

Molecular electrostatic potential and pattern recognition models to design potentially active pentamidine derivatives against *Trypanosoma brucei rhodesiense*

Potencial eletrostático molecular e modelos de reconhecimento de padrões para desenhar derivados da pentamidina potencialmente ativos contra *Trypanosoma brucei rhodesiense*

Potencial electrostático molecular y modelos de reconocimiento de patrones para diseñar derivados de pentamidina potencialmente activos contra *Trypanosoma brucei rhodesiense*

Received: 09/03/2021 | Reviewed: 09/10/2021 | Accept: 09/19/2021 | Published: 09/19/2021

Luã Felipe Souza de Oliveira

ORCID: <https://orcid.org/0000-0002-6875-9553>
Universidade Federal do Pará, Brazil
E-mail: oliveira.luansouza@gmail.com

Hérica Coelho Cordeiro

ORCID: <https://orcid.org/0000-0002-2527-3430>
Universidade Federal do Pará, Brazil
E-mail: herica.coelhocordeiro@gmail.com

Helieverton Geraldo de Brito

ORCID: <https://orcid.org/0000-0002-8038-1784>
Universidade Federal do Pará, Brazil
E-mail: helievertonbrito@yahoo.com.br

Ana Cecília Barbosa Pinheiro

ORCID: <https://orcid.org/0000-0002-0880-683X>
Universidade Federal do Pará, Brazil
E-mail: anacecipinheiro@gmail.com

Marcos Antonio Barros dos Santos

ORCID: <https://orcid.org/0000-0002-1424-4132>
Universidade do Estado do Pará, Brazil
E-mail: marcos@uepa.br

Heriberto Rodrigues Bitencourt

ORCID: <https://orcid.org/0000-0002-0003-2876>
Universidade Federal do Pará, Brazil
E-mail: heriberto.ufpa@gmail.com

Antonio Florêncio de Figueiredo

ORCID: <https://orcid.org/0000-0001-6218-6670>
Instituto Federal de Educação, Ciência e Tecnologia do Pará, Brazil
E-mail: antonio.florencio@ifpa.edu.br

Josué de Jesus Oliveira Araújo

ORCID: <https://orcid.org/0000-0001-5585-6250>
Universidade Federal do Pará, Brazil
E-mail: josuenietsche@gmail.com

Fábio dos Santos Gil

ORCID: <https://orcid.org/0000-0002-8277-8849>
Escola Estadual de Ensino Médio Deodoro de Mendonça, Brazil
E-mail: deodorom@yahoo.com.br

Márcio de Souza Farias

ORCID: <https://orcid.org/0000-0002-3498-500X>
Universidade Federal do Pará, Brazil
E-mail: toraqk@yahoo.com.br

Jardel Pinto Barbosa

ORCID: <https://orcid.org/0000-0002-1595-6713>
Universidade do Estado do Amapá, Brazil
E-mail: jardel.barbosa@ueap.edu.br

José Ciríaco Pinheiro

ORCID: <https://orcid.org/0000-0001-8376-3086>
Universidade Federal do Pará, Brazil
E-mail: josuenietsche@gmail.com

Abstract

Molecular electrostatic potential (MEP) and pattern recognition (PR) were used to draw potentially active pentamidine derivatives against *Trypanosoma brucei rhodesiense* (*T. b. rhodesiense*). PR models: Principal Component Analysis, PCA model; Hierarchical Cluster Analysis, HCA model; K-Nearest Neighbor, KNN model; Soft Independent Modeling of Class

Analogy, SIMCA model; and Stepwise Discriminant Analysis, SDA model, were built by reducing the dimensionality of a data matrix to twenty-eight pentamidine derivatives and allowed the compounds to be classified into two classes: more active and less active, according to their degrees of activity against *T. b. rhodesiense*. The study outlined that the properties HOMO (highest occupied molecular orbital) energy, VOL (molecular volume), and ASA_P (water accessible surface area of all polar ($|q_i| \geq 0.2$) atoms) are the most relevant for the construction of the models. The key structural features required for biological activity investigated through MEP were used as guidelines in the design of thirteen new compounds, which were evaluated by PR models as more active or less active against *T. b. rhodesiense*. The application of PR models indicated nine promising compounds (29, 30, 31, 32, 33, 36, 37, 39, and 40) for synthesis and biological assays.

Keywords: Molecular electrostatic potential; Pattern recognition models; Investigation of pentamidine derivatives; Design of pentamidine derivatives.

Resumo

Potencial eletrostático molecular (MEP) e reconhecimento de padrão (RP) foram usados para desenhar derivados de pentamidina potencialmente ativos contra *Trypanosome brucei rhodesiense* (*T. b. rhodesiense*). Modelos de RP: Análise de Componentes Principais, Modelo PCA; Análise de Agrupamento por Métodos Hierárquicos, Modelo HCA; K-ésimos Vizinhos mais Próximos, Modelo KNN; Modelagem Independente Suave de Analogia de Classe, Modelo SIMCA; e Análise de Discriminante por Etapas, Modelo SDA, foram construídos reduzindo a dimensionalidade de uma matriz de dados para vinte e oito derivativos de pentamidina e permitiram que os compostos fossem classificados em duas classes: mais ativos e menos ativos, de acordo com seus graus de atividade contra *T. b. rhodesiense*. O estudo mostrou que as propriedades energia do HOMO (orbital molecular ocupado mais alto), VOL (volume molecular) e ASA_P (área de superfície acessível à água de todos os átomos polares ($|q_i| \geq 0,2$) são as mais relevantes para a construção dos modelos. As principais características estruturais necessárias para a atividade biológica investigada através do MEP foram usadas como diretrizes no desenho de treze novos compostos, que foram avaliados pelos modelos de RP como mais ativos ou menos ativos contra *T. b. rhodesiense*. A aplicação dos modelos de RP indicou nove compostos promissores (29, 30, 31, 32, 33, 36, 37, 39 e 40) para síntese e ensaios biológicos.

Palavras-chave: Potencial eletrostático molecular; Modelos de reconhecimento de padrões; Investigação de derivados de pentamidina; Desenho de derivados da pentamidina.

Resumen

El potencial electrostático molecular (MEP) y el reconocimiento de patrones (RP) se utilizaron para diseñar derivados de pentamidina potencialmente activos contra *Trypanosome brucei rhodesiense* (*T. b. rhodesiense*). Modelos RP: Análisis de componentes principales, modelo PCA; Análisis de conglomerados por métodos jerárquicos, modelo HCA; Vecinos más cercanos K-th, modelo KNN; Modelado de Analogía de Clase Independiente Suave, modelo SIMCA; y Análisis discriminante de pasos, modelo SDA, se construyeron reduciendo la dimensionalidad de una matriz de datos a veintiocho derivados de pentamidina y permitieron clasificar los compuestos en dos clases: más activos y menos activos, según su grado de actividad contra *T. B. rhodesense*. El estudio mostró que las propiedades de energía HOMO (orbital molecular ocupado más alto), VOL (volumen molecular) y ASA_P (área de superficie accesible al agua de todos los átomos polares) ($|q_i| \geq 0,2$) son las más relevantes para la construcción de modelos. Las principales características estructurales necesarias para la actividad biológica investigada mediante el MEP se utilizaron como pautas en el diseño de trece nuevos compuestos, que fueron evaluados por los modelos PR como más activos o menos activos frente a *T. b. rodesiense*. La aplicación de modelos RP indicó nueve compuestos prometedores (29, 30, 31, 32, 33, 36, 37, 39 y 40) para síntesis y ensayos biológicos.

Palabras clave: Potencial electrostático molecular; Modelos de reconocimiento de patrones; Investigación de derivados de pentamidina; Diseño de derivados de pentamidina.

1. Introduction

Human African Trypanosomiasis (HAT) is endemic in 36 sub-Saharan countries: 24 of them experience *Trypanosoma brucei gambiense* (*T. b. gambiense*) transmission, leading to West African sleeping sickness, while in 13 countries *Trypanosoma brucei rhodesiense* (*T. b. rhodesiense*) is present, causing the most acute East African sleeping sickness syndrome, Uganda being the only country reporting both types (WHO, 2019; Selby, et al., 2019). HAT is maintained by an animal reservoir (Holmes, 2015; Franco, et al., 2018; Franco, et al., 2020) and, currently, its treatment requires a long period of parietal administration of drugs of high toxicity and poor efficacy (Doleželová, et al., 2018). Thus, the development of new drugs to treat sleeping sickness is necessary (Doleželová, et al., 2018; WHO, 2019).

Bakunova, et al., (2009) evaluated the diamidina (pentamidine) and some its derivatives in the standard screens used to identify its antiprotozoal drug candidates. Analysis of the structure-activity relationships showed high antitrypanosomal activity in

vitro.

The MEP (Bernardinelli, et al., 1994; Jefford, et al., 2000) and PR (Gangwal, et al., 2016) approaches have been used as independent strategies in the study of active compounds and in proposing new molecules for synthesis and biological testing (Bernardinelli, et al., 1994; Gangwal, et al., 2016). The combination of these two powerful tools to unravel the structure-activity relationship of bioactive compounds and, therefore, propose new molecules was shown by our research group on the designing of potentially active compounds against human hepatocellular carcinoma HepG2, *Plasmodium falciparum*, and *Trypanosoma cruzi*, respectively (Barbosa, et al., 2011; Cristino, et al., 2012; Santos, et al., 2020).

In this article, we present a study of the application of MEP and PR models for the investigation of pentamidine and derivatives with activity against *T. b. rhodesiense* reported in literature (Bakunova, et al., 2009), and the designing of new potentially active derivatives. In the first step of our study, MEP maps (Mehmood, et al., 2018; Aray, 2019; Rzesikowska, et al., 2019) were used in an attempt to identify key structural features of the compounds that are necessary for their activities, to investigate their probable interactions with a receptor through recognition in a biological process, and to propose the new derivatives. PR approaches (Varmuza, 1980; Varmuza, 2018; Beebe, et al., 1998) are used to build the models that will be applied later to the forecast set obtained with the MEP studies.

2. Computational Methodology

2.1 Molecular Electrostatic Potential in the Ligand-Receptor Biological Process

MEP is a very useful topological index in the study of hydrogen bond interactions as well as in the understanding of electrophilic and nucleophilic attack sites in a molecule obtained from electron density (Scrocco & Tomasi, 1979; Politzer & Murray, 2021). This property is also well suited for analyzing processes based on the “recognition” of one molecule by another, as in drug-receptor, and enzyme-substrate interactions (Scrocco & Tomasi, 1978; Politzer, et al., 1985). Because it is a physical property, MEP can be determined experimentally by diffraction or by computational studies (Politzer, et al., 2019). To investigate the reactive sites for pentamidine and its derivatives, the MEP was evaluated by using the HF/6-31G* approach (See Results and Discussion section). 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 electrons and nuclei of the molecule and a positive charge test (a proton) located at \vec{r} . For the studied molecules, the MEP values were computed using Eq. 1 (Politzer & Murray, 2021).

$$V(\vec{r}) = \sum_{i=1}^N \frac{Z_i}{|\vec{R}_i - \vec{r}|} - \int \frac{\rho(\vec{r}')d\vec{r}'}{|\vec{r}' - \vec{r}|} \quad (1)$$

where N is the number of nuclei with charges Z_i , located at position R_i 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.

2.2 PR approaches

PR approaches were introduced in the Chemistry in the 1970s to analyze various types of spectroscopic data (Kowalski & Brender, 1972). Since then, PR has been an excellent tool to aid in the interpretation of chemical data to obtain relevant information in different application sectors of chemical science (Varmuza, 1980; Brown, 2017; Varmuza, 2018).

In this section, we will make a brief presentation of the PR approaches used to build the models in this work. A deeper and detailed description of these matters can be found elsewhere (Varmuza, 1980; Varmuza, 2018; Beebe, Pell & Seasholtz, 1998).

2.2.1 PCA, HCA, KNN, SIMCA, and SDA approaches

PCA approach reduces the dimensionality of a data matrix consisting of a large number of interrelated variables, while retaining, as much as possible, the variation present in the data set. The final result of using the technique provides a small number of variables (molecular properties) that are believed to be best related to the dependent variable (Beebe, et al., 1998; Vidal, et al., 2016), in this case, biological activity against *T.b. rhodesiense*.

A two-dimensional dendrogram illustrating the fusions or divisions made at each successive step of the analysis represents the HCA approach. The purpose of using it is to display the data in such a way as to emphasize its natural clusters and patterns in this two-dimensional space. The similarity among the clusters is given by the length of their branches so that compounds presenting low similarity have long branches, whereas compounds of high similarity have short branches (Beebe, et al., 1998; Ghosal, et al., 2020).

The KNN approach (Beebe, et al., 1998) is based on the class dependency of an unknown sample. Classification by this method is related to the comparison of distance between samples after calculating multivariate Euclidean distances between each sample of the training set. The model is constructed and a set of tests to classify the sample into the predicted class, considering the multivariate distance from this sample to the K training set samples, is performed.

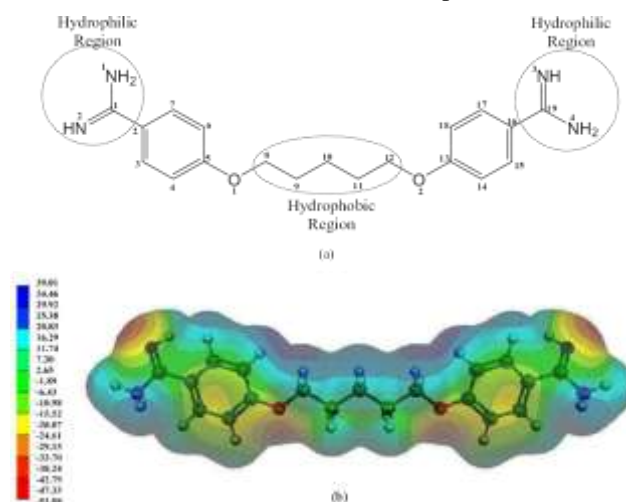
A model is constructed by means of the SIMCA approach for each class of the training set. In this method, a multidimensional box is built for each class, which means that shape and position of the samples in the classes are considered. The classification of a set sample is achieved by determining which space the sample occupies. It can be in the number of one, more than one or none of the classes (Beebe, et al., 1998).

The SDA approach (Johnson & Wichem, 1992; Zhang, et al., 2017; He, et al., 2020) allows you to separate samples from distinct populations and allocate new samples to previously defined populations. In this approach, at each step, the most powerful variable is inserted into the discriminant function. The approach is based on the F-test for the significance of the variables, that is, at each step, a variable is selected based on its significance and, after several steps; the most significant variables are extracted from the whole set investigated.

2.3 Pentamidine and studied derivatives

In Figures 1 and 2, the 2D structures of the pentamidine (Figure 1a) and derivatives (**1-28**), respectively, and the MEP map (kcal.mol⁻¹) of the pentamidine (Figure 1b) are shown – See Results and Discussion section.

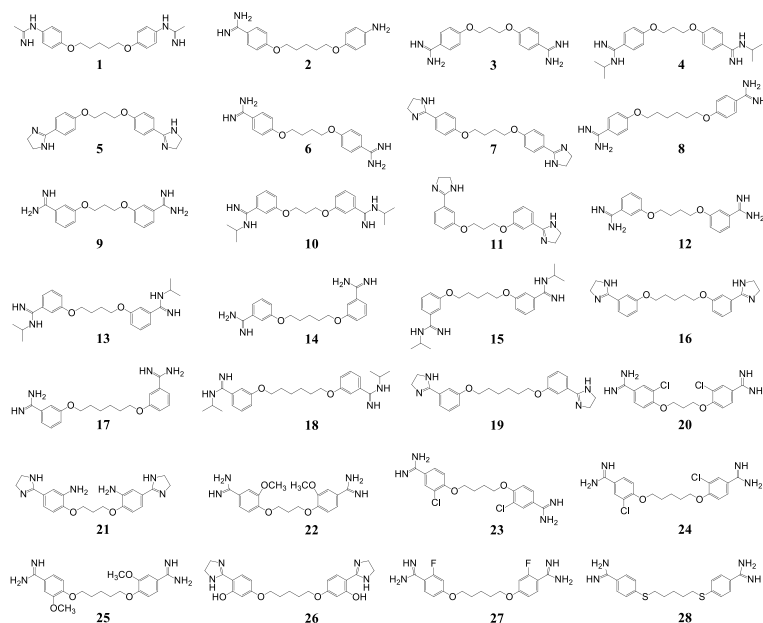
Figure 1 - (a) 2D Molecular structure and (b) MEP map (kcal.mol⁻¹) of the pentamidine.



Source: Authors (2020).

The atomic numbering of the compounds is the same adopted in pentamidine. The studied derivatives, from the point of view of the relationship experimental structure-activity, are well reported and documented on paper from literature (Bakunova et al., 2009). Here is a summary of the conclusions reached by the researchers for *T. b. rhodesiense*: unsubstituted amides were more active *in vitro* compared to the N-substituted derivatives; the introduction of 2,2'-dichloro substituents further improved antitrypanosomal properties; and the replacement of oxygen atoms in the alkyl chain with secondary amino groups improved activities of the resulting congeners against *T. b. rhodesiense*. In this work, the derivatives molecules were defined as more active against *T. b. rhodesiense*, when IC_{50} (μM) ≤ 0.075 , and as less active when IC_{50} (μM) > 0.075 .

Figure 2 - 2D molecular structure of the pentamidine derivatives with anti-*T. b. rhodesiense* activity.



Source: Bakunova, et al., (2009).

2.4 Molecular calculations

In the implementation of the calculations with pentamidine (Figure 1a) and derivatives (Figure 2) several theoretical approaches were used. For the quantum-chemical study, the methods available in the Gaussian program package (Frisch & Frisch, 1998) were used, namely: Austin model 1 (AM1) method (Dewar, et al., 1985), Parametric Method Number 3 (PM3) (Wu, et al., 2013), Hartree-Fock (HF) method (Roothaan, 1951), Becke's three-parameter hybrid method (Becke, 1993), the Lee-Yang-Parr (LYP) (Lee, et al., 1988), and standard basis sets (Hehre, et al., 1986) also available in the same program package. In pentamidine, geometry optimization was performed by AM1, PM3 approaches (Dewar, et al., 1985; Wu, et al., 2013) HF/3-21G*, HF/6-31G, HF/6-31G*, B3LYP/3-21G*, B3LYP/6-31G, and B3LYP/6-31G* approaches (Roothaan, 1951; Becke, 1993; Lee, et al., 1988) and basis sets (Hehre, et al., 1986). The calculations were carried out to find the approach and basis set that would present the best compromise between computational time and accuracy of the information relative to experimental data. The experimental structure of pentamidine was retrieved from the Cambridge Structural Database - CSD (Srikrishnan, et al., 2004). PCA and HCA approaches available in the Pirouette program (Infometrix, Inc, 2002) were used to compare the computed structures with different approaches of quantum chemistry with the experimental structure of pentamidine to identify the appropriate approach and the basis set for further calculations. The analyses were performed on an auto-scaled data matrix with dimension 57x9, where each row was related to 56 computed geometries and to 1 experimental geometry, and each column represented one of the 9 geometrical parameters of the pentamidine (bond lengths and bond angles). In order to compute all structures and to perform calculations to obtain the

molecular properties, the HF/6-31G* approach was selected (See Results and Discussion section), and the initial geometries of the derivatives (**1-28**) were built with the optimized geometry of the pentamidine selected by PCA and HCA approaches and subjected to the conformational search through optimization.

2.5 Molecular descriptors

In this study, all properties were obtained in the most stable conformation of each molecule and the descriptors were computed to give information about the influence of electronic, steric, hydrophilic, and hydrophobic features on the antitrypanosomal activity of the studied derivatives. Quantum-chemical descriptors such as total energy of molecules (TE), highest occupied molecular orbital (HOMO) energy, one-level below to highest occupied molecular orbital (HOMO-1) energy, lowest occupied molecular orbital (LUMO) energy, one-level above lowest occupied molecular (LUMO+1) energy, total dipole moment (μ), electronegativity of Mulliken (χ), atomic charges on the Nth atom (QN), molecular hardness (HD) and molecular softness (MS) were computed with the Gaussian 98 program (Frisch & Frisch, 1998). The QN descriptor was derived from the electrostatic potential as implemented in the Gaussian program. 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 (Williams & Yan, 1998; Chirlian & Franci, 1987). The charges derived from electrostatic potential present the advantage of being, in general, physically more satisfactory than the Mulliken charges (Singh & Kollman, 1984), especially regarding biological activity. Physicochemical descriptors such as total surface area (TSA), VOL, molecular refractivity (MR), molecule hydration energy (MHE), and coefficient of molecular partition octanol-water (log P) were obtained with the program ChemPlus (Hyperchem, Inc, 2008). In order to represent different sources of chemical information in terms of size, symmetry, and distribution of atoms in the molecule we have also included molecular holistic (MH) descriptors and these descriptors were obtained with the Dragon program (Todeschini & Consonni, 2009; Grisoni, et al., 2018).

3. Results and Discussion

3.1 Quantum chemical approach and basis set selection for the description of the geometry of pentamidine and derivatives

There is an advantage to using PCA and HCA at this stage of the study: all structural information is considered simultaneously and the correlations between them are considered. Table 1 shows the theoretical and experimental structural information (bond lengths and bond angles) of the geometry of the pentamidine. It was used with the aim to select using of PCA approach and HCA approach, which quantum-chemical approach and basis set give results closet to the experimental data (Srikrishnan, et al., 2004).

Table 1 - Experimental and theoretical structural parameters of the pentamidine.

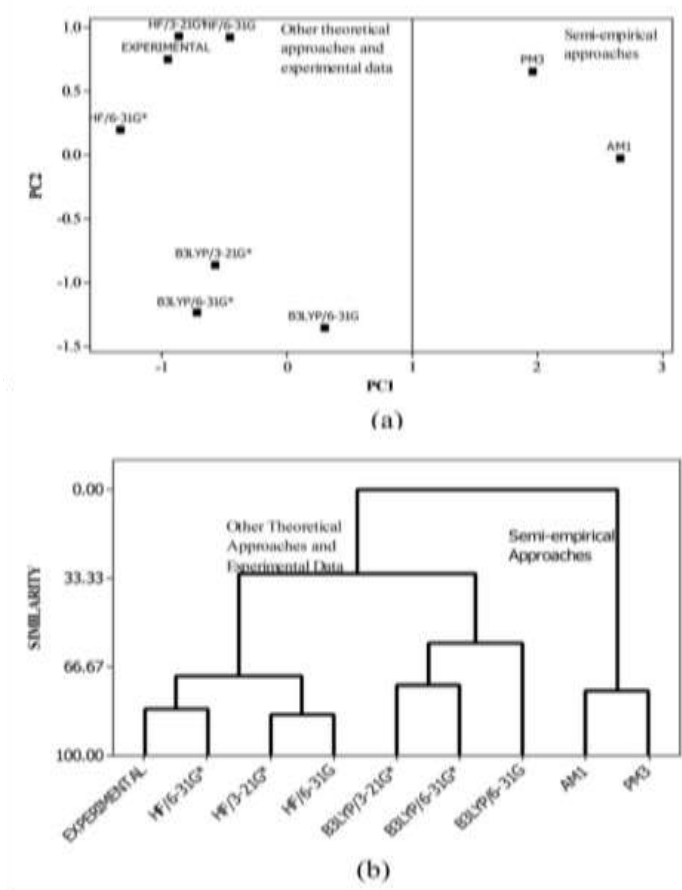
Geometric parameters	Method/Basis set								Exp ^a
	AM1	PM3	HF/3-21G*	HF/6-31G	HF/6-31G*	B3LYP/3-21G*	B3LYP/6-31G	B3LYP/6-31G*	
Bond length (Å)									
N1C1	1.299	1.295	1.265	1.267	1.261	1.290	1.298	1.285	1.297
N2C2	1.414	1.439	1.379	1.382	1.388	1.386	1.387	1.397	1.320
C1C2	1.497	1.481	1.489	1.489	1.491	1.493	1.490	1.490	1.481
C2C3	1.396	1.399	1.390	1.391	1.395	1.404	1.409	1.405	1.392
C3C4	1.395	1.386	1.376	1.377	1.378	1.388	1.393	1.389	1.374
C4C5	1.399	1.406	1.388	1.388	1.392	1.402	1.4025	1.402	1.377
C5C6	1.409	1.399	1.384	1.385	1.389	1.401	1.403	1.402	1.383
C6C7	1.388	1.381	1.382	1.383	1.385	1.391	1.396	1.395	1.394
C7C2	1.401	1.394	1.383	1.384	1.386	1.399	1.403	1.399	1.383
C5O1	1.379	1.376	1.368	1.373	1.346	1.384	1.389	1.363	1.366
O1C8	1.433	1.423	1.442	1.444	1.405	1.466	1.459	1.426	1.451
C8C9	1.519	1.527	1.525	1.525	1.518	1.529	1.524	1.523	1.521
C9C10	1.513	1.520	1.539	1.539	1.529	1.543	1.539	1.534	1.538
C10C11	1.513	1.520	1.539	1.539	1.529	1.542	1.539	1.534	1.536
C11C12	1.519	1.527	1.525	1.525	1.518	1.529	1.524	1.523	1.518
C12O2	1.433	1.423	1.442	1.444	1.405	1.466	1.459	1.426	1.441
O2C13	1.379	1.376	1.368	1.373	1.346	1.384	1.390	1.363	1.356
C13C14	1.409	1.404	1.388	1.388	1.392	1.402	1.402	1.402	1.405
C14C15	1.388	1.386	1.376	1.377	1.378	1.388	1.393	1.389	1.391
C15C16	1.401	1.399	1.390	1.391	1.385	1.404	1.409	1.405	1.383
C16C17	1.396	1.394	1.383	1.384	1.386	1.398	1.403	1.399	1.388
C17C18	1.395	1.391	1.382	1.383	1.385	1.391	1.396	1.393	1.402
C18C13	1.398	1.399	1.384	1.385	1.389	1.401	1.403	1.402	1.388
C16C19	1.497	1.481	1.489	1.489	1.491	1.493	1.490	1.490	1.483
C19N3	1.299	1.295	1.2645	1.267	1.261	1.290	1.298	1.285	1.293
C19N4	1.414	1.440	1.379	1.382	1.388	1.386	1.387	1.397	1.310
Bond angle (grade)									
N1C1N2	127.1	123.7	125.9	126.0	125.3	126.8	125.5	125.6	119.9
N1C1C2	118.5	119.5	118.9	118.9	119.7	118.2	118.6	119.3	120.3
N2C1C2	114.2	116.7	115.1	115.0	114.9	114.9	115.9	115.1	125.8
C1C2C3	120.2	120.1	122.5	122.3	122.3	122.9	122.8	122.5	121
C2C3C4	120.3	120.3	120.7	120.8	121.1	120.7	121.0	121.0	118.9
C3C4C5	119.3	119.5	120.2	120.0	120.2	120.4	119.9	120.2	122
C4C5C6	120.6	120.5	119.4	119.7	119.5	119.2	119.9	119.5	119.7
C5C6C7	119.3	119.2	119.9	119.8	119.7	120.0	119.5	119.7	118.7
C6C7C2	120.4	120.5	121.0	121.0	121.4	121.1	121.3	121.4	121.3
C7C2C3	119.9	119.9	118.6	118.7	118.2	118.6	118.3	118.2	119.7
C4C5O1	114.6	113.7	116.3	115.9	115.9	115.9	115.7	115.8	115.3
C6C5O1	124.8	125.8	124.3	124.9	124.6	124.8	124.4	124.7	125.1
C5O1C8	115.9	117.1	121.7	121.6	120.4	118.9	119.5	118.9	116.5
O1C8C9	105.7	106.2	105.9	105.8	107.6	105.6	106.9	107.6	104.8
C8C9C10	110.2	110.2	111.7	111.9	112.4	111.8	112.3	112.4	110.7
C9C10C11	110.7	111.0	111.9	111.9	112.4	111.9	112.4	112.6	110.8
C10C11C12	110.2	110.2	111.7	111.9	112.4	111.8	112.32	112.41	110.9
C11C12O2	105.7	106.2	105.9	105.9	107.6	105.6	106.9	107.6	105.8
C12O2C13	115.9	117.1	121.7	121.6	120.4	108.9	119.5	118.9	117.3
O2C3C14	114.6	113.7	116.3	115.97	115.9	115.9	115.7	115.8	114.5
O2C13C18	124.8	125.8	124.3	124.3	124.6	124.8	124.4	124.7	124.9
C13C14C15	119.3	119.5	120.2	120.0	120.2	120.4	119.9	120.2	119.3
C14C15C16	120.3	120.3	120.7	120.8	121.1	120.7	121	121.0	120.6
C15C16C17	119.9	119.9	118.6	118.7	118.2	118.6	118.3	118.2	119.7
C16C17C18	120.4	120.5	121.0	121.0	121.4	121.1	121.3	121.4	120.8
C7C18C13	119.3	119.3	119.97	119.8	119.7	120.0	119.6	119.7	118.9
C18C13C14	120.6	120.5	119.4	119.7	119.5	119.2	119.9	119.5	120.8
C15C16C19	120.2	120.1	122.5	122.3	122.3	122.9	122.8	122.5	120.9
C16C19N4	114.2	116.7	115.1	115.0	114.9	114.98	115.9	115.1	119.4
N3C19N4	127.1	123.7	125.9	126.0	125.3	126.8	118.6	125.6	120.5
C16C19N3	118.5	119.5	118.9	118.9	119.7	118.2	125.5	119.3	119.6

^a((Srikrishnan, et al., 2004). Source: Authors (2020).

The first two principal components retain 96.47% of the true information as follows: PC1 = 64.79 and PC2 = 28.68%. The

PC1 versus PC2 scores plot is shown in Figure 3a, from which it can be seen that the methods are discriminated into two classes, according to PC1. The semi-empirical approaches (AM1 and PM3) are on the right side, while the other theoretical (HF and B3LYP) approaches and experimental data are on the left side. Moreover, it is noticed that the HF/3-21G* and HF/6-31G* approaches/basis sets are the closest to the experimental data, indicating that either of them could be used in the development of this study.

Figure 3 - (a) Score plot of the two first PCs, PC1 and PC2, and (b) dendrogram obtained with the HCA for the separation of the methods/basis sets into two classes: semi-empirical approaches and other theoretical approaches and experimental data.



Source: Authors (2020).

Also, so as to investigate the most appropriate approach and basis set for further calculations, we used HCA approach. Figure 3b shows the dendrogram obtained with complete linkage method, in this figure one can 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 being, therefore, the most suitable to carry out this study.

3.2 The MEP of pentamidine

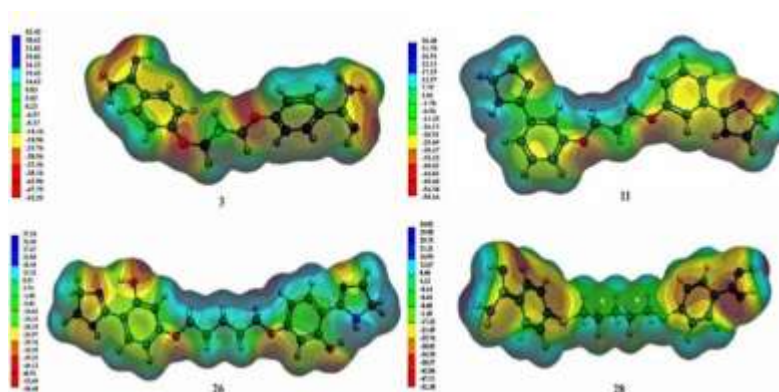
Figure 1a shows the 2D molecular structure of pentamidine, a bisbenzamidine with two aromatic nuclei and one unbranched saturated five-carbon chain inserted between two phenolic ether oxygens. Electronically, the molecule exhibits non-binding electron pairs in amidine groups, leading to the formation of a resonance effect between carbon and nitrogen atoms, π electrons in aromatic nuclei, and non-binding electron pairs in the atoms of oxygen from phenol ethers. We can still identify in the pentamidine structure two hydrophilic regions and a hydrophobic region. As one can see, these structural and electronic features presented by the molecule seem to be relevant in any biological recognition processes. Figure 1b shows the MEP map of the

pentamidine. In this figure, we notice that the molecule has contour surface, in the benzene rings, π electrons (yellow color), in the region between them, associated with the oxygen lone pairs (yellow and red colors), and at the ends of the molecular structure, associated with the lone pairs in amidine groups (yellow and red colors), corresponding to the regions indicated in Figure 1a, related to negative MEP, on which the lowest value for the electronic density is $-51.88 \text{ kcal.mol}^{-1}$ (red color), indicating an electronic density concentration due to the lone pairs of electrons in the oxygen (O_1 and O_2) and nitrogen (N_1 and N_3), also the difference of electronegativity between N_1 and N_3 , relative to carbon, provides the displacement of the bonding electrons to the proximity of these atoms. From MEP map, it would be predictable that an electrophile attack can occur in these negative regions of $V(\vec{r})$, and that they may be important in a possible biological recognition process, confirming the structural and electronic features presented by the molecular in Figure 1a. Importantly, the key to planning new active derivatives: the inclusion of substituents at positions C2, C16, O1, and O2 must maintain the structural and electronic features presented by the molecular structure in Figure 1a and evidenced by values of MEP in Figure 1b. The molecule also shows surface contour characterized by positive MEP (blue color), whose highest value corresponds the $+39.01 \text{ kcal.mol}^{-1}$ (blue color). It is expected that the positive regions de $V(\vec{r})$ will be preferred for nucleophilic processes.

3.3 MEP of the derivatives

In Figure 4, the MEP maps for four studied derivatives (**3** and **28** more active and **11** and **26** less active) are shown. In this figure, we can notice that the derivatives have similar MEP to pentamidine (Figure 1b) in regions of rings, between the rings and at the end of the molecular structures.

Figure 4 - MEP maps (kcal.mol^{-1}) for pentamidine derivatives (**3**, **28**, **11**, and **26**) with anti-*T. b. rhodesiense* activity



Source: Authors (2020).

It may also be noted that the inclusion of groups with electron-withdrawing function in the ring or the replacement of the oxygen atoms by sulfur atoms in the chain provides greater distribution of electron density in the molecule and increase the anti-*T. b. rhodesiense* activity of the derivatives. The contour surfaces corresponding to the lower MEP (red color), $V(\vec{r}) = -56.16 \text{ kcal.mol}^{-1}$, and higher electron density, yet are located around the oxygen or sulfur and nitrogen atoms. The lower values for MEP also stem from the existence of isolated electron pairs in atoms. Still, from MEP maps, it would be predictable that an electrophilic attack can occur in these regions (red color), and that they may be important in a possible biological recognition process. On the other hand, Figure 4 shows a positive region with $V(\vec{r}) = +43.42 \text{ kcal.mol}^{-1}$ (blue color) indicating that this site is probably susceptible to nucleophilic attacks.

3.4 PCA, HCA, KNN, SIMCA, and SDA models

To carry out the PR models, all variables were auto scaled as preprocessing so that they could be standardized, and they

could also have the same importance regarding the scale. The number of multivariate variables was reduced by using the Pearson coefficient (r). When two variables presented values of $r > 0.8$, one of the two was excluded from the matrix a , randomly, since they, theoretically, describe the same property (Ferreira, et al., 2002), that is, they also have a high correlation with anti-*T. b. rhodesiense* activity and only one of them is enough to be used as independent variable in the predictive models. Thus, properties that showed either the same values for most samples or a small correlation with the activity ($r < 3$) were eliminated too.

Three molecular descriptors were selected for PCA model. The molecular descriptors (*HOMO energy*, *VOL*, and *ASA_P*), the experimental data of activity IC_{50} (μM) for anti-*T. b. rhodesiense* derivatives and correlations including all data for 28 compounds can be seen in Table 2.

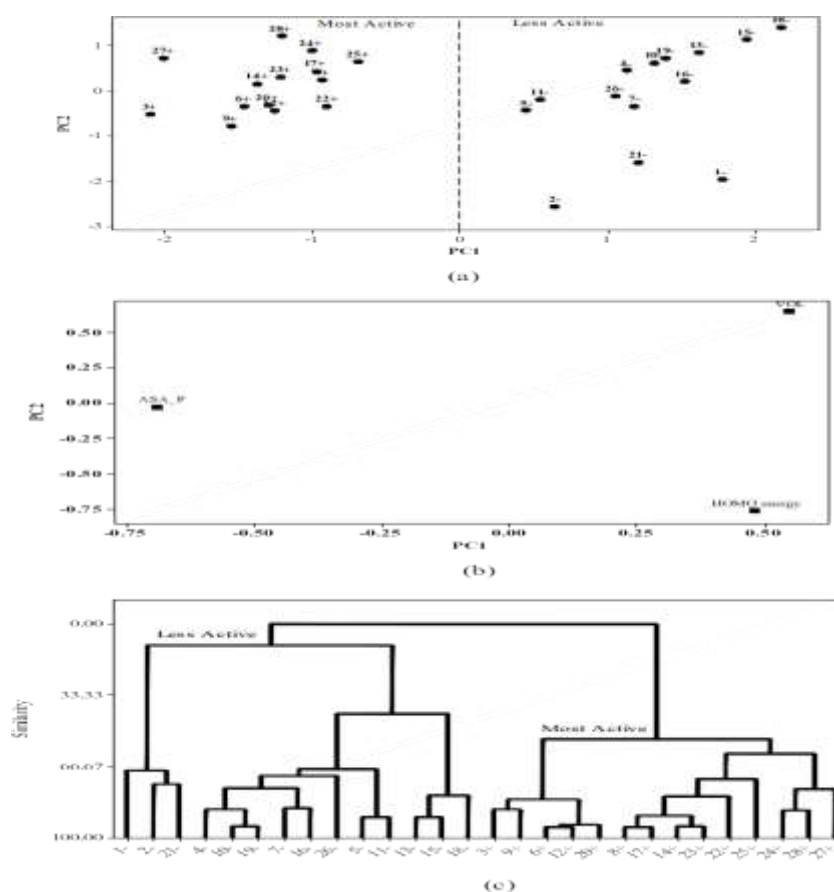
Table 2 - Pentamidine (for comparison) and derivatives parameters^a, experimental values of IC_{50} , and correlation matrix.

Derivatives	<i>HOMO energy</i> (kcal.mol ⁻¹)	<i>VOL</i> (Å ³)	<i>ASA_P</i> (Å ²)	IC_{50}^b (μM)
Pentamidine	-197.1	1073.6	202.4	0.007
1-	-169.4	1178.9	135.2	3.040
2-	-172.8	1007.4	146.9	1.310
3+	-198.1	959.74	233.5	0.007
4-	-190.4	1266.6	120.2	0.093
5-	-189.8	1111.3	120.2	0.110
6+	-196.5	1018.9	210.9	0.010
7-	-187.6	1162.9	89.77	0.244
8+	-196.8	1122.8	206.01	0.036
9+	-194.9	965.50	206.01	0.052
10-	-191.7	1273.2	100.7	0.543
11-	-192.2	1110.8	100.7	4.300
12+	-195.3	1019.9	200.4	0.056
13-	-190.9	1327.3	100.2	0.619
14+	-198.8	1073.7	213.6	0.065
15-	-191.2	1377.9	91.25	0.416
16-	-190.2	1218.9	69.68	1.490
17+	-198.6	1127.2	202.05	0.075
18-	-190.5	1439.3	98.65	0.258
19-	-192.8	1272.8	88.81	0.875
20+	-195.4	1043.0	212.05	0.004
21-	-178.1	1096.1	108.1	0.813
22+	-191.7	1105.1	221.2	0.023
23+	-198.6	1104.5	213.5	0.042
24+	-201.7	1158.4	201.0	0.044
25+	-195.7	1221.2	227.9	0.010
26-	-185.6	1243.1	140.5	3.700
27+	-204.4	1084.5	243.1	0.013
28+	-206.3	1140.2	189.1	0.005
<i>HOMO energy</i>		0.133	-0.542	
<i>VOL</i>			-0.644	

^aFor brevity, only the parameters which were used in best separation. ^b(Bakunova, et al., 2009).
 Source: Authors (2020).

The correlation between descriptors is less than 0.65. The first three principal components describe 100% of the true information for the 28 molecules. The PC1-PC2 (a) scores and (b) loadings plots and (c) dendrogram obtained with Hierarchical Cluster Analysis (HCA) for the separation of the derivatives into two classes: (+) more active and (-) less active against *T. b. rhodesiense* are shown in Figure 5. From Figure 5a, one can see that the derivatives are distributed into two distinct regions in PC1. The more active derivatives are on the left side (3, 6, 8, 9, 12, 14, 17, 20, 22, 25, 27, and 28) and the less active are on the right side (1, 2, 4, 5, 7, 10, 11, 13, 15, 16, 18, 19, 21, and 26). According to Figure 5b, the *ASA_P* descriptor is responsible for displaying more active derivatives on the left side, while the *HOMO energy* and *VOL* descriptors displayed less active analogs for the right side of this figure.

Figure 5 - (a) Scores, (b) loadings vector plots of the two first PCs, PC1 and PC2, for the three variables responsible for the separation and (c) dendrogram obtained with HCA for the separation of derivatives into two classes: (+) More active and (-) Less active against *T. b. rhodesiense*.



Source: Authors (2020).

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

$$PC1 = 0.845 (HOMO\ energy) + 0.552 (VOL) - 0.679 (ASA_P) \quad (2)$$

From this equation, one can see that more active pentamidine derivatives can be obtained when we have lower values for the *HOMO energy* combined with lower values for *VOL* and high values for *ASA_P*.

Table 3 - The loading vectors for PC1, PC2, and PC3.

Descriptors	PC1	PC2	PC3
HOMO energy	0.485	-0.768	0.419
VOL	0.552	0.640	0.535
ASA_P	-0.679	-0.029	0.734

Source: Authors (2020).

The results of the HCA model are displayed in the dendrogram in Figure 5c and are similar to those of PCA model. The pentamidine derivatives are fairly well grouped according to their activity. From this figure, one can see the two clusters: more active (3, 6, 8, 9, 14, 17, 20, 22- 25, 27, 28) and less active (1, 2, 4, 5, 7, 10, 11, 13, 15, 16, 18, 19, 21, 26).

Table 4 shows the results obtained with one (1NN) to four (4NN) nearest neighbors. To all results the percentage of correct information was 100%. We used the model 4NN because the larger the number of the nearest neighbors, the better reliability of the KNN method (Ferreira, 2015), and this was used for validation of the training set from Figure 2.

Table 4 - Classification matrix for application of the KNN method

Category	Number of compounds	Compounds incorrectly classified			
		1NN	2NN	3NN	4NN
More active	14	0	0	0	0
Less active	14	0	0	0	0
Total	28	0	0	0	0
Correct information (%)		100	100	100	100

Source: Authors (2020).

The SIMCA model was built with the same descriptors as PCA, HCA, and KNN models and used three (3) PCs in the modeling of the two classes: more active derivatives (3, 6, 8, 9, 12, 14, 17, 20, 22-25, 27, 28) and less active derivatives (1, 2, 4, 5, 7, 10, 11, 13, 15, 16, 18, 19, 21, 26). In Table 5, the results obtained for the SIMCA model are shown. In this case, one can see that the information percentage was also 100%.

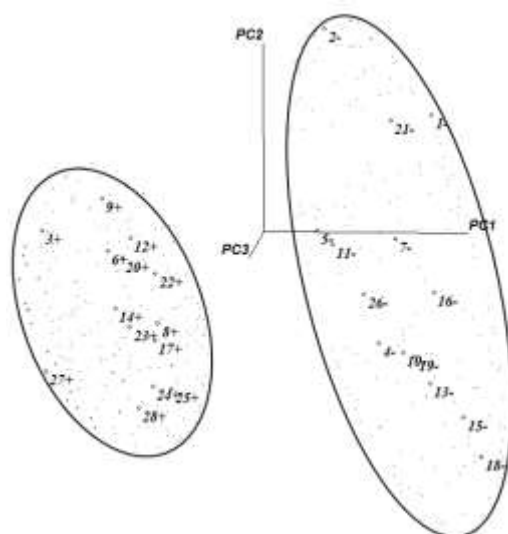
Table 5 - Classification matrix for application of SIMCA method.

Category	Number of compounds	SIMCA classification
More actives	14	14
Less actives	14	14
TOTAL	28	28
Correct information (%)		100

Source: Authors (2020).

Figure 6 shows the PC1-PC2-PC3 scores plot for the training set with the boundaries of the classes for SIMCA modeling. In this figure, one can see that SIMCA models have a good discriminating power, *i.e.*, high interclass distances and high between classes residuals.

Figure 6 - 3D scores plot showing boundaries classes for 28 pentamidine derivatives with anti-*T. b. rhodesiense* activity.



Source: Authors (2020).

In the construction of the SDA model, the discrimination functions for groups more active and less active, respectively, are given below:

$$\text{Group more active: } -4.94 (\text{HOMO energy}) - 1.20 (\text{VOL}) + 12.925 (\text{ASA}_P) - 7.89 \quad (3)$$

$$\text{Group less active: } 5.23 (\text{HOMO energy}) + 1.30 (\text{VOL}) - 13.65 (\text{ASA}_P) - 8.84 \quad (4)$$

From this discrimination functions, one can see that molecular descriptors *HOMO energy*, *VOL*, and *ASA_P* have high contribution in the construction of the model. According to the models by PCA, HCA, KNN, and SIMCA we can also notice that the descriptors are key properties for explaining the anti-*T. b. rhodesiense* activity of the derivatives the training set (Figure 2).

With the discriminant functions, Eqs 3 and 4, and the value of each descriptor for the compounds studied, the classification matrix for the training set in Table 6 is shown. According to this table, the derivatives were classified as more active and less active, with 100% certainty.

Table 6 - Classification matrix for application of SDA method.

Classified group	Number of compounds	True group	
		More active	Less active
Group (Class): more active	14	14	0
Group (Class): less active	14	0	14
TOTAL	28	14	14
Correct information (%)		100	100

Source: Authors (2020).

When activity against *T. b. rhodesiense* of new pentamidine derivative is investigated with the SDA model built in this report, one can use the rule, as follow: (a) calculate, initially, for the new derivative, 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 performed in

this work; (c) check which discrimination function (Eq. 3 or 4) presents the highest value. The new derivative is more active if it is related to discrimination function of group more active and vice-versa.

The reliability of the SDA model was assessed by cross-validation testing using the leave-one-out method. In this approach, one compound is omitted of the data set and the classification functions are built based on the remaining compounds. Afterwards, the omitted compound is classified according to the classification functions generated. In the next step, the omitted compound is included, and a new compound is removed, and the procedure goes on until the last compound is removed. Table 7 shows a summary of the results obtained with the cross-validation model.

Table 7 - Classification matrix for application of SDA method with cross-validation.

Classified group	Number of compounds	True group	
		More active	Less active
Group (Class): more active	14	14	0
Group (Class): less active	14	0	14
TOTAL	28	14	14
Correct information (%)		100	100

Source: Authors (2020).

As HOMO energy, VOL, and ASA_P properties were selected in the PCA, HCA, KNN, SIMCA, and SDA models as most important variables to describe the antitrypanosomal activity, some considerations about them may be relevant to the understanding of the behavior of more active pentamidine derivatives. The frontier molecular orbital (FMO) theory states that the transition state formation in a chemical reaction is due to the interaction between the frontier orbitals (HOMO e LUMO) of reactive species. According to the literature, these orbitals play a major role in governing many chemical and pharmacological processes and are also responsible for the formation of many charge-transfer complexes (Karelson, et al., 1996; Fukui, 1997; Bulat, et al., 2021). From Table 2, one can see that the HOMO energy for more active derivatives must present lower values than the less active ones, similar to pentamidine. This is an indication that biological processes occur allowing the electronic interaction between the most active derivatives and an eventual biological target, producing an increase of the antitrypanosomal activity. For the hypothesis of ligand/receptor complex formation due to nucleophilic attack of biological target in the ligand, we can verify a reaction with electrons transfer from the HOMO of the ligand molecule to the LUMO of the biological target.

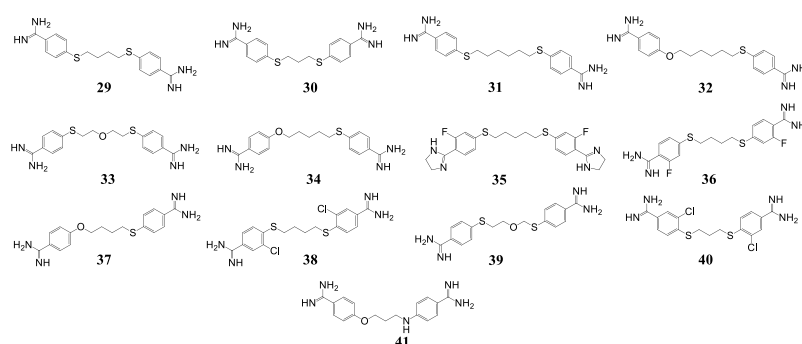
The VOL property represents steric interactions between compounds and the biological receptor. It is determined the transport characteristics of the molecules, such as intestinal absorption or blood-brain barrier penetration (Cruciani, et al., 2000). In Table 2, it can be seen that, in general, the more active derivatives present lower VOL than less active derivatives, similarly to pentamidine. This is an indication that biological processes occur allowing the steric interactions between the most active pentamidine derivatives and an eventual biological target, producing an increase of the antitrypanosomal activity.

The ASA is a hydrophilic property, by increasing the water-accessible surface of a molecule; one increases the drug's permeability to cells. The ASA_P property refers to the water accessible surface (in Å²) area of all polar atoms ($|q_i| \geq 0.2$) using a probe radius of 1.4 Å for molecule of water. Let q_i denote the partial charge of atom i . The ASA_P property depends on the storage of partial charges of the molecule and their conformation (Stanton & Jurs, 1990). The most active derivatives have hydrophilic regions constituted by the substituents -CNH₂NH at the ends of their chains. These regions accessible to the solvent may interact with the biological receptor through hydrogen-bridged interactions. In Table 2 derivatives with larger surface areas of all polar

atoms accessible to water are more active than those with these smaller areas, similar to pentamidine. This is an indication that biological processes occur allowing the hydrophilic interactions between the most active derivatives and an eventual biological target, producing an increase of antitrypanosomal activity.

Knowing the performance of the five PR models (PCA, HCA, KNN, SIMCA and SDA) for 28 studied derivatives, we decided to apply them to a series of thirteen compounds (Figure 7) designed to maintain the key structural features that are necessary for their activities evidenced by the MEP maps of the derivatives of the training set. The thirteen molecules proposed for the study of prediction of activity were drawn with the help of one of the collaborators of this work, belonging to the research group in organic chemistry of the Federal University of Pará, Brazil, and for the most promising syntheses in progress. In the future, antitrypanosomal tests with the most promising derivatives can be used to validate our PR models.

Figure 7 - 2D molecular structure of the pentamidine derivatives of the prediction set.



Source: Authors (2020).

The results obtained from the application of the PR models (PCA, HCA, KNN, SIMCA, and SDA) and the descriptors for the compounds of the prediction set are summarized in Tables 8 and 9, respectively.

Table 8 - Application of the PR models for pentamidine derivatives of the prediction set.

Compounds	PCA model	HCA model	KNN model	SIMCA model	SDA model
29	+	+	+	+	+
30	+	+	+	+	+
31	+	+	+	+	+
32	+	+	+	+	+
33	+	+	+	+	+
34	+	+	+	+	-
35	-	-	-	-	-
36	+	+	+	+	+
37	+	+	+	+	+
38	+	+	+	+	-
39	+	+	+	+	+
40	+	+	+	+	+
41	+	-	-	-	-

Source: Authors (2020).

In Table 8, the compounds **29**, **30**, **31**, **32**, **33**, **36**, **37**, **39**, **40** were predicted as more active against *T. b. rhodesiense* by the five models, while the compounds **35** and **41** were predicted as less active. On the other hand, only the SDA model did not classify the compounds **34** and **38** as more active. In this way, we can consider the derivatives **29**, **30**, **31**, **32**, **33**, **36**, **37**, **39**, **40** as potentially more actives.

Table 9 - Prediction pentamidine derivatives parameters.

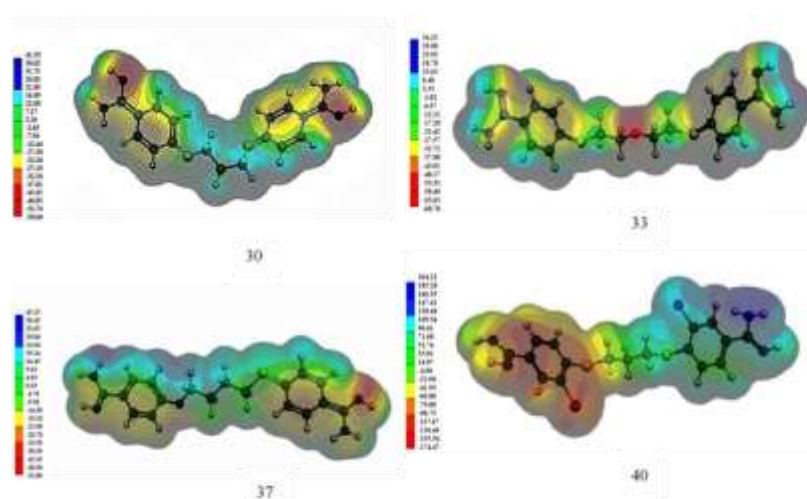
Compounds	HOMO energy (kcal.mol ⁻¹)	VOL (Å ³)	ASA_P (Å ²)
29	-205.114	1080.7	178.489
30	-204.763	1925.98	199.901
31	-204.825	1189.77	199.901
32	-189.301	1160.31	222.142
33	-206.407	1114.07	196.258
34	-190.273	1106.79	193.297
35	-190.342	1281.37	68.819
36	-192.376	1077.21	234.537
37	-185.73	1038.61	203.023
38	-191.303	1155.74	190.052
39	-187.563	1043.88	206.241
40	-189.935	1086.06	220.64
41	-172.722	967.49	190.511

Source: Authors (2020).

3.5 MEP maps for the most promising compounds of the prediction set

In Figure 8, the MEP maps for four of the most promising compounds (**30**, **33**, **37** and **40**) of the prediction set are shown. In the test of set compounds, similar to pentamidine and other compounds of the training set, we found minimum $V(\vec{r})$ values, with sites susceptible to electrophilic attacks, and maximum $V(\vec{r})$ values that indicate possibilities of attacks in nucleophilic processes. However, negative, and positive values for $V(\vec{r})$ in compound **40** of the test set result, respectively, from the substitution of the O atom by the S atom in the molecular structure of pentamidine and inclusion of the Cl atom at the *ortho* and *para* positions of the aromatic rings are highlights. For this compound, values of $V(\vec{r}) = -174.47$ and $V(\vec{r}) = +204.24$ kcal.mol⁻¹ indicate that these sites are likely to undergo electrophilic or nucleophilic attacks, respectively.

Figure 8 - MEP maps (kcal.mol⁻¹) for most promising pentamidine derivatives (**30**, **33**, **37** and **40**) of the prediction set.



Source: Authors (2020).

4. Concluding Remarks

MEP allowed us to identify the key structural features in pentamidine and derivatives that are necessary for their activities against *T. b. rhodesiense*. PR methods (PCA and HCA, KNN, SIMCA, and SDA) allowed the selection of variables (HOMO energy, VOL, and ASA_P) as the most important properties to describe the antitrypanosomal activity, to indicate the types of interaction between pentamidine and derivatives with a possible target in a biological process and it is interesting to notice that these properties represent three distinct classes of interactions between the molecules and a biological receptor: HOMO energy (electronic), VOL (steric), and ASA_P (hydrophilic). The key structural features of the compounds responsible for their biological activities, evidenced by the study with the MEP, were used to design 13 new pentamidine derivatives that, when submitted to the PR models, allowed us to infer that 9 of the proposed compounds (**29**, **30**, **31**, **32**, **33**, **36**, **37**, **39**, **40**) present promising potential for syntheses and biological evaluation, what in the future can be used to validate our PR models.

Acknowledgments

We thank Brazilian agencies Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for their financial support, the Virtual Computational Chemistry Laboratory (VCCLAB–Munich) for using the Drgon program, and the Swiss Center for Scientific Computing for using the Molekel program. We also employed computing facilities at the Laboratório de Química Teórica e Computacional (LQTC) at the Universidade Federal do Pará (UFPA).

This work is dedicated to Professor Ivan Alves da Silva (*in memory*). Professor Ivan Alves was member of the Department of Chemistry and contributed to the formation of many generations of chemists, physicists, geologists, pharmacists, and engineers at the College of Chemistry of the UFPA.

References

- Aray, Y. (2019). 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*, 123, 14421-14431.
- Bakunova, S.M., Bakunov, S. A., Patrick, D. A., Kumar, E. V. K. S., Ohemeng, K. A., Bridges, A. S., Wenzler, T., Barszcz, T., Jones, S. K., Werbovetz, K. A., Bun, R., & Tidwell, R. R. (2009). Structure-Activity Study of Pentamidine Analogues as Antiprotozoal Agents. *Journal of Medicinal Chemistry*, 52 (7), 2016-2035.
- Barbosa, J. P., Ferreira, J. E. V., Figueiredo, A. F., Almeida, R. C. O., Silva, O. P. P., Carvalho, J. R. C., Silva, O. P. P., Carvalho, J. R. C., Cristino, M. G. G., Ciriaco-Pinheiro, J., Vieira, J. L. F., & Serra, R. T. A. (2011). Molecular modeling and chemometric study of anticancer derivatives of artemisinin. *Journal of the Serbian Chemical Society*, 76 (9), 1263-1282.

- Becke, A. D. (1993). Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics*, 98 (7), 5648-5652.
- Beebe, K. R., Pell, R. J., & Seasholtz, M. B. (1998). *Chemometrics: A practical guide*. Wiley.
- Bernardinelli, G., Jefford, C. W., Marie, D., Thomson, C., & Weber, J. (1994). 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*, 21, 117-131.
- Brown, S. D. (2017). The chemometrics revolution re-examined. *Journal of Chemometrics*, 31 (1), e2856. doi.org/10.1002/cem.2856
- Bulat, F. A., Murray, J. S., & Politzer, P. (2021). Identifying the most energetic electrons in a molecule: The highest occupied molecular orbital and the average local ionization energy. *Computational and Theoretical Chemistry*, 1199, 113192.
- Chirlian, L. E., & Francl, M. M. (1987). Atomic charges derived from electrostatic potentials: A detailed study. *Journal of Computational Chemistry*, 8 (6), 894-905.
- Cristino, M. G. G., Meneses, C. C. F., Soeiro, M. M., Ferreira, J. E. V., Figueiredo, A. F., Barbosa, J. P., Almeida, R. C. O., Pinheiro, J. C., & Pinheiro, A. L. R. (2012). Computational Modeling of Antimalarial 10-Substituted Deoxoartemisinins. *Journal of Theoretical and Computational Chemistry*, 11 (2), 241-263.
- Cruciani, G., Crivori, P., Carrupt, P.-A., & Testa, B. (2000). Molecular Fields in Quantitative Structure-Permeation Relationships: The VolSurf approach. *Journal of Molecular Structure (Theochem)*, 503 (1-2), 17-30.
- Dewar, M. J. S., Zoebisch, E.G., Healy, E. F., & Stewart, J. J. P. (1985). Development and use of quantum mechanical molecular models. 76. AMI: a new general purpose quantum mechanical molecular model. *Journal of the American Chemical Society*, 107 (13), 3902-3909.
- Doleželoš, E., Terán, D., Gahura, O., Kotrbová, Z., Procházková, M., Keough, D., Špaček, P., Hocková, D., Guddat, L., & Zíková, A. (2018). Evaluation of the *Trypanosoma brucei* 6-oxopurine salvage pathway as a potential target for drug discovery. *PLoS Neglected Tropical Diseases*, 12 (2), e0006301. doi.org/10.1371/journal.pntd.0006301
- Ferreira, M. M. C., Montanari, C. A., & Gaudio, A. C. (2002). Seleção de variáveis em QSAR. *Quim Nova*, 25 (3), 439-448.
- Ferreira, M. M. C. (2015). *Químimetria: Conceitos, Métodos e Aplicações*. Campinas: Editora UNICAMP.
- Franco, J. R., Cecchi, G., Priotto, G., Paone, M., Diarra, A., Grout, L., Simarro, P. P., Zhao, W., & Argaw, D. (2018). Monitoring the elimination of human African trypanosomiasis: Update to 2016. *PLoS Neglected Tropical Diseases* 12 (12), e0006890. doi.org/10.1371/journal.pntd.0006890
- Franco, J. R., Cecchi, G., Priotto, G., Paone, M., Diarra, A., Grout, L., Simarro, P. P., Zhao, W., & Argaw, D. (2020). Monitoring the elimination of human African trypanosomiasis at continental and country level: Update to 2018. *PLoS Neglected Tropical Diseases* 14 (5), e0008261. doi.org/10.1371/journal.pntd.0008261
- Frisch, A., & Frisch, M. J. (1998). *Gaussian 98 User 'S Reference, revision A. 7*. Gaussian, Inc.
- Fukui, K. (1997). *Frontier Orbitals and Reaction Paths*. Singapore: World Scientific.
- Gangwal, R. P., Damre, M. V., & Sangamwar, A. T. (2016). Overview and recent advances in QSAR studies. In A. G. Mercader, P. R. Duchwicz & P. M. Sivakumar (Eds.), *Chemometrics Applications and Research. QSAR in Medicinal Chemistry* (pp. 1-32).: Apple Academic Press.
- Ghosal, S., Bhattacharyya, R., & Majumder, M. (2020). Impact of complete lockdown on total infection and death rates: A hierarchical cluster analysis. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 14 (4), 707-711.
- Grisoni, F., Consonni V., & Todeschini R. (2018). Computational Chemogenomics: Methods in Molecular Biology. In J. Brown (Ed.), *Impact of Molecular Descriptors on Computational Models* (pp. 171-209). Humana Press.
- He, H., Han, Na., Ji, C., Zhao, Y., Hu, S., Kong, Q., Ye, J., Ji, A., & Sun, Q. (2020). Identification of five types of forensic body fluids based on stepwise discriminant analysis. *Forensic Science International: Genetics*, 48, 102330. doi.org/10.1016/j.fsigen.2020.102337
- Hehre, W. J., Radom, L., Schleyer, P. v. R., & Pople, J. Á. (1986). *Ab Initio Molecular Theory*. Wiley.
- Holmes, P. (2015). On the Road to Elimination of Rhodensiense Human African Trypanosomiasis: First WHO Meeting of Stakeholders. *PLoS. Neglected Tropical Diseases*, 9 (4), e0003571. 10.1371/journal.pntd.0003571
- Hyperchem, Inc. (2008). *ChemPlus: Modular Extensions to HyperChem Release 8.06, Molecular Modeling for Windows*. Gainesville.
- Infometrix, Inc (2002) *Pirouette 3.01*. Woodinville.
- Jefford, C. W., Grigorov, M., Weber, J., Lüthi, H. P., & Troncher, J. M. J. (2000). Correlating the Molecular Electrostatic Potentials of Some Organic Peroxides with Their Antimalarial Activities. *Journal of Chemical Information and Computer Sciences*, 40 (2), 354-357.
- Johnson, R. A., & Wichem, D. W. (1992). *Applied Multivariate Statistical Analysis*. Prentice-Hall.
- Karelson, M., Lobanov, V. S., & Katrizky, A. R. (1996). Quantum-Chemical Descriptors in QSAR/QSPR Studies. *Chemical Reviews*, 96 (3), 1027-1042.
- Kowalski, B. R., & Brender, C. F. (1972). Pattern Recognition. A Powerful Approach to Interpreting Chemical Data. *Journal of the American Chemical Society* 94 (16), 5632-5639.
- Lee, C., Yang, W., & Parr, R.G. (1988). Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B*,

37 (2), 785–789.

Mehmood, A., Jones, S. I., Tao, P., & Janesko, B. J. (2018). An orbital-overlap complement to ligand and binding site electrostatic potential maps. *Journal of Chemical Information and Modeling*, 58 (9), 1836-1846.

Politzer, P., Laurence, P. R., & Jayasuriya, K. (1985). Molecular electrostatic potentials: an effective tool for the elucidation of biochemical phenomena. *Environmental Health Perspectives*, 61, 191-202.

Politzer, P., Murray, J. S. & Clark, T. (2019). Explicit inclusion of polarizing electric fields in σ - and π -hole interactions. *The Journal of Physical Chemistry A*, 123 (46), 10123-10130.

Politzer, P., & Murray, J. S. (2021). Electrostatic potentials at the nuclei of atoms and molecules. *Theoretical Chemistry Accounts* 140 (7). doi.org/10.1007/s00214-020-02701-0

Politzer, P. & Murray, J. S. (2021). Chemical Reactivity in Confined Systems: Theory, Modelling and Applications. In P. K. Chattaraj & D. Chakraborty (Eds.), *Molecular Electrostatic Potentials: Significance and Applications* (pp. 113-134): Wiley.

Roothaan, C. C. (1951). New developments in molecular orbital theory. *Reviews of Modern Physics*, 23 (2), 69-89.

Rzesikowska, K., Krawczuk, A., & Kalinowska-Tluscik, J. (2019). Electrostatic potential and non-covalent interactions analysis for the design of selective 5-HT7 ligands. *Journal of Molecular Graphics and Modelling*, 91, 130-139. doi.org/10.1016/j.jmgm.2019.06.007

Santos, M. A. B., Oliveira, L. F. S., Figueiredo, A. F., Gil, F. S., Farias, M. S., Bitercourt, H. R., Lobato, J. R. B., Ferreira, R. D. P., Alves, S. S. S., Aquino, E. L. C., & Ciriaco-Pinheiro, J. (2020). Molecular Electrostatic Potential and Chemometric Techniques as Tools to Design Bioactive Compounds. In A. Stefaniu, A. Rasul, & G. Hussain (Eds.), *Cheminformatics and its Applications* (pp. 1-27). London: IntechOpen.

Scrocco, E., & Tomasi, J. (1978). Electronic Molecular Structure, Reactivity and Intermolecular Forces: An Euristic Interpretation by Means of Electrostatic Molecular Potentials. *Advances in Quantum Chemistry*, 11, 115–193.

Selby, R., Wamboga, C., Erphas, O., Mugenyi, A., Jamonneau, V., Waiswa, C. Torr, S. J., & Lehane, M. (2019). Gambian human African Trypanosomiasis in North West Uganda. Are we on course for the 2020 target? *PLoS Neglected Tropical Diseases*, 13 (8), e0007550. doi. org/10.1371/journal.pntd.0007550

Singh, U. C., & Kollman, P. A. (1984). An approach to computing electrostatic charges for molecules. *Journal of Computational Chemistry*, 5 (2), 129-145.

Srikrishnan, T., De, N. C., Alam, A. S., & Kapoor, J. (2004). Crystal and molecular structure of pentamidine diisethionate: an anti-protozoal drug used in AIDS related pneumonia. *Journal of Chemical Crystallography*, 34 (11), 813-818.

Stanton, D. & Jurs, P. (1990). Development and Use of Charged Partial Surface-Area Structural Descriptors in Computer-Assisted Quantitative Structure-Property Relationship Studies. *Analytical Chemistry* 62 (21), 2323–2329.

Todeschini, R., & Consonni, V. (2009). *Molecular Descriptors for Chemoinformatics*. Wiley-VCH.

Varmuza, K. (1980). *Pattern Recognition in Chemistry*. Springer-Verlog.

Varmuza, K. (2018). Methods for multivariate data analysis. In: T. Engel, & J. Gasteiger (Eds). *Chemoinformatics - Basic Concepts and Methods* (pp. 339-437). Wiley-VCH. Weinheim.

Vidal, R., Ma, Y., & Sastry, S. S. (2016). *Generalized Principal Component Analysis*. Springer.

Williams, D. E., & Yan, J. M. (1998). Point-Charge Models for Molecules Derived from Least-Squares Fitting of the Electric Potential. *Advances in Atomic and Molecular Physics*, 23, 87-130.

World Health Organization. (2019). Human African Trypanosomiasis. http://www.who.int/trypanosomiasis_african/en/

Wu, X., Thiel, W., Pezeshki, S., & Lin, H. (2013). Specific Reaction Path Hamiltonian for Proton Transfer in Water: Reparameterized Semiempirical Models. *Journal of Chemical Theoretical and Computational*, 9 (6), 2672-2686.

Zhang, L.-X., Sun, Y., Zhao, H., Zhu, N., Sun, X.-D., Jin, X., Zou, A.-M., Mi, Y., & Xu, J.-R. (2017). A Bayesian Stepwise Discriminant Model for Predicting Risk Factors of Preterm Premature Rupture of Membranes: A Case-control Study. *Chinese Medical Journal*, 130 (20), 2416-22. 10.4103/0366-6999.216396