \tag{31}$$

It is concluded that the Fock operator acts on the wave function and generates an energy value multiplied by the orbital. Thus, the Hartree-Fock equations are pseudo eigenvalues and allow to calculate the orbitals and also the energy levels using the self-consistent field (SCF), through an iterative process. To start such a process a guess is needed to improve the iteration as it is performed and due to this fact the HF equations are not useful for molecular calculations as they do not give good initial guesses for the MO wave functions.

2.4. Basis sets

A numerical method can be used to solve the Hartree-Fock equations, but it is usually better to expand the unknown MOs in terms of known wave functions. Therefore, each MO (ϕ) is expanded in terms of the so-called atomic orbitals (basis functions, χ) which are called LCAO (Linear Combination of Atomic Orbitals) since the expansion is linear.

$$\phi = \sum_{\alpha}^{M\ basis} C_{\alpha i} \chi_{\alpha} \ , \tag{32}$$

where the parameter $C_{\alpha i}$ called expansion coefficient is a key aspect since its optimization will imply the optimization of all calculations.

This expansion is introduced in the HF equations ($F_i \phi_i = \varepsilon_i \phi_i$) and after multiplying by the complex conjugate of a given orbital and integrating out Eq.32 by a specific basis function, the Roothaan-Hall equations are obtained

$$\boldsymbol{F_i} \sum_{\alpha}^{M\ basis} C_{\alpha i} \chi_{\alpha} = \varepsilon_i \sum_{\alpha}^{M\ basis} C_{\alpha i} \chi_{\alpha} \ , \tag{33}$$

$$\boldsymbol{FC = SC\varepsilon} \ , \tag{34}$$

$$F_{\alpha\beta} = \langle \chi_\alpha | \boldsymbol{F} | \chi_b \rangle \qquad \text{and} \qquad S_{\alpha\beta} = \langle \chi_\alpha | \chi_\beta \rangle \ . \tag{35}$$

These are the HF equations in the atomic orbital basis and all the M_{basis} equations are collected in matrix notation. Thus, the matrix F corresponds to the Fock elements, C contains the unknown expansion coefficients, S is the overlap matrix between the basis functions and ε is the diagonal matrix for the orbital energies.

The Roothaan-Hall equation provide, after diagonalization of the Fock matrix, the unknown MO coefficients ($C_{\alpha i}$). The problem is that the Fock matrix is only known if all the $C_{\alpha i}$ are known and the process will start with a guess of the coefficients that form this F matrix and their diagonalization. The new set of coefficients is then used to calculate a new Fock matrix and the procedure ends when the coefficients used for the construction of the F matrix are equal to those resulting from the diagonalization. And the final set of coefficients determines a self-consistent field solution.

The concept of basis set was mentioned in the Roothaan-Hall equation where the basis function is defined as a linear combination of functions and the basis set is the group of specific basis functions used in a particular calculation. The size and type of basis set used influences the accuracy of the computations and the smaller the basis, the poorer the representation. Knowing that the computational effort of ab initio methods formally scales at least as M_{basis}^4, it is of course of paramount importance to make the basis set as small as possible, without compromising accuracy.

Some properties of an ideal basis are:

- It must reflect the nature of the problem.
- It should be available for all atoms in the periodic table, for atom-dependent basis functions.
- It should allow calculations for a number of other properties a part of the energy and energy-related quantities.

But the computational efficiency imposed is sometimes incompatible with these properties. Thus, a basis capable of giving good accuracy for a variety of properties will always be greater than one that is only valid for a specific purpose.

2.5. Types of basis sets

Slater-type orbitals (STO) and Gaussian-type orbitals (GTO) are the two types of basis functions, also called atomic orbitals (AO), used in electronic structure calculations. Slater-type orbitals are in the form of

$$\chi_{\zeta,n,l,m}(r,\theta,\varphi) = NY_{l,m}(\theta,\varphi)r^{n-1}e^{-\zeta r} \ , \tag{36}$$

where N is the normalization constant, $Y_{l,m}$ are spherical harmonic functions, ζ is the orbital exponent and n the principal quantum number. Here, an exponential dependence on the distance between the nucleus and the electron is seen to reflect the exact orbitals for the hydrogen atom. This type of basis functions do not have radial nodes, but are introduced by making linear combinations of STOs. They are the most used for atomic and diatomic molecules where a great precision is required since at the moment that three and four center integrals appear they become inefficient.

On the other hand, Gaussian-type orbitals can substitute for GTO functions and solve such limitations. GTOs are written in terms of polar or Cartesian coordinates of the form

$$\chi_{\zeta,n,l,m}(r,\theta,\varphi) = NY_{l,m}(\theta,\varphi)r^{2n-2-1}e^{-\zeta r^2} \ , \tag{37}$$

$$\chi_{\zeta,l_x,l_y,l_z} = Nx^{l_x}y^{l_y}z^{l_z}e^{-\zeta r^2} \ , \tag{38}$$

and the type of orbital is determined by the sum of l_x, l_y, l_z.

Gaussian-type orbitals are the ones used in most ab initio methods and in Cartesian coordinates, since molecular integrals are easy to derive and encode from them.

The dependence on r^2 in GTOs makes them inferior to STOs because they have two problems. First, they do not adequately represent the behavior near the nucleus since GTO has a zero slope at the nucleus, as opposed to STO which has a discontinuous derivative. And the other problem is that the tail of the GTO wave function is poorly represented as a consequence of early dips from the core compared to STOs. Three times as many GTOs as STOs are needed to obtain the same level of accuracy. Although

the number of basis functions increases, the ease of calculations causes the trade-off to decline to GTO in practical electronic system calculations.

2.5.1. Classification of Gaussian Type Basis Sets

Once the basis set (GTOs) is chosen, which is defined only by the number of functions of each type (s-, p-, d-, etc.), the value of the exponent for each function and the way to combine the functions in fixed linear combinations, these three parameters are combined to obtain large degrees of freedom. This is reflected in the wide range of basis sets reported in the literature.

The combination of the complete set of basis functions, known as primitive GTOs (PGTOs), into a smaller set of functions forming linear combinations is called contraction of the basis set and the obtained functions are called contracted GTOs (CGTOs).

Next, according to the number of functions used to describe the AOs, a classification is made. The smallest possible number of functions is a minimal basis set and only enough functions are used to contain all the electrons of the neutral atom(s). These are called single Zeta (SZ) and for hydrogen or helium it means a single s-function (1s), for the first row of the periodic table it implies two s-functions (1s and 2s) and a set of p-functions ($2p_x$, $2p_y$ and $2p_z$) and so on.

As an improvement, all the basis functions can be duplicated in one set, producing the Double Zeta (DZ) which uses two s-functions (1s and 1s') for hydrogen and helium, four s-functions (1s, 1s', 2s and 2s') and two sets of p-functions for the elements in the row. This allows a better description of the electron distribution. Further increasing the basis set yields a Triple Zeta (TZ) containing three times as many functions as the minimum basis. The names Quadruple Zeta (QZ) and Quintuple or Pentuple Zeta (PZ or 5Z) are also used for the next levels of basis sets. Increasing the number of functions makes it possible to describe the bond in all directions and environments.

As opposed to the core electrons, which are not related to chemical bonding and are treated with minimal basis sets, the valence electrons are more affected by the chemical environment. Chemical bonding occurs between the valence orbitals and this causes the valence electrons to be treated with larger basis sets, such as DZ or TZ. Thus, the split valence basis set is used to reduce the time required for large molecule calculations.

In addition, polarization functions appear as a tool to increase the size of the basis set. They are used on the orbitals of higher angular momentum and allow polarizing the atomic electron density to obtain a better representation of the electron density of molecules. Taking hydrogen as an example, which should contain an s (spherical) orbital when bonded to an atom with a markedly lower electronegativity such as lithium, the electron density is no longer spherical and the addition of a p orbital function will change the electron density in the direction of the bond, i.e. its polarization. The most commonly used polarization functions are p-orbitals, as well as d-orbitals or f-orbitals and they polarize the previous

functions, i.e. p-orbitals polarize s-orbitals, d-orbitals polarize p-orbitals and so on. Thus, the new way of naming is Double Zeta plus Polarization (DZP) type basis for a DZ basis with a single set of polarization functions added, for example. As a note, there is a variant in which the polarization functions are only added to the non-hydrogen atoms, assuming a "passive" role of the hydrogens. Since hydrogen usually represents a large number of atoms in the system, a saving of three basis functions for each hydrogen is significant.

On the other hand, the diffuse functions are added to the basis set to represent a very wide electron distribution. They are necessary whenever there are loosely bound electrons, anions or excited states.

2.5.2. Contracted Basis Functions

Contracted functions apply linear combinations of primitive basis functions in what is known as a contraction in order to improve computational efficiency in performing calculations on any system of molecules. But this type of functions, on the other hand, reduce the accuracy and as a key point in contraction of basis sets, their function is to obtain the highest computational savings without losing the accuracy of the primitive basis set.

If a set of *M* primitive Gaussian primitive basis functions (PGTOs) is contracted, the final contracted functions are

$$\chi(CGTO) = \sum_i^k \alpha_i \chi_i(PGTO) \ , \tag{39}$$

where α_i is the contraction coefficient and requires optimization as the orbital exponents in the functions.

Contraction is useful for orbitals describing inner (core) electrons, since they require a relatively large number of functions to represent the cusp of the wave function near the core. Contraction of a basis set will always increase the energy, as it restricts the number of variational parameters and makes the basis set less flexible, but it will also reduce the computational cost significantly.

The basis set in terms of primitive and contracted functions is specified with the following notation

$$(10s4p1d/4s1p) \rightarrow [3s2p1d/2s1p],$$

where the left side, written in parentheses, is divided into the number of primitives with heavy atoms before the bar and hydrogen after. And on the right side, written in brackets, is the number of contracted functions. This notation only allows to know the size of the final basis without further information on how the contraction is done.

Furthermore, there are two ways of contracting basis sets: segmented contraction and general contraction. In a segmented contraction the set of PGTOs is divided into smaller sets of functions that become CGTOs by determining appropriate coefficients. Thus, each primitive only contributes to one

contracted function and the exponent and the contraction coefficient are optimized at the same time by variational optimization. In general contraction, all primitive functions are allowed to contribute to all contracted functions, but with different contraction coefficients.

In most cases, a combination of segmented and general contraction is used in the basis sets [3]. The inner shell orbitals are described by primitive functions with large exponents and the outer shell orbitals are described mainly by primitive functions with smaller exponents.

Some basis sets that are often used in routine computations are

- Pople-style basis sets (3-21G, 6-31G, 6-31G*, 6-311G, 6-311G*).
- The Dunning-Huzinaga basis sets (cc-pVDZ, cc-pVTZ, cc-pVTZ, cc-pVQZ) determined energy-optimized basis sets with no contraction up to (10s6p) for the first row elements and were the first published high-quality basis sets. Subsequently, Duijneveldt extended them to (14s9p) and Partridge to (18s13p). Finally, Dunning has used Huzinaga's primitive OWGs to derive several contraction schemes, known as Dunning-Huzinaga (DH)-type basis sets [3].
- MINI, MIDI and MAXI basis sets: Tatewaki and Huzinaga optimized minimal basis sets for a large part of the periodic table at the HF level. MINI-n are minimal basis sets with three PGTOs in the CGTOs 2s, and a variable number of PGTOs in the CGTOs 1s and 2s. MIDI-n are identical to MINI-n, but the outer valence function is deconcentrated. And the MAXI-n basis sets employ four PGTOs for the CGTOs 2s and five to seven PGTOs for the CGTOs 1s and 2p [1]. The MAXI-n basis sets have a variable number of PGTOs in the CGTOs 1s and 2s.
- Ahlrichs-type basis sets: SVP, TZ and QZ-quality designed basis sets for elements up to Kr. They are explain in more detail in the results discussion.

2.6. Post Hartree-Fock methods

As it is mentioned above, the HF calculations do not include electron correlation and this is one of the most important limitations it has because it takes the electron repulsion effect as an average and does not consider explicit electron-electron interactions. With the central field approximation that defines the HF method it is said that the probability of finding an electron somewhere around atoms is determined by the distance to the nucleus but not by the distance to other electrons, something that is not physically true.

To solve this limitation, in most of systems HF calculation are done which are then corrected to include electron correlation. Thus, the three main methods employed are: Configuration Interaction (CI), Coupled Cluster (CC) and Many-Body Perturbation Theory (MBPT). Such correlation is important to improve the computational accuracy of energies or molecular geometries.

3. DENSITY FUNCTIONAL THEORY AND THE ELECTRON DENSITY

Since the last century, a key problem in the application of quantum mechanics to chemistry has been the explanation of the chemical bond. It was Berlin [4] who initiated a very satisfactory line of reasoning that links the redistribution of charge that takes place when a chemical bond is formed in terms of an electronic transfer from the retronuclear to the internuclear regions. Subsequently, the Hellmann-Feynman theorem was used to justify the connection between the electron density and the forces exerted upon the nuclei. It was found that in diatomics there was a two dimensional where these forces cancel out. Thus, two spatial regions appear: the bonding region where electrons contribute to nuclear attraction and the antibonding region, where they contribute to nuclear repulsion.

The Density Functional Theory (DFT) determines the electronic energy of the ground state by the electron density (ρ) and this implies the existence of a one-to-one correspondence between the electron density of a system and its energy. Three important arguments justify this assertion:

- The cusps of the densities define the position of the nuclei.
- The heights of the cusps define the corresponding nuclear charges.
- The integral of the density defines the number of electrons.

To illustrate this theorem proposed by Hohenberg-Kohn, it is compared with the wave function approximation. The electron density depends always on three variables, regardless the size of the system, whereas the wave function depends on a growing number of parameters as the number of electrons increases. The goal of the DFT method is to create functionals that connect each of the different densities with the corresponding ground state energy to solve the problem of the relationship between these two quantities. So, the success of DFT methods is due to the calculation of the electron kinetic energy from an auxiliary set of orbitals representing the electron density. Thus, in this method the only unknown functional should be the exchange-correlation energy, a very small term within the total energy that does not influence the accuracy of the computational models.

3.1. HF and DFT comparison

The Hartree-Fock (HF) and density functional (DFT) methods are most often used to solve the same problem or task and good results are obtained with both, but it is important to know the differences that make the right choice of method for the analysis of molecules:

- HF is a wave mechanics method that employs an exact Hamiltonian operator, but with the use of approximations to the wave functions. Whereas DFT is a density-based method that performs approximations on the energy functional but allows free variation of the electron density ρ .
- Both are conceptually and computationally similar, but DFT provides better results, which make it the more popular method.

- The main problem with DFT is the inability to improve results and the inability to describe certain features such as Van der Waals interactions (which are also not recovered in HF).

4. COMPUTATIONAL METHODOLOGY

In the development of this thesis, the analysis of the chosen molecules is through computational programs of electronic structure [5]. At present, there are a large number of programs available with a varying degree of complexity and within this group some of the most widely used are Gamess [6], Dalton [7], Octopus [8], Gaussian [9] or Orca [9]. Common to all of them is the creation of an input file where the necessary instructions are given to run the program and the characteristics to be analyzed. Once the software performs all the calculations, the final results and the desired molecular properties are reflected in the output file.

The structure of the input file is the same for each molecule and contains the applied method (HF or DFT), the optimization geometry, the chosen basis set and the coordinates of the molecule. On the other hand, the output file contains much more data and may vary from one system to another, but the general structure will be:

- The input, previously designed.
- The calculation of the basis sets.
- The preparation of initial guesses.
- The SCF calculations.
- Additional calculations (post-CRL calculations and geometry configuration for stationary points).
- Finally, all properties of the calculations.

The development of the thesis is carried out with three different and free computer programs available for any type of operating system. Each of them has a specific function:

- AVODADRO: it is a very intuitive program that allows to draw any desired molecule and allows to generate the input from the drawn molecule which is necessary to carry out the ORCA program. It is an easy tool for people not used to create input files.
- ORCA: is the basis and the main program of the thesis. It describes itself as an ab initio, DFT and semi-empirical electronic package. Its levels of theory are Hartree-Fock to DFT and multi-reference model.
- AIMALL: is the program used to finally extract the desired information of the molecule systems from the ORCA output file. It is also a very intuitive tool and allows different visualizations of the molecule depending on the analyzed property.

4.1. Method and how to choose a basis

The Hartree-Fock theory is the method chosen for this thesis. It is an ab initio method that follows the Born-Oppenheimer approximation, does not consider relativistic effects and uses linear combinations of atomic orbitals, as explained in the theoretical section of this work. It is an easy method for computational development calculations and its use is reflected in the input file with the acronym HF.

Once the method is chosen, the next step is to find the right basis set for the analysis with a compromise between accuracy and computational efficiency as well as wave function. As a general trend, the more sophisticated the methods, the larger basis sets are required. Since the HF method is not a high performance method the use of DZP or TZP quality basis sets would be sufficient for the systems chosen in this thesis.

The three types of contracted bases selected in this thesis are Pople, Dunning and Alhrichs and the main difference between them will be the optimization in exponents and shrinkage coefficients. The analysis focuses on the different properties of the molecules using at least three basis sets for each of the groups described in more detail below:

- Pople basis sets [10]

Pople basis sets are the most popular in ab initio molecular calculations because of their ease of manipulation for software and the wide range of basis functions available. As shown in Table X, all the basis sets used are split valence basis sets in which minimal basis functions are used for the core orbitals, but the valence orbitals are split into two or three parts. Then, variational procedures allow the parameters to be optimized and, as a constraint, the same exponent is used for the s and p valence functions, which decreases the flexibility of the basis sets and increases computational efficiency [11].

The nomenclature used in this type of bases is k-nlmG, where k corresponds to the number of PGTOs describing the core orbitals and nlm indicates how many functions the valence orbitals are divided into, as well as the number of PGTOs used to describe them. A double-zeta base implies only two numbers in the name (nl) and a triple-zeta means three values (nlm). The most common Pople bases are: 3-21G, 6-31G and 6-311G and to all of them polarization and diffuse functions can be added. To enhance them with polarization functions, a * is used for one set of d-type polarization functions added to each non-hydrogen atom of the molecule and ** implies the addition of a d-type polarization functions to the non-hydrogen and a set of p-type polarization functions is added to the hydrogen. On the other hand, diffuse functions are indicated with + sign for one set of sp-type diffuse basis functions added to non-hydrogen atom and ++ which means that one set of sp-type diffuse functions is added to each non-hydrogen and one s-type diffuse function is added to hydrogen atoms.

- Dunning basis sets

The correlation-consistent (Dunning) basis sets come from an update of the Dunning basis sets (DZ, TZ...) and are made to recover the correlation energy of the valence electrons. The correlation energy is the true non-relativistic energy that is in error with the total energy of an ab initio model HF wave function. And consistent correlation involves functions that contribute similar amounts of correlation energy at the same stage, regardless of the type of function. The nomenclature is cc-pVNZ where the cc corresponds to correlation consistent, the p to polarized, the V indicates valence basis sets and N refers to the level of multiplicity (D=double. T=triple...) [12].

- Alhrichs Basis Sets

Alhrichs basis sets were developed as a second generation, called def2 basis sets. There are four basis sets def2-SVP, def2-TZVP, def2-TZVPP and def2-QZVP and the different abbreviations denote: SVP for split valence polarized, TZP stands for triple zeta and so on. The quality of these basis sets is for elements up to Kr and their optimization was done by gradient techniques. The def2 basis sets are designed to provide consistent accuracy across the entire periodic table, whereas this is not true for the more commonly used People and Dunning basis sets. The Alhrichs basis functions were designed as a segmented contraction scheme and the best basis of this group is the one that does not have the longest computational time in trade-off with the best accuracy representation of the systems.

The Pople and Alhrichs basis sets were designed from the segmented contraction scheme, while the Dunning basis sets were designed from the general contraction scheme. And in the end, the basis functions chosen from each group are the ones that do not have high computational time to process and provide a good representation of each group. In total, eleven basis functions are selected and some characteristics of the different basis are listed in Table 1 as a summary [13]. Thus, the basis functions are Double, Triple or Quadruple-Zeta and polarization or diffuse functions can be added to them. The contraction scheme of each basis function still depends on whether the elements are from the s- or p-block.

Table 1. Description of the basis functions used in the thesis.

Quality	Group	Name	Polarization or diffuse function	Composition	
				s-Block elements	p-Block elements
SV	Alhrichs	Def2-SVP	Polarization	(7s4p)$\rightarrow$ [3s2p]	(7s4p1d)$\rightarrow$[3s2p1d]
DZ	Pople	3-21G	None	(6s3p)$\rightarrow$[3s2p]	(6s3p)$\rightarrow$[2s2p]
DZ	Pople	6-31G	None	(10s4p)$\rightarrow$[3s2p]	(10s4p)$\rightarrow$[3s2p]
DZ	Pople	6-31G*	Polarization	(10s4p1d)$\rightarrow$ [3s2p1d]	(10s4p1d)$\rightarrow$[3s2p1d]
DZ	Dunning	cc-pVDZ	Polarization	(9s4p1d)$\rightarrow$[3s2p1d]	(9s4p1d)$\rightarrow$[3s2p1d]
TZ	Alhrichs	Def2-TZVP	Polarization	(11s4p1d) $\rightarrow$ [5s3p1d]	(11s6p2d1f) $\rightarrow$ [5s3p2d1f]
TZ	Pople	6-311G	None	(11s5p2d)$\rightarrow$[4s3p2d]	(11s5p2d)$\rightarrow$[4s3p2d]
TZ	Pople	6-311G*	Polarization	(11s5p2d1f)$\rightarrow$[4s3p2d1f]	(11s5p2d1f)$\rightarrow$[4s3p2d1f]
TZ	Dunning	cc-pVTZ	Polarization	(11s5p2d1f)$\rightarrow$[4s3p2d1f]	(10s5p2d1f)$\rightarrow$[4s3p2d1f]
QZ	Alhrichs	Def2-QZVP	Polarization	(15s7p2d1f)$\rightarrow$ [7s4p2d1f]	(15s8p3d2f1g)$\rightarrow$[7s4p3d2f1g]
QZ	Dunning	cc-pVQZ	Polarization	(12s6p3d2f1g)$\rightarrow$[5s4p3d2f1g]	(12s6p3d2f1g)$\rightarrow$[5s4p3d2f1g]

4.2. Systems under scrutiny

In this work seven diatomic systems and six polyatomic systems corresponding to the second period of the periodic table are studied, which in terms of computational chemistry are related as elements of the first row. Only different hydride molecules are analyzed and most of them are heteronuclear diatomic molecules (LiH, BeH_2, BH_3, B_2H_4, B_2H_6, CH_4, C_2H_2, C_2H_4, NH_3, N_2H_4, H_2O, H_2O_2, and HF).

In the following Table 2 a general description of each system is set out, these data are available in the NIST webbook and it is an interesting tool to get a first idea of them [14] . Then, in the following section two molecules will be explained in more detail and the rest will be treated as comparison data of properties along the period of the periodic table.

Table 2. General properties of systems.

SYSTEM	BOND ORDER	ROTATIONAL CONSTANT (cm^{-1})	INTERNUCLEAR DISTANCE (Å)	POINT GROUP
LiH	0.875	7.513	1.595	$C_{\infty v}$
BeH$_2$	2	4.753	1.326	$D_{\infty h}$
BH$_3$	3	7.874 ; 3.879	1.190	D_{3h}
B$_2$H$_6$	0.5	2.656 ; 0.606 ; 0.557	1.763; 1.200; 1.320	D_{2h}
CH$_4$	4	5.241	1.087	T_d
C$_2$H$_2$	3	1.177	1.063 ; 1.203	$D_{\infty h}$
C$_2$H$_4$	2	4.828 ; 1.001 ; 0.828	1.339 ; 1.086	D_{2h}
NH$_3$	3	9.444 ; 6.196	1.012	C_{3v}
N$_2$H$_4$	3	0.809	1.446 ; 1.016	C_2
H$_2$O	2	27.877 ; 14.512 ; 9.285	0.958	C_{2v}
H$_2$O$_2$	1	10.356 ; 0.866 ; 0.827	1.475 ; 0.950	C_{2h}
HF	1	-	2.780	C_s

5. RESULTS AND DISCUSSION

The software offers a wide range of properties that can be calculated, from the total energy of the system to the vibration or rotation frequencies. Since all the systems analyzed have the variational parameters optimized for the total energy, this property can be studied in more detail. The chosen method is Hartree-Fock, so the total energy is also called Hartree-Fock energy. And other properties such as the internuclear distance or the rotational constant are noted later as next steps to discuss the accuracy of the different basis functions with respect to the mentioned properties.

5.1. Detailed analysis for LiH and HF

Lithium hydride (LiH) and hydrogen fluoride (HF) are the two extreme molecules corresponding to elements in the first row of the periodic table and, in order to see the behaviour of diatomic molecular hydride in this period, these two systems will be analyzed in detail. Using electronic structure software, many properties of these molecules can be calculated and the bond strength of the systems can also be concluded.

In a first step, the total energy of each of the systems is studied by changing the basis function and taking into account the optimization of the geometry in both systems (Appendix 8.4). Three types of basis sets are chosen: Pople, Dunning and Alhrichs, and within each group the size of the basis functions is increased starting from SV (in Alhrichs) or DZ (in Pople and Alhrichs) up to QZ, as seen in Table 1.

The geometries selected to start the optimization were obtained from the Avogadro's computer programme where a molecule can be drawn and with the tool "generate ORCA input", a pre-design file is created with guess data (Appendix 8.3); the values used in all the systems for optimized geometries are listed in Appendix 8.4. The main reason for choosing this method in all systems is because it will give better results as it finds the most stable atomic arrangement for the molecule and molecules are more stable when their energy is lower.

However, the truth is that there is no criterion for choosing the most efficient basis functions, although there are some theoretical concepts that can conclude which will give the most accurate results. For example, as the size of the basis set (SV, DZ, TZ, etc.) increases, better results are obtained. Thus, some conclusions can be drawn based on theory, before going into context:

- The main trend is that the smaller the basis, the less accurate energies are obtained, with the smallest Pople basis set giving the worst energy values.
- The worst basis function among all systems is 3-21G.
- Pople basis functions can be improved with polarization functions.
- The best basis function is def2-QZVP or cc-pVQZ, although there are some exceptions as seen in Appendix 8.5.
- The difference between the minimal basis function (SV) and a double basis function (DZ) is greater than the difference in DZ and triple basis functions (TZ).

(As a note, in Appendix 8.1 there is a table with all the different units used throughout the thesis and the necessary conversion. Also, all the tables containing the data analyzed in this and the following sections of the thesis can be found in Appendixes 8.4 and 8.5).

5.1.1. Total energy analysis

The LiH system is analyzed for the total energy by applying the different basis sets described in Table 1 under the optimization geometry. The graphical representation in Figure 1 shows the high difference in the energy value of the 3-21G basis and a peak in the def2-QZVP basis function implies the lowest energy value for that system, the rest of the basis functions have energy values in the same range. The 3-21G values give a very bad result even for a very simple diatomic system like LiH. The def2-QZVP basis function gives the best energy value since it is a polarization function with the largest size. All the basis functions without 3-21G are plotted in Figure 2 and the energy changes between them are more clearly seen.
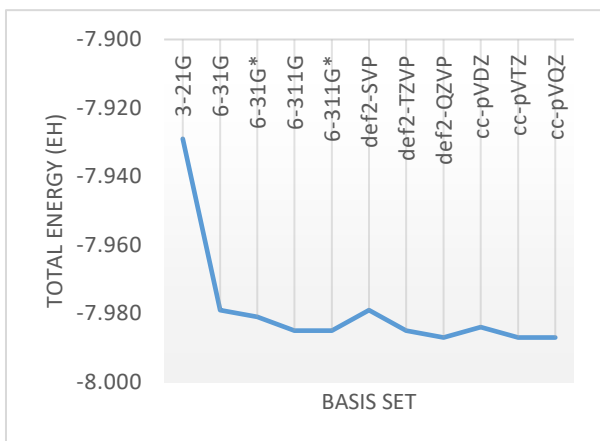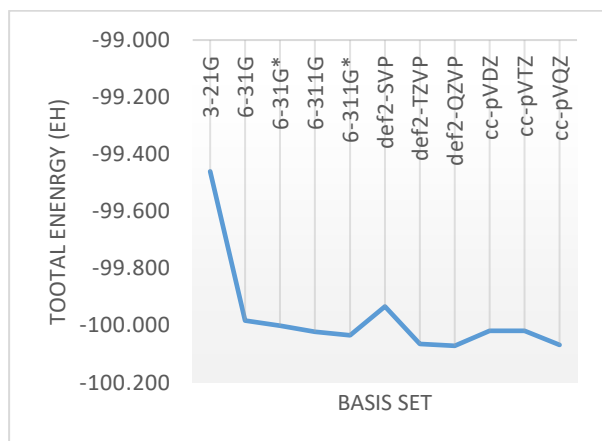
Figure 1. Evolution of the total energy of LiH with the set of basis functions explored at optimized geometries.
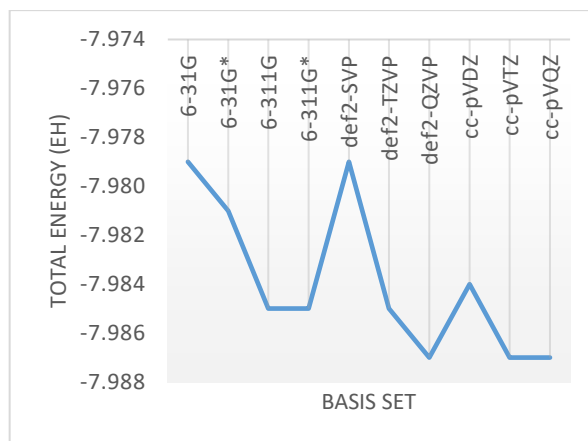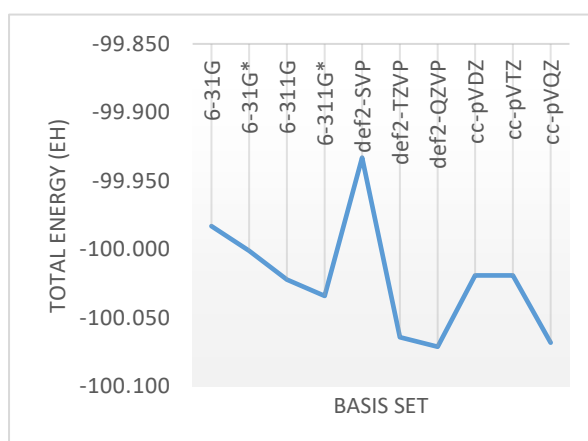

Figure 2. Evolution of the total energy of LiH with the set of basis functions explored at optimized geometries and without 3-21G basis function.

Next, the evolution of the total energy for the HF system with the different basis sets, under optimized geometries is realized in Figure 3 and again the lowest energy value corresponds to the def2-QZVP basis function, which is in agreement with the theory explained above. It is also observed that def2-SVP shows a very bad result according to Table 1 because this basis function has a small contraction scheme which implies that the number of primitive functions used is smaller and this type of behaviour is also obtained in other systems such as the LiH system. For the HF system the same situation as above is repeated with respect to the 3-21G basis function and in Figure 4 this basis function is removed to see in more detail the energy changes between the rests of the basis functions.

Figure 3. Evolution of the total energy of HF with the set of basis functions explored at optimized geometries.
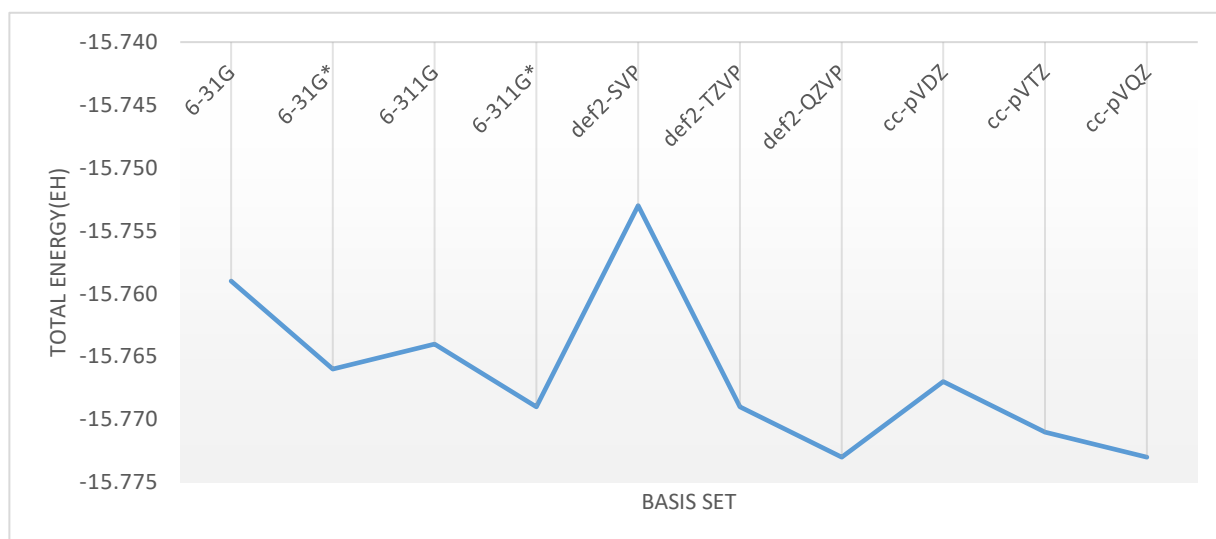

Figure 4. Evolution of the total energy of HF with the set of basis functions explored at optimized geometries without 3-21G basis.

5.2. Discussion

To see the behavior through the second period of the periodic table in each of the chosen systems, the total energy is analyzed and all data are collected in Appendix 8.5. The graphical representation of all of them shows that the basis function def2-QZVP or cc-pVQZ gives the lowest energy value for the optimized geometry according to the theoretical basis detailed above and depending on the system analyzed.

The analysis of the BeH_2 system is shown in Figure 5 and reflects the same trend as before. This trend implies that each time the basis set family is changed a peak appears in the plot because the size of the bases decreases again. This is observed, for example, in the 6-311G* basis function, which is equivalent to a triple valence polarization basis function and gives a lower energy value than def2-SVP, which is equivalent to a single valence polarization basis function. Thus, all the graphs have the basis set ordered from worst to best family starting from the left side of all the figures. And within each family, as the size of the basis function increases, the energy value will decrease and in the end, the largest basis in each family of basis sets will converge to the same or very similar energy value.
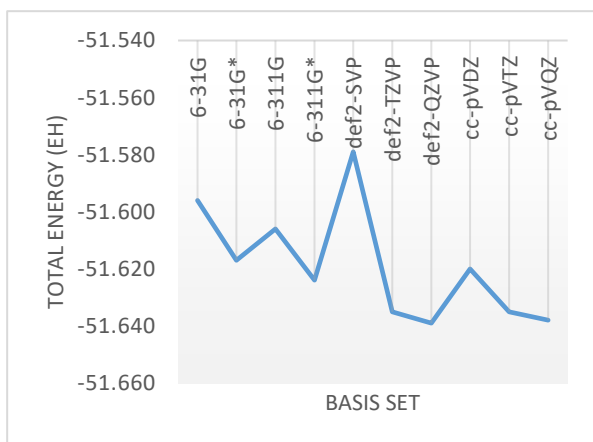

Figure 5. Evolution of the total energy of BeH_2 with the set of basis functions explored at optimized geometries.

In the case of boron, three different systems are considered, the diatomic one, BH_3 and two other heretoatomic ones, B_2H_4 and B_2H_6. The BH_3 system has a general scheme similar to that of BeH_2, as shown in Figure 6, since the two characteristic peaks that occur when changing the family of basis functions are appreciated. A first peak is also observed to the left of the figure, less pronounced corresponding to the jump from a polarization to a non-polarization basis function, always within the Pople family of basis functions. The lowest energy value is obtained with the def2-QZVP and cc-pVQZ basis functions converging to the same value, as explained above.
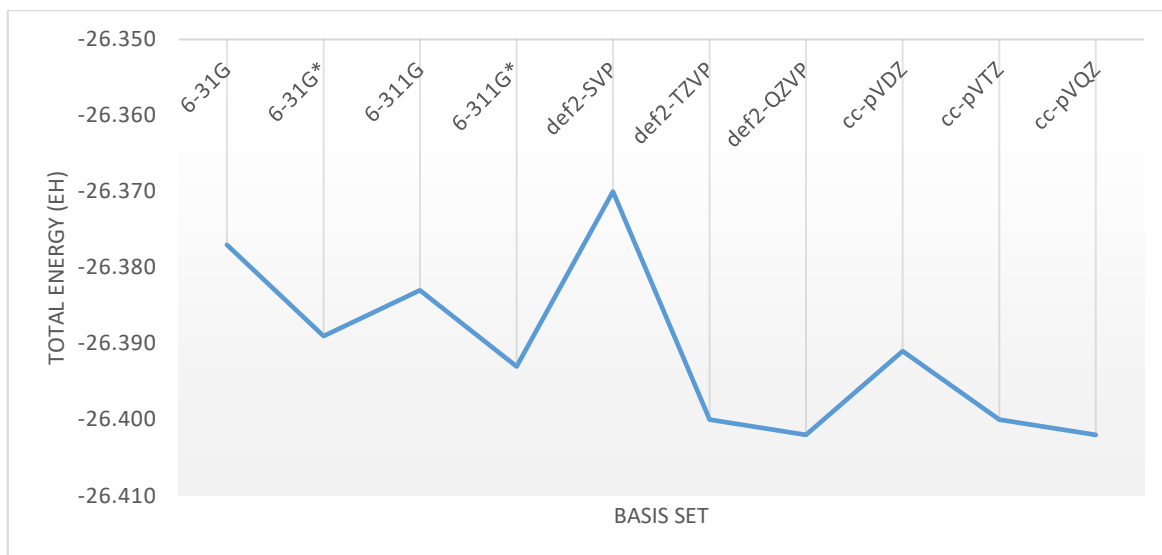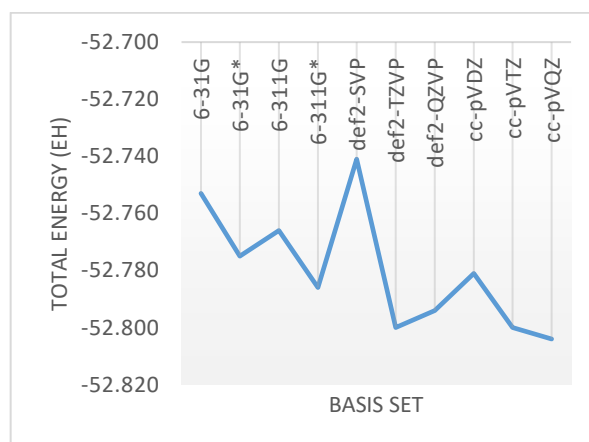
Figure 6. Evolution of the total energy of BH_3 with the set of basis functions explored at optimized geometries.

On the other hand, the B_2H_4 system presents more pronounced peaks (Figure 7) when a jump to the next basis set family occurs and the lowest energy value is given by def2-QZVP. Finally, the B_2H_6 system is represented in Figure 8 where it is observed that the general trend is followed by this system and the lowest energy value is obtained for the basis function cc-pVQZ. It is observed from these last two boron systems that they are not diatomic molecules and consequently have boron-boron bonds and boron-hydrogen bonds. In addition, they have a greater number of electrons and, therefore, more nuclei, which causes the energy values found to be much lower for both systems.

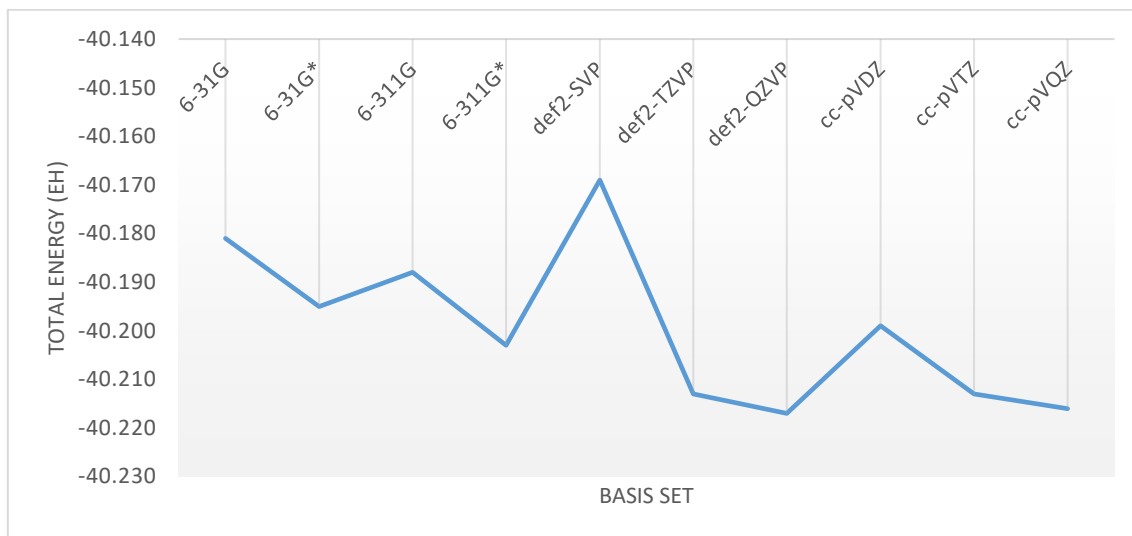**Figure 7.** Evolution of the total energy of B_2H_4 with the set of basis functions explored at optimized geometries.

Figure 8. Evolution of the total energy of B_2H_6 with the set of basis functions explored at optimized geometries.

The next atom of the period is carbon and three systems are analyzed: CH_4, C_2H_4 and C_2H_2. Methane (CH_4) is the first one chosen and will follow the general scheme for the energy values obtained as seen in Figure 9. The lowest energy value is given by the basis functions def2-QZVP and cc-pVQZ as they

are the two largest basis functions within their respective families, Alhrichs and Dunning. Therefore, the results obtained are in agreement with the theoretical explanation made in previous sections.
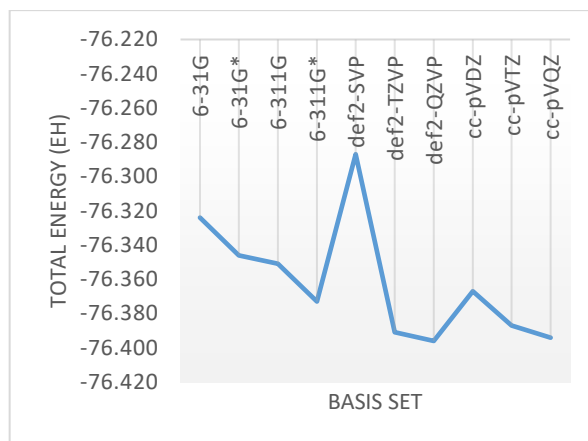

Figure 9. Evolution of the total energy of CH_4 with the set of basis functions explored at optimized geometries.

The energy values of two heteroatomic carbon systems, C_2H_4 and C_2H_2, are described below, where some variations are observed (Figures 10 and 11, respectively). In C_2H_4 there is a carbon-carbon double bond and carbon-hydrogen bonds, but the increase in the negative energy values (which means a large decrease in energy) is due to the incorporation of a larger number of electrons in the system itself. Consequently, there is an increase in the number of nuclei and the lowest energy value is obtained using the def2-QZVP basis function. It is also observed that all Alhrichs basis functions give similar energy values, even def2-SVP which is not common and the change of basis set family from Pople to Alhrichs is not clear.

And to conclude with carbon systems, the C_2H_2 molecule is analyzed. This contains a triple bond between the two carbons which is not the cause of the decrease in the energy obtained, but as with the C_2H_4 system, there is also an increase in electrons. The lowest energy value is obtained by the def2-QZVP basis function. In this case, the peaks caused by the change of basis set family are quite well observed, so the graph (Figure 11) follows the general trend of other diatomic molecules described in the thesis.
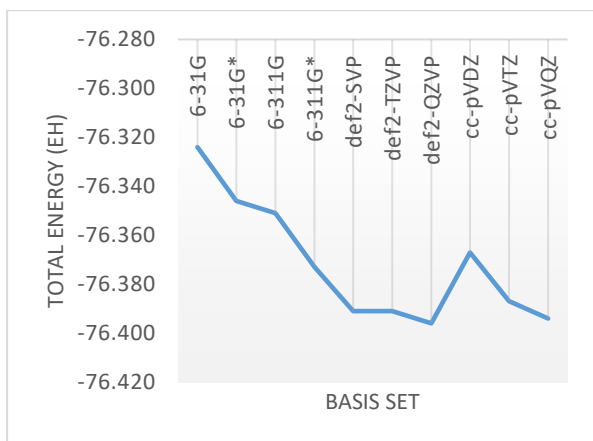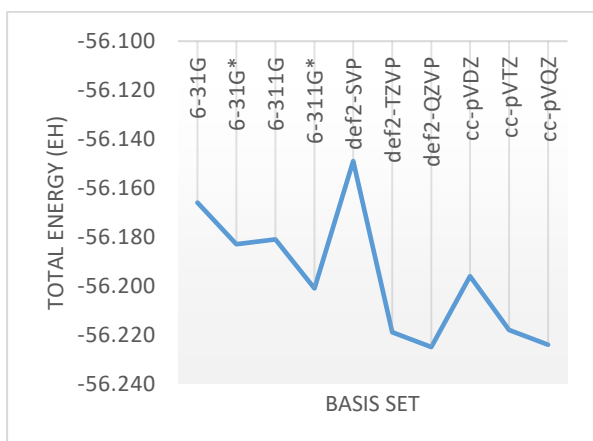
Figure 10. Evolution of the total energy of C_2H_4 with the set of basis functions explored at optimized geometries.
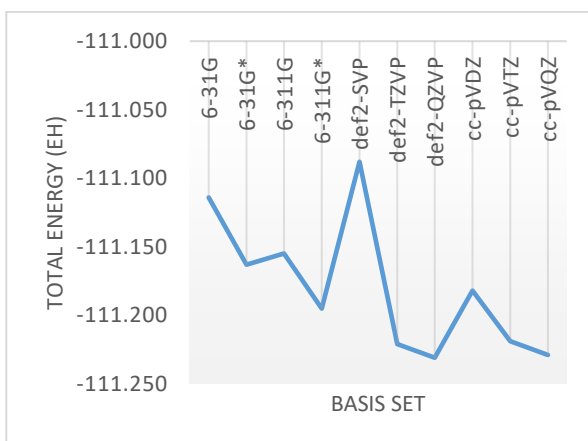

Figure 11. Evolution of the total energy of C_2H_2 with the set of basis functions explored at optimized geometries.

Turning to nitrogen, only two different systems are analyzed, the diatomic (NH_3) and one heteroatomic (N_2H_4, hydrazine). For the ammonia system, there is no characteristic that stands out, since it is a diatomic molecule that follows the general trend. The lowest energy value is obtained with the def2-QZVP and cc-pVQZ basis functions (Figure 12). On the other hand, hydrazine, which has a nitrogen-nitrogen bond and nitrogen-hydrogen bonds, gives much lower energy values due solely to electron gain. Despite this, it also copies the general trend of the basis sets followed in diatomic molecules (Figure 13).
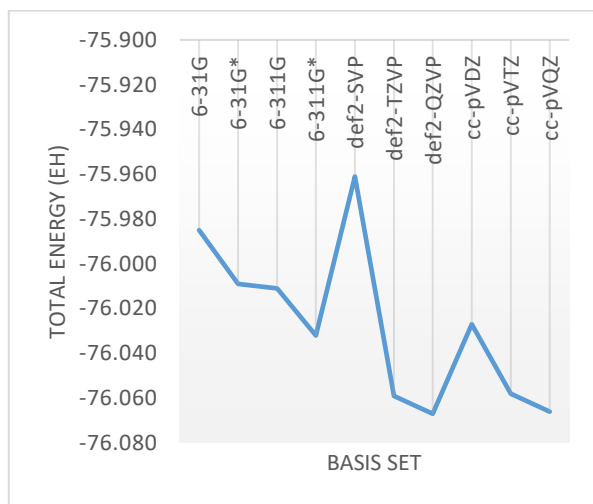

Figure 12. Evolution of the total energy of NH_3 with the set of basis functions explored at optimized geometries.

Figure 13. Evolution of the total energy of N_2H_4 with the set of basis functions explored at optimized geometries.

Finally, the last two systems described are related to one of the most electronegative and largest atoms of the whole period, oxygen. The water molecule, a diatomic system, is plotted in Figure 14 for the total energy of the whole basis set and the lowest energy value is obtained with the def2-QZVP basis function,

so it more or less follows the general trend of diatomic systems for the total energy. Furthermore, if one looks at Figure 15, the H_2O_2 system, which has an oxygen-oxygen bond, has very low energy values that can be predicted because there is an increase in the number of electrons as well as in the number of nuclei. And as in the previous systems, the smallest energy value is obtained with the basis function def2-QZVP.

Figure 14. Evolution of the total energy of H_2O with the set of basis functions explored at optimized geometries.

Figure 15. Evolution of the total energy of H_2O_2 with the set of basis functions explored at optimized geometries.

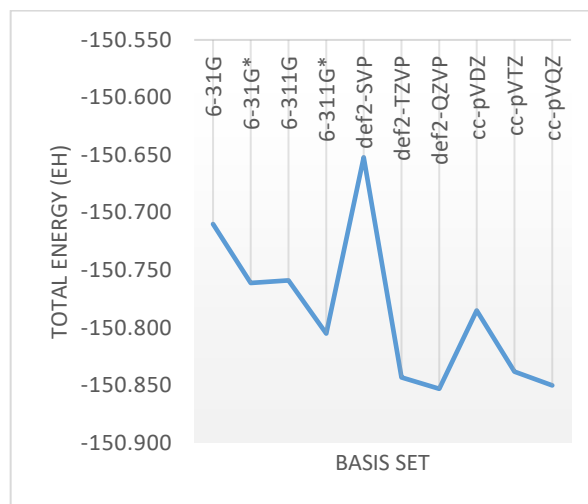To conclude the analysis carried out in this thesis, a molecule as well-known as water is chosen to analyze the behavior of the different basis functions, but this time with respect to size. Thus, taking from Appendix 8.5 all the energy values calculated before for the H_2O molecule are collected in a different order, from the smallest to the largest basis function and regardless of which basis set family it corresponds to. These values are plotted in Figure 16 and, as expected, a different behavior is observed. It is again shown that the 3-21G basis function gives the highest energy value and as the basis size increases, the energy value decreases and the basis functions converge to a more concrete value. The best energy values are found on the right side of the graph where the basis functions are of the Quadruple Zeta type and therefore have a larger size. The importance of polarization basis functions that offer better values can also be appreciated, since when observing one of the Pople's basis functions that do not have polarization such as 6-311G, a small peak appears in the graph indicating that its adjuncts, def2-TZVP and 6-311G, achieve better results because they have polarization functions.
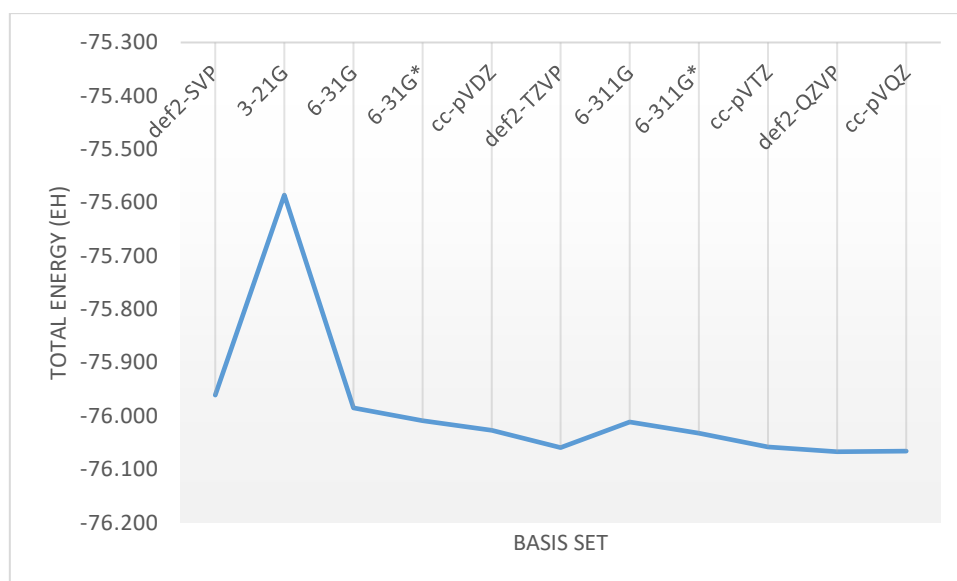
Figure 16. Evolution of the **t**otal energy of H_2O system with basis functions at optimized geometries and ordered by size.

6. CONCLUSIONS

In any electronic structure calculation it is important to analyze and compare different basis sets to choose a reasonable compromise between accuracy and computational cost. The most relevant aspect to select the right basis function is the accuracy, but it is not the only one. In addition, the cost of the calculation can be another feature to take into account when performing any calculation. Thus, in this thesis it is studied the performance of some basis functions of the Pople, Dunning and Alhrichs groups.

In order to state whether one basis function is better than another, it is necessary to compare more than one property since these basis functions depend on many factors such as the method used for the calculation, the type of system analyzed or the variational parameters applied during the procedure. Thus, with all the data analyzed through the thesis, the following conclusions can be drawn:

Looking at the eleven basis functions in terms of the total energy with the optimized geometry in all of them and applying the Hartree-Fock method, it is observed that the results improve as the size of the basis functions is increased. Moreover, if polarization functions are added, the accuracy of the obtained values improves substantially and the increase of the number of functions describing the orbitals also contributes to a better analysis of the chosen systems. Finally, a classification can be made where the Pople family of basis functions is the worst and the Dunning family will give the best energy values for the systems analyzed.

7. BIBLIOGRAPHY

[1] F. Jensen, Introduction to Computational Chemistry. John Wiley & Sons. Chichester (UK), 1999.

[2] T. Born-oppenheimer, "Crude Born-Oppenheimer Approximation," pp. 4–6, 2021.

[3] F. Jensen, "Contracted basis sets for density functional calculations: Segmented versus general contraction," J. Chem. Phys., vol. 122, no. 7, 2005, doi: 10.1063/1.1844313.

[4] A. Martín Pendás, "Análisis de la densidad electrónica", unpublished, pp. 22–34, 2005.

[5] "List of quantum chemistry and solid-state physics software - Wikipedia." https://en.wikipedia.org/wiki/List_of_quantum_chemistry_and_solid-state_physics_software (accessed Jul. 08, 2021).

[6] M. S. Gordon and M. W. Schmidt, "Advances in electronic structure theory: GAMESS a decade later," Theory Appl. Comput. Chem., pp. 1167–1189, 2005, doi: 10.1016/B978-044451719-7/50084-6.

[7] K. Aidas et al., "The Dalton quantum chemistry program system," Wiley Interdiscip. Rev. Comput. Mol. Sci., vol. 4, no. 3, pp. 269–284, 2014, doi: 10.1002/wcms.1172.

[8] N. Tancogne-Dejean et al., "Octopus, a computational framework for exploring light-driven phenomena and quantum dynamics in extended and finite systems," J. Chem. Phys., vol. 152, no. 12, 2020, doi: 10.1063/1.5142502.

[9] Gaussian 16, Revision C.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2016

[10] M. J. Frisch, J. A. Pople, and J. S. Binkley, "Self-consistent molecular orbital methods 25. Supplementary functions for Gaussian basis sets," J. Chem. Phys., vol. 80, no. 7, pp. 3265–3269, 1984, doi: 10.1063/1.447079.

[11] W. J. Hehre, R. F. Stewart, and J. A. Pople, "Self-consistent molecular-orbital methods. I. Use of gaussian expansions of slater-type atomic orbitals," J. Chem. Phys., vol. 51, no. 6, pp. 2657–2664, 1969, doi: 10.1063/1.1672392.

[12] T. H. Dunning, "Gaussian basis functions for use in molecular calculations. III. Contraction of (10s6p) atomic basis sets for the first-row atoms," J. Chem. Phys., vol. 55, no. 2, pp. 716–723, 1971, doi: 10.1063/1.1676139.

[13] "Basis Set Exchange (BSE)," 2019. https://www.basissetexchange.org/ (accessed Jul. 08, 2021).

[14] The National Institute of Standards and Technology (NIST), "Computational Chemistry Comparison and Benchmark DataBase." https://cccbdb.nist.gov/ (accessed Jul. 08, 2021).

[15] E. R. Cohen et al., "Quantities , Units and Symbols in Physical Chemistry The first and second editions were prepared for publication by," Int. Union Pure Appl. Chem., vol. 233, no. 3, pp. 11–13, 2007.

[16] U. Relative, "Fundamental Physical Constants — Non-SI units Fundamental Physical Constants — Non-SI units," vol. 1, pp. 1–2, 1998.

8. APPENDIX
8.1. Table of different units and constants

There are very common conversion units required though all the thesis in order to do the calculations, such us: [15]

- $1Eh = 27.2113824$ eV $= 1$ a.u. $= 627.51$ Kcal/mol $= 2.1947 \times 10^5$ cm^{-1}.
- 1 Kcal/mol $= 4.184$ KJ/mol

And the most relevant constants and variables are also collected in the following table: [16]

SYMBOL	NAME	a.u. UNITS	SI UNITS	
m_e	Electron mass	1	$9,195 \times 10^{-31}$	Kg
e	Electron charge	1	$1,602 \times 10^{-19}$	C
$\hbar$	Reduced Plank's constant	1	$1,055 \times 10^{-34}$	Js
a_0	Bohr radius	1	$5,292 \times 10^{-11}$	m

8.2. Input and output ORCA file

The input file has the general structure:

#comments → whatever can be written without altering the calculations.

! Keywords → Method, basis sets…

*xyz Charge. Mult

Atom A   0.000   0.000   0.000

Atom B   0.000   0.000   0.000

*

Keywords can be given in any order and using more than one line. Coordinates can be expressed in Cartesian or with z-matrix formalism. In Table 3, they are expressed in a.u. units (Bohr). Two of the keywords used in all ORCA inputs are: LARGEPRINT which allows to generate plot orbitals diagrams which can be visualize in the Avogadro's program, and AIM which permit to transfer the output file to the AIMALL program.

Table 3. Input files for the LiH and HF molecules systems.

| LiH system	HF system
# avogadro generated ORCA input file # Basic Mode # ! RHF OPT def2-TZVP LARGEPRINT AIM * xyz 0 1 Li -1.63828 1.94862 0.00000 H -0.04828 1.94862 0.00000 *	# avogadro generated ORCA input file # Basic Mode # ! RHF OPT def2-TZVP LARGEPRINT AIM * xyz 0 1 F -3.46473 2.89970 0.94674 H -2.58473 2.89970 0.94674 *
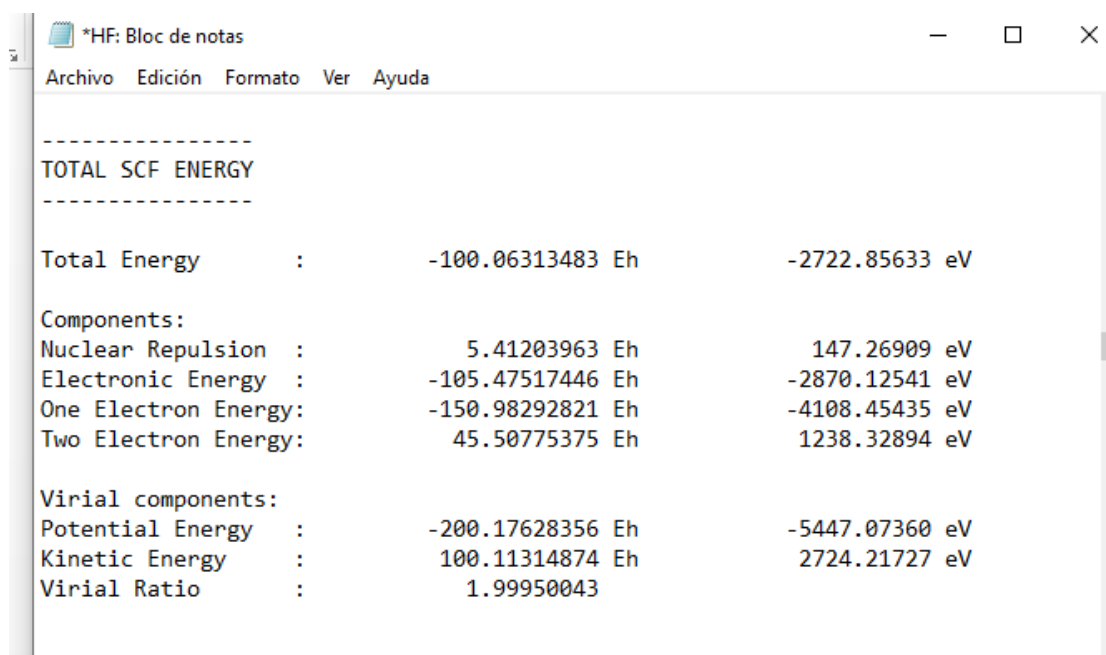
Figure 17. Example of a part of the output file of the HF system.

8.3. Operation of the Avogadro program

In order to create an input file, the Avogadro program is a very useful program for non-specialists. Therefore, if an input of a molecule is needed for any calculation, once the molecule is drawn, the option *Orca>Generate Orca* input appears in extension. Then the window shown in the image below opens and the input is generated with the possibility of choosing the method, the calculation procedure or the basis set, among others. Finally, the input is saved on the computer and is ready to be used in the ORCA program.
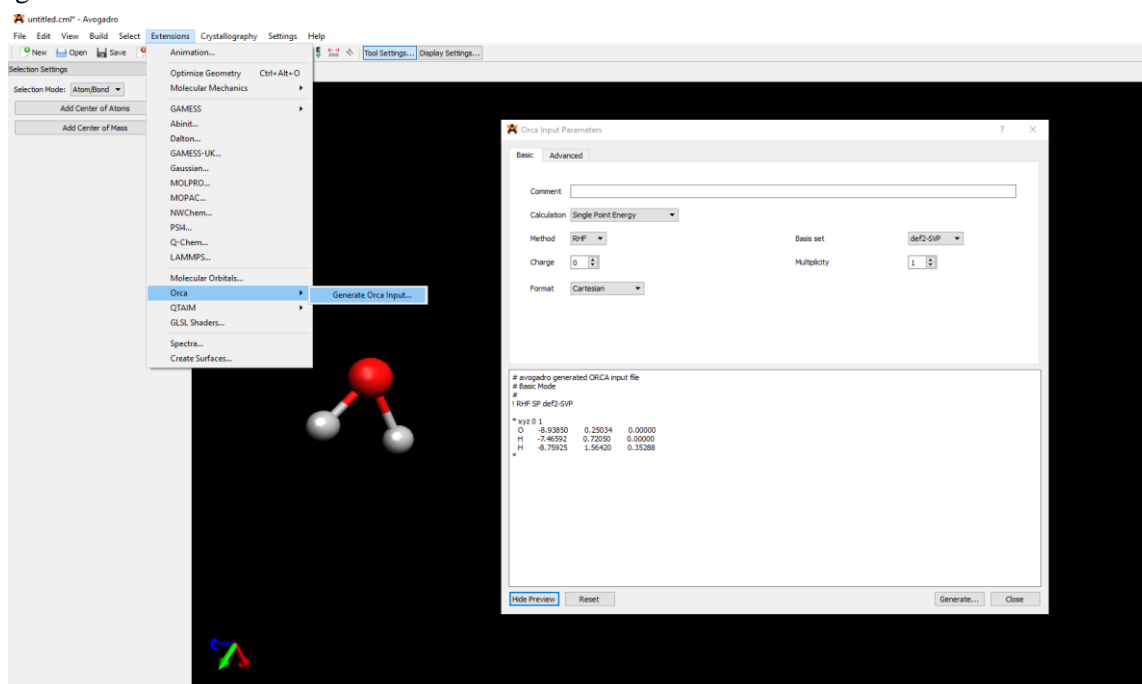
Figure 18. Screenshot of the Avogadro program where it is seen how to create an input file.

8.4. Optimized geometries

Table 4. Optimized geometries for the def2-TZVP basis function used in all systems analyzed.

SYSTEM	GEOMETRY (xyz) in a.u.		
LiH	Li -1.63828	1.94862	0.00000
	H -0.04828	1.94862	0.00000
BeH₂	Be 0.61130	0.40937	0.00000
	H 1.78530	0.40937	0.00000
	H -0.56270	0.40937	0.00000
BH₃	B -2.33924	1.40825	0.00000
	H -1.23124	1.40825	0.00000
	H -2.89324	0.45888	0.13945
	H -2.89324	2.35762	-0.13945
B₂H₄	B -7.32743	1.81762	0.00000
	B -4.81703	0.93337	0.00000
	H -8.14274	1.17046	-0.37961
	H -7.55718	2.83289	0.37961
	H -4.58728	-0.08190	-0.37961
	H -4.00172	1.58053	0.37961
B₂H₆	B -5.19080	0.95570	0.54299
	H -5.74480	0.04569	0.84735
	H -5.74480	1.86571	0.23864
	B -1.21299	1.27701	0.49323
	H -0.10499	1.27701	0.49323
	H -1.76699	2.01544	1.10599
	H -3.22316	0.29479	-0.05317
	H -3.38002	2.09204	0.38919
CH₄	C 2.36929	1.75677	0.89574
	H 3.43929	1.75677	0.89574
	H 2.01263	1.33406	-0.02023
	H 2.01262	2.76138	0.98764
	H 2.01262	1.17488	1.71981
C₂H₄	C 1.22943	2.19721	0.75198
	C 2.91751	1.47735	0.94979
	H 0.49090	1.65721	0.27449
	H 1.02414	3.13969	1.11887
	H 3.12280	0.53487	0.58290
	H 3.65604	2.01735	1.42728
C₂H₂	C -6.65908	1.08075	0.00000
	C -3.75745	1.06437	0.00000
	H -7.65306	1.08636	0.00000
	H -2.76346	1.05876	0.00000

NH₃	N	0.84441	0.46641	0.17059
	H	1.86441	0.46641	0.17059
	H	0.50442	-0.49525	0.17197
	H	0.50441	0.94605	-0.66292
N₂H₄	N	3.45436	2.97627	-0.27312
	N	2.37279	2.01639	-0.16216
	H	3.31898	3.58845	0.52669
	H	3.49346	3.31708	-1.10741
	H	2.31780	1.57258	0.75458
	H	2.49227	1.30120	-0.87953
H₂O	O	0.20611	2.16372	1.23478
	H	1.17611	2.16372	1.23478
	H	-0.11722	1.48386	1.84646
H₂O₂	O	1.28549	2.24256	0.77401
	O	-0.13635	2.85501	0.60877
	H	1.47224	1.65262	0.02702
	H	-0.32310	3.44495	1.35576
HF	F	-3.46473	2.89970	0.94674
	H	-2.58473	2.89970	0.94674

8.5. Total energy values

Table 5. Total energy values (in E_h) with optimized geometry for all systems analyzed with the different basis functions.

BASIS SETS	LiH	BeH$_2$	BH$_3$	B$_2$H$_4$	B$_2$H$_6$	CH$_4$	C$_2$H$_4$	C$_2$H$_2$	NH$_3$	N$_2$H$_4$	H$_2$O	H$_2$O$_2$	HF
3-21G	-7.929	-15.673	-26.237	-51.320	-52.475	-39.977	-75.924	-75.924	-55.872	-110.546	-75.586	-149.946	-99.460
6-31G	-7.979	-15.759	-26.377	-51.596	-52.753	-40.181	-76.324	-76.324	-56.166	-111.114	-75.985	-150.710	-99.983
6-31G*	-7.981	-15.766	-26.389	-51.617	-52.775	-40.195	-76.346	-76.346	-56.183	-111.163	-76.009	-150.761	-100.001
6-311G	-7.985	-15.764	-26.383	-51.606	-52.766	-40.188	-76.351	-76.351	-56.181	-111.155	-76.011	-150.759	-100.022
6-311G*	-7.985	-15.769	-26.393	-51.624	-52.786	-40.203	-76.373	-76.373	-56.201	-111.195	-76.032	-150.805	-100.034
def2-SVP	-7.979	-15.753	-26.370	-51.579	-52.741	-40.169	-76.391	-76.287	-56.149	-111.088	-75.961	-150.652	-99.933
def2-TZVP	-7.985	-15.769	-26.400	-51.635	-52.800	-40.213	-76.391	-76.391	-56.219	-111.221	-76.059	-150.843	-100.064
def2-QZVP	-7.987	-15.773	-26.402	-51.639	-52.794	-40.217	-76.396	-76.396	-56.225	-111.231	-76.067	-150.853	-100.071
cc-pVDZ	-7.984	-15.767	-26.391	-51.620	-52.781	-40.199	-76.367	-76.367	-56.196	-111.182	-76.027	-150.785	-100.019
cc-pVTZ	-7.987	-15.771	-26.400	-51.635	-52.800	-40.213	-76.387	-76.387	-56.218	-111.219	-76.058	-150.838	-100.058
cc-pVQZ	-7.987	-15.773	-26.402	-51.638	-52.591	-40.216	-76.394	-76.394	-56.224	-111.229	-76.066	-150.850	-100.068