Reviews in Theoretical Science
Vol. 2, pp. 1–25, 2014
(www.aspbs.com/rits)

Molecular Modeling: Origin, Fundamental Concepts and Applications Using Structure-Activity Relationship and Quantitative Structure-Activity Relationship

Cleydson Breno Rodrigues dos Santos[1,2,3,*], Cleison Carvalho Lobato[3], Marcos Alexandre Costa de Sousa[3], Williams Jorge da Cruz Macêdo[4], and José Carlos Tavares Carvalho[1,2,3]

[1]*Postgraduate Program in Biotechnology and Biodiversity-Network BIONORTE, Federal University of Amapá, Rod JK km 2, Macapá–Amapá, 68902-280, Brazil*
[2]*Laboratory of Modelling and Computational Chemistry, Federal University of Amapá, Rod JK km 2, Macapa, Amapá 68902-280, Brazil*
[3]*Laboratory of Drug Research, School of Pharmaceutical Sciences, Federal University of Amapá, Rod JK km 2, Macapá–Amapá, 68902-280, Brazil*
[4]*Secretary of State for Public Health, Central Laboratory of the State of Pará—LACEN/PA, Rod Augusto Montenegro, km 10, Icoaraci, 66823-010, Belém-Pará-Amazon-Brazil*

Molecular modeling is an important tool to aid the understanding of the fundamental concepts of structure-activity relationships, and to elucidate the mechanism of action of drugs (drug-receptor interaction), used in the teaching-research-extension. The physico-chemical properties as well as three-dimensional visualization of electronic and steric molecular properties elucidation of the interaction between drugs and macromolecules target can be calculated and/or suggested by molecular modeling programs. In this work we show that studies of structure-activity relationships are of great importance in modern chemistry, biochemistry, molecular biology, and other fields of knowledge of health sciences. In order to obtain a significant correlation, it is essential that the descriptors are used appropriately. Thus, the quantum chemical calculations are an attractive source of new molecular descriptors that can, in principle, express all the geometric and electronic properties of molecules and their interactions with biological receptor.

KEYWORDS: Molecular Modeling, Quantum Chemical Methods, Quantum Chemical Descriptors, Multivariate Analysis, Structure-Activity Relationships, Artemisinin.

CONTENTS

*Author to whom correspondence should be addressed.
Email: breno@unifap.br
Received: 30 March 2013
Accepted: 17 April 2013

Cleydson Breno Rodrigues dos Santos Has graduation in Bachelors Chemistry in Federal University of Pará–UFPA (2006), Masters in Chemistry from the UFPA (2008) and Doctorate by Bionorte Network in Biotechnology in Federal University of Amazonas–UFAM. Is professor Level II and is part of the research group on Drugs and Applied Computational Chemistry, Federal University of Amapá–UNIFAP. Performs works in the area of Computational Chemistry and Molecular Modeling, aiming to simulate drug-receptors interactions with aim of clarifying the possible mechanisms of action of new drugs of biodiversity Brazilian Amazon. Has experience in the area of Theoretical Chemistry, Computational Chemistry and Molecular Modeling, with an emphasis on Teaching-Research-Extension.

Cleison Carvalho Lobato

Marcos Alexandre Costa de Sousa

Williams Jorge da Cruz Macêdo

José Carlos Tavares Carvalho

1. ORIGIN OF COMPUTATIONAL CHEMISTRY

The name of computational chemistry can be understood, in general, the field of computational methods applied chemistry and related fields. Since there is a large number of computational procedures, computational chemistry presents itself as an interdisciplinary field, by branching out into different areas that, traditionally, it is customary to divide the chemical (Physical Chemistry, Organic Chemistry, Pharmaceutical Chemistry, Biochemistry, Inorganic Chemistry, Analytical Chemistry, chemical Technology and other knowledge areas of human health, as well as the Molecular Biology and Computer Science). Over the past 25 years many application areas of computational chemistry suffered a significant development due to the emergence of new computers, sophisticated software and a better understanding of the basic principles.[1]

The potential offered by current technology hardware and software led to the development of a variety of techniques for numerical and symbolic computations. These methods have opened many application areas and the spectacular increase the potential of this technology to make computational chemistry one of the most promising interdisciplinary toward the twenty-first century.[1]

The Computational Chemistry is a vector in the direction of unification, by identifying himself as an interdisciplinary field based on a ubiquitous laboratory tool—the computer—it is a tool capable of treating both quantum and classical modeling, geometry and chemical information. This essential characteristic and perhaps only introduced a new scientific community based on the ability of the computer to solve chemical problems and the meeting of scientists from different fields traditionally separated. The interaction between them has led to scientific and technological achievements of great importance as well as the encouragement of new generations of young researchers.[1]

Improving the understanding of the students about the concepts of chemistry has been a major goal of researchers in the Teaching of Chemistry (and Science in general) during the last decades. A resource that has been used since the decade of 60 as a tool for learning is the computer, as can be seen in the pioneering work of Atkinson (1968),[2] Suppes Morningstar (1968).[3] The possibilities of using this technology are very large, and over the years it has evolved and changed, as we can see in some articles selected to investigate the presence of this theme in the world of chemistry.[4]

At the beginning of the decade 70, at the University of Lancaster, England, a course in Quantum Chemistry has been carefully organized by chemical Duke[5] in order to regain lost motivation due to failed attempts of computer scientists that by introducing computational techniques in teaching chemistry, not taking matters relevant to chemistry. This placement reveals that, in 1972, the use of computers in chemical education was already the focus of research, but was not being conducted in a way to encourage students. In their experiment, Duke a program used to calculate the properties of aromatic compounds by the molecular orbital method.

Fortunately the framework of dissatisfaction was overcome notes as Duke (1972)[5] "This practical experiment proved successful in that most students learned a substantial body of knowledge on the application of molecular orbital theory to chemistry organic and seems rather interested." Years later, the same topic was addressed in the University Chemical Laboratory, Cambridge, UK, where microcomputers were used as a teaching tool for Molecular Orbital Theory. "We believe that such programs are of great help in teaching theoretical chemistry."[6]

The Royal Swedish Academy of Sciences awarded the Nobel Prize in Chemistry 1998 researchers: Walter Kohn (University of California, Santa Barbara, California, USA) for his contribution to the development of Density Functional Theory and Pople (North Western University, Evanston, Illinois, USA) for his contribution to the development of computational methods in quantum chemistry. With these developments that were started from the 1960s, the chemistry reaffirms itself as an exact science, computable.[7]

The IUPAC defines computational chemistry as follows: "Molecular aspects of research as made practical by the use of computers."[8] Since then, progress in development of software and hardware combined with a steady reduction of cost of materials informatics, computational chemistry makes one of the most promising areas of this new century. More recently supramolecular chemistry (compounds formed by several molecules) was approached simultaneously by researchers from the Université Louis Pasteur in France and Novosibirsk State University in Russia in 2000, through a course-based model CAI, instruction aided by computer (Computer Aided Instruction), which, according to the researchers Varnek et al.[9] "allows the visualization of complex structures and performing calculations modest occurs while concurrently reading the text, and computer-assisted courses are easily upgradeable, which is especially important for fields (science) that expand and develop rapidly."

In the same year, the Department of Chemistry and Biochemistry Drigham Young University in Utah, USA, has developed a method of teaching that included a package with animations for understanding Molecular Orbital in organic reactions. The purpose of the use of simulations was to facilitate the visualization and understanding of this topic. With this method the authors Fleming, Hart and Savage (2000)[10] concluded that "students may benefit from three-dimensional computer representations of chemical events."

The development and subsequent use of software in the classroom helps solving chemical problems, and versatility

of computational chemistry not only allows its application in teaching chemistry as well as in research and development laboratories and industries. The molecular modeling, for example, is an important tool in developing pharmaceuticals and can be used in rational design of new drugs.

According to Rodrigues (2001)[11] "Molecular modeling provides important information for the process of drug discovery. It allows to obtain specific properties of a molecule that can influence the interaction with the receptor."

The development of software for computational chemistry followed two separate paths, with the current complement of a third. Initially, we worked with the development of codes capable of solving the equations of quantum mechanics to atomic systems strictly based on "first principles." These calculations involve the resolution of a large amount of integrals, even after applying the usual simplifications, such as the Born-Oppenheimer approximation (roughly speaking can be seen as the separation of electronic and nuclear motion) and the absence of relativistic effects, among others. From this, we have reached an initial commitment resulting in a method called SCF (Self-Consistent Field), virtually omnipresent in all computational chemistry programs. This method leads to a new approximation that ignores the electron–electron interaction (which is "re-introduced" in several ways, the most popular is via perturbation theory, methods abbreviated MP (n), where the electron correlation effects (interaction between electrons) is introduced via TP–RS (Perturbation Theory Rayleigh-Schrödinger).

The second development path followed not from first principles, but the approaches and results of calculations parameterized by atomic groups in order to save the computational steps. Aiming to "embrace" fast calculation of larger systems, and therefore greater chemical interest. The semi-empirical methods parameterize several calculations that the *ab initio* methods necessarily repeated for each new system. So make up scales of power calculation, with a natural loss of accuracy.

Among the most popular methods, are the INDO, CNDO, AM1 and PM3 methods longer used. Finally, in recent decades, a new scale of calculation was introduced, with the MD simulation methods (Molecular Dynamics) and Monte Carlo,[12] which basically does not treat the chemical system as obeying rules of quantum mechanics, but classics like atomic systems using classical potentials to model the behavior of these systems. Applies to systems with thousands of atoms, typically for modeling systems of biological interest or solvation, where this approach is justified by its chemical interest. It is common for software currently existing computational chemistry methods or methods using these products, which we will not discuss here. Some of these methods can be used both for research but also for teaching chemistry.

2. COMPUTATIONAL METHODS USED IN THE CALCULATION OF MOLECULAR PROPERTIES

The methods of quantum chemistry can be applied to quantitative structure-activity relationships (QSAR) for the direct derivation of the electronic descriptors from the molecular wave function. In general, the more rigorous theoretical treatment does not use empirical parameters and is called *ab initio*. Although this type of method provides relatively accurate information about the electronic behavior, he is, in operational terms, slower and more expensive. Therefore, several semi-empirical methods have been developed which are based on certain assumptions which serve to simplify the calculations and use certain parameters obtained from experimental data. Note that the accuracy of these methods is related to the error associated with the selected basis set and the level of treatment of electron correlation. Therefore, there are several computational methods used in the calculations of molecular properties, among them are: the semi-empirical, *ab initio* and Density Functional Theory (DFT).

2.1. Semi-Empirical Methods

The semi-empirical methods using the same formalism as mechanical-like employing basis sets including only the electrons of the valence shell of the system.

The reason behind this approach is that the electrons involved in chemical reactions and other phenomena are the intermolecular electron from the valence shell.[13]

Thus, the great advantage of the semi-empirical methods compared to *ab initio* methods is the higher processing speed, since the calculations are simplified, reducing the cost of memory and computational time. Unlike *ab initio* methods, methods employing semi-empirical is empirical parameters, i.e., derived from experimental data such as geometry of equilibrium heat of formation, molecular dipole moment and ionization potentials, or previously calculated by the Schrödinger equation, allowing some integrals present in the *ab initio* method are not calculated simplifying computations.[14, 15]

The first method using this approach is the CNDO (Complete Neglect of Differential Overlap), in which the atomic orbitals are considered in evaluating spherically symmetrical electron repulsion integral. Other methods also use these approaches such as INDO (Intermediate Neglect of Differential Overlap), and NDDO (Neglect of Diatomic Differential Overlap).[13]

The semi-empirical methods most commonly used are AM1 (Austin Model1)[16] and PM3 (Parametric Method 3),[17] both methods incorporate approaches very similar, but differ in the parameterization.

Recently, the AM1 method was subjected to a re-parameterization for the atoms of H, C, N, O, P, S, F, Cl, Br and I, resulting in method RM1 (Recife Model1), with minor miscalculations than those generated by AM1 and PM3.[18]

2.2. *Ab Initio* Methods

The term Latin means *ab initio* "from the top" or "from the fundamental principles," i.e., calculations are performed from fundamental physical constants using exact equations, involving a total electronic population of molecule without the use experimental parameters and without additional approaches. The first method for calculating the electronic structure was the Hartree-Fock (HF) which employs the full Schrödinger equation to treat all the electrons in a chemical system.[14]

This model employs sets of basis functions (basis set) calculations such as the functions of the Slater Type (STO) and Gaussian Functions (GTO 3-21G, 6-31G). These bases have several deficiencies minimum and to enhance them is the inclusion of polarization function (i.e., *p* orbitals represented by *).[13]

Thus, 6-31G* refers to basis set 6-31G with polarization function to heavy atoms (i.e., atoms different from hydrogen), 6-31G** refers to the inclusion of function for the polarization atoms hydrogen and helium. The 6-31G** basis is particularly useful where there are hydrogen bonds. Basis functions with partial polarization have also been developed, for example, 3-21G* which is the same minimum 3-21G basis functions with partial polarization.[13]

Although the *ab initio* methods give a quantitative prediction of high quality for a wide variety of systems, they are time consuming and of high cost computation. A resource is commonly used to optimize the geometry with a set of base simplest and then perform calculations to "single point" with a more complete set of base allowing to determine the energy and other properties of a molecular system, using a more sophisticated calculation basis.[13]

2.3. Density Functional Theory

The Density Functional Theory (DFT) is a very successful formalism, where the main objective is to replace the wave function, used to write the electrons in methods like Hartree-Fock, the electron density. The HF calculations consider an average electron density, since the DFT calculations consider instant interactions of pairs of electrons with opposite spins.[19] It is an approach based on the theory of Hohenberg and Kohn which states that all properties of a system are functions of the charge density.

Thus, the Hohenberg-Kohn theorem allows to write the total electronic energy as a function of the electron density ρ:

$$E(\rho) = E_{KE}(\rho) + E_C(\rho) + E_H(\rho) + E_{xc}(\rho) \qquad (1)$$

where $E_{KE}(\rho)$ is the kinetic energy, $E_C(\rho)$ is the interaction term nucleus-electron, $E_H(\rho)$ is the Coulomb energy and $E_{xc}(\rho)$ contains the contributions of exchange and correlation.

The molecular orbital calculations of density functional are usually written as a linear expansion of atomic orbitals (ie basis functions) that can be represented using Gaussian type functions, Slater orbitals or orbital numeric.[19] Functional models density as well as the Hartree-Fock models are applicable in molecules of 50–100 atoms.[15] The exact function is not known, so there is a varied range of different functional that can provide different results for the same problem. The method B3LYP (Becke, Lee, Yang and Parr) is a hybrid method widely applied, where part of the functional is obtained by quantum mechanics (HF combines energy exchange with DFT exchange term) and part is parameterized (adds functional correlation).[20]

3. THE EVOLUTION OF COMPUTATIONAL CHEMISTRY

One of the most important advances in the design and discovery of new drugs has been the use of Molecular Modeling (MM). Currently, the MM is an indispensable tool not only in the process of drug discovery, but also in optimizing existing prototypes and the rational design of drug candidates.[14–23]

According to IUPAC, the MM is the investigation of molecular structures and properties by the use of computational chemistry and graphical visualization techniques, aiming to provide a three-dimensional representation, under a given set of circumstances.[14] The nature of the molecular properties used and the extent to which they describe the structural characteristics of the molecules may be related to biological activity, which is an important part of any QSAR study.

The great development of MM in recent years is due largely to the advancement of computational resources in terms of hardware (calculation speed) and software (computer programs), in addition to advances in computational chemistry, nuclear magnetic resonance, ray crystallography X-biochemistry and molecular biology. This allowed large contribution in the discovery of drug candidates, leading to rapid progress in research and attracting the interest of both academic environment and pharmaceutical industries.[21,23]

Silva et al.[24] used density functional theory (DFT) calculations (B3PW91/DGDZVP) to determine ^{13}C and ^{1}H nuclear magnetic resonance (NMR) chemical shifts for the two dihydrochalcones: 3,4,5-trimethoxydihydrochalcone and 2,3,4,4-tetramethoxydihydrochalcone. The experimental and theoretical NMR data were analyzed by simple linear regression and the more relevant parameters were selected. In additional, other statistical parameters (correlation coefficients, significance and predictability) were available to judge the quality of the calculations. Finally, the statistical analysis show good correlation experimental and theoretical NMR data with high predictive power.

The computers increase considerably the possibilities for scientific research in drug discovery and thereby can allow chemists collect, store, manipulate, analyze and visualize data. QSAR analysis may be used for planning of biological properties such as potency, efficacy, selectivity and bioavailability of a drug.[25]

Due to recent advances in computational area and the development of efficient algorithms for calculation, a great advance was also verified in the development of quantum chemical calculations. The *ab initio* and semi-empirical methods quantum chemical molecular parameters provide realistic in a short period of time. Quantum-chemical calculations are great source of molecular descriptors that can, in principle, express many geometric and electronic properties of molecules and their interactions. In fact, several recent studies on SAR and QSAR employing quantum-chemical descriptors alone or combined with conventional descriptors.[26, 27] The Quantum Chemistry provides a more accurate and detailed description of the electronic effects when compared to empirical methods.

Ibrahim et al.[28] used Semiemperical molecular modeling technique is applied to assess the interaction of amino acids (alanine, asparagines, aspartic, arginine, cysteine, glutamine, glycine and tryptophan) with chitosan. Results indicate the selectivity of chitosan furthermore we introduce the site whereas amino acid could interact with chitosan. Chitosan is interacting with amino acid through NH_2 group. It is concluded that chitosan is acting with amino acids like protein interaction which dedicate chitosan for many applications in the biological system.

Bayuelo et al.[29] proposed methodology to the alternative way to determining of local reactivity indexes using MQS based on the Hirshfeld partitioning. In addition contribution was postulated in this news perspectives in the field such as chemical reactivity, chemical potential, hardness and electrophilicity relative, alternatives to the traditional (chemical potential, hardness and electrophilicity) chosen in the conceptual DFT which allowed us to report the local reactivity indexes proposed with the global that considering the pursuit of local descriptors of reactivity on supported ideas of MQS in Cycloaddition Reactions.

Methods of Quantum Chemistry and molecular modeling techniques allow the definition of a large number of atomic and molecular properties characterizing properties related to reactivity, form and mode of binding fragments and molecular substituents. Due to the large information content contained in many molecular descriptors, the use of quantum-chemical descriptors in QSAR studies has two main advantages: the compounds and their substituents and several fragments can be directly characterized based only on their molecular structures; mechanism proposed action can be justified directly in terms of the chemical reactivity of the compounds studied. Consequently, the obtained QSAR models include information about the nature of intermolecular forces involved in determining the biological activity of the compounds under study.[30]

3.1. Design and Conformational Analysis

3.1.1. Design and Three-Dimensional Visualization

Several programs two-dimensional of design of molecules are available and easy to use, as ChemWindow, Isis Draw, ChemDraw[31] and Chem3D.[32] They allow the preparation of figures and diagrams with desired quality and accuracy and facilitate the documentation and scientific communication.

The software ChemSketch 12.00[33] is an advanced design that provides chemical molecular properties, optimization and 3D visualization, ability to name the molecules, as IUPAC, and still has a large database of chemical structures and laboratory materials. The software automatically calculates the valence of each atom and restricts the construction of the molecule based on the octet rule, unless instructed to do this restriction. Then is possible to request the construction of 3D spatial form of the species studied, which triggers another window where the academic can rotate tridimensionally the species studied, in addition to observing these species in different visualizations with possibility to visualize bonds and spatial arrangement of species prominently in each of these representations.

Some programs allow the calculation and representation of various molecular properties, including formula and molecular mass, exact mass and elementary theoretical analysis. More complete programs such as ChemDraw Ultra,[34] provide additionally the correct chemical name (IUPAC) of chemical compounds and can predict the corresponding chemical shifts of ^{1}H and ^{13}C NMR, melting points and freezing, log P, molar refractivity and heat of formation.[35]

The design and visualization of 3D drugs, with steric factors relevant to biological activity, are important for analysis of the size, volume and shape of the molecules.[35]

The Molekel is a free software multiplatform molecular visualization. It was originally developed at the University of Geneva by Flükiger in the 1990s for Silicon Graphics computers. In 1998, Stefan Portmann took responsibility and released version 3.0. The version 4.0 was almost one version of the platform independent. Other developments lead version 4.3, before Stefan Portman moved and stopped developing the codes. In 2006, the Swiss National Supercomputing Centre (CSCS) restarted the project and version 5.0 was released on December 21 of the same year (FLUKIGER, 2001).[36]

3.1.2. Conformational Analysis and Energy Minimization

To obtain the conformational analysis and energy minimization, we can cite: the program Chem3D[32] widely used in studies with this objective.[35] However, other programs like Molecular Modeling Pro,[37] ChemSite (ChemSW),[38] Alchemy, Sybyl, ChemX, cache and WebLab Viewer are also available.

In the area of molecular modeling, graphics construction and projects of drugs, the program Hyperchem[39] for being a tool specializing in 3D structures of interest to the medical, pharmaceutical and organic chemistry. The program lets you design complicated molecules. This software

is also an alternative in the field of spectroscopy, which besides the ability to simulate a priori by the NMR spectra quantum methods, contains a database of approximately 10.000 molecules applicable to macromolecules as well as small molecules. The software also includes animations, and quantum chemical calculations and molecular mechanics.

Conformational analysis of a molecule is performed by rotating a binding with parallel change of torsional angles, and calculation corresponding of the steric energy, due to spatial overlap of atoms unlinked and barriers torsional rotation.[40]

Molecules designed three-dimensionally are not necessarily the most stable conformation. During the generation of a particular structure, distortion occurs in the molecule, with formation unfavorable of lengths, bond of angles and torsional angles. Atoms do not interact also-linked in the same region of space and cause steric demand and electrostatic. To correct these distortions molecules are optimized by energy minimization process, from two mathematical models (i) molecular mechanics or (ii) quantum mechanics. Interactions occur unpredictably related to overlapping molecular orbital, the electron density distribution or steric interference can be solved by computational methods. The energy minimization and conformational analysis are used interactively to optimize the geometry of a molecule.[40]

The choice of method for energy minimization depends on factors related to the size of the molecule, parameters of availability and stored data and computational resources. Molecular models generated by the computer are the result of mathematical equations that estimate the positions and properties of the electrons and nuclei, the calculations exploit experimentally, the characteristics of a structure, providing a new perspective on the molecule.[40]

The three-dimensional structural representation of drugs into computer programs and the construction of molecular models are important for learning the geometric characteristics and molecular essential for the biological activity of some drug classes. In the stage of design and three-dimensional visualization should be gather fundamental knowledge of organic chemistry such as stereochemistry, nomenclature and reactivity to understanding the structure-activity relationship of drugs.

In quantum chemistry the softwares more used are Gaussian and GaussView that uses the laws of quantum mechanics to predict the energies, structures and properties and vibrational frequency of molecular systems (FRISCH, 2003).[41]

The GaussView 5.0 is a program that can work on Windows and responsible for building the structures under study, by viewing these as well as for generating the input of the species under study for the program calculations–Gaussian 03W. This includes an advanced molecular modeler, which can be used for construction and molecular dimensions of the three test.[41]

The Gaussian 03W is a program that can work on Windows and Linux that performs computations used in the study of reaction mechanisms, equilibrium geometries of neutral molecules, radicals and ions, and the determination of physicochemical parameters. Appreciates structure, reactivity, thermodynamic properties, energy barriers (transition states), conformational analysis, employing the optimization of molecules and theoretical calculations of vibrational spectra. From the optimization is obtained the most appropriate structure to the molecule, whereas the lengths and bond angles and power stabilization calculated by E(RB = HF-LYP) ua.[41]

The density functional theory (DFT) is based on the electron density. She gives the distribution of charges in a molecule to assume an approximate Hamiltonian with interaction between pairs exclusively under the Born-Oppenheir approximation and after neglect relativistic effects. The DFT has great computational speed while providing good accuracy.[19]

There are several functional in the DFT, in this work we used the B3LYP, which is a functional hybrid. In 1988, the model BLYP was proposed by Becke–Lee–Yang, but with the increasing complexity of the compounds to be studied today, the trend is the use of B3LYP, this is a model that reflects the combination of term LYP correlation with the density functional exchange Becke, B3. In this introduction there are three parameters calculated by Becke and determined by settings using the Hartree-Fock method, which is based on corrections heat of formation for a series of molecules.[20]

The DFT reputation as conjecture that the exact function is not known, therefore the calculated energy is only a good approximation. Nevertheless, the existing functional not give satisfactory results in cases of weak interactions, anions (situations with scattering electronic cloud) and electron delocalization.

4. MOLECULAR DESCRIPTORS

Obtaining properties (descriptors) depends on the molecular level theory and method, and represents a means of chemical information contained in the molecular structure of the compound studied. This information is transformed and encoded for lots of problems chemical, pharmacological and toxicological studies on the relationship between quantitative structure-activity and structure-property (QSPR and QSAR). The molecular properties take into account different aspects of chemical information, this information can be through experiments or theoretical calculations simple counting, consider the entire molecule, fragments or functional groups, knowledge of the 3D structure of the molecule or molecular graphics his or her simply formula, information defined by scalar values, vectors or scalar fields.[42]

Recently Santos et al. (2013),[43] published in the Journal of Computational and Theoretical Nanoscience,

a prestigious journal in the interdisciplinary area, where the impact factor is 0.932 (2011) an article entitled "Validation of computational methods applied in molecular modeling of artemisinin with antimalarial activity." Where we propose a combination of chemical quantum methods and multivariate analysis to study the geometric parameters of artemisinin in the region endoperoxide of the ring (1, 2, 13–trioxane), in order to be effective in selecting the method and level of theory when compared with data crystallographic, aiming to classify and correlate. The most important geometrical parameters selected by principal component analysis (PCA) were O13C12, O1O2C3, C3O13C12C12a and C12C12aO1O2. The results of PCA showed that the model was built with three main components (3PCs), explains 97.0861% of the total variance. The level of theory HF/6-31G** show high similarity with the experimental data assuming that the combination of *ab initio* method can be used for modeling the molecular and for futures calculations of molecular properties structure of artemisinin and its derivatives antimalarial drugs with mechanism of action in the ring region endoperoxide.

4.1. Map of Molecular Electrostatic Potential

The charge density of a chemical species describes the distribution of electrons responsible for the chemical behavior of each species. To estimate the magnitude of this charge is not an easy task, considering that a molecule is a dynamic system and not a simple arrangement of protons and electrons with positive charges with negative charges. Thus, classical mechanics is unable to explain this type of system. There are several methods of calculating atomic charges available computer programs.[44] The difficulty lies in the fact that calculation of the loads are not obtained directly from the wave function.[44] The method of calculation of atomic charges most popular is the Mulliken population analysis,[45] but it is a method for designating arbitrary loads, since, for performing these calculations, the charge density between two atoms is split evenly, not taking into account the electronegativity of these atoms. Another method to evaluate the distribution of the load is to adjust the molecular electrostatic potential, which is a property directly obtained from a calculation SCF (Self Consistent-Field) at a series of points located at the centers atoms. Therefore, we define a set of points around the molecule for calculating the electrostatic potential and a further adjustment is made to model the point loads. From this reasoning, there are many methods derived from the electrostatic potential for the calculation of atomic charges.[44] One method used to calculate the loads arising from the electrostatic potential was developed by Chirlian and Franci.[46] According to this method, the potential is determined for a selected number of dots arranged spherically around the molecule. The interactions between ligand and receptor are closely linked to biological mechanisms electrostatic—attraction, repulsion, load transfers. Thus, charges are calculated at important positions and substituent groups in the molecules.

The MEP is one of the descriptors used in most studies and aims to reveal the size and location of the total molecular electrostatic potential in the molecule. The surfaces of three-dimensional maps of molecular electrostatic potential (MEPs) after superimposition is generated in the molecule of a positively charged particle that under the contact surface of the molecule van der Waals repulsion shows a region representing a potential positive, blueness and the region of negative potential in the molecule, represented by red color.

The electronic parameters are one of the main factors that govern drug-receptor interaction, in this sense, the map of molecular electrostatic potential (MEP) may be an alternative approach in order to understand the electrostatic contribution of these derivatives for biological activity.

To construct the MEP requires three steps: the construction of the surface electron density of the molecule, the construction of the surface electrostatic potential and applying colors to denote the surface obtained potential values.

One of the frequent topics of theoretical chemistry is research to improve methods to elucidate the behavior of molecules and other reactive chemical species. Among the numerous existing reactivity indices the molecular electrostatic potential $V(r)$ that is generated around a molecule by its nuclei and electrons, is known for being a real physical property can be determined experimentally by diffraction methods, as well as computationally.[25]

The MEP at a given point (x, y, z) in the vicinity of a molecule is defined in terms of the interaction energy between the electrical charge generated from the molecule's electrons and nuclei and a positive test charge (a proton) located at r. For the studied compounds, the $V(r)$ values were calculated as described previously using Eq. (2)

$$V(r) = \sum_A \frac{Z_A}{|R_A - r|} - \int \frac{\rho(r')}{|r' - r|} dr' \qquad (2)$$

where Z_A is the charge of nucleus A, located at R_A, $\rho(r')$ is the electronic density function of the molecular, and r' is the dummy integration variable.

The molecular electrostatic potential (Fig. 1) has been an important tool for analyzing processes of recognition of a molecule by other types of interactions such as drug-receptor and enzyme-substrate because of its potential by being a kind that sees another in a process biological recognition.

In Figure 1 the MEP intended to identify and key features evaluate the compound from qualitative comparisons of the form of molecular electrostatic potential in the region of the ring 1, 2, 4-trioxane artemisinin (red color). Since the geometric shape of the electrostatic potential in the region of the ring 1, 2, 4-trioxane is similar for

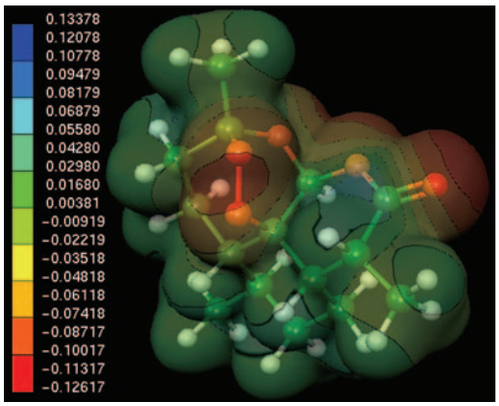

Fig. 1. Electrostatic Potential Map of artemisinin was calculated using the Hartree-Fock (HF) method and HF/6-31G** level of theory. The MEP was realized by the Molekel program.[38]

all active compounds, that is characteristic according to the literature.[47] Therefore, compounds having some structural similarity may have electrostatic potentials that allow a being recognized by another, with similar biological activities.[25]

Structure-activity relationship (SAR) indicates the molecular structure modifications that increase the drug effectiveness. In general, reports show that these modifications are made throughout small changes in the leading compound structure, followed by trials in laboratory to quantify the variations in the biological activity due to changes in the molecular structure.[48]

4.2. Energy Frontier Orbital (HOMO and LUMO)

A second category of quantum-chemical descriptors widely used in studies SAR/QSAR is related to energy of the frontier orbitals (HOMO and LUMO). The reason for this relates to the fact that these properties provide information about the character electron donor and/or electron–acceptor compound and thereby forming a charge transfer complex (CTC).[49]

The energy of Molecular Orbital Highest Occupied Energy (HOMO) and the Molecular Orbital Energy of Lowest Unoccupied (LUMO) are quantum-chemical descriptors, which play an important role in chemical reactions and the formation of several complexes of charge transfer.

In Figure 2 we can see the region bounded by the HOMO orbital measuring the character electron–donor compound, and the LUMO measuring the electron acceptor character. From these definitions, two important features can be observed: the higher the energy of the HOMO, the greater the electron–donor capacity and the lower the energy of the LUMO lower the resistance to accept electrons. In this figure we observe that the HOMO is located in the region of the ring which is the trioxane pharmacophore of the active molecules. When the rings are aromatic substituent or have high electron density, such as

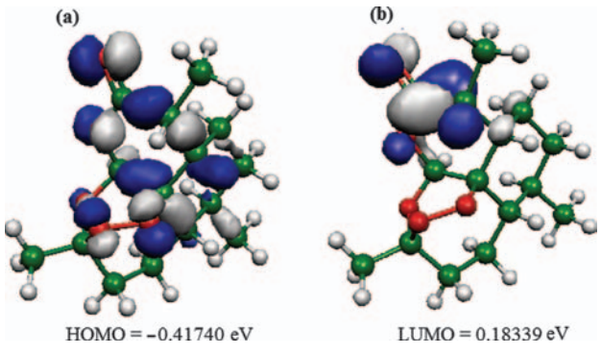

Fig. 2. Orbital energy Homo (a) and Lumo (b) were calculated using the Hartree-Fock (HF) method and HF/6-31G** level of theory. The Homo and Lumo orbitls were realized by the Molekel program.[38]

in carbonyls, amines and amides, the most pronounced HOMO will be strongly influenced to conduct stereo electronics side effects that might impair the pharmacological activity of the compound.

The energies of HOMO and LUMO have been used for some decades as indices of chemical reactivity and are commonly correlated with other indices, such as electron affinity and ionization potential.[50–54]

The energy of HOMO is directly related to the ionization potential of the compound and characterizes the ability of the molecule to perform nucleophilic attacks. The LUMO energy is directly related to electron affinity, characterized by the susceptibility of the compound in relation to attacks by nucleophiles.[55] The difference between the orbital energies of HOMO–LUMO (gap) is an important indicator of molecular stability. Molecules with low band gap value are generally reactive, while molecules with a high value of gap indicating high stability of the molecule in the sense of low reactivity in chemical reactions.[56]

$$\text{gap} = E_{\text{HOMO}} - E_{\text{LUMO}} \tag{3}$$

Lobato et al. (2012),[57] computational calculations performed with the aim of studying the reactivity and stability of isomeric products in reactions of halidrification in alkenes by analyzing the orbital border (HOMO and LUMO), index of chemical reactivity, affinity electronics and ionization potential. Where the product more stable, among the investigated was 2-Iodo-2-methylpropane having a more stable variation between normal and branched chain 1.067% stability.

4.3. Descriptors Polarizability (α), Hardness (η) and Molecular Softness (S)

Descriptors are very important in studies of SAR/QSAR, it can be correlated with lipophilicity, molar volume and impediments stereos, aiding the interpretation of the mechanisms of interaction between a compound and their respective biological receptor. The concept of chemical hardness and softness (or softness) molecular was formulated in accordance with the concept of acids and Lewis

bases.[58] For a basic "soft" (or softer) the donor atom has high polarizability and low electronegativity, and can be easily oxidized, or is associated with occupied orbitals (HOMO) of high energy, making it more effective interaction with the LUMO of soft acids (the energy difference between the HOMO and LUMO is small).

In the interaction between a basic "hard" and an acid "hard" there is a large energy difference between the frontier orbitals HOMO and LUMO, making the electronic transition. For both categories of acids, have the following characteristics: the acids "soft" (or soft), the acceptor atom has small positive charge, a large volume, and various other excitable electrons easily, in acids "hard," the atom acceptor has high positive charge, small size and has no other electrons easily excitable. This classification was made according to the following rule: acids "soft" with bases "soft" (character covalent interactions) and acids "hard" with bases "hard." The difficulty of the principle of acids and bases (hard and soft) is in relation to quantification. For this reason we developed a way to calculate molecular hardness (η) and softness (S) from measurements of ionization potential (IP) and electron affinity (difference of total energy between a neutral and anion species, AE) or from the energies of HOMO and LUMO:[59]

$$\eta = \frac{1}{2}(PI - AE) \tag{4}$$

$$\eta = \frac{1}{2}(E_{\mathrm{LUMO}} - E_{\mathrm{HOMO}}) \tag{5}$$

$$S = \frac{1}{\eta} \tag{6}$$

The ionization potential of an atom is a measure of the force with which an electron is bonded to an atom. The first ionization potential (IP) of an atom is the energy needed to remove an electron from that atom at an infinite distance ($A \rightarrow A^{+} + e^{-}$). Low ionization potential values for active compounds may indicate possible mechanisms for transferring charges in the ligand-receptor interaction, and may also indicate that the ionic form of the substance it shows biological activity.[60]

The electronegativity of an element is a measure of the strength of an atom to attract electrons to itself involved in a binding which this atom is also involved.[45] This property can be used to estimate the ability of a molecule to attract electrons to another, when there is an interaction between these two molecules. Mulliken electronegativity (χ) Calculated as:[48]

$$\chi = \frac{1}{2}(-E_{\mathrm{LUMO}} - E_{\mathrm{HOMO}}) \tag{7}$$

It is possible therefore estimate the hardness or softness of a molecule. These values are expressed in terms of the ionization energy of the atom and its neutral anion. Therefore, molecules that have a high ionization potential and high electronegativity have high absolute hardness, and the higher the hardness, the lower the smoothness of the molecule. Thus, it can be said that the hardness represents the resistance of a molecule to deformation and softness represents the ease with which a molecule is deformed. The smaller the higher the hardness or softness, the lower the amount of energy required for the transition of an electron of the HOMO to LUMO.[61]

Lobato et al. (2011)[62] carried out studies of isomeric products in chain reactions of halidrification in alkenes using computational methods at the HF/3-21G level of theory for the determination of molecular properties, including: polarizability (α), hardness (η), molecular softness (S), bond length (C2-X), total energy and the construction of maps of molecular electrostatic potential of the substrates, reagents and products. Aiming to analyze the stability of isomeric products studied.

4.4. Molecular Dipole Moment (μ)

This is a property that measures the magnitude of charge when displaced atoms of different electronegativity are interconnected. The direction of the dipole moment of a molecule is based on the relative electronegativities of the atoms of this molecules and value is obtained by the vector resultant of the dipole moments of each bond present in the molecule. The presence of substituents with different electronegativity alters molecular properties as acidity and basicity of a compound so that the dipole moment can answer questions about the same reactivity.[63, 64]

The polarity of a molecule is important for various physicochemical properties and thus, whereas the drug and receptor interaction occurs because of the differences in charges with opposite values, many descriptors have been proposed to quantify the effects of polarity, among which the dipole moment of the molecule (which reflects only the overall polarity thereof) is the most used. The electric dipole moment (m) calculated the $\mu = |\mu|$, where μ is given by:

$$\mu = \int \rho(r) r dr \tag{8}$$

and $r(r)$ stands for electrical charge density.

4.5. Principal Quantum-Chemical Descriptors Utilized in Studies SAR/QSAR

The quantum-chemical descriptors are fundamentally different from measures obtained experimentally, although there is some overlap naturally. A basic disadvantage of quantum-chemical descriptors is the failure to reproduce stereo effect.[30] A summary of the principal terms used in quantum-chemical studies SAR/QSAR is presented in Table I.

The energy parameters such as electron energy, total energy and heat of formation, are also widely used for correlating structure and activity. The electronic energy is determined by means the Born-Oppenheimer

Table I. Summary of the principal quantum-chemical descriptors used in studies SAR/QSAR.

Nome	Definição
Q_N	Liquid atomic charge on atom N
$Q_{\min}$, $Q_{\max}$	Atoms charges more negative and more positive
Q_{AB}	Total charge of the group containing the atoms A and B
$\sum q_A^2$	Sum of the squares of the charge densities in atoms of type A
$q_{E,A}$ $q_{N,A}$	Nucleophilic and electrophilic electronic charge calculated from the occupied and unoccupied orbitals
Q_B, Q_A	Sum of absolute values of the charges of all the atoms in a given molecule or functional group
Q_B^2, Q_A^2	Sum of squares of the charges of all the atoms in a given molecule or functional group
Q_m	Mean of the absolute values of the charges on all atoms
$\varepsilon_{\text{HOMO}}$, $\varepsilon_{\text{LUMO}}$	Occupied molecular orbital energy of the highest energy ($\varepsilon_{\text{HOMO}}$) and unoccupied molecular orbital energy of the lowest energy ($\varepsilon_{\text{LUMO}}$)
$\Delta\eta = \eta_R - \eta_T$	Hardness activation states R and T represent the transition state and reagents
$q_{N\sigma}$, q_{Np}	Electron density σ and π in atom A
$Q_{A,\text{HOMO}}$, $Q_{A,\text{LUMO}}$	Electron density HOMO/LUMO in the atom A
$f_r^E = \sum(C_{\text{HOMO},n})^2$	Electron density of electrophilic frontier atomic, $C_{\text{HOMO},n}$ are the coefficients of the atomic orbitals X_n in HOMO
$f_r^N = \sum(C_{\text{LUMO},n})^2$	Electron density of nucleophilic frontier atomic, $C_{\text{LUMO},n}$ are the coefficients of the atomic orbitals X_n in LUMO
$F_r^E = f_r^E/\varepsilon_{\text{HOMO}}$ $F_r^N = f_r^N/\varepsilon_{\text{LUMO}}$	Índices of electron density of frontier
$\sum S_{E,A}$, $\sum S_{N,A}$	Sum of nucleophilic and electrophilic super displacement
$\sum \pi_{AA}$	Sum of auto polarizabilities atomic
A	Molecular polarizability
$\alpha = 1/3(a_{XX} + a_{YY} + a_{ZZ})$	Mean polarizability of the molecules
$\beta^2 = 1/2[(\alpha_{XX} - \alpha_{YY})^2 + (\alpha_{YY} - \alpha_{ZZ})^2 + (\alpha_{ZZ} - \alpha_{YY})^2]$	Anisotropy of the polarizability
$p = \sum_{\lambda=1}^{N} \lvert Q_A \rvert / N$	Polarization of the molecule, the sum of the atomic charges of all atoms in the molecule
μ_{char}, μ_{hybr}	Components charge and hybridization dipole moment
D_x, D_y, D_z	Components of the dipole moment along the axes
Δ	Polarity submolecular parameters (greater difference in charge of electrons between two atoms)
$D = \sum_{AB} \lvert Q_A - Q_B \rvert / N_{AB}$	Indices dipole local sum over all pairs of bonded atoms
T	Quadrupole moment tensor
E_T	Total energy
$E_b = \sum_i^N E_{Ni} - E_T$	Bond energy
ΔH_f°	Enthalpy of formation
$\Delta(\Delta H_f^\circ)$	Heat of formation relative
E_{prot}	Protonation energy (energy difference between the protonated form and the sum of the energies of form neutral and of separate proton)
E_{Hidrat}	Hydration energy (energy related to the stability of different molecular conformations in aqueous solution)
RM	Molecular refractivity (property that depends on the structure of the bioactive substance and expresses the character lipophilic and electronic of substituent groups present in the molecule)

approximation, i.e., assuming a fixed position of the nuclei, the Schrödinger equation is resolved in order to find the electronic energy of the molecule. This procedure is repeated for different configurations of fixed of the nuclei (through interactions SCF). The nuclear configuration which corresponds to the minimum value of the energy geometry is the of equilibrium geometry of the molecule.[65] The total energy is often utilized to estimate the stability of a chemical species, and corresponds to the sum of repulsion energy nuclear with the electronic energy. As the total energy, heat of formation of a molecule is also used to estimate the chemical stability. From the total energy of the system, we calculate the energy of atomization from subtracting the total energies of the atoms in their stoichiometric ratios. The heat of formation (ΔH_f) can be calculated using the enthalpies of atomization of the atoms and the energy of atomization.[66]

5. ORIGIN OF QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIP

In 1963, Hansch and Fujita[67] observed that the biological activity of some series of compounds correlated with lipo-hydrophilicity of the molecules, then the technique developed was expanded to other classes of compounds,[68] exhibiting equally elevated correlation. These studies resulted in the quantitative analysis of chemical structure and biological activity (QSAR-Quantitative Structure-Activity Relationship), whose greatest interest is providing the rational development of a novel compound, particularly a better drug, avoiding random synthesis and biological testing of new onerous molecules. A mathematical model developed by Free and Wilson[69] also contributed to the QSAR studies of that era. Currently, most of the models constructed based on QSAR descriptors (parameters that correlate with the biological activities of molecules)

threedimensional which encode the molecular properties, such as steric and electrostatic, based on the spatial structure of a class counterpart of molecules.

Studies of quantitative relationships between chemical structure and biological activity (QSAR), or between chemical structure and some type of physicochemical property (QSPR), are of great importance in modern chemistry and biochemistry. The central objective in studies of QSAR/QSPR is rationalizing the search for compounds with desired properties using chemical intuition and experience in a way mathematically quantified and computerized. Since it the correlation between the structure/property and activity is found, large numbers of compounds, including those that have not yet been synthesized, can be readily examined on the computer in order to select structures with desired properties. Thus, it is possible to select the most promising compounds for synthesis and testing laboratories. It can be said that the studies involving QSAR/QSPR are considered excellent tools to accelerate and succeed in development of new molecules to be used as pharmaceuticals, materials, additives and other purposes. As is not easy to find correlations structure-activity-property, the exponential growth in the number of papers involving studies QSAR/QSPR clearly demonstrates the rapid progress in this area.[70–75] To obtain a significant correlation, it is critical that appropriate descriptors are employed, whether theoretical, empirical or derived from experimental data. Many descriptors reflect simple molecular properties and can provide information on the physico-chemical nature of the activity/property under study.[76]

Many studies employing quantum-chemical descriptors are realized in the area of QSAR more than QSPR, ie the descriptors have been correlated with biological activities such as inhibition of enzymatic activity and/or hallucinogenic activity.[77, 78] In part this is because, historically, the search for quantitative relationships with chemical structures began with the development of theoretical methods for drug discovery. The employment of quantum-chemical descriptors has great utility for correlating the reactivity of organic compounds, partition coefficient octanol/water chromatographic retention indices and various other physical properties of molecules.[76]

In recent decades, many significant advances were observed in the area of QSAR, stand out: developing methods dimensional (2D QSAR)[79, 80] three-dimensional (3D QSAR),[81] in addition to methods that involve more advanced molecular information (4D and 5D QSAR), with the main difference between these methods is related to the type of molecular information used for building of quantitative model.[82, 83] Analyzes using more advanced techniques such as QSAR 3D and 4D, require knowledge of the molecular conformations for three dimensional alignment, assuming that the biological response is directly associated with the interactions between the bioactive molecule and biological receptor, these being described interactions and represented by their stereos fields, electrostatic and other three-dimensional fields. Thus, obtaining quantum-chemical properties is important to the fields of molecular generating and hence the development of QSAR models 3D and 4D of good quality.

The methods CoMFA (Comparative Molecular Field Analysis)[84] and CoMSIA (Comparative Molecular Similarity Indices Analysis)[85] stand out as 3D-QSAR methods due to the numerous published papers using these methodologies. These largely accepted methods that are the most widely discussed in 3D-QSAR studies and require three dimensional alignment of the ligands and therefore there is a need structural similarity for overlapping of the chemical structures, therefore, must meet a series congener. Formalisms 4D,[86] 5D[87] and 6D[88] have been applied to incorporate new degrees of freedom (dimensions), so that a more refined analysis on the adaptation of the active site of an enzyme to the topology of the ligand, and vice versa, can be better represented. However, molecular descriptors 2D, usually physico-chemical descriptors referred to in classical QSAR analyzes not have been shown to be lower than those 3D descriptors being extremely potent as the convenience and simplicity of the calculations.[89] In fact, the need for a conformation of the linker scanning and a comprehensive three-dimensional alignment of structures that may not correspond to the forms of bioactive molecules, reflects the main disadvantages of the techniques associated with the methodology nD, so are an approximation.

A predictive approach equally, but much faster, cheaper and simpler to operate, was developed in 2005 and named MIA-QSAR (Multivariate Image Analysis applied to QSAR).[90] The MIA descriptors have been successfully applied not only to correlate chemical structures with biological activities,[91–96] but also with physical properties such as boiling temperatures,[97] chemical shifts[98] and electrophoretic profiles.[99] The method is based on using pixels as image descriptors; how pixels can be processed numerically as binary, white color digit equals 765 and the black pixels digit 0, according to the RGB color system. In MIA-QSAR, the images correspond to the chemical structures drawn by some program to design molecules such as ChemDraw or ChemSketch. Structural modifications or changes in position of the substituents on a series congener molecules correspond to changes in the coordinates of the pixels of the image, and these changes explain the variance in the Y block, the block corresponding to the dependent variables (biological activities, for example).

There aren't advices of some rare manuscripts and reviews of examination boards show some skepticism about the existence of meaning physicochemical descriptors for MIA and therefore on them can be correlated with some property chemical, physical or biological. Some even linked the results of an MIA-QSAR analysis by chance that the correlation could exist between students' scores on a test with the alphabetical order of their names,

which would be a completely arbitrary assumption. We reinforce the assertion that MIA descriptors may encode the chemical, physical and biological, the chemical and physical description must be incorporated in any manner in which substituents are represented. For example, the MIA descriptors may encode steric effects (substituents of organic molecules occupying a large area in the space devoted to the design of structures), stereogenic centers (wedge or dashed lines to represent links for either forward or backward relative to a chiral carbon).[100]

6. MULTIVARIATE ANALYSIS METHODS

When measurements are made on a number of objects, the results are usually arranged in a matrix, which is called data matrix. Measures (in our study the molecular descriptors and biological activity) are placed in columns, and objects (in our case the compounds studied) are associated with the lines. For a matrix of multidimensional data, multivariate statistical methods are needed to understand these data in its entirety.

After obtaining a given number of molecular parameters, it becomes necessary to use methods that allow the simultaneous analysis of all parameters obtained since, initially, the main factors responsible for the biological ligand-receptor interaction are not known. Thus, the multivariate methods of analysis are very useful tools in studies of this type, i.e., analysis of data sets with a high number of properties, which makes data interpretation.

Some methods which often use multivariate data analysis are: principal components analysis (PCA) has been widely used in chemical and biological problems, and the main aim of this analysis is to show the data in a multidimensional space of low dimensionality with the minimum of loss total information;[101] hierarchical cluster analysis (HCA), partial least squares regression (PLS) method of *K*-Nearest Neighbor (KNN), stepwise discriminant analysis (SDA) and independent models of similarity using principal component (SIMCA–Soft Independent Modeling of Class Analogy).[102–114]

6.1. Analysis of Principal Components (PCA)

Principal component analysis is a method of data compression based on the correlation between the variables, this data compression generates a small set of variables which are called principal components and these are orthogonal between Themselves, so that the correlation between variables does not limit its application, unlike the multiple linear regression that is sensitive to the presence of highly correlated variables, because it makes the regression coefficients unstable and without significance.[115, 116]

From the mathematical point of view the model obtained with principal component analysis is written as:

$$y_{ij} = \sum_{k=1}^{p} P_{ik} a_{kj} + \sum_{k=p+1}^{m} P_{ik}^{(0)} a_{kj}^{(0)} \quad (9)$$

where P_{ik} are called principal components (PCs), and also are referred to as "scores" and correspond to the characteristics of the compounds. P_{ik} are orthogonal vectors, and are determined so that the data matrix is replicated. a_{kj} represents the weight, i.e., a measure of the contribution of the k PC with the j variable. A high value of a_{kj} demonstrates a high importance of k PC for the j variable. Thus the j variable has a high contribution to the k PC, so a_{kj} are also called "loadings."[101]

The number of PCs that can be extracted from the data matrix equals the number of original variables and with this number of components the data matrix can be reproduced exactly, but this is not the desired result, since it does not lead to reduction of dimensionality the matrix. What aims is to find a number of PCs so that the original variables are represented with the minimum of loss of relevant information.[101]

6.1.1. Staggering

A way of treat all variables with the same importance is standardizing them by autoscale according to the following equation:

$$Y'_{jm} = \frac{(y_{jm} - \bar{y}_j)}{s_j} \quad (10)$$

where $\bar{y}_{jm}$ is the mean of variable j to the object m. Variables autoscaled have an average equal to zero and unit variance represented by Eqs. (11) and (12) respectively.

$$\bar{y}_j = \frac{\sum_{j=1}^{n} y'_{jm}}{n} = 0 \quad (11)$$

$$s_j^2 = \frac{\sum_{k=1}^{n} (y'_{jm} - y'_i)}{n-1} = 1 \quad (12)$$

6.2. Hierarchical Cluster Analysis (HCA)

Another method of multivariate analysis is of great importance hierarchical cluster analysis, its aim is to show the data in such a way to accentuate their natural groupings and standards. As the principal component analysis, the results of hierarchical cluster analysis are qualitative, being arranged in the form of a dendrogram thus allowing displaying samples (the compounds studied in our case) or variables (molecular descriptors in our study) in a space two-dimensional.[115]

In hierarchical cluster analysis the distance between samples (the compounds studied here) is calculated and transformed into a similarity matrix S, whose elements are the similarity index. So for two samples m and n, the similarity index is written as:

$$S_{mn} = 1 - \frac{d_{mn}}{d_{\max}} \quad (13)$$

Being S_{mn} an element of S, $d_{\max}$ is the maximum distance for a pair of samples, d_{mn} comes to be the Euclidean distance between samples m and n calculated as:

$$d_{mn} = [(x_{m1} - x_{n1})^2 + (x_{m2} - x_{n2})^2 + \cdots + (x_{mh} - x_{nh})^2]^{1/2} \quad (14)$$

With x_{ij} being a matrix element of the original data.[115]

6.3. Method of Partial Least Squares (PLS)

One of the most promising methods of multivariate analysis is the partial least squares (PLS), because hundreds or thousands of independent variables (the bloc X, in our study the molecular descriptors) may be correlated with one or several dependent variables (bloc Y in our case the biological activity). In the PLS analysis the resulting vectors are slightly displaced from their original positions, so that the correlation of corresponding vectors derived from bloc X and Y be optimized.[117]

Through the PLS method is possible obtain a description of the dependent variable Y (biological activity) as a linear combination of molecular descriptors, through the principal components being they are not correlated. To ensure that the principal components obtained are important for biological activity, the bloc Y (biological activity in our study) is used for finding the standard within the bloc X (molecular descriptors in our case) that is correlated with the Y bloc, or is the principal components are optimized to best describe the relationship between the bloc X and Y both being the principal components now called latent variables and are used to model the bloc Y.[116]

The PLS model finds new variables, K latent variables, which are also called "scores" and are represented as t_k ($k = 1, 2, 3, \ldots, K$). The "scores" are linear combinations of original variables j_s weighting coefficients W^*_{vk} ($k = 1, 2, 3, \ldots, K$).

$$t_{ik} = \sum_k W^*_{jk} x_{ik} \tag{15}$$

The "scores" are good predictors of the bloc Y, then the PLS model can be written as:

$$Y_{im} \sum a C_{ma} t_{ia} + f_{im} \tag{16}$$

in matrix form has the following form:

$$Y = TC' + F \tag{17}$$

combining Eqs. (15) and (16) we obtain the model as a regression model:

$$y_{im} = \sum a C_{ma} \sum k W^*_{jk} + f_{im} \tag{18}$$

where the term f_{im} represents the deviation between the Observed and predicted data (in our study the biological activity).[118]

In PLS analysis results can be transformed into regression coefficients, also called vector regression, the variables of the bloc X (the molecular descriptors in our study), these indicate what the most important descriptors in model building.[116, 117]

After building the PLS model is necessary to evaluate it and this review is to ascertain if the specification of the model fits the data adequately predicted. This review is divided into three parts: (1) evaluation of the degree of fit, (2) evaluation of degree of significance and (3) evaluation the degree of predictability.[119]

6.3.1. Evaluation of Degree of Fit of the Model PLS

The evaluation of the degree of adjustment is done by calculating the following statistical parameters: (a) correlation coefficient R, (b) correlation coefficient adjusted R^2_{adjusted}, (c) standard deviation and (d) analysis of the waste, with mathematical expressions are shown respectively below:

$$R^2 = 1 - \frac{\sum (y_{\text{predicted}} - y_{\text{Experimental}})^2}{\sum (y_{\text{predicted}} - y_{\text{mean}})^2} \tag{19}$$

$$R^2_{\text{adjusted}} = R^2 - \left(\frac{k-1}{n-k}\right)(1 - R^2) \tag{20}$$

where n, in Eqs. (20) and (21), is the number of samples included in the model (in our case the number of compounds included), k is the number of variables (molecular descriptors in our study) included in the model.

$$s^2 = \frac{\sum (y_{\text{Experimental}} - y_{\text{predicted}})^2}{n-2} \tag{21}$$

$$(y_{\text{Experimental}} - y_{\text{predicted}}) \tag{22}$$

For a QSAR model to be accepted coefficient R must be greater than 0.9 for tests of biological activity *in vitro* and greater than 0.8 for testing *in vivo*.[117] The correlation coefficient adjusted which considers the number corrections to the number of variables and number of compounds used, must be the greatest possible and it is expected that the standard deviation and the waste is closest to zero.[119] But the standard deviation s cannot be greater than the standard deviation of biological data, around 0.3 which is the mean error of many biological data whereas for *in vivo* biological test this value should be less.[117]

6.3.2. Evaluation of Degree of Significance of the PLS Model

The degree of significance is assessed by performing validation tests, in our case we tested only the statistical significance of the correlation coefficient R^2 through the hypothesis test called F test that checks how much of the variability of the Y bloc (biological activity) can be explained by the bloc X (molecular descriptors included in PLS model). Order to validate the correlation coefficient through the test F, the mathematical expression being shown below, it is necessary to compare the value of F obtained in constructing the PLS model with the tabulated value.[119]

$$F_{(k,\,n-k-1)} = \left[\frac{R^2(n-k-1)}{k(1-R^2)}\right] \tag{23}$$

where n in Eq. (23) corresponds to the number of samples included in the model (in our case, the number of compounds included), k is the number of variables used to construct the model (in our study, molecular descriptors). The value of F shall be four to five times the value tabulated, thus demonstrating that the model is statistically significant and useful for predictive purposes.[120]

6.3.3. Evaluation the Degree of Predictability of the PLS Model

The cross-validation procedure should be used to evaluate the degree of predictability of the QSAR model, and select the model that has greater predictive ability. The cross-validation is performed so that one or several objects (compound studied) are eliminated from the set of data, either randomly or systematically.[117] The cross-validation is done in the following steps:

(1) delete one of the model compounds,
(2) reconstruct the model without this compound,
(3) use the model to calculate the value of the biological activity of the compound excluded,
(4) obtain the residue between the experimental value of the biological activity and the predicted value for the compound excluded;
(5) remake steps (1)–(4) for the other compounds, one at a time.[119]

The evaluation of the degree of predictability is done by calculating some statistical parameters that show the predictive quality of the PLS model: sum of squares of deviations (*PRESS*), standard deviation of the cross-validation (S_{PRESS}), standard error of prediction (*SEP*) e o correlation coefficient of the cross-validation (Q^2), with the respective equations are shown below:

$$PRESS = \sum (y_{\text{Experimental}} - y_{\text{predicted}})^2 \tag{24}$$

it is ideal that the value of the PRESS not increase with the number of latent variables.[121]

$$S_{PRESS} = \frac{(PRESS)^{1/2}}{n-k-1} \tag{25}$$

where n in Eq. (25) is the number of samples included in the model (compounds) and K is the number of variables used to construct the model (molecular descriptors), S_{PRESS} is obtained as a criterion for the optimum number of latent variables, the S_{PRESS} lowest value indicates the optimal number of latent variables,[117] a model with excellent predictive capability submit S_{PRESS} close to zero.[119]

$$SEP = \sqrt{\frac{PRESS}{n}} \tag{26}$$

in Eq. (26), n is the number of samples (compounds) included in the QSAR model, the *SEP* is different of S_{PRESS} because not consider the degree of freedom in the calculation.[117] It is also used as the main criterion for checking the degree of PLS prediction model.[116]

$$Q^2 = \frac{PRESS}{\sum (y_{\text{Experimental}} - y_{\text{mean}})^2} \tag{27}$$

This parameter is used to evaluate the statistical quality of the PLS model,[115] it describes the amount of variance in y can be predicted, its value ranges from zero to one (0 to 1), where one (1) signifies a model perfect, and zero (0) a model without relevance.[122] To validate a latent variable is necessary that the value Q^2 does not decrease with increased latent variables, but increase with each latent variable.[121] A PLS model with high predictability for objects (compounds) are not included in the model should display the value of Q^2 nearest one (1) as possible,[119] but a value above 0.3 ($Q^2 > 0.3$) is generally satisfactory.[123]

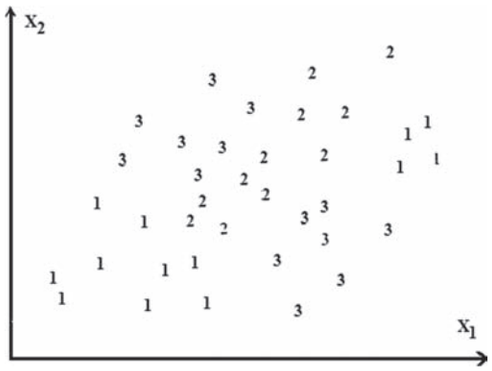

Fig. 3. Classification of K near neighbors belonging to three classes.

6.4. K-Nearest Neighbor (KNN)

An unknown object can be classified[124] by investigating objects of their close neighbors of K which class the group is known. To find a close neighbor of the object is necessary to compute the distance (usually Euclidean distance) to every object in the dataset. An unknown object is assigned to the class that has the most among the K neighbors (Fig. 3). Frequently used values for K are between one and ten. This approach is similar to a spectral library search, although more complicated (nonlinear). Similar measures are used for research in the library and usually point to identify the object in place of the standings. The method KNN can also be compared as a product for estimating the location probability density for the class.

Moreover, this mathematical simplicity of KNN classification has some advantages: no ordering on the distribution of data is required, the class of objects need not be linearly separated, he is a multiclass method, and it is not necessary to classify the training.

A KNN classification consists for any vector object assembly training, this, however, requires a computational time rather large even for data set size medium, the KNN classification is often used as a reference method.

6.5. Stepwise Discriminant Analysis (SDA)

Stepwise discriminant analysis (SDA) is a method that can be used for discrimination (recognition or classification) and prediction samples. Its main purpose is to determine discriminant functions, which represent linear combinations of variables calculated.[125] The procedure used in the

SDA method is to construct the discriminant functions (a function for active compounds and one for inactive) adding one variable each time until obtaining the end discriminant functions based on the set of variables which best discriminates between the groups of compounds. This method is useful to select the most relevant variables to separate the compounds into different groups (often referred to as the discriminating power of the variables), since constructs discriminant functions using a variable time until the best discriminant function is obtained. After statistical validation of the model through this procedure, the discriminant functions can be used to make predictions with unknown compounds.

6.6. Soft Independent Modeling of Class Analogy (SIMCA)

It is a method for classification[124] that considers information of population distribution estimates a confidence level of classification and may provide new samples as belonging to one or more classes or no classes. To make the classification the SIMCA uses the space of the principal components of each class. Thus, the n class shall be represented by Eq. (28).

$$X_n = \bar{T}_n \bar{P}_n^t + E_n \tag{28}$$

where $\mathbf{X}_n$ are the class data, $\mathbf{T}_n$ the matrix containing coordinates in the principal component of the class n, $\mathbf{P}_n$ the linear transformation matrix and $\mathbf{E}_n$ the matrix of residues. In the construction of the SIMCA classification model calculates for each class separately, the standard deviation of residues. For the space described by the principal components are computed variances of samples in each axis. These two parameters are used for classification of new samples. The objective of SIMCA is to create a limited space for each class. This can be better understood for a class described by two principal components. In geometric terms, the residues of this class correspond to the distances of the samples to the plane of the main components. Thus, the calculation of the standard deviation of residues originates two planes parallel to these components, i.e., one above and one below. Considering the variance in each main component and planes, for the standard deviation of residues, one can say that the class is bounded by a box, a hyper box in case of three or more components and a cylinder for a main component.

The classification of a new sample is made by its projection on the main components of each class, where the variances are calculated and its residue. These two are compared by F tests with those already determined for constructing the model. Thus, those class where the residue is less than or equal to, the same is true for the variances, the sample is classified positively. With this, the sample may be placed in one or more classes. Otherwise, larger deviation or variance, the sample is classified as belonging to the class not.

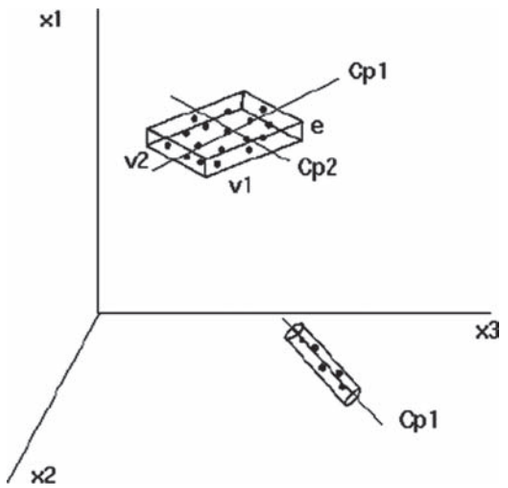

Fig. 4. Graphical representation of a matrix with three columns and divided into two classes. x_1, x_2 and x_3 represent the columns of the matrix, the first principal component PC1, PC2, the second principal component v_1 variance in PC1, v_2 variance in PC2, the standard deviation of residuals is represented by e.

The Figure 4 shows an example of a matrix with three columns, divided into two classes. The first is represented by two major components, where v1 is the variance in the first principal component PC1, v2 variance in the second component PC2 and the standard deviation of residues are represented by:

7. CASE STUDY ON MALARIA

Malaria is a potentially serious infectious disease caused by parasites (protozoan of the genus *Plasmodium*) that are transmitted from one person to another by the bite of female mosquitoes of the genus *Anopheles*. The transmitters of human malaria are insects of the order diptera, family *Culicidae* and the genus *Anopheles*. This genus comprises about 400 species, of which only few have importance for the epidemiology of malaria in each region. Among the different species of malaria, malaria caused by *Plasmodium falciparum* and *Plasmodium vivax* are the most widespread. The malaria produced by *Plasmodium vivax*, are rarely fatal, however, symptoms tend to recur periodically, even after long periods of treatment (BRASIL, 2012).

The malaria is one of the most common diseases in tropical and subtropical countries with more than 300 million infections and millions of deaths from malaria occur worldwide each year. The rapid spread of resistance to quinoline antimalarial made malaria a serious global problem, so it is essential to seek new drugs against malaria and understand its mechanism of action for the treatment of patients.[127] Where, pathways rule, institutions do not have the economic resources that can be used to access high technology which provides the adequate development of research aimed at solving the problem.[128]

7.1. Design and Development of New Antimalarial Drugs

Despite several studies that have been done many years ago, today does not exist a vaccine that confers protection against malaria satisfactory. It is common to confuse the vaccine against yellow fever as though it were against the malaria (antimalarial), but only the first exists, is effective and fundamental to take it when you travel to yellow fever endemic areas, which generally overlaps with the malarial areas (BRASIL, 2012).

The processes for obtaining new drugs have changed much over the years. Until approximately the decade eighties, new drugs were discovered through tests black–box type in cells or models–animals. This traditional method of drug development, also called blind triage, consists of testing several micromolecules randomly in biological assays, without any knowledge of the mechanisms of action and/or interaction of the ligand molecule. Although most of the drug therapeutically useful available today have been discovered through this process gradually this methodology proved to be inefficient due to the ever smaller probability of finding a new drug, being necessary to test millions of compounds in some cases. This fact resulted in a great increase in the time and costs involved in the discovery and development of new drugs.[129, 130]

The impossibility of removing the vector of malaria transmission requires the necessity of new agents that may have a novel mechanism of action. Therefore, the search for new compounds with effective action to combat *Plasmodiumfalciparum* has high priority due to its severity becomes an urgent mission of global research programs.[131]

In order to plan a drug applies one of the following strategies: the rational design and molecular modification.[132] The molecular modification consists in considering a compound of known chemical structure and biological action proven as a model or prototype, synthesize and test new compounds which are congeners, homologues or structural analogues of the drug matrix.[132] This method when used requires high time and high financial investment in obtaining a new drug.

The implementation of rational design requires information from different areas of human knowledge, especially those related to electronic levels of the drug, biological activity, physicochemical parameters, such as: hydrophobic, steric and electronic, related to biological activity.[133]

The Figure 5 shows a flowchart with several steps of Rational Drug Design Based on Structure (RDBE). The first step constitutes the appropriate choice of therapeutic target. The target molecular bioreceptor or may be a protein related disease for which is desired develop a chemotherapy treatment, where function shall be blocked or activated. After identifying the molecular target, the three-dimensional molecular structure of the bioreceptor needs to be obtained. The molecular structures

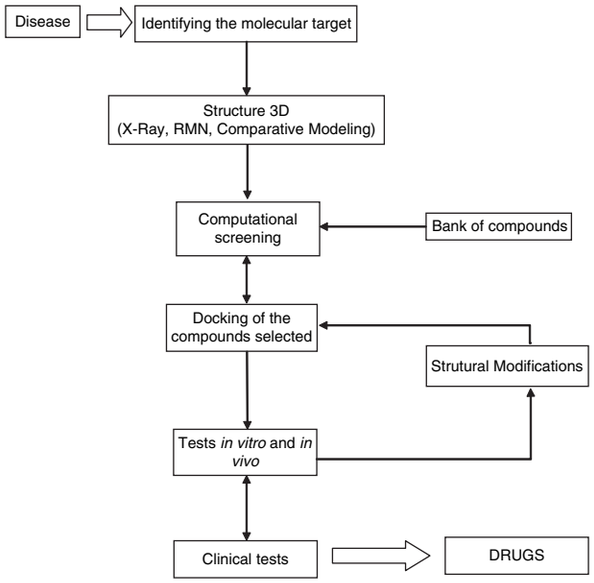

Fig. 5. Steps of rational drug design based on structure.

may be obtained using experimental techniques, such as X-ray diffraction in crystals and nuclear magnetic resonance (NMR), or by theoretical methods as comparative modeling.[134] In this step, the docking methodologies (drug-receptor interaction) need to be faster than those used for the refinement of compounds prototypes. Large banks of molecular structures of ligands may contain millions of molecules are tested against molecular target with the use these methods in order to identify compounds prototypes. When a promising compound is found, the medicinal chemist analyzes the modifications that can be made in the molecule, so that the desired biological response is potentiated. In this phase, enter the scene more accurate docking methodologies, which aim to identify both binding conformation as quantify more precisely the binding affinity receptor–ligand, aiming an optimization of promising molecules selected. In vitro and *in vivo* tests are conducted to validate and guide the optimization of compounds prototypes and so that other characteristics such as toxicity, are analyzed.[134]

From among the various techniques used in rational drug design, we can relate: design technique with the aid of computing, principally using physicochemical parameters involved in biological activity and methods of quantum chemistry to determine the most promising compounds in a series.[48, 133, 135, 136]

7.2. The Therapy of Malaria

In medicinal chemistry, the term "structure-activity relationship" comprises studying the effects that the chemical structure of a compound (ligand) may cause during their interaction with the biological receptor, and consequently rationalize the main factors that govern this interaction. The interactions of a drug with its biological receptor

are determined by intermolecular forces, i.e., interactions lipophilic, polar, electrostatic and steric. Therefore, substances showing therapeutic properties when interacting with a specific target (an enzyme, a receptor, an ion channel, a nucleic acid or other biological macromolecule) should have a three dimensional structure so that the provisions of their functional groups promote greater complementarities to the binding site. This can be summed up as follows: how much better the "Docking" and the complementarity of the surface properties of a drug, greater its affinity and greater may be its biological activity.[117, 137]

To describe the types of interactions between a biological receptor and its ligand may be used an extensive set of molecular properties, since these properties are directly related to intermolecular forces involved in ligand-receptor interaction, even as are related to transport properties and distribution of drugs.

The molecular descriptors represent an important tool for predicting properties of substances, classify chemical structures or search for similarities between them.[138–140] Different descriptors have been introduced in recent years and the number continues growing, since it is believed that with this increase, significant problems in studies on structure-activity relationships (SAR) would be solved.[141]

The treatment of malaria was employed by traditional Chinese medicine for more than 2000 years. The medicine used artemisinin (qinghaosu) is extracted from the plant *Artemisia annua* L, used in combatting of 52 species of diseases in People's Republic of China.[142]

Artemisinin (qinghaosu) has a single structure (Fig. 6) having a lactone stable endoperoxide (1, 2, 4-trioxane) sesquiterpene totally different from the previous antimalarial in its structure and mode of action, being isolated from *Artemisia annua* is a compound of remarkable life and economic antimalarial effective against Plasmodium falciparum and cerebral malaria.[127]

Artemisinin and its derivatives induce a rapid reduction in the number of parasites when compared with other known drugs. They are therefore of particular interest for severe malaria. The first reduction in the number of parasites is also beneficial for combination therapies. This led to an enormous interest in the mechanism of action, chemical[143] and drug development[144] of a new class of antimalarials.

The group endoperoxide is essential for antimalarial activity[142] and is mediated by active oxygen (superoxide, H_2O_2 and/or radicals hidroxis) or carbon free radicals.[145]

The artemisinin has a broad and extraordinary activity against parasites in asexual form, killing at all stages in malaria caused by Plasmodium falciparum. The artemisinin also kills the gametocytes, including the four stages of gametocytes, which are sensitive only to primaquine.[146]

In vitro biological assays indicated that some amount of iron must be added artemisinin to show antimalarial activity. In humans, the heme compound (Fig. 7), which is a product originating from one digestive process the RBCs must be the source of iron for artemisinin. In human malaria parasites digest over 70% of the hemoglobin within the red blood cells infected, since the globin and heme as products, the protein is hydrolyzed, generating aminoacids that are utilized in protein synthesis by the parasite, since heme undergoes a polymerization process having as a product hemozoína.[147–149]

The polymerization of heme is a target of some drugs antimalarials such as chloroquine, which inhibits this process. Strains of chloroquine-resistant Plasmodium berghei that for want of hemozoín are caused by the non-occurrence of polymerization of heme. This reinforces the view that the inhibition of heme polymerization is the mode of action of artemisinin, thus indicating that artemisinin very possibly interacts with the free heme and for this reason inhibits the polymerization process.[149]

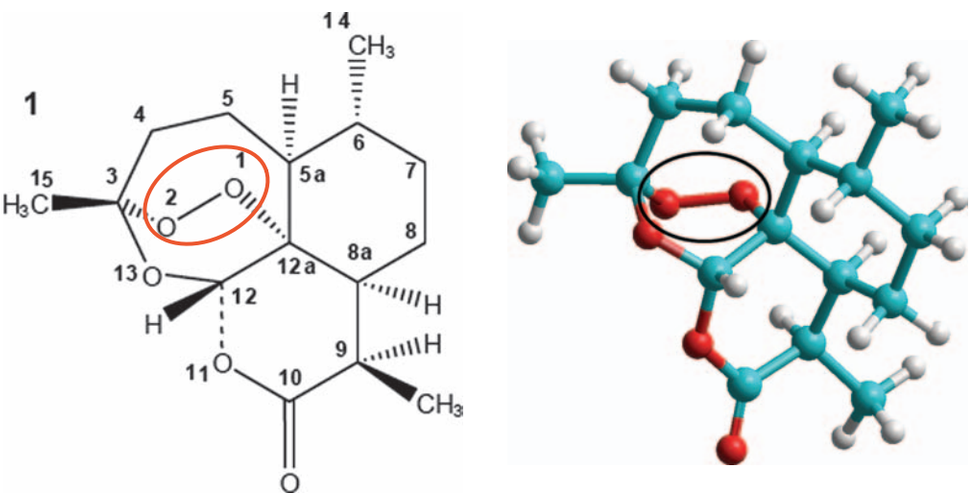

Fig. 6. Artemisinin (structure) and the region essential for expression of the biological activity (pharmacophore). The Structures were visualized using the ChemSketch 12.00[33] and Hyperchem[39] softwares, respectively.

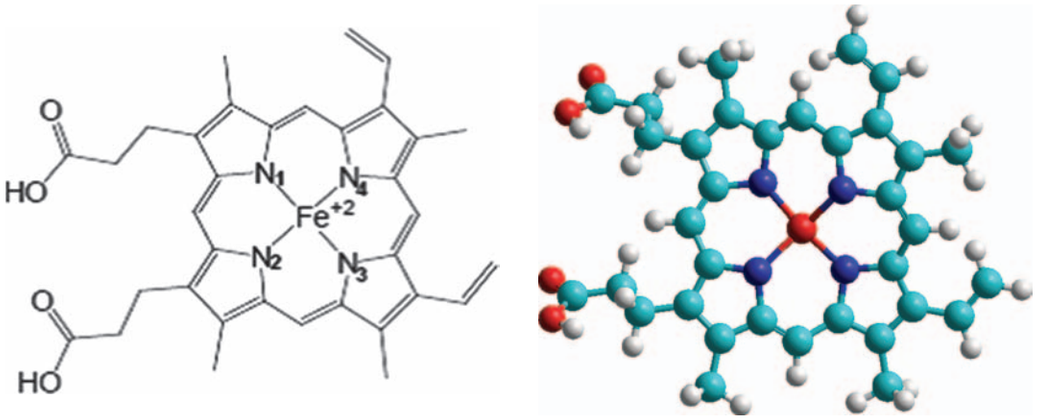

Fig. 7. The numbering the heme group is systematic for the purpose of determining the geometrical parameters. This structure was visualized by means of ChemSketch 12.00[33] and Hyperchem[39] softwares, respectively.

The numbering of the heme is systematic for purposes of determination of geometrical parameters: bond length, bond angle and torsion angle between artemisinin and derivatives (drugs) with heme (receptor).

The mechanism for antimalarial action of these compounds is still without conclusion, however investigations give great significance to the endoperoxide group in artemisinin for biological activity.

There are two possibilities for the heme iron attach in artemisinin derivatives. Posner et al. (1995)[150] have proposed the attack of the iron to the compound, occurs by the position O2 and produces radial free at position O1 with subsequent rearrangement to a free radical C4. This radical formed was suggested to be an important substance for biological activity. The compound formed is changed to another compound by a reaction of beta cleavage. Subsequently, the new compound is rearranged to form an epoxide compound. Alternatively, a direct intramolecular formation for the radical form another compound. This compound is capable of alkylated protein specific malarial parasite and possibly cause damage to parasites.[151] Moreover, Jefford et al.[152] believe that the attack of the iron the compounds takes place at the position O1 and produces a free radical at position O2. After the C3—C4 bond is broken to give a carbon radical C4. This radical can also be very harmful to the parasite. In the discovery and development of drugs, knowing the mechanism of action may help the development of new and more effective drugs.

In artemisinin peroxide binding in 1 and 2 oxygens (O1 and O2) is contained pharmacophoric structure of trioxane ring of artemisinin, is essential for the expression of antimalarial activity. Peroxides are known by suffer reductive cleavage transition metal of low valency, generating free radicals oxygens. These free radicals oxygens are potent agents of abstraction of hydrogen, can also generate free radicals carbon by abstracting intramolecular hydrogen atoms, these radicals carbons are assumed to react with biomolecules.[153]

The consideration of molecular flexibility of receptor and ligand implies in treatment hundreds of thousands of degrees of freedom, on the part of algorithms of "Docking." The molecular recognition is a dynamic process and highly complex, involving a large number of intermolecular interactions between the ligand, the receptor molecule and solvent. Due to its complexity, the problem of "docking" is generally divided into two subproblems: (i) development of an algorithm that investigate of way effective a complex energy hypersurface to predict the conformation and orientation of a ligand molecule relative to the active site of the receptor; (ii) predicting the binding affinity of a receptor complex–ligand, i.e., the development of a model for assessing binding free energy (usually called in the literature function of "scoring"), it is computationally viable to correctly discriminate between different binding modes of the same binder and/or to determine between two different ligands, those with the greatest binding affinity for a given receptor.[154]

The first program of "docking" using Genetic Algorithm (GA) was implemented by Judson (1993).[155] The program "docking" GOLD uses a genetic algorithm to evolve multiple subpopulations of ligands, where migration between populations is permitted.[156] The program AutoDock has implemented a Lamarckian Genetic Algorithm (LGA). The LGA is a hybrid GA with local search (LS). The programs "docking" most widely used today, and that are generally used for comparison with new proposed methods are: DOCK, FLEXX, AutoDock and GOLD.

Currently, there large banks of molecular structures for public access, for example, the Protein Data Bank (PDB), where various structures are deposited and can be obtained, and several banks of molecular structures of ligands, as the Maybridge and the Cambridge Structural Database.[157] In the problem of docking protein–ligand the purpose is to find and quantify the mode of correct binding a ligand molecule in the active site of a macromolecule receptor such that the receptor function can be enabled or inhibited.[137]

Cheng et al. (2002)[158] performed a study of molecular docking and 3D-QSAR, quantitative relationship three-dimensional structure activity, in order to understand the antimalarial mechanism and the relationship between physicochemical properties and antimalarial activity of artemisinin analogues by molecular docking simulations to probe the interactions of these analogues with heme. The 3D-QSAR was based docking models employing comparative molecular field force analysis (CoMFA) and comparative molecular similarity indices analysis (COMSIA). The subsequent analysis of partial least squares (PLS) indicated that the required energies of the calculation correlated well with the experimental values of the activity. The CoMFA and COMSIA models were active conformations that have demonstrated a good predictive ability. In turn combining together the results of docking.

The studies on the mechanism of action artemisinin have been conducted in search of providing guides for synthesis of new derivatives with improved efficiency and stability in combat malaria *falciparum*. In these studies the activities of artemisinin and its derivatives appear to be mediated by its interaction with iron in hemoglobin, see Figure 8, through endoperoxide function.[159]

The endoperoxide binding (C—O—O—C) of artemisinin is believed to be the key to the mode of action of the drug. The iron in the +2 oxidation state (Fe^{+2}) catalyzes the breaking of this bond, resulting in highly reactive free radicals these free radicals from artemisinin derivatives modify and inhibit a variety of parasite molecules, causing them to die. A rich source of intracellular Fe^{+2} is the heme that is an essential component of hemoglobin, and also responsible for activating artemisinin within the parasite.[160]

Malaria parasites in humans degrade hemoglobin and red blood cell within the heme and globin. Subsequently, globin is hydrolyzed to produce amino acids as the source for protein synthesis. The toxic portion of heme will be principally detoxify by the process of hemozoina polymerization.[147, 148]

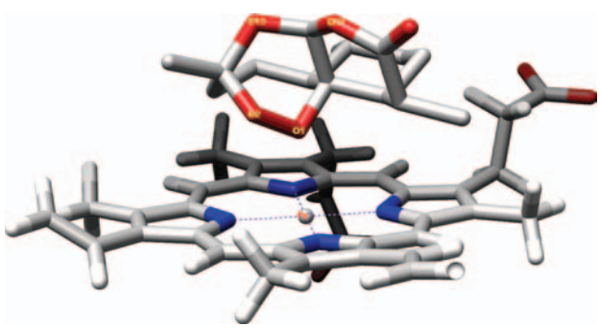

Fig. 8. The molecular docking can be performed using the GOLD software.[156] The structure of heme to the realization of interaction drug–receptor/artemisinin-heme (Molecular Docking) was obtained from the Protein Data Bank (PDB).[157]

8. SOME EXAMPLES OF USE QUANTUM-CHEMICAL PARAMETERS IN STUDIES SAR/QSAR

Martins et al. related the geometric and electronic descriptors derivatives rutaecarpine analogues with biological activity against cancer of the central nervous system (CNS), where calculations of quantum chemistry used was the molecular level B3LYP/6-31 (d) and statistical analyzes were performed for 21 rutaecarpine derived analogues. Of (86) calculated molecular descriptors, (05) were selected for constructing the model of principal component analysis (PCA). The component PC1, which responds by 46.11% of the total variance, was able to completely discriminate compounds in two classes: active and inactive. All molecular descriptors selected by PCA model were electronic parameters. The hierarchical cluster analysis (HCA) was also applied to the descriptors selected by PCA model. Based on (05) descriptors selected were possible to suggest new derivatives assets of the rutaecarpine to be synthesized. Furthermore, a model of partial least squares for discriminant analysis (PLS-DA) was built supervised and successfully applied in discriminating similar to rutaercarpine, which was validated using an independent set of compounds.[48]

One of the first studies of structure-activity relationship of artemisinin and antimalarial molecules using molecular electrostatic potential maps was performed by Thomson, Cory and Zerner (1991), through the theoretical study of the structure of artemisinin and its derivatives using semi-empirical methods (AM1 and PM3) and Hartree-Fock (3-21G, 6-31G). The MEP obtained with AM1 and PM3, when compared showed distinct differences between active and inactive molecules, such as 8-deoxyartemisinin because active molecules have a wide band around the negative potential of the molecule containing endoperoxide binding.[161]

Bernardinelli (1994) used maps of molecular electrostatic potential aiming to identify the key features that are necessary for the antimalarial activity of artemisinin and some derivatives. The MEP showed that the active molecules have a region of negative potential similarly near the trioxane ring, but this region is displaced in inactive compounds. The MEP were used to make predictions of quality new and more effective antimalarial molecules, thereby Bernardinelli et al. concluded that any further active molecule will have MEP similar to artemisinin.[47]

Pinheiro, Ferreira and Romero (2001) techniques combined quantum chemical (Hartree-Fock 3-21G) and multivariate analyzes methods (PCA, HCA, KNN and SIMCA) to study and propose diidroartemisinin derivatives. Through the technique PCA and HCA selected seven (7) descriptors that were responsible for the classification of compounds into two distinct classes, and with construction of qualitative models KNN and SIMCA proposed two

(2) compounds of a set of twelve (12) tested predicted as high activity.[162]

Pinheiro et al. (2003) planned artemisinin derivatives with antimalarial activity with the help of quantum chemistry and partial least squares method (PLS). They built a QSAR model based on five (5) molecular descriptors used to predict the antimalarial activity of ten (10) compounds with unknown activity, and of these one (1) compound was predicted to be more active than the compounds studied. Also performed constructing molecular graphics and molecular docking studies between artemisinin and heme. Proposed by molecular docking of artemisinin and some derivatives with the hemoglobin A, a view of the binding mode between artemisinin-heme.[159]

Ferreira et al. (2010) studied artemisinin and 18 derivatives with antimalarial activity against strains of Plasmodium falciparum W-2, through quantum chemical and multivariate analysis. The optimization of the geometry of the structures was carried out with the theory of Hartree-Fock (HF) and HF/3-21G** basis set. Maps of molecular electrostatic potential (MEP) and molecular docking were used to investigate the interaction between the ligands and the receptor (heme). The principal component analysis (PCA) and hierarchical cluster analysis (HCA) were used to select the most important descriptors related activity.[163]

Leite and colleagues (2010) conducted studies of 18 natural compounds brazilian flora, which have the peroxide group and presumably act in heme protein, leading to a reduction of binding of peroxide and producing radicals which can kill the etiological agent of malaria (Plasmodium falciparum strains), and may show antimalarial activity. These facts motivated to study the interaction of the 18 natural peroxides, initially performing a conformational search using the MM3 method for each molecule. The most stable conformers were optimized by PM3(tm) method. Then there was a docking between peroxide and of the heme group, followed again by a conformational search. In conclusion, the results showed that four of the compounds (10, 13, 14 and 15) may be desirable antimalarial activity.[164]

Barbosa et al. (2011) performed molecular modeling and chemometric studies involving artemisinin and 28 derivatives with anticancer activity against human hepatocellular carcinoma HepG2. The calculation of the studied compounds were performed B3LYP/6-31G** level. The electrostatic potential maps were used in an attempt to identify key structural features of artemisinin and its derivatives, that are required for its activities, and to investigate its interaction with transferrin. The chemometric method (PCA, HCA, KNN, SIMCA and SDA) were used to reduce the dimensionality and investigate which subset of variables could be more effective for the classification of compounds according to their degree of anticancer activity. Furthermore, the molecular docking was used to investigate the interaction between the ligands and receptor. The results showed that the approximation of the ligands to the receptor is through endoperoxide binding.[165]

Recently Figueiredo et al. (2011) developed studies antimalarial compounds with biological activity against Plasmodium falciparum K1. These studies have led to obtaining multivariate models for artemisinin derivatives and a series of dispiro-1,2,4-trioxolanes. A predictive model was generated by PLS method, with three latent variables explaining 99.8% of the total variance, $Q^2 = 0.87$, $R^2 = 0.85$, obtained for 16/4 molecules in the training/external validation set. The descriptors selected for the model were the binding free energy, logarithm of octanol-water partition coefficient (log$P$) and molecular volume. The application of these models has enabled the prediction of activities compounds designed with the information obtained from studies developed. Moreover, the studies of new series of antimalarial compounds are found in the study phase which also can be inferred about the activities of new compounds to be designed.[166]

Carvalho et al. (2011) studied with B3LYP/6-31G** level of theory the artemisinin and 31 analogues with anti leishmanicidal activity against *Leishmania donovani,* and proposed a set of 13 artemisinins, 7 less active and 6 that have not been tested, and of these six, one was expected to be more active against *L. donovani.* In this study, maps Electrostatic Potential (MEP) were used in an attempt to identify key structural features of artemisinin and analogs, and mode of interaction with its receptor (heme). The chemometric methods: PCA, HCA, SDA, KNN and SIMCA were used to reduce dimensionality and investigate which subset of descriptors are responsible for the classification of the activity anti leishmanicidal as larger and smaller.[167]

Recently Cristino e colleagues (2012) used the B3LYP/6-31G* theory to model artemisinin and Nineteen 10-substitued deoxoartemisinin derivatives, with different degrees activity against strains of *Plasmodium falciparum* D-6 of Sierra Leone. The chemometric methods (PCA), (HCA), (KNN), (SIMCA) and (SDA) were employed to reduce the dimensionality and investigate which subset of descriptors are responsible for classification between more and less active deoxoartemisinin antimalarial. The chemometric methods: Principal Component Analysis (PCA), Hierarchical Cluster Analysis (HCA), K-Nearest Neighbor (KNN), Soft Independent Modeling of Class Analogy (SIMCA) and discriminant analysis (SDA) were used to reduce the dimensionality and investigate which subset of descriptors are responsible for classification between more and less active.[168]

Recent studies of 51 peroxides were made to find correlations between *in silico* parameters and experimental data for identifying new antimalarials from natural sources. The interaction of heme was studied by molecular docking refinement followed by conformational analysis using semiempirical parametric method 6 (PM6). The results indicated that compounds 5 and 24 are promising antimalarials.[169]

Zhao (2013) conducted a study about the natures general of Dynamic of proteins and thermodynamics which were analyzed based in the theory irreversible for protein folding and in the theory protein structure thermodynamic. The main contents included:
(1) basic concepts of irreversible thermodynamics,
(2) the irreversible thermodynamic theory to the folding of proteins that reveals the fundamental rules of movement coupled to a protein,
(3) thermodynamic theory of the structure of protein
(4) the dynamic nature and thermodynamics of proteins conformational change,
(5) the role of protein dynamics in enzymatic reaction,
(6) the thermodynamic relationship between the protein and biological function.

Many problems of enzymology and protein science are discussed. The analysis of Zhao showed that the properties and function of the proteins could be well explained by the application of concepts dynamics and thermodynamics proteins. Dynamic of proteins and thermodynamics are fundamental to all biological processes, among them we can mention the molecular docking simulations of antimalarial compounds with heme protein.[170]

In recent years we can realize that studies and research related to computational quantum chemistry and molecular modeling have been growing in an exponential manner. Khrennikov has reported about a series of recent debates about the bases-quantum and incompletion of quantum mechanics and a new classical model which reproduces the bases quantum averages and correlations, classical field theory prequantum statistics.[171]

This article reinforces that theoretical studies contribute significantly to new discoveries and research involving various areas of human knowledge, as we mention:
(1) Chang et al. (2013) has investigated theoretically the encapsulation of amino acids of a carbon nanotube single wall zigzag, and the results revealed the stability of amino acids along the inner wall of the cavity. The essential structural features were observed about the modification of the properties of encapsulated molecules as well as the surface properties. The carbon nanotube experienced has associated geometric distortions with corresponding internal molecular structure.[172]
(2) Kala et al. (2013) computational calculations performed using the semi-empirical method that is computationally flexible to capture various quantitative descriptions of molecules of moderate size, and these were investigated the transport properties for the quantum heterocyclic molecules such as pyrrole, furan and thiophene, based on Green's function nonequilibrium (NEGF) formalism combined with Extended Huckel theory (EHT). In these systems, molecular transport properties are strongly influenced by the geometry of the molecule, the chain length of the molecules and their bonding strength to the electrodes attached to the side.[173]
(3) Li (2013) developed a three-dimensional model to investigate the formation of nanoparticles incorporated by ion beam implantation. The nucleation and growth process dynamics, including the known maturation Ostwald, were successfully reconstructed by a theoretical model. This theoretical model gives a remarkable insight into the formation mechanism, and makes it possible to control and optimize fully the nanostructure through ion implantation technology.[174]

9. FINAL CONSIDERATIONS

This study aimed to demonstrate the variety of types of molecular descriptors based on quantum-chemical calculations derived from the wave functions and charge distribution, and how these descriptors have been used to explain the physico-chemical and/or biological activities in studies of structure-activity relationships.

The selection of the best set of descriptors for the activity under study can be achieved by careful selection and combination among many descriptors. In many cases the quantum-chemical descriptors have physical meaning and are used to unravel the complex mechanisms details involved in intra and intermolecular interaction. It should be highlighted that the quantum-chemical descriptors may be calculated from the structure of the molecules that is, starting from a geometry optimized or determined experimentally (for example, X-rays and nuclear magnetic resonance). However, these descriptors are not completely universal, because they are dependent on the structures and systems studied. Even though it is based on a minimum of energy, the calculated descriptors have values very close to their respective empirical values and indicate trends electronic systems under study. Based on the aspects highlighted in this work is evident, therefore, that the quantum-chemical descriptors have a wide range of applications in SAR and QSAR studies, as well as in many areas of integration of fundamental knowledge of Organic Chemistry, Biochemistry, Molecular Biology, Pharmacology and Pharmaceutical Chemistry.

Thus it is justified to study and develop new derivatives with more potent biological activity using quantum chemical methods (SAR and QSAR), multivariate analysis (PCA, HCA, PLS) analysis of pattern recognition KNN, SDA and SIMCA and Molecular Docking (interaction drug-receptor), using the resulting information as a guide to obtaining of new derivatives active most promising to be synthesized.

Acknowledgments: We gratefully acknowledge the final support of the Brazilian Agency Conselho Nacional de Desenvolvimento Científico e Tecnológico. We also thank the Instituto de Ciências da Saúde of Federal University of Pará for the use of the GaussView and Gaussian softwares. The authors would like to thank the Postgraduate Program in Biotechnology and Biodiversity—Network

BIONORTE and to the Laboratório de Modelagem e Química Computacional, Federal University of Amapá for the computational support.

References and Notes

1. F. M. S. S. Fernandes, Química Computacional: Uma seta na direção do século XXI, edited by H. V. Caetano and M. G. Santos, Cadernos Didácticos de Ciências, Lisboa (**2001**), Chap. 4, pp. 55–74.
2. R. C. Atkinson and P. Suppes, Program in computer-assisted intruction: Final report, Washington, DC. Office of Education (DGEW), edited by M. R. Simonson and A. Thompson, Educational Computing Foundations, Prentice Hall, Upper Saddle River, New Jersey (**1968**).
3. P. Suppes and M. Morningstar *Science* 166, 343 (**1968**).
4. A. A. Ribeiro and I. M. Greca, *Quim. Nova* 26, 542 (**2003**).
5. B. U. Duke, *J. Chem. Educ.* 49, 703 (**1972**).
6. S. Colwell and N. Handy, *J. Chem. Educ.* 65, 21 (**1988**).
7. L. C. Freitas, *Quim. Nova na Escola* 8, 3 (**1998**).
8. IUPAC, *Glossary of Terms Used in Medicinal Chemistry* Iupac recommendations (**1998**).
9. A. Varnek, *J. Chem. Educ.* 77, 222 (**2000**).
10. S. A. Fleming, G. R. Hart, and P. B Savage, *J. Chem. Educ.* 77, 790 (**2000**).
11. C. Rodrigues, *Cadernos Temáticos de Química Nova na Escola* 3, 43 (**2001**).
12. M. P. Allen and D. J. Tildesle, Computer Simulation of Liquids, Oxford University Press (**1987**).
13. A. Leach, Molecular Modelling–Principles and Applications, Pearson Education Limited (**2001**).
14. C. M. R. Sant'Anna, *Quim. Nova* 25, 505 (**2002**).
15. W. J. Hehre, A Guide to Molecular Mechanics and Quantum Chemical Calculations, Wavefunction, Inc. (**2003**).
16. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart, *J. Am. Chem. Soc.* 107, 3902 (**1985**).
17. J. J. P. Stewart, *J. Comput. Chem.* 10, 209 (**1989**).
18. G. B. Rocha, R. O. Freire, A. M. Simas, and J. J. P. Stewart, *J. Comput. Chem.* 27, 1101 (**2006**).
19. A. R. Leach, Molecular Modelling Principles and Applications, Longman, Essex (**1996**).
20. J. B. Foresman and A. E. Frisch, Exploring Chemistry with Electronic Structure Methods: A Guide to Using Gaussian, Gaussian, Inc, USA (**1993**).
21. N. C. Cohen, Guidebook on Molecular Modeling in Drug Design, Academic Press, San Diego (**1996**).
22. I. Carvalho, A. D. L. Borges, and L. S. C. Bernardes, *J. Chem. Educ.* 82, 588 (**2005**).
23. C. G. Wermuth, The Practice of Medicinal Chemistry, 3rd edn., Academic Press, London (**2009**).
24. S. O. Silva, M. J. C. Corrêa, H. R. Bitencourt, W. R. Monteiro, J. Lameira, L. S. Santos, G. M. S. P. Guilhon, and D. S. B. Brasil, *J. Comput. Theor. Nanosci.* 9, 953 (**2012**).
25. P. Politzer and J. S. Murray, *Theor. Chem.* 108, 134 (**2002**).
26. K. C. Weber, K. M. Honório, A. T. Bruni, A. D. Andricopulo, and A. B. F. Silva, *Struct. Chem.* 17, 307 (**2006**).
27. M. Fernández and J. Caballero, *J. Mol. Model* 13, 465 (**2007**).
28. M. A. Ibrahim and A. E. A. Gawad, *J. Comput. Theor. Nanosci.* 9, 1120 (**2012**).
29. A. M. Bayuelo and R. V. Reyes, *J. Math. Chem.* 51, 125 (**2013**).
30. A. Arroio, K. M. Honório, and A. B. F. Silva, *Quim. Nova* 33, 694 (**2010**).
31. Cambridge Soft Corporation, CS ChemDraw 5.0, Chemical Structucture Drawing Standard, Cambridge (**1998**).
32. Cambridge Soft Corporation, CS Chem3D 5.0, Molecular Modeling and Analysis, Cambridge (**1998**).
33. ACD/Chemsketch Freware, version 12.00, Advanced Chemistry Development, Inc, Toronto, ON, Canada (**2010**).
34. Cambridge Soft Corporation, CS Chem3D Ultra 5.0, Ultimate Modeling, Visualization and Analysis, Cambridge (**1998**).
35. I. Carvalho, M. T. Pupo, A. D. L. Borges, and L. S. C. Bernardes, *Quim. Nova* 26, 428 (**2003**).
36. P. Flukiger, Molekel 4.2, Swiss Center for Scientific Computing, Mano, Switzerland (**2002**).
37. Molecular Modeling ProTM4.0, Computational Chemistry Program, ChemSW®, Inc., Fairfield (**2001**).
38. ChemSite: Interative 3D Molecular Modeling, ChemSW® Inc., Fairfield (**2001**).
39. Chemplus, Modular Extensions for HyperChem Release 6.02, Molecular Modeling for Windows, HyperClub, Inc, Gainesville (**2000**).
40. G. L. Patrick, An Introduction to Medicinal Chemistry, 2nd edn., Oxford University Press, Oxford (**2001**).
41. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr, T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, Gaussian 03, Revision C.02, Gaussian, Inc. Wallingford, CT (**2004**).
42. R. Todeschini and P. Gramatica, *Perspectives in Drug Discovery and Design* 355 (**1998**).
43. C. B. R. Santos, J. B. Vieira, A. S. Formigosa, E. V. M. Costa, M. T. Pinheiro, J. O. Silva, W. J. C. Macêdo, and J. C. T. Carvalho, *J. Comput. Theor. Nanosci.* No Prelo (**2013**)
44. R. E. Bruns, P. H. Guadagnini, and A. A. Souza, *Quim. Nova* 19, 148 (**1996**).
45. R. S. Mulliken, *J. Chem. Phys.* 2, 782 (**1934**).
46. L. E. Chirlian and M. M. Francl, *J. Comput. Chem.* 8, 894 (**1987**).
47. Bernardinelli, *Int. J. Quant. Chem* 21, 117 (**1994**).
48. G. R. Martins, H. B. Napolitano, L. T. F. M. Camargo, and A. J. Camargo, *J. Braz. Chem. Soc*. 23, 2183 (**2012**).
49. K. M. Honório and A. B. F. Silva, *Int. J. Quant. Chem.* 95, 126 (**2003**).
50. R. Contreras, L. R. Domingo, J. Andrés, P. Pérez, and O. Tapia, *J. Phys. Chem.* 103, 1367 (**1999**).
51. B. W. Clare, *J. Mol. Struct.* (*Theochem*) 331, 63 (**1995**).
52. B. W. Clare, *J. Mol. Struct. (Theochem)* 337, 139 (**1995**).
53. B. W. Clare, *Theor. Chim. Acta* 87, 415 (**1994**).
54. C. A. Heaton, A. K. Miller, and R. L. Powell, *J. Fluorine Chem.* 107, 1 (**2001**).
55. G. H. Grant and W. G. Richards, Computational Chemistry, Oxford Science Publications (**1996**).
56. G. Zhang and C. B. Musgrave, *J. Phys. Chem.* 111, 1554 (**2007**).
57. C. C. Lobato, E. M. Silva, J. B. Vieira, W. J. C. Macêdo, E. Viana, J. C. T. Carvalho, and C. B. R. Santos, *Estudo computacional da reatividade dos produtos isoméricos em reações de halidrificação através da análise dos orbitais de fronteiras—HOMO e LUMO*. In 52° Congresso Brasileiro de Química—Química e Inovação: Caminho Para a Sustentabilidade, Recife-PE, Brasil (**2012**).
58. D. F. Shriver, P. W. Atkins, and C. H. Langford, Inorganic Chemistry, 2nd edn., Oxford University, Oxford (**1994**).

59. R. G. Parr and P. K. Chattaraj, *J. Am. Chem. Soc.* 113, 1854 (**1991**).
60. A. V. Bunge, Introdução à Química Quântica, Edgard Blucher Ltda, São Paulo (**1977**).
61. R. G. Pearson and W. E. Palke, *J. Phys. Chem.* 96, 3283 (**1992**).
62. C. C. Lobato, R. D. Trindade, J. E. M. Diniz, and C. B. R. Santos, *Análise Computacional dos Produtos Isoméricos de Cadeia da Reação de Halidrificação em Alcenos*. In 51° Congresso Brasileiro de Química–Meio Ambiente e Energia, São Luís-MA, Brasil (**2011**).
63. J. A. Hirsch, Concepts in Theoretical Organic Chemistry Allynand Baco, Boston (**1975**).
64. E. J. Lien, Z. R. Guo, and C. T. Su, *J. Pharm. Sci.* 71, 641 (**1982**).
65. I. N. Levine, Quantum Chemistry, 4th edn., Prentice Hall, New York (**1991**).
66. A. Szabo and N. S. Ostlund, Modern Quantum Chemistry: Introduction to Advanced Electronics Structure Theory, Macmillan, New York (**1982**).
67. C. Hansch, R. M. Muir, T. Fujita, P. P. Maloney, C. F. Geiger, and M. J. Streich, *J. Am. Chem. Soc.* 85, 2817 (**1963**).
68. C. Hansch and T. Fujita, *J. Am. Chem. Soc*. 86, 1616 (**1964**).
69. S. M. Free and J. W. Wilson, *J. Med. Chem.* 7, 395 (**1964**).
70. F. A. Pasha, H. K. Srivastava, and P. P. Singh, *Bioorg. Med. Chem.* 13, 6823 (**2005**).
71. K. M. Honório, R. C. Garratt, I. Polikarpov, and A. D. Andricopulo, *J. Mol. Graph. Model* 25, 921 (**2007**).
72. K. C. Weber, K. M. Honório, A. D. Andricopulo, and A. B. F. Silva, *Med. Chem.* 4, 328 (**2008**).
73. J. Padmanabhan, R. Parthasarathi, V. Subramanian, and P. K. Chattaraj, *Bioorg. Med. Chem.* 14, 1021 (**2006**).
74. J. Ghasemi, S. Saaidpour, and S. D. Brown, *J. Mol. Struct.* (*Theochem*) 805, 27 (**2007**).
75. A. A. Toropov, A. P. Toropova, and E. Benfenati, *Chem. Phys. Lett.* 461, 343 (**2008**).
76. M. Karelson and V. S. Lobanov, *Chem. Rev.* 96, 1027 (**1996**).
77. K. C. Weber, K. M. Honório, S. L. Silva, R. Mercadante, and A. B. F. Silva, *Int. J. Quantum Chem.* 103, 731 (**2005**).
78. F. A. Molfetta, A. T. Bruni, K. M. Honório, and A. B. F. Silva, *Eur. J. Med. Chem.* 40, 329 (**2005**).
79. K. M. Honório, R. C. Garratt, and A. D. Andricopulo, *Bioorg. Med. Chem. Lett.* 15, 3119 (**2005**).
80. K. M. Honório, R. C. Garratt, I. Polikarpov, and A. D. Andricopulo, *Lett. Drug Des. Disc.* 3, 261 (**2006**).
81. K. M. Honório, R. C. Garratt, I. Polikarpov, and A. D. Andricopulo, *J. Mol. Graph. Model* 25, 921 (**2007**).
82. A. K. Debnath, *Mini-Rev. Med. Chem.* 1, 187 (**2001**).
83. A. Vedani and M. Dobler, *J. Med. Chem.* 45, 2139 (**2002**).
84. R. D. Cramer, D. E. Patterson, and J. D. Bunce, *J. Am. Chem. Soc.* 110, 5959 (**1988**).
85. G. Klebe, U. Abraham, and T. Mietzner, *J. Med. Chem.* 37, 4130 (**1994**).
86. A. J. Hopfinger, S. Wang, J. S. Tokarski, B. Jin, M. Albuquerque, P. J. Madhav, and C. Duraiswami, *J. Am. Chem. Soc.* 119, 10509 (**1997**).
87. A. Vedani and M. Dobler, *J. Med. Chem.* 45, 2139 (**2002**).
88. A. Vedani, M. Dobler, and M. A. Lill, *J. Med. Chem.* 48, 3700 (**2005**).
89. F. Tian, P. Zhou, and Z. Li, *J. Mol. Struct.* 871, 7140 (**200**).
90. M. P. Freitas, S. D. Brown, and J. A. Martins, *J. Mol. Struct.* 738, 149 (**2005**).
91. M. P. Freitas, *Org. Biomol. Chem.* 4, 1154 (**2006**).
92. M. P. Freitas, *Curr. Comput.-Aid. Drug Des.* 3, 235 (**2007**).
93. M. P. Freitas, *Chemom. Intell. Lab. Sys.* 91, 173 (**2008**).
94. M. Goodarzi and M. P. Freitas, *Chemom. Intell. Lab. Sys.* 96, 59 (**2009**).
95. M. Goodarzi and M. P. Freitas, *Mol. Simul.* 36, 267 (**2010**).
96. R. A. Cormanich, M. P. Freitas, and R. Rittner, *J. Braz. Chem. Soc.* 22, 637 (**2011**).
97. M. Goodarzi and M. P. Freitas, *J. Phys. Chem. A.* 112, 11263 (**2008**).
98. M. Goodarzi, M. P. Freitas, and T. C. Ramalho, *Spectrochim. Acta Part. A.* 74, 563 (**2009**).
99. M. Goodarzi and M. P. Freitas, *Separ. Purif. Technol.* 68, 363 (**2009**).
100. R. A. Cormanich, C. A. Nunes, and M. P. Freitas, *Quim. Nova* 35, 1157 (**2012**).
101. R. Franke and A. Gruska, *Chemometrics Methods in Molecular Design* 115 (**1996**).
102. A. J. Camargo, R. Mercadante, K. M. Honório, C. N. Alves, and A. B. F. Silva, *J. Mol. Struct.* (*Theochem*) 583, 105 (**2002**).
103. J. Souza, R. H. D. Santos, M. M. C. Ferreira, F. A. Molfetta, A. J. Camargo, K. M. Honório, and A. B. F. Silva, *Eur. J. Med. Chem.* 38, 929 (**2003**).
104. J. Souza, F. A. Molfetta, K. M. Honório, R. H. A. Santos, A. B. F. Silva, *J. Chem. Inf. Comput. Sci.* 44, 1153 (**2004**).
105. K. M. Honório and A. B. F. Silva, *J. Mol. Mod.* 11, 200 (**2005**).
106. A. K. Calgarotto, S. Miotto, K. M. Honório, A. B. F. Silva, S. Marangoni, J. L. Silva, M. Comar, K. M. T. Oliveira, and S. L. Silva, *J. Mol. Struct.* (*Theochem*) 808, 25 (**2007**).
107. K. C. Weber, K. M. Honório, A. T. Bruni, and A. B. F. Silva, *J. Mol. Mod.* 12, 915 (**2006**).
108. A. Arroio, K. M. Honório, and A. B. F. Silva, *J. Mol. Struct.* (*Theochem*) 709, 223 (**2004**).
109. F. A. Molfetta, K. M. Honório, C. N. Alves, and A. B. F. Silva, *J. Mol. Struct.* (*Theochem*) 674, 191 (**2004**).
110. A. J. Camargo, K. M. Honório, R. Mercadante, F. A. Molfetta, C. N. Alves, and A. B. F. Silva, *J. Braz. Chem. Soc.* 14, 809 (**2003**).
111. V. R. S. Malta, A. V. Pinto, F. A. Molfetta, K. M. Honório, C. A. Simone, M. A. Pereira, R. H. A. Santos, and A. B. F. Silva, *J. Mol. Struct.* (*Theochem*) 634, 271 (**2003**).
112. A. Arroio, E. F. Lima, K. M. Honório, and A. B. F. Silva, *Struct. Chem.* 20, 577 (**2009**).
113. C. N. Alves, J. C. Pinheiro, A. J. Camargo, M. M. C. Ferreira, and A. B. F. Silva, *J. Mol. Struct.* (*Theochem*) 530, 39 (**2000**).
114. C. N. Alves, J. C. Pinheiro, A. J. Camargo, M. M. C. Ferreira, R. A. F. Romero, and A. B. F. Silva, *J. Mol. Struct.* (*Theochem*) 541, 81 (**2001**).
115. M. M. C. Ferreira, *J. Braz. Chem. Soc.* 13, 742 (**2002**).
116. M. M. C. Ferreira, C. A. Montanari, and A. C. Gaudio, *Quim. Nova* 25, 439 (**2002**).
117. H. Kubinyi, *Methods and Principles in Medicinal Chemistry* 1, 21 (**1993**).
118. S. Wold, *Chemometrics Methods in Molecular Design* 202 (**1996**).
119. A. C. Gaudio and E. Zandonade, *Quim. Nova* 25, 658 (**2001**).
120. B. Barros Neto, I. S. Scarminio, and R. E. Bruns, Planejamento e Otimização de Experimentos, 2nd edn., Campinas, UNICAMP (**1996**), pp. 152–298.
121. D. T. Stanton, *J. Chem. Inf. Comput. Sci.* 43, 1423 (**2003**).
122. M. Sjöström and L. Ricksson, *Chemometrics Methods in Molecular Design* 68 (**1996**).
123. G. L. Patrick, *An Introduction to Medicinal Chemistry* 3, 293 (**2006**).
124. R. G. Brereton, Chemometrics: Data Analysis for the Laboratory and Chemical Plant, John Wiley & Sons, Ltd. (**2003**).
125. W. P. Gardiner, Statistical Analysis Methods for Chemists, Royal Society Chemistry, Cambridge (**1997**).
126. Brasil, Ministerio da Saude, Secretaria de Vigilancia em Saude, Departamento de Vigilancia Epidemiologica, *Guia pratico de tratamento da malaria no Brasil*, Brasília (**2012**).
127. M. Srivastava, H. Singh, and P. K. Naik, *Curr. Res. J. Biol. Sci.* 2, 83 (**2010**).
128. A. R. Butler, and Y.-L. Wu, *Chem. Soc. Rev*. 21, 85 (**1992**).
129. E. Sun and F. E. Cohen, *Gene* 137, 127 (**1993**).
130. J. G. Lombardino and J. A. Lowe, III, *Nature Reviews–Drug Discovery* 3, 853 (**2004**).

131. Bowman, D. Lawson, D. Bashan, D. Brown, T. Chillingworth, C. M. Churcher, A. Craig, R. M. Davies, K. Deulin, T. Feltwell, S. Genthes, R. Gwilliam, N. Ham, D. Harnis, S. Holroyd, T. I-Iornsby, P. Horrocks, K. Jageis, B. Jassal, S. J. Kyes, S. McLean, S. Moule, K. Mungail, L. Murphy, K. Oliver, M. A. Quali, M.-A. Rajadream, S. Rutter, J. Skelton, R. Squares, S. Squares, J. E. Su, S. Whitehead, J. R. Woodward, C. Nawbold, and B. G. Barreil, *Nature* 400, 532 **(1999)**.

132. A. Korolkovas, *Ciênc. e Cult.* 41, 528 **(1989)**.

133. N. C. Cohen, J. M. Blaney, C. Humblet, P. Gund, and D. C. Bany, *J. Med. Chem*. 33, 883 **(1990)**.

134. A. Sali and T. L. Blundell, *J. Mol. Biol.* 234, 779 **(1993)**.

135. M. A. Avery, *J. Med. Chem*. 36, 4264 **(1993)**.

136. S. K. Kokpol, *Analitycal Sciences* 4, 565 **(1998)**.

137. E. J. Barreiro and C. A. M. Fraga, Química Medicinal: As bases moleculares da ação dos fármacos, Artmed, Porto Alegre **(2001)**.

138. R. A. Toledo, M. C. Santos, K. M. Honório, A. B. F. Silva, E. T. G. Cavalheiro, and L. H. Mazo, *Anal. Lett.* 39, 507 **(2006)**.

139. R. A. Toledo, M. C. Santos, K. M. Honório, A. B. F. Silva, E. T. G. Cavalheiro, and L. H. Mazo, *Quim. Nova* 28, 456 **(2005)**.

140. H. B. Suffredini, M. C. Santos, D. Souza, L. Codognoto, P. Homem-de-Mello, K. M. Honório, A. B. F. Silva, S. A. S. Machado, and L. A. Avaca, *Anal. Lett.* 38, 1587 **(2005)**.

141. P. Labute, *J. Mol. Graphics Model* 18, 464 **(2000)**.

142. D. L. Klayman, *Science* 228, 1049 **(1985)**.

143. R. K. Haynes and S. C. Vonwiller, *Tetrahedron. Lett.* 37, 253 **(1996)**.

144. M. Jung, *Curr. Med. Chem.* 1, 35 **(1994)**.

145. J. N. Cumming, P. Ploypradith, and G. H. Posner, *A. Pharmacol* 37, 253 **(1997)**.

146. World Health Organization, *Guidelines for the Treatment of Malaria* **(2006)**.

147. S. Tonmunphean, V. Parasuk, and S. Kokpol, *Quant. Struct-Act. Relat.* 19, 475 **(2000)**.

148. S. Tonmunphean, V. Parasuk, and S. Kokpol, *Quant. Struct-Act. Relat.* 19 **(2000)**.

149. S. Tonmunphean, V. Parasuk, and S. Kokpol, *J. Mol. Model* 7, 26 **(2001)**.

150. G. H. Posner, J. N. Cummings, P. Ploypradith, and. C. O. Oh, *J. Am. Chem. Soc.* 117, 5885 **(1995)**.

151. W. Asawamahasakda, I. Ittarat, Y.-M. Pu, H. Ziffer, and S. R. Meshnick, *Antimicrob. Agents Chemother.* 38, 1854 **(1994)**.

152. C. W. Jefford, *Curr. Med. Chem.* 8, 1803 **(2001)**.

153. P. L. Olliaro, *Trends in Parasitology* 17, 122 **(2001)**.

154. N. H. Morgon and K. Coutinho, *Métodos de química teórica e modelagem molecular*, Livraria da física **(2007)**, pp. 491–527.

155. R. S. Judson, E. P. Jaegar, A. M. Treasurywala, and M. L. Peterson, *J. Comput. Chem.* 14, 1407 (1**993)**.

156. G. Jones, *J. Mol Biol.* 267, 727 **(1997)**.

157. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, *The Protein Data Bank. Nucleic Acids Research* 28, 235 **(2000)**.

158. F. Cheng, *Bioorg. Med. Chem.* 10, 2883 **(2002)**.

159. J. C. Pinheiro, *QSAR Comb. Sci.* 22, 830 **(2003)**.

160. R. G. Ridley, *Science* 424, 887 **(2003)**.

161. C. Thomson, M. Cory, and M. Zerner, *Int. J. Quant. Chem.* 18, 231 **(1991)**.

162. J. C. Pinheiro, M. M. C. Ferreira, and O. A. S. Romero, *J. Mol. Struct.* (*Theochem*) 572, 35 **(2001)**.

163. J. E. V. Ferreira, A. F. Figueiredo, J. P. Barbosa, M. G. G. Cristino, W. J. C. Macêdo, O. P. P. Silva, B. V. Malheiros, R. T. Serra, and J. C. Pinheiro, *J. Serb. Chem. Soc.* 75, 1533 **(2010)**.

164. F. H. A. Leite, A. G. Taranto, M. C. Santos Junior, A. Branco, M. T. Araujo, and J. W. M. Carneiro, *Int. J. Quant. Chem*. 110, 2057 **(2010)**.

165. J. P. Barbosa, J. E. V. Ferreira, A. F. Figueiredo, R. C. O. Almeida, O. P. P. Silva, J. R. C. Carvalho, M. G. G. Cristino, J. C. Pinheiro, J. L. F. Vieira, and R. T. A. Serra, *J. Serb. Chem. Soc*. 76, 1263 **(2011)**.

166. A. F. Figueiredo, J. E. V. Ferreira, J. P. Barbosa, W. J. C. Macedo, M. G. G. Cristino, M. S. Lobato, J. C. Pinheiro, R. T. A. Serra, *J. Comput. Theor. Nanosci*. 8, 1847 **(2011)**.

167. J. R. C. Carvalho, J. E. V. Ferreira, J. P. Barbosa, M. S. Lobato, C. C. F. Meneses, M. M. Soeiro, M. S. Farias, R. C. O. Almeida, K. C. Ventura, J. C. Pinheiro, and A. L. R. Pinheiro, *J. Comput. Theor. Nanosci*. 8, 2193 **(2011)**.

168. M. G. G. Cristino, C. C. F. Meneses, M. M. Soeiro, J. E. V. Ferreira, A. F. Figueiredo, J. P. Barbosa, R. C. O. Almeida, J. C. Pinheiro, and A. L. R. Pinheiro, *J. Theor. Comput. Chem.* 11, 241 **(2012)**.

169. F. H. A. Leite, J. W. M. Carneiro, M. T. Araujo, M. Comar, and A. G. Taranto, *Int. J. Quant. Chem*. 112, 3390 **(2012)**.

170. Q. Zhao, *Reviews in Theoretical Science* 1, 83 **(2013)**.

171. A. Khrennikov, *Reviews in Theoretical Science* 1, 34 **(2013)**.

172. C. M. Chang, H. L. Tseng, A. de Leon, A. Posada-Amarillas, and A. F. Jalbout, *J. Comput. Theor. Nanosci*. 10, 521 **(2013)**.

173. C. P. Kala, P. A. Priya, and D. J. Thiruvadigal, *J. Comput. Theor. Nanosci*. 10, 213 **(2013)**.

174. K. D-. Li, *J. Comput. Theor. Nanosci*. 10, 629 **(2013)**.