We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Chapter

Molecular Electrostatic Potential and Chemometric Techniques as Tools to Design Bioactive Compounds

Marcos Antônio B. dos Santos, Luã Felipe S. de Oliveira,
Antônio Florêncio de Figueiredo, Fábio dos Santos Gil,
Márcio de Souza Farias, Heriberto Rodrigues Bitencourt,
José Ribamar B. Lobato, Raimundo Dirceu de P. Farreira,
Sady Salomão da S. Alves, Edilson Luiz C. de Aquino
and José Ciríaco-Pinheiro

Abstract

In this chapter, firstly, we briefly review aspects of the approximation of quantum chemistry, molecular electrostatic potential (MEP), and chemometrics techniques, which are accredited as important tools in the development of chemical science and are frequently used in the study and design of bioactive compounds. Ultimately, we use MEP and pattern recognition (PR) techniques as tools to design nitrofuran compounds with biological activity against *Trypanosoma cruzi* (*T. cruzi*). PR models (PCA, HCA, KNN, SDA, and SIMCA) were constructed and demonstrated that 23 nitrofurans can be classified into two classes or groups: more active and less active according to their degrees of activity against *T. cruzi*. Properties such as charge on the N atom of the nitro group (QN1); the difference between the highest occupied molecular orbital (HOMO) energy and the lowest unoccupied molecular orbital (LUMO) energy (GAP energy); molecular representation of structure based on electron diffraction code of signal 5, unweighted (Mor05u); and Moriguchi water–octanol partition coefficient (MlogP) are responsible for the classification into more active and less active studied nitrofurans. It is interesting to notice that these properties represent three distinct classes of interactions between the nitrofurans and the biological receptor: electronic (QN1 and GAP energy), steric (Mor05u), and hydrophobic (MlogP). The results of the application of PR models on the validation set evidenced two nitrofuran compounds (compounds **25** and **30**) as more promising for synthesis and biological assays, which in the future can be used to validate our PR models.

Keywords: molecular electrostatic potential, chemometric techniques, pattern recognition techniques, chemoinformatics, design of bioactive compounds

1. Introduction

Reports of theoretical bases of MEP and the development of efficient computational methods state that MEP has become an important reactivity index in studies of a large variety of molecular interactions [1]. The usefulness of this theoretical approach in studies and interpretation of chemical, biochemical, and related phenomena is well documented [2–18].

Chemometrics is a discipline that collects mathematical, statistical, information theory, and computer science tools to deal with complex chemical data [19–22]. PR techniques were introduced in the chemistry, at the beginning of the 1970s, to analyze various types of spectroscopic data. Since then, PR became part of chemometrics and has been an excellent tool to aid in the interpretation of chemical data to obtain relevant information in different application sectors of chemical science [19, 20]. PR techniques are especially useful for the classification of objects into discrete classes on the basis of measured features. A set of characteristic features of an object is considered as an abstract pattern that contains information about a not directly measured property of the object [19].

The MEP and PR techniques have been used as independent strategies in the study of active compounds and lead to the proposal of new molecules for synthesis and biological testing. The joint applications of these powerful tools were described carefully, to unravel the structure-activity relationship of bioactive compounds, consequently proposing new molecules. Therefore, a more intense exploration of its potentials is needed in order to design biologically active compounds.

The design of molecules with a desired property is one of the objectives of chemoinformatics. In this chapter, we present a study of the application of MEP and PR techniques to design nitrofuran compounds with potential activity against *T. cruzi*. In the first step of our study, MEP maps will be used in an attempt to identify the key structural features of nitrofuran compounds that are necessary for their activities and investigate their probable interactions with a molecular receptor through recognition in a biological process. Subsequently, PR techniques are used to construct models that will be applied later to a forecast set constructed with the accumulated perceptions in the MEP studies.

2. MEP and chemometrics techniques as tools for the design of bioactive compounds: a brief review

According to the literature, MEP [1, 3] has been a tool of quantum chemistry used by researchers for several decades to study and understand the relationships between structure and activity of molecules. Among the papers that point out the importance of this tool in the matter, and consequently in the planning of bioactive compounds, we can mention those reported by Bernardinelli et al. [23] and by Jefford et al. [24].

Another tool, in the form of a set of techniques has been used emphatically over the years in the understanding of the structure-activity relationship of molecules is Chemometrics [25–27]. This set of techniques has also enables the planning of new biologically active compounds, and most of the developed research is focused on the construction of QSAR (quantitative structure-activity relationship) models.

The combination of MEP and chemometrics as tools for designing new bioactive compounds has almost always been focused on the elaboration of quantitative models, for example, the CoMFA methodology [28]. This methodology was developed in the late 1980s by Cramer et al. [29]. Its application is richly extensive and recently it has been used in several studies of structure–activity relationships of bioactive

compounds. Chatbar et al. conducted a study of triazine morpholino derivatives as mTOR inhibitors for the treatment of breast cancer [30]. Pourbasheer et al. performed 3D-QSAR and 2D-QSAR analyses on the series of compounds hepatitis C virus NS5B polymerase inhibitors [31]. Cramer applied the CoMFA methodology for a large majority of 116 biological targets and obtained acceptable 3D-QSAR models [32]. Cramer et al. introduced in the literature a novel alignment methodology for training or test set structures in 3D-QSAR [33]. Dong et al. performed QSAR analyses of aromatic heterocycle thiosemicarbazone analogues for finding novel tyrosinase inhibitors [34]. Dong et al. built 3D-QSAR models of dabigatran analogues as thrombin inhibitors [35]. Ding et al. performed 3D-QSAR models of 6-aryl-5-cyano-pyrimidine derivatives to explore the structure requirements of LSD1 inhibitors [36].

Applications of MEP to investigate the key features of compounds that are necessary for their biological activities and thus proposing new derivatives as well as the construction of chemometric models as indicative of the most promising among the new derivatives for syntheses and biological assays were reported by us in literature [37–43]. Pinheiro et al. stated the use of MEP and partial least squares regression (PLS) method in the design of new artemisinin derivatives with activities against *Plasmodium falciparum* [37]. Cardoso et al., using MEP maps and multivariate QSAR, designed new artemisinin derivatives with antimalarial activity [38]. Ferreira et al., through MEP maps and multivariate analysis, designed antimalarial artemisinins [39]. Figueiredo et al. designed new derivatives of dispiro-1,2,4-trioxolones with activity against falciparum malaria [40]. Carvalho et al., through maps of MEP and pattern recognition methods, proposed new artemisinin derivatives with activity against *Leishmania donovani* [41]. Barbosa et al. used MEP maps and pattern recognition techniques to plan new derivatives of artemisinin anticancer HepG2 [42]. Cristino et al. proposed new derivatives of 10-substituted Deoartemisinis with activity against *P. falciparum* [43] through the use of MEP maps and pattern recognition techniques.

3. MEP and PR techniques as tools to design nitrofuran compounds with biological activity against *T. cruzi*

3.1 Computational

3.1.1 Biological recognition process ligand/receptor through the molecular electrostatic potential

The MEP is also suitable for analyzing processes based on the "recognition" of one molecule by another as in drug-receptor and enzyme-substrate interactions, because it is through their potentials that the two species first "see" each other [2, 3, 44–46].

MEP for the electronic density is a very useful property for understanding the site of electrophilic attack and nucleophilic reactions as well as the hydrogen bonding interactions [46]. 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 charge test (a proton) located at $\vec{r}$. Being a real physical property, MEP can be determined experimentally by diffraction or by computational tools [3]. For the studied nitrofuran molecules, the MEP values were computed through Eq. (1) [45]

$$V(\vec{r}) = \sum_{j=1}^{K} \frac{Z_j}{\left|\vec{R_j} - \vec{r}\right|} - \int \frac{\rho(\vec{r'})d\vec{r'}}{\left|\vec{r'} - \vec{r}\right|} \tag{1}$$

where K is the number of nuclei with charges Z_j, located at position R_j and $\rho \, (\vec{r})$ is the electronic charge density. The first term on the right side of Eq. (1) represents the contribution of the nuclei, which is positive; the second term brings in the effect of the electrons, which is negative. In the investigation of the reactive sites of nitrofuran compounds, the MEP was evaluated through of the HF/6-31G method.

3.1.2 RP techniques

In this section, we will make a brief presentation of the PR techniques used in this chapter. A deeper and detailed description of these matters can be found elsewhere [47–66].

3.1.2.1 Principal component analysis (PCA) technique

When computing large multivariate data, it is mandatory to find and reduce unknown data trends using exploratory tools. The main idea of the PCA technique is to reduce the dimensionality of a data set consisting of large numbers of inter-related variables while retaining the variation present in the data set as much as possible. This can be achieved by transforming them into a new set of variables, the PCs, which are uncorrelated and ordered so that the first few retain most of the variation present in all of the original variables. As the final result, the PCA technique performs the selection of a small number of variables (molecular properties) considered better related to the dependent property or feature [67], in this study, the biological activity against *T. cruzi*.

3.1.2.2 Hierarchical cluster analysis (HCA) technique

This technique has become, together with PCA, another important tool in pattern recognition [67]. The purpose of using it is to display the data in such a way as to emphasize its natural clusters and patterns in a two-dimensional space. The results are presented as dendrograms. In HCA technique, the distances between objects or variables are calculated and computed through the similarity index which ranges from zero, that is, no similarity and large distance among objects, to one, for identical objects.

3.1.2.3 K-nearest neighbor (KNN) technique

The KNN technique [67] classifies the objects based on distance comparison among them. The multivariate Euclidean distances between every pair of objects with known class membership are calculated. The closest K objects are used to build the model. The optimal K is determined by cross-validation applied to the training set objects. The classification of a test object is determined based on the multivariate distance of this object with respect to the K objects in the training set. In this technique no assumption is made about the size and shape of the training set classes.

3.1.2.4 Stepwise discriminant analysis (SDA) technique

This technique separates objects from distinct populations and allocates new objects into populations previously defined. It uses a stepwise procedure in which, at each step, the most powerful variable is entered into the discriminant function. The SDA technique is anchored in the F-test for the significance of variables and at each step selects a variable based on its significance, and, after several steps, the most significant variables are extracted from the set in question [20, 68].

3.1.2.5 Soft independent modeling of class analogy (SIMCA) technique

This SIMCA technique develops principal component models for each training set category. Its main objective is the reliable classification of new samples. When a prediction is made with the SIMCA technique, new samples insufficiently close to the PC space of a class are considered nonmembers. Furthermore, the technique requires that each training sample be pre-assigned to one of Q different categories, where Q is typically greater than one. It provides three possible outcome predictions: the sample fits only one pre-defined category, the sample does not fit any of the pre-defined categories, and the sample fits into more than one pre-defined category [67].

3.1.3 Computers, software, compounds, and molecular descriptors

For the present chapter, we performed molecular calculations on an AMD PHENOM 955 X4 2.2 GHz processor with 4 Gb of RAM with the Gaussian 98 program package [69]. The MEP was computed from the electronic density, and the maps were displayed using the MOLEKEL software [70], while the PR models were carried out on a PC Pentium machine with the Pirouette program [71].

Figure 1 shows the 2D structure of the 5-nitrofuran-2-aldoxim molecule [72] used in the selection of method/basis set (see Section 3.1.3.1). In **Figures 2** and **3** the 2D structures of the nitrofuran compounds from the training [73–75] and prediction sets are displayed, respectively. In this work, the nitrofuran molecules were defined as more active against *T. cruzi*, when in vitro *growth rate inhibition (GR) T. cruzi* ≥ 75, and as less active when in vitro *growth rate inhibition T. cruzi* < 75.

In general, the structure–activity relationship shows that for the compounds **1–6**, the increase in the carbon chain improves the activity against *T. cruzi*. The comparison between compounds **3** and **2** evidences increased activity by the substitution of the N atom by O. We can also notice that increasing the number of unsaturations and returning the nitrogen to the chain will lead to a decrease in biological activity (**7, 8**). Still in relation to compound **1**, increasing the unsaturations, returning the atom of O, and increasing the carbon chain length (**9–12**) substantially increase the activity against *T. cruzi*. On the other hand, in compounds **13** and **14**, returning to an unsaturation in the main chain and introducing electron-withdrawing groups and more electronegative atoms, there is a decrease in chagasic activity. This evidence can also be verified for compounds **16, 17, 19–22**.

The molecular descriptors were obtained for the most stable conformation of each compound. These descriptors were computed to give information about the influence of electronic, steric, hydrophilic, and hydrophobic features on the anti-trypanosomal activity of the studied nitrofurans. The atomic charges in this work were derived from the electrostatic potential obtained with HF/6-31G method/basis
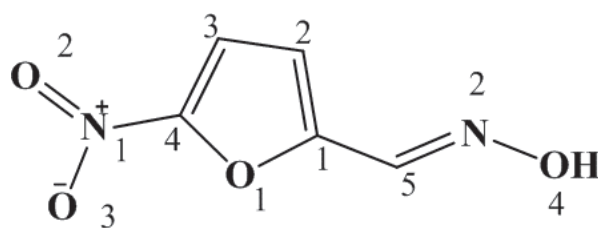

Figure 1.
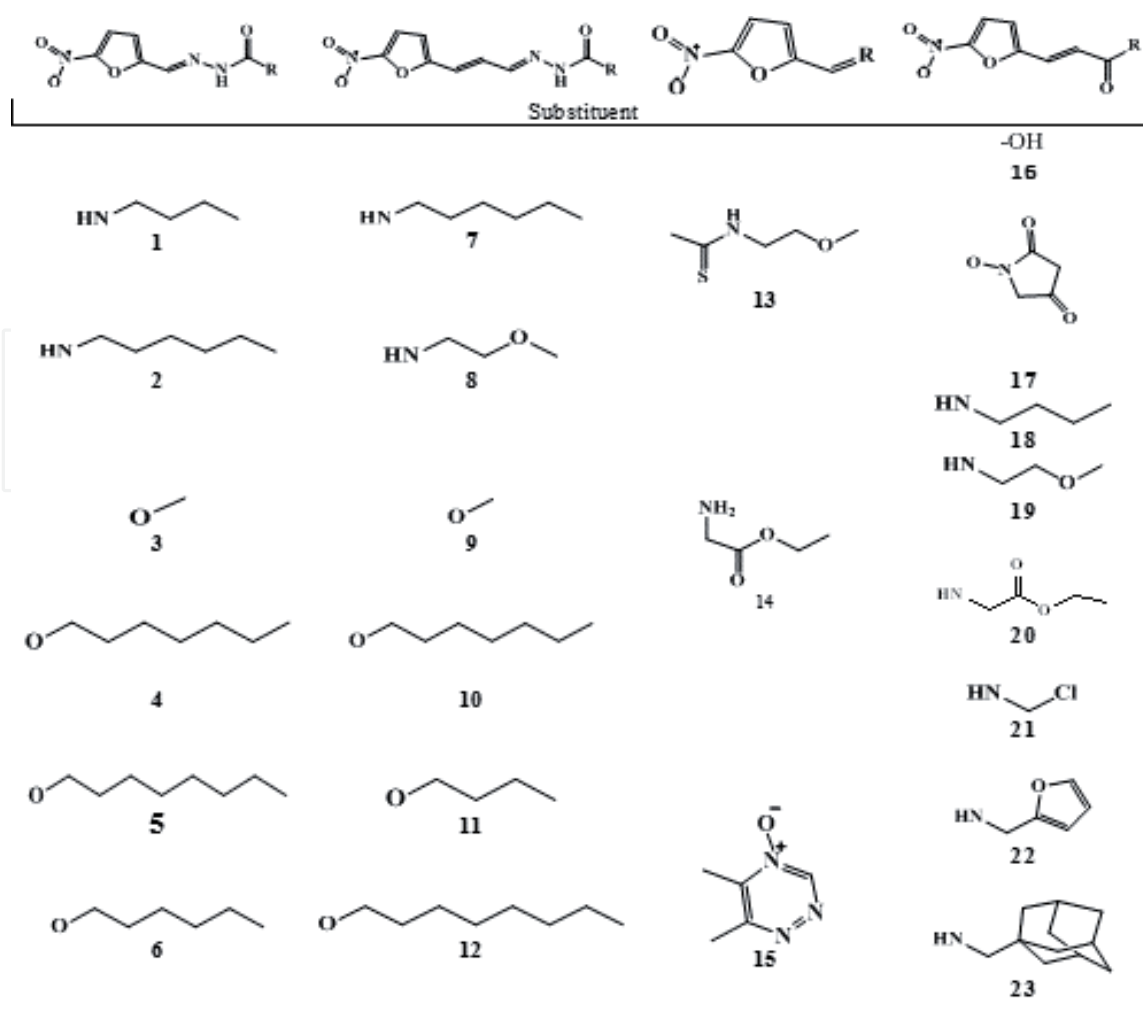*2D molecular structure for 5-nitrofuran-2-aldoxime.*

Figure 2.
2D molecular structure for nitrofurans (training set).

set as implemented in the Gaussian program package. The electrostatic potential is obtained through the calculation of a set of punctual atomic charges so that it represents the possible best quantum molecular electrostatic potential for a set of points defined around the molecule [76, 77]. The charges derived from electrostatic potential present the advantage of being, in general, physically more satisfactory than the charges of Mülliken [78], especially with regard to biological activity.

The quantum–chemical descriptors employed and obtained with the Gaussian 98 program package [69] were total energy of molecules (TE), highest occupied molecular orbital (HOMO) energy, one level below to highest occupied molecular orbital (HOMO–1) energy; lowest unoccupied molecular orbital (LUMO) energy, one level about lowest unoccupied molecular orbital (LUMO+1) energy, HOMO energy–LUMO energy (gap energy), total dipole moment (μ), Mulliken's electro-negativity (χ), atomic charges on the Nth atom (QN), molecular hardness (HD), and molecular softness (MS).

The physicochemical descriptors obtained with ChemPlus module [79] were total surface area (TSA), molecular volume (VOL), molecular refractivity (MR), and molecule hydration energy (MHE).

Molecular holistic (MH) descriptors were included with the purpose of repre-senting different sources of chemical information in terms of molecular size, sym-metry, and distribution of atoms in molecules. Also, we include topologic indices, connectivity indices, geometric descriptors, 3D-MoRSE descriptors, and Moriguchi octanol–water partition coefficient (MlogP). These descriptors were calculated with the Dragon software [80].
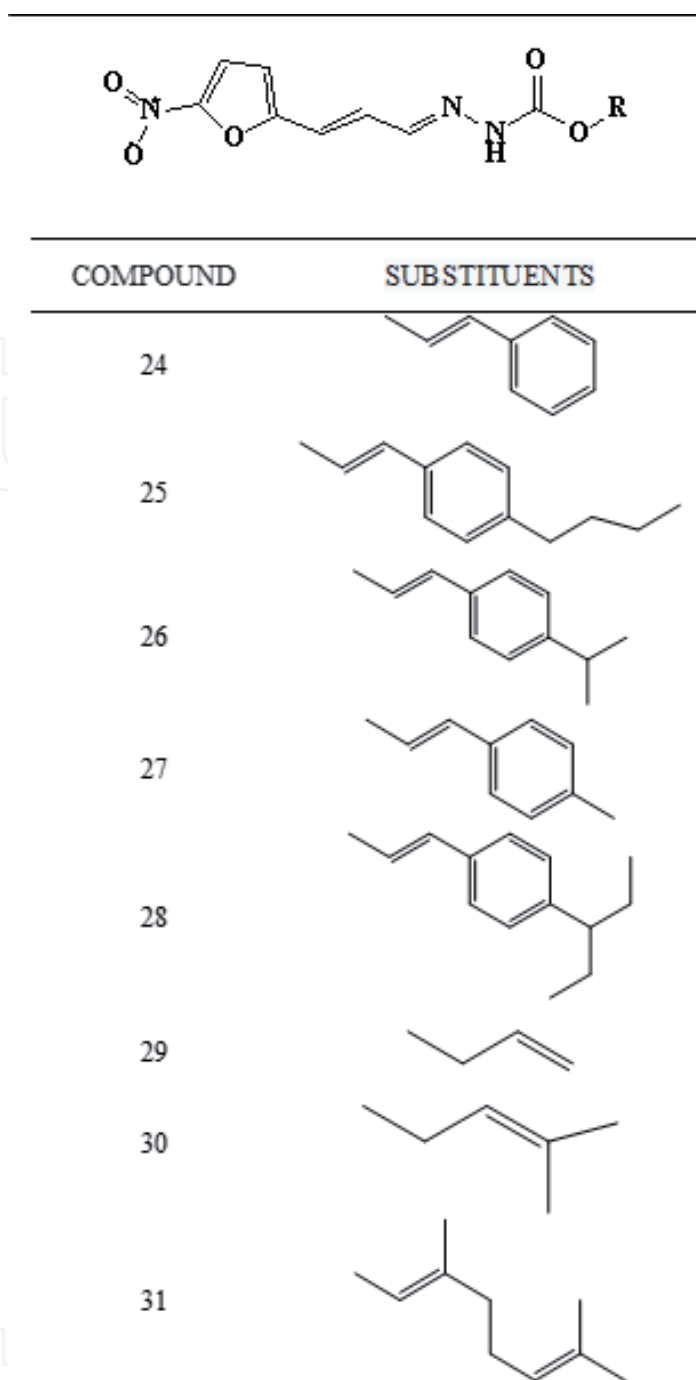
Figure 3.
2D molecular structures for nitrofurans for the prediction set.

3.1.3.1 Theoretical approach and basis set used in the molecular calculations

In the calculations with the nitrofuran compounds (**Figure 1**), quantum–chemical approaches were used [81–87]. We use Becke's three-parameter hybrid methods [81], the Lee-Yang-Parr (LYP) correlation functional [82], B3LYP and Becke's 1988 functional (BLYP) [83], Hartree-Fock (HF) method [84], Austin model 1 (AM1) method [85], Parametric Method Number 3 (PM3) [86], and standard basis sets [87] available in the Gaussian program package. In 5-nitrofuran-2-aldoxim, geometry optimization was carried out by B3LYP/6-21G, B3LYP/6-21G*, B3LYP/6-31G, B3LYP/6-31-G*, BLYP/6-21G, BLYP/6-21G*, BLYP/6-31G, BLYP/6-31G*, HF/6-21G, HF/6-21G*, HF/6-31G, and HF/6-31G* approaches [81–84] and basis sets [87] and AM1 and PM3 approaches [85, 86] . The calculations were performed to find the approach and basis set that would present the best compromise between

computational time and accuracy of the information relative to the experimental data. The experimental structure of 5-nitrofuran-2-aldoxim molecule was retrieved from the Cambridge Structural Database CSD [72]. PCA and HCA techniques were used to compare the computed structures with different methods/basis sets of quantum chemistry with the experimental structure of 5-nitrofuran-2-aldoxim molecule to identify the appropriate method and the basis set for further calculations. The analyzes were carried out on an auto-scaled data matrix with dimension 26 × 5, where each row was associate 26 computed and 1 experimental geometry, and each column represented one of 5 geometrical parameters of the 5-nitrofuran-2-aldoxim molecule (bond lengths and bond angles). In order to compute all structures and perform calculations to obtain the molecular properties, the HF/6-31G method has selected (see Results and discussion section); the initial geometries of the nitrofurans (**Figures** 2 and **3**) were built with the optimized geometry of the 5-nitrofuran-2-al-doxim molecule selected by PCA and HCA techniques. A conformational analysis for each compound was carried out with the MM⁺ algorithm [79], and the lowest energy conformation was submitted to a conformational search with the Gaussian program.

3.2 Results and discussion

3.2.1 Quantum–chemical approach and basis set selection for the description of the geometries of nitrofurans

The advantage in using the PCA and HCA techniques in this step was that all structural information are considered simultaneously and it takes into account the correlations among them. **Table 1** shows the theoretical and experimental structural information (bond lengths and bond angles) of the geometry of the 5-nitrofuran-2-aldoxim molecule. It was used with the aim to select using PCA and HCA techniques, which quantum–chemical approach and basis set give results closest to the experimental data [72].

The first two principal components explain 86.02% of the original information as follows: PC1 = 58.01% and PC2 = 28.02%. The PC1 versus PC2 scores plot is shown in **Figure 4**, from which it can be seen that the methods are discriminated into two classes according to PC2. The semiempirical approaches (AM1 and PM3) are at the top of the graph, while the other theoretical (HF, BLYP, and B3LYP) approaches and experimental data are at the bottom. Moreover, it can be seen that the HF/6-31G approach/basis set is the closest to the experimental data, indicating that they should be used in the development of this work.

Also, to investigate the most appropriate approach and basis set for further calculations, we used HCA. **Figure 5** shows the dendrogram obtained with complete linkage method; from this figure, we conclude that the theoretical approaches are distributed in a similar way as in PCA, i.e., HCA confirmed the PCA results. Moreover, we can observe that the HF/6-31G approach/basis set is closer to the experimental data therefore being the most suitable to carry out this work.

3.2.2 MEP maps for compounds of the training set

Figure 6 shows the MEP maps for the nitrofurans in the training set. The analysis of these maps reveals that the most active compounds, in general, have the following characteristics:

(i) Compounds with an unsaturation and presenting O atom neighboring the carbonyl in the carbonic chain present greater electron density in the proximities of the furan ring with the decrease of the chain size. In these compounds (**4, 5,** and **6**), MEP

Approaches/basis set															
Geometric parameters	B3LYP/6-21G	B3LYP/6-21G*	B3LYP/6-31G	B3LYP/6-31G*	BLYP/6-21G	BLYP/6-21G*	BLYP/6-31G	BLYP/6-31G*	HF/6-21G	HF/6-21G*	HF/6-31G	HF/6-31G*	AM1	PM3	Exp [72]
Bond length (Å)															
C_2C_3	1.42	1.42	1.42	1.42	1.43	1.47	1.43	1.42	1.43	1.43	1.43	1.43	1.43	1.43	1.41
C_4C_5	1.36	1.36	1.37	1.37	1.38	1.38	1.39	1.38	1.34	1.39	1.34	1.34	1.40	1.39	1.34
C_1C_2	1.38	1.38	1.38	1.38	1.39	1.39	1.40	1.39	1.35	1.35	1.35	1.35	1.33	1.38	1.36
C_1O_1	1.40	1.37	1.39	1.36	1.42	1.39	1.42	1.38	1.37	1.39	1.37	1.33	1.34	1.37	1.37
C_4O_1	1.38	1.35	1.38	1.35	1.41	1.37	1.40	1.37	1.36	1.37	1.35	1.33	1.40	1.38	1.35
C_4N_1	1.41	1.43	1.41	1.43	1.43	1.44	1.43	1.49	1.40	1.43	1.40	1.42	1.45	1.48	1.42
N_1O_2	1.41	1.43	1.41	1.43	1.43	1.44	1.43	1.48	1.40	1.43	1.41	1.42	1.46	1.47	1.42
N_1O_3	1.28	1.23	1.26	1.23	1.31	1.25	1.29	1.25	1.24	1.19	1.22	1.19	1.19	1.21	1.22
C_1C_5	1;29	1.23	1.27	1.23	1.32	1.26	1.30	1.26	1.26	1.20	1.23	1.20	1.20	1.22	1.22
C_5N_2	1.43	1.44	1.43	1.44	1.44	1.45	1.44	1.45	1.45	1.46	1.45	1.46	1.45	1.45	1.45
N_2O_4	1.29	1.28	1.29	1.28	1.32	1.31	1.31	1.30	1.26	1.25	1.26	1.25	1.31	1.29	1.27
O_4H_1	1.47	1.40	1.44	1.39	1.50	1.42	1.47	1.41	1.44	1.37	1.40	1.36	1.31	1.39	1.38
Bond angle (°)															
$C_1O_1C_4$	105.3	105.6	106.0	106.1	104.7	105.3	105.5	105.8	105.4	106.3	105.8	106.9	105.3	106.3	104.5
$O_1C_1C_2$	109.5	110.2	109.2	110.1	109.5	110.0	109.2	109.9	109.1	109.6	110.7	109.4	105.2	106.0	104.8
$O_1C_1C_5$	119.3	118.7	119.8	119.5	119.2	118.8	119.7	119.6	119.5	119.4	118.5	119.8	110.6	110.7	110.2
$C_5C_1C_2$	131.2	131.0	130.9	130.4	131.3	131.1	131.1	130.5	131.3	130.9	130.6	130.9	119.5	120.4	114.1
$C_5N_2O_2$	121.2	120.6	121.8	121.3	121.2	120.7	121.9	121.4	122.0	120.9	120.1	121.7	129.7	128.8	135.6
$C_1O_1C_4$	107.9	110.0	109.5	110.6	107.1	109.2	108.5	109.8	108.8	109.6	111.2	111.4	122.8	120.4	127.8
$N_2O_4H_1$	100.7	100.8	103.6	102.7	99.3	99.8	102.0	101.6	102.4	103.7	102.1	106.9	115.2	116.7	112.2

$C_1C_2C_3$	107.5	106.6	107.5	106.6	107.7	106.8	107.7	106.9	107.8	106.9	106.0	106.9	104.2	101.6	106
$O_1C_4C_3$	111.5	112.3	111.2	112.0	111.5	112.2	111.2	111.9	111.1	111.4	112.8	111.1	105.9	106.0	105.1
$C_3C_4N_1$	130.2	129.9	130.6	130.2	130.1	130.0	130.9	130.5	130.9	130.3	129.5	130.4	111.1	110.6	113.2
$O_1C_4N_1$	118.4	117.8	118.1	117.6	118.3	117.7	117.9	117.5	117.8	118.2	117.5	118.4	131.4	131.4	129.8
$C_4N_1O_2$	115.0	114.9	115.8	115.6	114.6	114.6	115.9	115.5	115.7	115.4	115.0	116.0	117.4	117.8	116.9
$C_4N_1O_3$	117.7	117.2	118.9	118.1	117.6	117.2	115.6	118.2	118.9	118.1	117.2	119.2	117.3	117.5	116.3
$O_2N_1O_3$	127.3	127.9	125.3	126.2	127.7	128.1	118.8	126.2	125.2	126.4	127.6	126.4	119.6	120.1	118.8

*Refers to the base sets cited in the corresponding references.

Table 1.
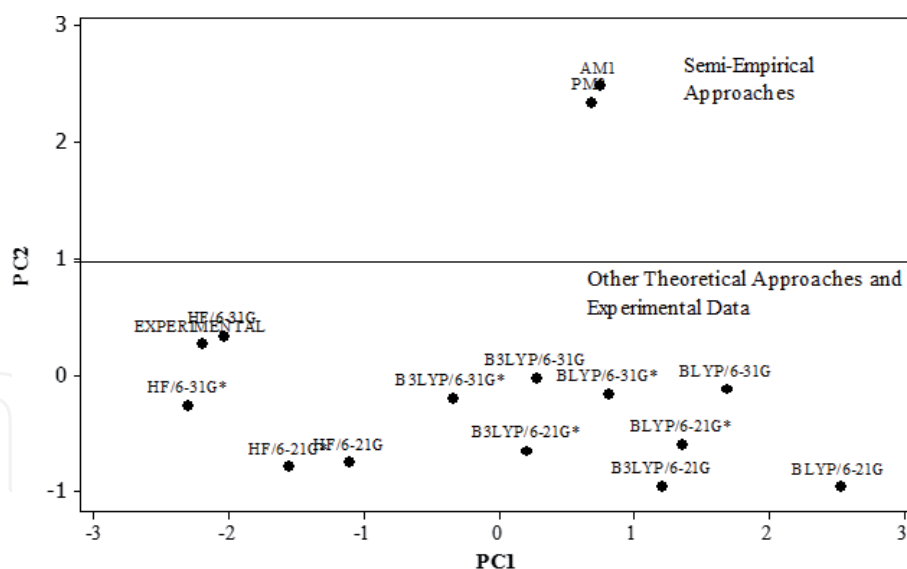*Experimental and theoretical structural parameter of the 5-nirofuran-2-aldoxime.*

Figure 4.
Score plots of the two first PCs, PC1 and PC2, for the separation of the approaches basis sets into classes: semiempirical and semiempirical not.
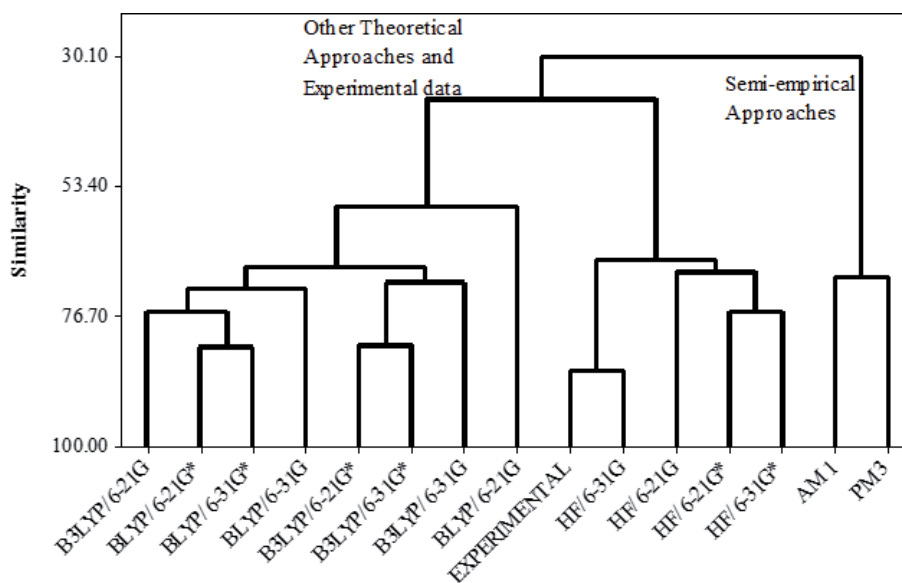

Figure 5.
Dendrogram obtained with HCA technique for the separation of the approach basis set into two classes: semiempirical and semiempirical not.

maps show negative regions ranging from −82.99 to −4.87 kcal/mol. In the most active compound (**6**), as can be seen, the most negative values are in the nitro group, the O atom of the furan ring and the O atoms of the ester group (red and yellow). Also, the MEP maps of these compounds exhibit positive regions between the +4.54 and + 76.96 kcal/mol values (green and blue). Compounds with double unsaturation, containing N atom next to the carbonyl, raise the electronic density with the increase of the carbonic chain. In the most active compound (**7**), the MEP map shows a region of negative values between −77.74 and − 1.31 kcal/mol, with the electron density concentrating mainly on the atoms of the nitro group, on the O atom of the furanic ring and on the N and O atoms of the amide group (red and yellow). According to the MEP map, these compounds present positive MEP between +5.64 and 61.21 kcal/mol (green and blue).

(ii) Compounds with double unsaturation, containing O atom neighboring the carbonyl, raising the carbon chain, increase the electron density in the atoms of the
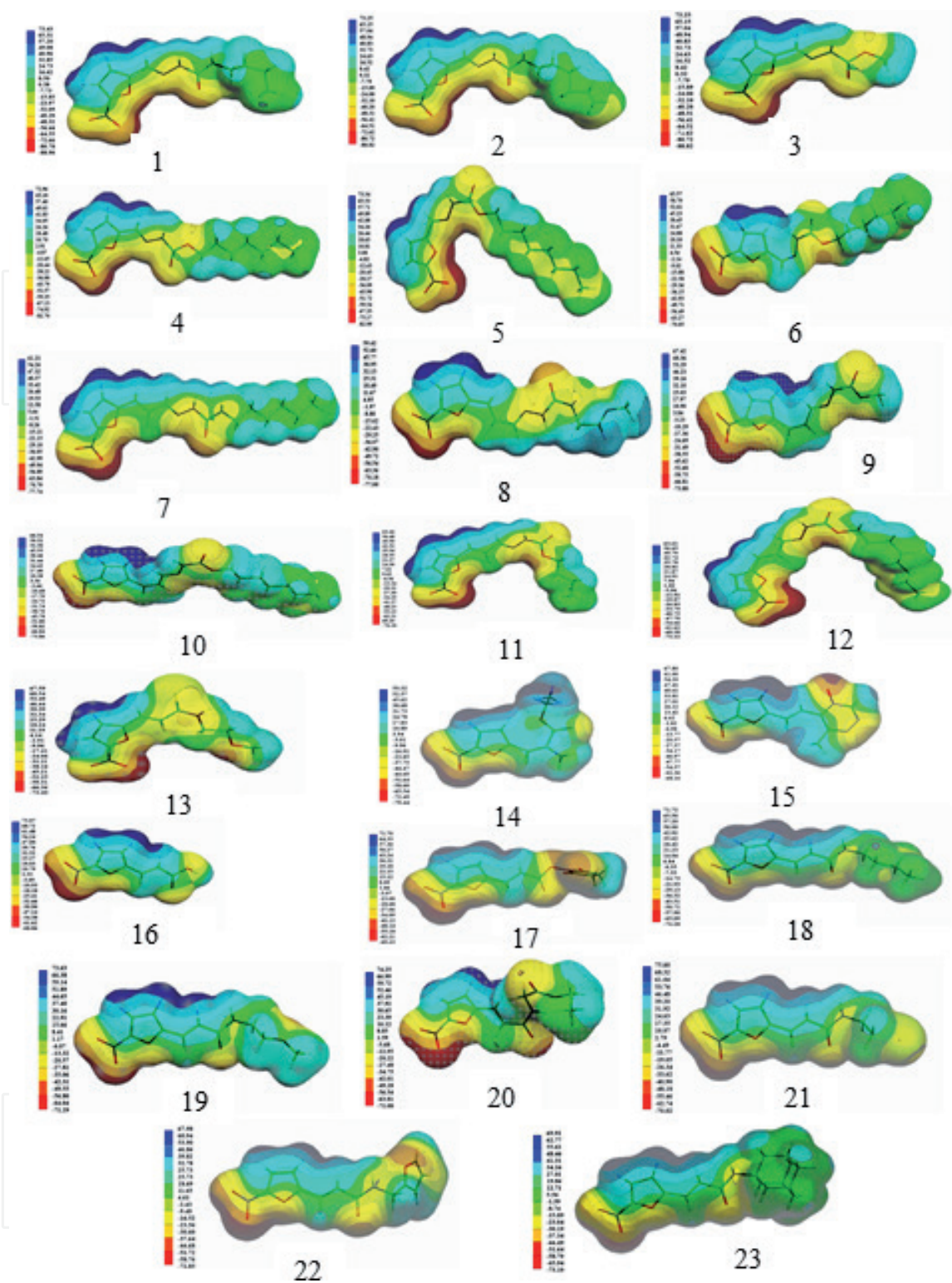
Figure 6.
MEP maps (kcal/mol) for nitrofurans (training set).

nitro group, extending through the O atom of the furan ring to the O atoms of the ester group following the unsaturated chain. In these compounds (**10–12**), the MEP maps exhibit more negative values between -76.18 and -6.36 kcal/mol (red and yellow). They exhibit positive MEP in the range of $+0.63$ to 67.42 kcal/mol (green and blue)

(iii) Compound with an unsaturation, N atom neighboring the carbonyl in the carbonic chain and bulky substituents, has higher electron density in the vicinity of the furan ring and in the N and O atoms of the amide group. In this compound

(**23**), the MEP map shows a negative region (red and yellow) between −73.10 and − 1.59 kcal/mol on the mentioned atoms and positive region between +5.56 and 69.91 kcal/mol (green and blue). The electron density around the nitro group, the O atom of the furan ring, and other atoms may induce the nitrofurans to show anti-trypanosomal activity, suggesting the complexation in those regions with the active site of the receptor in a biological recognition process.

From the above discussion, as a rule, to plan more active nitrofurans, we can assume we resort to one of the basic structures of the most active compounds and introduce groups of atoms or substituents electron donors enhancing the key structural features that are necessary for their activities.

3.2.3 Chemometric modeling

To perform the chemometric modeling, all variables were auto-scaled as pre-processing so that they could be standardized and so they could have the same importance regarding the scale. Furthermore, given a large quantity of multivariate data available, it was necessary to reduce the number of variables. Thus if any two descriptors had a high Pearson correlation coefficient ($r > 0.8$), one of the two was excluded from the matrix at random, since theoretically they describe the same property [88]; they also have a high correlation with antitrypanosomal activity, and only one of them is enough to be used as independent variable in a predictive model.

3.2.3.1 PCA model

Four molecular descriptors were selected for PCA model. The molecular descriptors (QN1, gap energy, Mor05u, and MlogP), in vitro *T. cruzi* growth inhibition (experimental data), and activity and correlation matrix including all data for 23 nitrofurans can be seen in **Table 2**. The correlation between descriptors is less than 0.786. The first three principal components (PCs) describing 96.48 of the original information for the 23 are as follows: 45.70, 30.91, and 19.87%. PC1-PC2 scores for the samples are shown in **Figure 7**. From this figure, we can see that the nitrofurans are distributed into two distinct regions in PC1. The more active compounds are on the left side (**4–7, 10–12, 18,** and **23**) and the less active on the right side (**1–3, 8, 9, 13–17,** and **19–22**). According to **Figure 8**, the MlogP descriptor is responsible for displaying more active compounds on the left side, while the gap energy, QN1, and Mor05u descriptors displayed fewer active compounds for the right side from this figure.

Table 3 shows the loading vectors for PC1, PC2, and PC3. According to this table, PC1 can be expressed through the following equation:

$$PC1 = 0.20 \, (QN1) + 0.06 \, (Gap \ energy) + 0.71 \, (Mor05u) - 68 \, (MlogP). \quad (2)$$

From this equation, more active nitrofurans, in general, can be obtained when we have lower values for the QN1 combined with lower values for Gap energy and Mor05u and higher values for MlogP.

3.2.3.2 HCA model

The results of the HCA model are displayed in the dendrogram in **Figure 9** and are similar to those of PCA model. The nitrofurans are fairly well grouped according to their activity. From this figure, the two clusters (+ and −) mirror the same two classes displayed by PCA model (**Figure 7**).

Nitrofurans	QN1	Gap energy (kcal/mol)	Mor05u	MlogP	% in vitro *T. cruzi* growth inhibition[a,b]	Activity[c]
1−	0.201	220.9	−3.966	1.135	30	LA
2−	0.201	220.9	−2.938	1.708	20	LA
3−	0.165	220.9	−2.723	0.181	32	LA
4+	0.165	226.5	−6.869	1.980	92.7	MA
5+	0.165	225.3	−7.439	3.155	83.7	MA
6+	0.169	229.7	−0.016	1.708	96.2	MA
7+	0.164	208.3	−7.439	1.889	81.9	MA
8−	0.164	205.2	−4.854	0.334	26.7	LA
9−	0.166	215.9	−3.292	0.478	58	LA
10+	0.166	215.9	−7.470	2.146	90	MA
11+	0.164	208.3	−5.674	1.354	87.4	MA
12+	0.164	208.3	−8.435	3.307	92.3	MA
13−	0.167	195.2	−4.338	0.751	12	LA
14−	0.161	203.3	−2.872	0.501	3	LA
15−	0.167	208.3	−4.217	0.411	30	LA
16−	0.167	225.3	−2.373	0.609	20	LA
17−	0.167	225.9	−4.054	1.063	6	LA
18+	0.167	225.3	−6.339	2.001	75	MA
19−	0.166	225.3	−4.145	0.398	31	LA
20−	0.167	226.5	−4.786	0.667	35	LA
21−	0.167	225.3	−3.398	1.157	23	LA
22−	0.166	218.4	−3.876	0.802	14	LA
23+	0.166	224.6	−6.314	3.014	90.5	MA
Gap energy	−0.171					
Mor05u	0.27	−0.006				
MlogP	0.026	−0.184	−0.785			

[a]Inhibitor concentration of 5 μM. [b]Growth inhibition ≥ 75, more active (MA)[c], and growth inhibition < 75, less active (LA)[c].

Table 2.
Values for the four most important descriptors which classify the studied nitrofuran compounds, in vitro T. cruzi growth inhibition (experimental data), activity, and correlation matrix.

3.2.3.3 KNN model

Table 4 shows the results for the KNN models obtained with the KNN technique and constructed with one (1NN) to four (4NN) nearest neighbors. To all models the percentage of correct information was 100%. We used the model 4NN because the greater the number of the nearest neighbors, the better the reliability of the KNN technique, and the same was used for validation of the training set from **Figure 2**.

3.2.3.4 SDA model

In the construction of the SDA model, the discrimination functions for groups more active and less active, respectively, are given below:
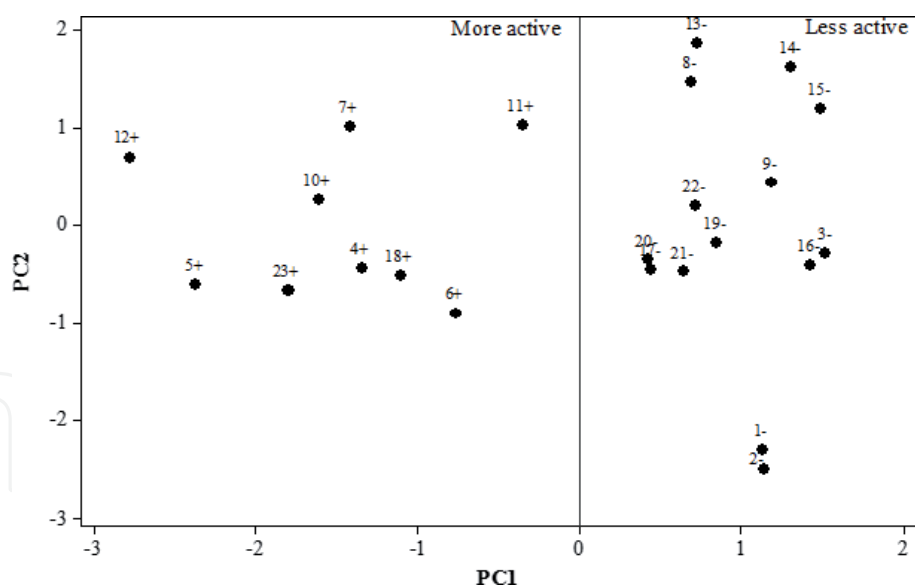
Figure 7.
Score plots of the two first PCs, PC1 and PC2, responsible for the separation of the 23 nitrofurans (training set) into two classes: (+) more active and (−) less active against T. cruzi.
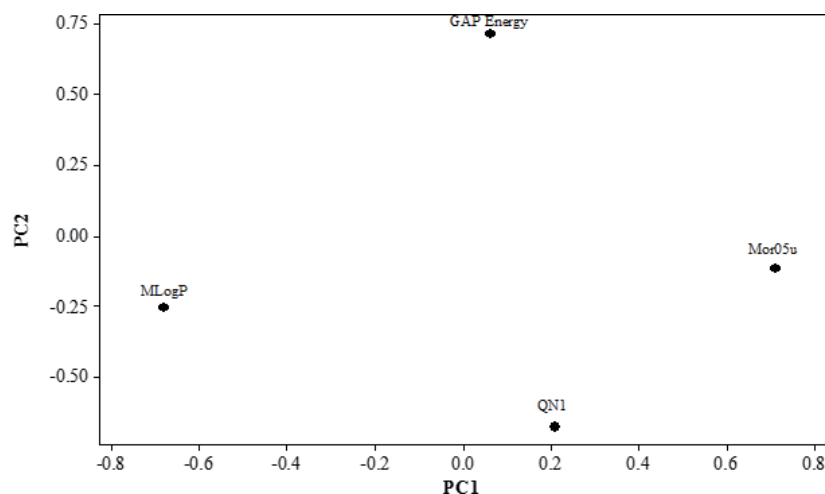

Figure 8.
Loading vector plots of the first PCs, PC1 and PC2, for four variables responsible for the separation of the 23 nitrofurans (training set) into two classes: (+) more active and (−) less active against T. cruzi.

Group MA (more active):

$$0.51(QN1) + 0.43\text{Gap energy} + 3.05\text{Mor05u} - 1.5\text{MlogP} - 0.62 \qquad (3)$$

Group LA (less active):

$$-0.80QN1 - 0.67\text{Gap energy} - 4.75\text{Mor05u} + 2.34\text{MlogP} - 3.92 \qquad (4)$$

Also, through the discrimination functions, Eqs. (3) and (4), and of the value of each descriptor for the nitrofurans, we obtain the classification matrix by using all compounds from the training set (**Table 5**). The classification error was 0.00% resulting in a satisfactory separation of more active and less active compounds. From SDA model, the allocation rule was derived when the activity against *T. cruzi* of new nitrofurans is investigated: (a) initially calculate, for the new compound, the value of the most important descriptors obtained in the construction of the SDA model, (b) put these auto-scaled values in the two discrimination functions

Variable	PC1	PC2	PC3
QN1	0.20	0.66	0.69
Gap energy	0.06	−0.70	0.70
Mor05u	0.71	0.11	−0.10
MlogP	−0.68	0.26	0.17

Table 3.
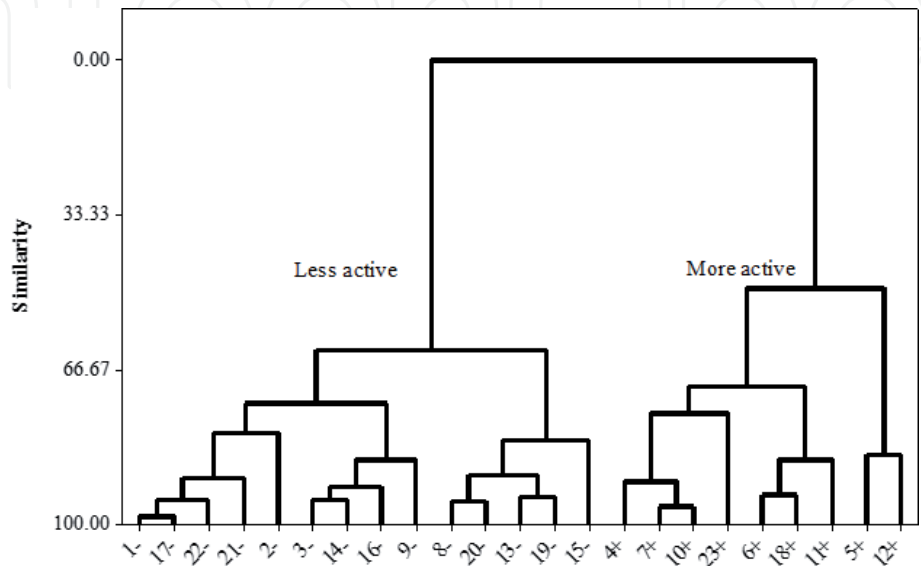*Variables matrix for the first three principal components.*

Figure 9.
Dendrogram obtained with HCA technique for the separation of the nitrofurans into two classes: (+) more active and (−) less active against T. cruzi.

Category	Number of compounds	Compounds incorrectly classified			
		1NN	2NN	3NN	4NN
Class:more active	9	0	0	0	0
Class: less active	14	0	0	0	0
Total	23	0	0	0	0
% Correct information		100	100	100	100

Table 4.
Classification obtained with the KKN technique.

performed in this work, and (c) check which discrimination function, Eq. (3) or Eq. (4), presents higher value. The new compound is more active if it is related to discrimination function of group more active and vice versa.

In order to check the reliability of the model, the "leave-one-out technique" was employed. One nitrofuran compound is excluded from the data set, and the remaining compounds are used in building the classification functions.

Subsequently, the removed analogue is classified according the generated classification functions. In the further step, the omitted compound is included, and a new nitrofuran is removed, and the procedure goes on until the last compound is removed. In **Table 6** the results obtained with the cross-validation model are summarized.

		True group	
Classification group or class	**Number of compounds**	**More active**	**Less active**
Group (Class): more active	9	9	0
Group (Class): less active	14	0	14
Total	23	9	14
% Correct information	—	100	100

Table 5.
Classification matrix obtained using SDA technique.

		True group	
Classification group or class	**Number of compounds**	**More active**	**Less active**
Group (class): more active	9	9	0
Group (class): less active	14	0	14
Total	23	9	14
% correct information	—	100	100

Table 6.
Classification matrix obtained by using SDA technique with cross-validation technique.

3.2.3.5 SIMCA model

The SIMCA model were built with the same descriptors as PCA, HCA, KNN, and SDA models and used two (2) PCs in the modeling of the two classes: more active nitrofurans (**4–7, 10–12, 18,** and **23**) and less active (**1–3, 8, 9, 13–17,** and **19–22**) nitrofurans. In **Table 7**, the obtained results for the SIMCA model are shown. In this case, the information percentage was also 100%. According to the PCA, HCA, KNN, SDA, and SIMCA models, we can also notice that the QN1, gap energy, Mor05u, and MlogP descriptors are key properties for explaining the anti-*T. cruzi* activity of the nitrofurans training set (**Figure 2**).

As QN1, gap energy, Mor05u, and MlogP properties were selected in the chemometric modeling as the most important characteristics to describe the antitrypanosomal activity, some considerations about them may be relevant to the understanding of the behavior of more active nitrofurans. According to classical chemical theory, chemical interactions can be classified in two categories: electrostatic (polar) or orbital (covalent). Electrical charges in the molecule are indubitably the impelling cause of electrostatic interactions. It has been demonstrated that local electron densities or charges are important in many chemical reactions, physicochemical properties, and ligand–receptor interactions [89, 90]. Thus, charge-based parameters have been widely employed as chemical reactivity indices or as measures of weak intermolecular interactions. Many quantum–chemical descriptors are derived from the partial charge distribution in a molecule or from the electron densities on particular atoms [91]. From **Table 2,** we can observe that, in general, QN1 for more active analogues must present lower values than the less active ones. This is an indication that biological processes can occur through electrostatic interactions between the more active nitrofurans and an eventual biological receptor.

Gap energy is an important stability index. A large gap energy implies high stability for the molecule in the sense of its lower reactivity in chemical reactions.

Category	Number of compounds	Correct classification
Class: more active	9	9
Class: less active	14	14
TOTAL	23	
% correct information		100

Table 7.
Classification obtained by using SIMCA technique.

It is an approximation of the lowest excitation energy of the molecule and can be used for the definition of absolute and activation hardness [89, 90]. In **Table 2**, we can observe that, in general, the more active nitrofurans present lower gap energy than the less active ones. This indicates that the more active nitrofurans have a great probability of interacting with the biological receptor through a charge transfer mechanism.

Mor05u is a 3D-MoRSE descriptor based on the idea of obtaining information from 3D atomic coordinates through the transformed used in electrons diffraction studies [91] and is strictly related to the stereochemistry of the compounds [92]. According to **Table 2**, the more active nitrofurans present lower values of Mor5u. This may be, in general, an indication of the importance of the stereochemical properties of the more active nitrofurans in a possible mechanism of action of its own.

MlogP is an important hydrophobic descriptor in diverse biochemical, pharmacological, and toxicological processes involved in drug absorption [93]. As identified in **Table 2**, the more active reported nitrofurans exhibit the higher MlogP values. This is an indication that in processes involving nitrofurans and a biological receptor, hydrophobic interactions may be important in the mechanism of action of these compounds.

Knowing the performance of the RP models constructed for the 23 studied nitrofurans, we decided to apply them to a series of eight compounds (**Figure 3**) designed to maintain the key structural features that are necessary for their biological activities evidenced by the MEP maps of the compounds of the training set. The basic nucleus of these compounds corresponds to that of the most active nitrofurans with double unsaturation, containing vicinal O atom to carbonyl (see compounds **10**–**12**). The eight molecules proposed for the study of prediction of activity were drawn with the help of one of the collaborators of this work, who belong to the research group in organic chemistry of the Federal University of Pará, Brazil, and the most promising syntheses are in progress. In the future, antitrypanosomal tests with the most promising nitrofurans can be used to validate our RP models.

The results obtained of the application of the PR models (PCA, HCA, KNN, SDA, and SIMCA) and the descriptors for the compounds of the prediction set are summarized in **Tables 8** and **9**, respectively. In **Table 8**, the compounds **25** and **30** were predicted as more active against *T. cruzi* with the five models. Only the KNN model predicted compound **26** as the most active. Meanwhile, only the PCA and HCA models predicted compound **31** as the most active. On the other hand, all models, except the SDA model, predicted compounds **24**, **27**, and **28** as the most active. In turn, the SIMCA model did not classify compounds **29** and **31** into any of the two classes. Thus, we can consider nitrofurans **25** and **30** as potentially more active in a future test against *T. cruzi*. For the values reported for compounds **25** and **30** (**Table 9**), it can be shown that in order to design more active nitrofurans we must combine smaller values for the descriptors QN1, gap energy, and Mor05u with higher value for the descriptor MlogP.

Nitrofuran	PCA model	HCA model	KNN model	SDA model	SIMCA model
24	MA	MA	MA	LA	MA
25	MA	MA	MA	MA	MA
26	LA	LA	MA	LA	LA
27	MA	MA	MA	LA	MA
28	MA	MA	MA	LA	MA
29	MA	MA	MA	MA	0
30	MA	MA	MA	MA	MA
31	MA	MA	LA	LA	0

Table 8.
Results of application of chemometric models for the nitrofurans of the prediction set.

Nitrofuran	QN1	Gap energy (kcal/mol)	Mor05u	MLogP
24	0.165	205.2	−6.352	3.155
25	0.165	203.3	−7.332	2.250
26	0.165	204.6	−5.835	1.146
27	0.169	203.9	−6.164	2.508
28	0.166	203.9	−7.146	1.875
29	0.164	229.7	−8.201	3.854
30	0.164	229.7	−6.421	3.373
31	0.164	223.4	−5.525	2.167

Table 9.
Values for descriptors for the prediction set.

3.2.4 MEP maps for compounds of the prediction set

Figure 10 shows the MEP maps for the most active nitrofurans in the validation set (**25** and **30**). Also, in these compounds, as can be seen, raising the carbon chain increases the electron density in the atoms of the nitro group, extending through the O of the furan ring to the O atoms of the ester group accompanying the unsaturated chain. In these compounds, the MEP maps show more negative values between −74.27 and − 1.76 kcal/mol (red and yellow). They exhibit positive MEP in the range + 4.84 to +57.58 kcal/mol (green and blue).

The negative MEP region of compounds **25** and **30**, similar to the more active compounds in the training set, is susceptible to attack in a biological recognition process.

3.3 Concluding remarks

MEP and chemometric techniques in the last decades have become efficient tools in the study of the structure–activity relationships of bioactive molecules. The use of such tools has occurred through the inherent principles of each or combining their potentials to more efficiently unravel information about the structure–activity relationships of pharmacologically interesting compounds. This chapter is circumscribed in this second possibility. MEP maps were constructed for 23 nitrofurans with activity against *T. cruzi* reported in the literature. The key structural features
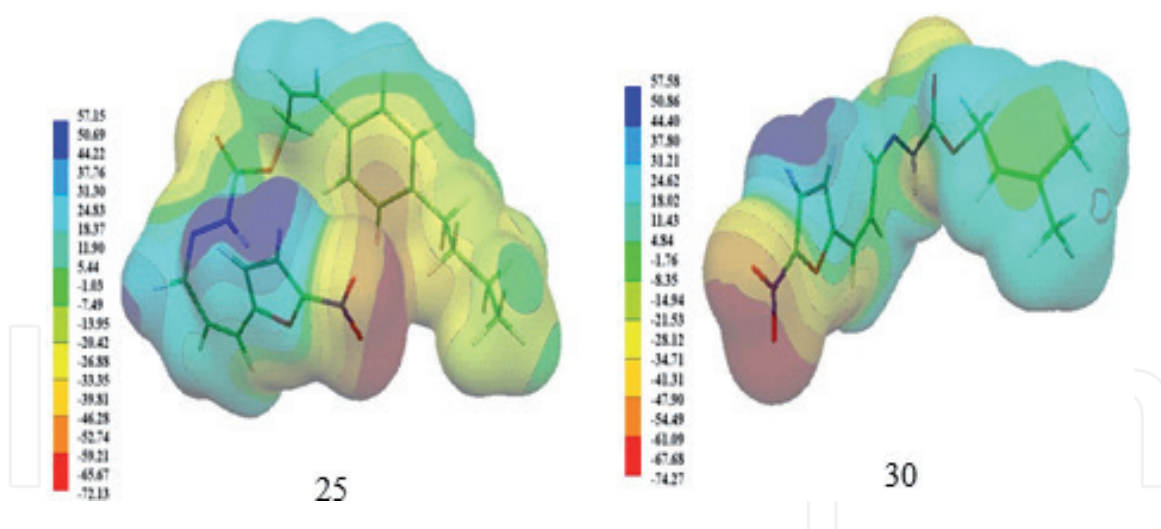
Figure 10.
MEP maps (kcal/mol) for most promising nitrofurans in the prediction set against T. cruzi.

required for antitrypanosomal activity, along with chemical intuition, allowed the introduction of substituents in one of the most active nitrofurans in the training set to obtain eight new derivatives.

PR models (PCA, HCA, KNN, SDA, and SIMCA) were constructed and demonstrated that 23 nitrofurans can be classified into two classes or groups: more active and less active according to their degrees of activity against *T. cruzi*. The properties QN1, gap energy, Mor05u, and MlogP are responsible for the classification into more active and less active studied nitrofurans. It is interesting to notice that these properties represent three distinct classes of interactions between the nitrofurans and the biological receptor: electronic (QN1 and gap energy), steric (Mor05u), and hydrophobic (MlogP). Here it is important to mention that Paulino et al., studying the influence of molecular parameters on the activity of 5-nitrofurans against *T. cruzi*, reported the importance of electronic properties and molecular hydrophobicity as well as the variation of the nitrofurans electronic structure to explain the greater activity of these compounds as inhibitors of the growth of this protozoan [94].

The results of the application of PR models on the validation set evidenced two nitrofurans (**25** and **30**) as more promising for synthesis and biological assays, which in the future can be used to validate our PR models.

Acknowledgements

Author details

Marcos Antônio B. dos Santos[1], Luã Felipe S. de Oliveira[2],
Antônio Florêncio de Figueiredo[3], Fábio dos Santos Gil[2],
Márcio de Souza Farias[2], Heriberto Rodrigues Bitencourt[4], José Ribamar B. Lobato[2],
Raimundo Dirceu de P. Farreira[2], Sady Salomão da S. Alves[3],
Edilson Luiz C. de Aquino[2] and José Ciríaco-Pinheiro[2]*

1 University of the State of Pará, Pará, Brazil

2 Computational and Theoretical Chemistry Laboratory, Federal University of Pará, Pará, Brazil

3 Federal Institute of Education, Science and Technology, Pará, Brazil

4 Group of Organic Chemistry, Federal University of Pará, Pará, Brazil

*Address all correspondence to: ciriaco@ufpa.br

IntechOpen

References

[1] Bonnacorsi RR, Scrocco E, Tomasi J. Molecular SCF calculations for the ground state of some three-membered ring molecules: (CH2)3, (CH2)2NH, (CH2)2NH2+, (CH2)2O, (CH2)2S, (CH)2CH2, and N2CH2. The Journal of Chemical Physics. 1970;**52**:5270-5284. DOI: 10.1063/1.1672775

[2] Scrocco E, Tomasi J. The electrostatic molecular potential as a tool for the interpretation of molecular properties, in: New concepts II. Topics in Current Chemistry. 1973;**42**:95-170. DOI: 10.1007/3-540-06399-4

[3] Politzer P, Truhlar G, editors. Chemical Applications of Atomic and Molecular Electrostatic Potentials. New York: Plenum Press; 1981. ISSN: 978-4757-9634-6

[4] Rangel NL, Seminario JM. Molecular electrostatic potential devices on graphite and silicon surfaces. The Journal of Physical Chemistry A. 2006;**110**:12298-12302. DOI: 10.1021/jp064766i

[5] Müller JJ, Lapko A, Ruckpaul K, Heinemann U. Modeling of electrostatic recognition processes in the mammalian mitochondrial steroid hydroxylase system. Biophysical Chemistry. 2003;**100**:281-292. DOI: 10.1016/S0301-4622(02)00286-7

[6] Kotsikorou E, Sharir H, Shore DM, Hurst DP, Lynch DL, Madrigal KE, et al. Identification of the GPR55 antagonist binding site using a novel set of high-potency GPR55 selective ligands. Biochemistry. 2013;**52**:9456-9469. DOI: 10.1021/bi4008885

[7] Ford KA. Role of electrostatic potential in the in silico prediction of molecular bioactivation and mutagenesis. Molecular Pharmaceutics. 2013;**10**:1171-1182. DOI: 10.1021/mp3004385

[8] Politzer P, Murray JS. Quantitative analyses of molecular surface electrostatic potentials in relation to hydrogen bonding and co-crystallization. Crystal Growth & Design. 2015;**15**:3767-3774. DOI: 10.1021/acs.cgd.5b00419

[9] Lande DN, Gejji SP. Cooperative hydrogen bonding, molecular electrostatic potentials, and spectral characteristics of partial thia-substituted calix [4] arene macrocycles. The Journal of Physical Chemistry A. 2016;**120**:7385-7397. DOI: 10.1021/acs.jpca.6b07568

[10] Anjali BA, Sayyed FB, Suresh CH. Correlation and prediction of redox potentials of hydrogen evolution mononuclear cobalt catalysts via molecular electrostatic potential: A DFT study. The Journal of Physical Chemistry A. 2016;**120**:1112-1119. DOI: 10.1021/acs.jpca.5b11543

[11] Mehmood A, Jones SI, Tao P, Janesko BJ. An orbital-overlap complement to ligand and binding site electrostatic potential maps. Journal of Chemical Information and Modeling. 2018;**58**:1836-1846. DOI: 10.1021/acs.jcim.8b00370

[12] Liu L, Miao L, Li L, Li F, Lu Y, Shang Z, et al. Molecular electrostatic potential: A new tool to predict the lithiation process of organic battery materials. The Journal of Physical Chemistry Letter. 2018;**9**:3573-3579. DOI: 10.1021/acs.jpclett.8b01123

[13] Scilabra P, Murray JS, Terraneo G, Resnati G. Chalcogen bonds in crystals of bis(o-anilinium)diselenide salts. Crystal Growth & Design. 2019;**19**:1149-1154. DOI: 10.1021/acs.cgd.8b01634

[14] Pramanik S, Dey T, Mukherjee AK. Five benzoic acid derivatives: Crystallographic study using X-ray powder diffraction, electronic structure and molecular electrostatic potential calculation. Journal of Molecular Structure. 2019;**1175**:185-194. DOI: 10.1016/j.molstruc.2018.07.090

[15] Salluma LO, Vaza WF, Borgesa NM, Campos CEM, Bartoluzzib AJ, Francoc CHJ, et al. Synthesis, conformational analysis and molecular docking studies on three novel dihydropyrimidine derivatives. Journal of Molecular Structure. 2019;**1192**:274-287. DOI: 10.1016/j.molstruc.2019.04.100

[16] Rzesikowska K, Krawczuk A, Kalinowska-Tluscik J. Electrostatic potential and non-covalent interactions analysis for the design of selective 5-HT7ligands. Journal of Molecular Graphics and Modelling. 2019;**91**:130-139. DOI: 10.1016/j.jmgm.2019.06.007

[17] Aray Y. Nature of the active sites of molybdenum-based catalysts and their interaction with sulfur- and nitrogen-containing molecules using the quantum theory of atoms in molecules and the molecular electrostatic potential. The Journal of Physical Chemistry C. 2019;**123**:14421–14431. In press. DOI: 10.1021/acs.jpcc.9b01951

[18] Cruz JC, Hernández-Esparza R, Vázquez-Mayagoitia A, Vargas R, Garza J. Implementation of the molecular electrostatic potential over GPUs. Journal of Chemical Information and Modeling. 2019;**59**:3120–3127. in press. DOI: 10.1021/acs.jcim.8b00951

[19] Varmuza K. Pattern Recognition in Chemistry. 1980. Springer-Verlag, Berlin. DOI: 10.1002/bbpc.19810850930

[20] Johnson RA, Wichem DW. Applied Multivariate Statistical Analysis. New Jersey: Prentice-Hall; 1992. ISBN: 0-130-41146-9

[21] Brown SD, Sum ST, Despagne F, Lavine BK. Chemometrics. Analytical Chemistry. 1996;**68**:21R-61R. DOI: S0003-2700(96)00005-4

[22] Brown SD. The chemometrics revolution re-examined. Journal of Chemometrics. 2017;**31**:e2856. DOI: 10.1002/cem.2856

[23] Bernardinelli G, Jefford CW, Maric D, Thomson C, Weber J. Computational studies of the structures and properties of potential antimalarial compounds based on the 1,2,4-Trioxane ring structure. I. Artemisinin-like molecules. International Journal of Quantum Chemistry: Quantum Biology Symposium. 1994;**21**:113-131. DOI: 10.1002/qua.560520703

[24] Jefford CW, Grigorov M, Weber J, Lüthi HP, Troncher JMJ. Journal of Chemical Information and Computer Sciences. 2000;**40**:354-357 ISSN: 0095-2338

[25] Kubinyi H. QSAR: Hansch analysis and related approaches. In: Mannhold R, Krogsgaard-Larsen P, Timmerman H, editors. Methods and Principles in Medicinal Chemistry, Vol. 1. Weinheim: VHC; 1993. ISBN: 987-3527300358

[26] van de Waterbeemd H. Chemometric Methods in Molecular Design. New York: VHC; 2008. ISBN: 978-3-527-61544-5

[27] Gangwal RP, Damre MV, Sangamwar AT. Overview and recent advances in Qsar studies. Mercader AG, Duchwicz PR, Sivakumar PM, editors. CHEMOMETRICS: Applications and Research. QSAR in Medicinal Chemistry. Canada: Apple Academic Press; 2016. p. 1-32. ISBN: 978-1771-8811-35

[28] Kubinyi H, Folkers G, Martin YC, editors. 3DQSAR in Drug Design, Vols. 2 and 3. Dordrecht, The Netherlands: Kluwer; 1998. DOI: 978-0-7923-4791-0

[29] Cramer RD, Petterson DE, Brunce JD. Comparative molecular field analysis (CoMFA) 1. Effect of shape binding of steroids to carrier proteins. The Journal American of Chemical Society. 1988;**110**:5959-5967. DOI: 10.1021/ja00226a005

[30] Chhatbar DM, Chaube UJ, Vyas VK, Bhatt HF. CoMFA, CoMSIA, Topomer CoMFA, HQSAR, molecular docking and molecular dynamics simulations study of triazine morpholino derivatives as mTOR inhibitors for the treatment of breast cancer. Computational Biology and Chemistry. 2019;**80**:351-363. DOI: 10.1016/j.compbiolchem.2019.04.017

[31] Pourbasheer E, Aalizadeh R, Tabar SS, Ganjali MR, Norouzi P, Shadmanesh J. 2D and 3D quantitative structure–activity relationship study of hepatitis C virus NS5B polymerase inhibitors by comparative molecular field analysis and comparative molecular similarity indices analysis methods. Journal of Chemical Information and Modeling. 2014;**54**:2902-2914. DOI: 10.1021/ci500216c

[32] Cramer RD. Template CoMFA applied to 116 biological targets. Journal of Chemical Information and Modeling. 2014;**54**:2147-2156. DOI: 10.1021/ci500230a

[33] Cramer RD, Wendt B. Template CoMFA: The 3D-QSAR grail? Journal of Chemical Information and Modeling. 2014;**54**:660-671. DOI: 10.1021/ci400696v

[34] Dong H, Liu J, Liu X, Yu Y, Gao S. Molecular docking and QSAR analyses of aromatic heterocycle thiosemicarbazone analogues for finding novel tyrosinase inhibitors. Biooganic Chemistry. 2017;**75**:106-117. DOI: 10.1016/j.bioorg.2017.07.002

[35] Dong MH, Chen HF, Ren YJ, Shao FM. Molecular modeling studies, synthesis and biological evaluation of dabigatran analogues as thrombin inhibitors. Bioorganic & Medicinal Chemistry. 2016;**24**:73-84. DOI: 10.1016/j.bmc.2015.11.025

[36] Ding L, Wang ZZ, Sun XD, Yang J, Ma CY, Li W, et al. 3D-QSAR, molecular docking and molecular dynamics simulations study of 6-aryl-5-Cyano-Pyrimidine derivatives to explore the structure requirements of LSD1 inhibitors. Biooganic & Medicinal Chemistry Letters. 2017;**27**:3521-3528. DOI: 10.1016/j.bmcl.2017.05.065

[37] Pinheiro JC, Kiralj R, Ferreira MMC, Romero OAS. Artemisinin derivatives with antimalarial activity against plasmodium falciparum designed with the aid of quantum chemical and partial least squares methods. QSAR & Combinatorial Science. 2003;**22**:830-842. DOI: 10.1002/qsar.200330829

[38] Cardoso FJB, Figueiredo AF, Lobato MS, Miranda RM, Almeida RCO, Pinheiro JCP. A study on antimalarial artemisinin derivatives using MEP maps and multivariate QSAR. Journal of Molecular Moldeling. 2008;**14**:39-49. DOI: 10.1007/s00894-007-0249-9

[39] Ferreira JEV, Figueiredo AF, Barbosa JP, Cristino MGG, Macedo WJC, Silva OPP, et al. A study of new antimalarial artemisinins through molecular modeling and multivariate. Journal of the Serbian Chemical Society. 2010;**75**:1533-1548. DOI: 10.2298/JSC100126124F

[40] Figueiredo AF, Ferreira JEV, Barbosa JP, Macedo WJC, Cristino MGG, Lobato MS, et al. A computational study on antimalarial dispiro-1,2,4-trioxolanes. Journal of Computational and Theoretical Nanoscience. 2011;**8**:1-10. DOI: 10.1166/jctn.2011.1892

[41] Carvalho JRC, Ferreira JEV, Barbosa JP, Lobato MS, Meneses CCF,

Soeiro MM, et al. Computational modeling of artemisinins with antileishmanial activity. Journal of Computational and Theoretical Nanoscience. 2011;**8**:1-11. DOI: 10.166/jctn.2011.1943

[42] Barbosa JP, Ferreira JEV, Figueiredo AF, Almeida RCO, Silva OPP, Carvalho JRC, et al. Molecular modeling and chemometric study of anticancer derivatives of artemisinin. Journal of the Serbian Chemical Society. 2011;**76**:1263-1282. DOI: 10.2298/JSC111227111B

[43] Cristino MGG, Meneses CCF, Soeiro MM, Ferreira JEV, Figueiredo AF, Barbosa JP, et al. Computational modeling of antimalarial 10-substituted deoxoartemisinis. Journal of Theoretical and Computational Chemistry. 2012;**11**:241-263. DOI: 10.1142/S0219633612500162

[44] Politzer P, Laurence PR, Jayasuriya K. Molecular electrostatic potentials: An effective tool for the elucidation of biochemical phenomena. Environmental Health Perspectives. 1985;**61**:191-202. DOI: 10.1289/ehp.8561191

[45] Politzer P, Murray JS. The fundamental nature and role of the electrostatic potential in atoms and molecules. Theoretical Chemistry Accounts. 2002;**108**:134-149. DOI: 10.1007/s00214-002-0363-9

[46] Scrocco E, Tomasi J. Electronic molecular structure, reactivity and intermolecular forces: An euristic interpretation by means of electrostatic molecular potentials. Advances in Quantum Chemistry. 1979;**1978**(11):115-193. DOI: 10.1016/S0065-3276(08)60236-1

[47] Bishop CM. Pattern Recognition and Machine Learning. 2006. Springer, Singapore. ISBN: 978-0-387-31073-2

[48] Sebestyen GS. Decision-Making Processes in Pattern Recognition. New York: Academic Press; 1962

[49] Fu KS. Sequential Methods in Pattern Recognition and Machine Learning. New York: Academic Press; 1968. ASIN: B001QC5IZS

[50] Watanabe S. Methodologies of Pattern Recognition. New York: Academic Press; 1969. DOI: 978-1-4832-3093-1

[51] Mendel JM, Fu KS. Adaptive, Learning, and Pattern Recognition Systems; Theory and Applications, Vol. 66. 1st ed. New York: Academic Press; 1970. ISBN: 9780080955759

[52] González AG. Critical aspects of supervised Pattern recognition methods for interpreting compositional data. In: Varmuza K, editor. Chemometric in Applications Practical. Shanghai: In Tech; 2012. DOI: 978-953-51-0438-4

[53] Li Y, Wang S, Tian Q, Ding X. Feature representation for statistical-learning-based object detection: A review. Pattern Recognition. 2015;**48**:3542-3559. DOI: 10.1016/j.patcog.2015.04.018

[54] Jain AK. Data clustering: 50 years beyond K-means. Pattern Recognition Letters. 2010;**31**:651-666. DOI: 10.1016/j.patrec.2009.09.011

[55] Jurs PC, Kowalski BR, Isenhour TL. Investigation of combined patterns from diverse analytical data using computerized learning machines. Analytical Chemistry;**1969**:41, 1949-1953. DOI: 10.1021/ac50159a027

[56] Kowalski BR, Jurs PC, Isenhour TL. Computerized learning machines applied to chemical problem. Interpretation of infrared spectrometry. Analytical Chemistry. 1969;**41**:1945-1949. DOI: 10.1021/ac50159a026

[57] Kowalski BR, Reilly CA. Nuclear magnetic resonance spectral interpretation by Pattern recognition. The Journal Physical Chemistry. 1971;**75**:1402-1411. DOI: 10.1021/j100680a008

[58] Wangen LE, Isenhour TL. Semiquantitative analysis of mixed gamma-ray spectra by computerized learning machines. Analytical Chemistry. 1970;**42**:737-743. DOI: 10.1021/ac60289a005

[59] Sybrandt LB, Perone SP. Computerized learning machine applied to qualitative analysis of mixtures by stationary electrode polarography analytical chemistry. Analytical Chemistry. 1971;**43**:382-388. DOI: 10.1021/ac60322a009

[60] Isenhour TL, Jurs PC. Some chemical applications of machine intelligence. Analytical Chemistry. 1971;**43**:20. DOI: 10.1021/ac60304a037

[61] Kowalski BR, Brender CF. Pattern recognition. A powerful approach to interpreting chemical data. The Journal of American Chemical Society. 1972;**94**:5632-5639. DOI: 10.1021/ja00771a016

[62] Koskinen JR, Kowalski BR. Interactive pattern recognition in the chemical laboratory. Journal of Chemical Information and Computer Sciences. 1975;**15**:119-123. DOI: 10.1021/ci60002a010

[63] Kryger L. Interpretation of analytical chemical information by pattern recognition methods-A survey. Talanta. 1981;**28**:871-887. DOI: 10.1016/0039-9140(81)80223-8

[64] Danzer K, Singer R. Application of pattern recognition methods for the investigation of chemical homogeneity of solids. Mikrochimica Acta. 1985;**85**:219-226 ISSN: 0026-3672

[65] von Waterbeemd H, Tayar NE, Carrupt PA, Testa B. Pattern recognition study of QSAR substituent descriptors. Journal of Computer-Aided Molecular Design. 1989;**3**:111-132 ISSN: 0920-654X

[66] Laplante JP, Pemberton M, Hjelmfelt A, Ross J. Experiments on pattern recognition by chemical kinetics. The Journal Physical Chemistry. 1995;**99**:1063-10065. DOI: 10.1021/j100025a001

[67] Beebe KR, Pell RJ, Seasholtz MB. Chemometrics: A Practical Guide 1998. New York: Wiley & Sons; 1998. ISBN-10: 0471124516

[68] Mardia KV, Kent JT, Bibby JM. Multivariate Analysis. New York: Academic Press; 1979. ISBN: 9780124712522

[69] Frisch A, Frisch MJ. Gaussian 98 User 'S Reference, Revision a.7. Pittsburgh: Gaussian, Inc; 1998

[70] Flukiger P, Luth HP, Portmann S, Weber J. MOLEKEL 4.3. Manno, Switzerland: Swiss Center for Scientific Computing; 2000-2001

[71] Infometrix, Inc. Pirouette 3.01 2002, Woodinville

[72] Olszak TA, Peeters OM, Blaton NM, Ranter CJ. 5-Nitrofuran-2-aldoxime. Acta Crystallographica C. 1995;**51**:1304-1306. DOI: 10.1107/S0108270194008425

[73] Aguirre G, Cabrera E, Cerecetto H, Di Maio R, González M, Seoane G, et al. Design, synthesis and biological evaluation of new potent 5-nitrofuryl derivatives as anti-*Trypanosoma cruzi* agents. Studies of trypanothione binding site of trypanothione reductase as target for rational design. European Journal of Medicinal Chemistry. 2004;**39**(5):421-431. DOI: 10.1016/j.ejmech.2004.02.007

[74] Cerecetto H, Di Maio R, Ibarruri G, Seoane G, Denicola A, Peluffo G, et al. Synthesis and anti-trypanosomal activity of novel 5-nitro-2-furaldehyde and 5-nitrothiophene-2-carboxaldehyde semicarbazone derivatives. Il Farmaco. 1998;**53**:89-94. DOI: 10.1016/S0014-827X(97)00011-6

[75] Cerecetto H, Di Maio R, González M, Risso M, Sagrera G, Seoane G, et al. Synthesis and antitrypanosomal evaluation of E-isomers of 5-nitro-2-furaldehyde and 5-nitrothiophene-2-carboxaldehyde semicarbazone derivatives. Structure–activity relationships. European Journal of Medicinal Chemistry. 2000;**35**:343-350. DOI: 10.1016/S0223-5234(00)00131-8

[76] Williams DE, Yan JM. Point-charge models for molecules derived from least-squares fitting of the electric potential. Advances in Atomic and Molecular Physics. 1998;**23**:87-130. DOI: 10.1016/S00065-2199(08)60106-2

[77] Chirlian LE, Francl MM. Atomic charges derived from electrostatic potentials: A detailed study. Journal of Computational Chemistry. 1984;**8**:894-905. DOI: 10.1002/jcc.540080616

[78] Singh UC, Kollman PA. An approach to computing electrostatic charges for molecules. Journal of Computational Chemistry. 1984;**5**:129-145. DOI: 10.1002/jcc.5400502004

[79] Hyperchem 8.0.6, Inc. ChemPlus: Modular Extensions to HyperChem Release 8.06. Molecular Modeling for Windows 2008. Gainesville

[80] Tetko IV, Gasteiger J, Todeschini R, Mauri A, Livingstone D, Ertl P, et al. Virtual computational chemistry laboratory-design and description. Journal of Computer-Aided Molecular Design. 2005;**19**:453-463. DOI: 10.1007/s10822-005-8694-y

[81] Becke AD. Density-functional thermochemistry. III. The role of exact exchange. The Journal of Chemical Physics. 1993;**98**:5648-5652. DOI: 10.1063/1.464913

[82] Lee C, Yang W, Parr RG. Development of the colic-salvetti correlation-energy formula into a functional of the electron density. Physical Review B. 1988;**37**:785-789. DOI: 10.1103/PhysRevB.37.785

[83] Becke AD. Density-functional exchange-energy approximation with correct asymptotic behavior. Physical Review A. 1988;**38**:3098-3100. DOI: 10.1103/PhysRevA.38.3098

[84] Roothaan CC. New developments in molecular orbital theory. Reviews of Modern Physics. 1951;**23**:69-89. DOI: 10.1103/RevModPhys.23.69

[85] Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP. Development and use of quantum mechanical molecular models. 76. AM1: A new general purpose quantum mechanical molecular model. Journal of the American Chemical Society. 1985;**107**:3902-3909. DOI: 390910.1021/ja00299a024

[86] Wu X, Thiel W, Pezeshki S, Lin H. Specific reaction path hamiltonian for proton transfer in water: Reparameterized semiempirical models. Journal of Chemical Theory and Computation. 2013;**9**:2672-2686. DOI: 10.1021/ct400224n

[87] Hehre WJ, Radom L, PvR S, Pople JA. Ab Initio Molecular Theory. New York: Wiley; 1986. DOI: 10.1002/jcc.540070314

[88] Ferreira MMC, Montanari CA, Gaudio AC. Seleção de variáveis em QSAR. Química Nova. 2002;**25**:439-448. DOI: 10.1590/S0100-40422002000300017

[89] Todeschini R, Consonni V. In: Mannhold R, Kubinyi H, Timmerman H, editors. Molecular Descriptors for Chemoinformatics. Vol I & II. Weinheim: Wiley-VCH; 2009. ISBN: 978-3-527-31852-0

[90] Karelson M, Victor S, Lobanov A, Katritzky R. Quantum-chemical descriptors in QSAR/QSPR studies. Chemical Reviews. 1996;**96**:1027-1043. DOI: 10.1021/cr950202r

[91] Gosav S, Praisler M, Dorohoi DO. ANN expert system screening for illicit amphetamines using molecular descriptors. Journal of Molecular Structure. 2007;**834**:188-194. DOI: 10.1016/j.molstruc.2006.12.059

[92] Scotti M, Fernandes MA, Ferreira MJP, Esmereciano VP. Quantitative structure–activity relationship of sesquiterpene lactones with cytotoxic activity. Bioorganic & Medicinal Chemistry. 2007;**15**:2927-2934. DOI: 10.1016/j.bmc.2007.02.005

[93] Moriguchi I, Hirano S, Liu Q, Nakagome I, Matsushita Y. Simple method of calculating ocatanol/water partition coefficient. Chemical and Pharmaceutical Bulletin. 1992;**40**:127-130 ISSN: 1347-5223

[94] Paulino-Blumenfeld M, Hansz M, Hikici N. Electronic properties and free radical production by nitrofuran compounds. Free Radical Research Communications. 1992;**16**:207-215. DOI: 10.3109/10715769209049174