ISSN: 3067-3151

INTERDISCIPLINARY JOURNAL OF PHYSICS, CHEMISTRY, AND BIOLOGY

A CHEMOMETRIC STUDY ON THE ANALGESIC POTENTIAL OF OPIOIDS: IMPLICATIONS FOR CHEMISTRY EDUCATION

[1]Aisha Modupe Salihu and [2]Emmanuel Okechukwu Eze

[1]Department of Chemistry, David Umahi Federal University of Health Sciences, Uburu, Ebonyi State, Nigeria.

[2]Chemical Sciences Department, Godfrey Okoye University, Enugu State, Nigeria.

DOI: https://doi.org/10.5281/zenodo.15744492

ABSTRACT: This study explores the use of chemometric techniques to analyze opioids with varying degrees of inhibitory activity, aiming to support chemistry education while identifying potential analgesic compounds. The crystallographic structure of mitragynine was retrieved from the literature and fully optimized using multiple computational methods and basis sets. Through Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA), the B3LYP/6-31G** method was identified as the most suitable for further computational investigation.

The optimized structure of mitragynine served as the basis for constructing 3D models of nine additional opioids. These compounds underwent full geometric optimizations and molecular property calculations at their lowest energy conformations. The molecular descriptors derived were subjected to Exploratory Data Analysis using PCA and HCA, effectively grouping the ten opioids (training set) into two categories: the most active opioids (MAO – compounds 1, 2, 5, 6, 9, and 10) and the least active opioids (LAO – compounds 3, 4, 7, and 8).

To validate these groupings, classification methods including the K-Nearest Neighbor (KNN) method and Stepwise Discriminant Analysis (SDA) were applied, confirming the distinct separation of the MAO and LAO classes. Based on these findings and chemical reasoning, nine new mitragynine-derived opioids were proposed. Evaluation through the established chemometric models indicated that six of these novel compounds hold significant analgesic potential, making them candidates for synthesis and biological testing.

This approach demonstrates the power of combining computational chemistry and chemo metrics in drug discovery and educational settings, offering an accessible platform for students while contributing to the search for effective analgesics.

Keywords: Chemo metric Analysis, Opioid Activity, Mitragynine Derivatives, Molecular Modeling, Computational Chemistry

1. INTRODUCTION

The word pain comes from the Greek word *poine*, meaning penalty. Physiologists distinguish pain from nociception. Nociception refers to signals arriving in the central nervous system resulting from the activation of specialized sensory receptors called nociceptors, which provide information about tissue damage. Pain is the unpleasant emotional experience that usually accompanies nociception [1].

The term opioids were proposed by Acheson to designate drugs with an action similar to that of morphine, but with a different chemical structure. However, the concept of opioids has evolved and now includes all natural, semi-synthetic or synthetic substances that react with opioid receptors, either as agonists or antagonists [2].

Morphine and mitragynine are opioids with analgesic properties [3, 4]. Morphine is a substance extracted from the green seed pods of the opium poppy [5]. Mitragynine is extracted from Mitragyna speciosa [3]. Figure 1 shows the 2D-structures of morphine (a) and mitragynine (b). As can be seen in this figure, mitragynine has structural aspects similar to morphine, both have three functional groups, extremely important in analgesic activity, namely: the anionic site, the phenolic site and the hydrophobic region.

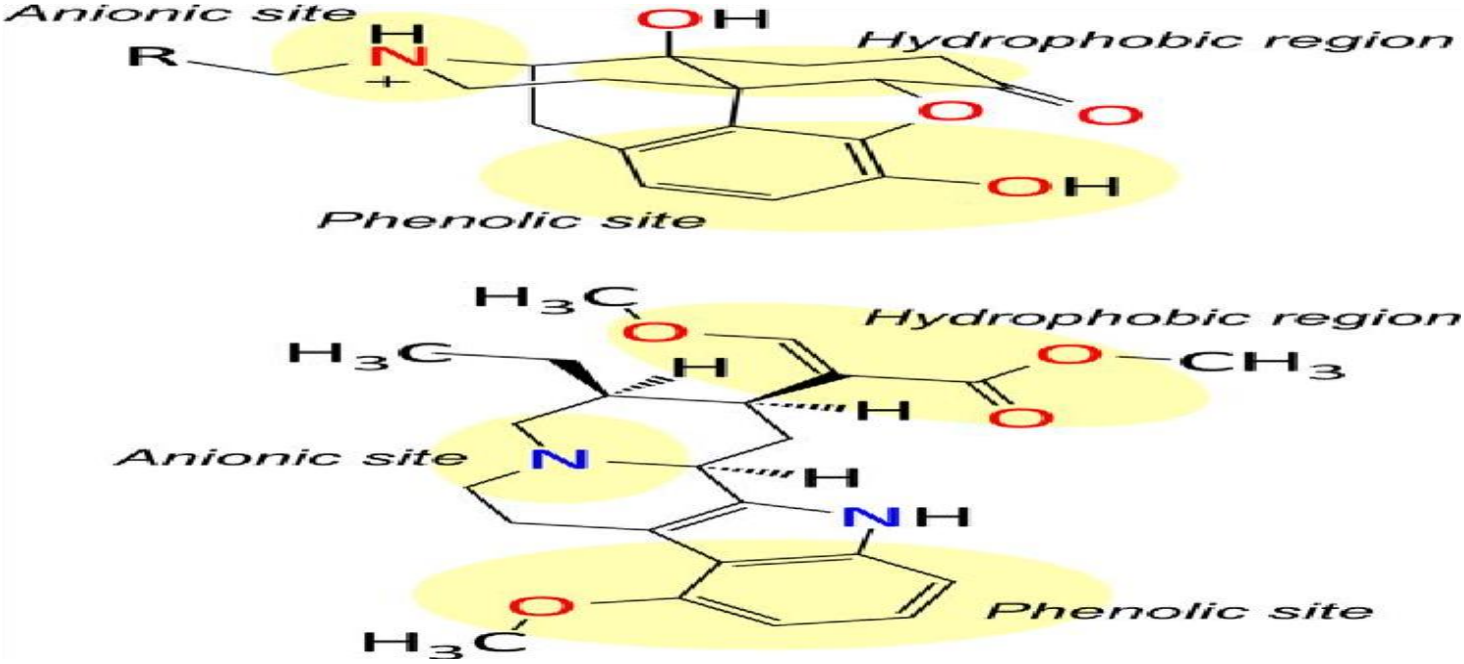


Figure 1. 2D structure of mitragynine and morphine with highlighting, color yellow, to the functional groups: anionic site, phenolic site, and hydrophobic region, respectively, important in analgesic activity.

The discipline of Chemometrics combines mathematical, statistical, information theory and computer science knowledge to handle complex chemical data; its techniques were introduced in chemistry in the first half of the 1970s to unravel various types of spectroscopic data, and since then they have been excellent tools to aid in the interpretation of chemical data and obtain relevant information in different areas of application of chemical science. These techniques are especially useful for classifying objects into discrete classes based on their experimentally measured characteristics. A set of features of an object is considered an abstract pattern that contains information about a non-directly measured property of the object. For a detailed examination of this matter see references [6 - 10].

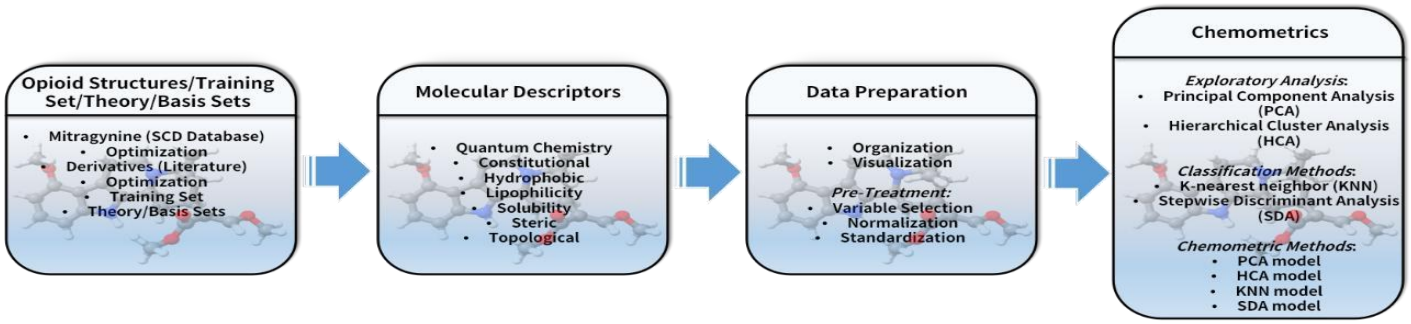


Figure 2. Schematic representation of the chemometric experiment

In this work, a chemometric experiment, Figure 2, involving opioids (mitragynine and derivatives) with different activities [3, 4] is reported. Exploratory data analysis techniques (Principal Component Analysis [10-13] and Hierarchical Cluster Analysis [10 - 13]) and classification methods (K-Nearest Neighbor Method [10 - 13] and Stepwise Discriminant Analysis [10 - 13]) were used as tools to develop the referred experiment, which can be reproduced at all levels of Chemistry Education, serving as a stimulating and motivating element for the appropriation of relevant techniques in the treatment of various aspects of systems of chemical interest.

2. EXPERIMENTAL PROCEDURES

2.1 Opioid Structures/Training Set/Theory/Basis Set

According to the experiment scheme, Figure 2, anchored in Table 1 [10-13], the 3D-structure of mitragynine, Figure 1a, was taken from the Cambridge Structure Database (CSD), Code REFCODE:BOTYUF [14], and completely optimized with different Theories/Basis Sets to establish the most suitable theory/basis set for the development of the experiment: semiempirical theories:
AM1 [15] and PM3 [16]; Hartree-Fock (HF) theory: HF/3-21G (HF321G); HF/6-31G (HF631G); HF-6-31G* (HF631G1); HF/6-31G** (HF631G2) [17, 18]; HF CEP-31G (HFCEP31G) [17, 19]; and HF/STO-3G (HFSTO3G) [17, 18]; density functional theory (DFT): BLYP/STO-3G
(BLYPSTO3G); BLYP/3-21G (BLYP321G); BLYP/6-31G (BLYP631G); BLYP/631G* (BLYP631G1); BLYP/6-31G** (BLYP631G2); BLYP/CEP-31G (BLYPCEP31G) [19-21]; B3LYP/STO-3G (B3LYPSTO3G); B3LYP/3-21G (B3LYP321G); B3LYP/6-31G
(B3LYP631G); B3LYP/6-31G* (B3LYP/6-31G1); B3LYP/6-31G** (B3LYP631G2) [18, 20, 21]; and B3LYP CEP-31G (B3LYPCEP31G) [19-21].

Table 1. Topic and Subtopics anchoring the Chemometric experiment planned to serve Chemistry Education students

Topics	Subtopics
Introduction to Chemometrics [10-13]	Data Preparation for Analysis: Organization; Visualization; Pretreatment: Variable Selection; Normatization; Standardization. Exploratory Analysis: Principal Components Analysis (PCA) and Hierarchical Cluster Analysis (HCA); Classification Methods: Knearest neighbor (KNN) Method; Stepwise Discriminant Analysis (SDA) Method

The calculated and experimental structural parameters, bond lengths, bond angles and dihedral angles of mitragynine in the most stable conformation were used in the construction of a data matrix of dimension 66 x 21 (66 rows corresponding to the structural parameters and 21 columns constructed with the twenty (20) theories and the Experimental (EXP). The resulting matrix was subjected to scrutiny with Exploratory Analysis [10-13]: Principal Components Analysis (PCA) and Hierarchical Cluster Analysis (HCA). The use of these techniques allows to simultaneously analyze data in a multidimensional space instead of point-by-point comparison. The result of this procedure indicated the theory/basis set HF/6-31G** (HF631G2) as the most suitable to develop the experiment (see Section *3.1 Selection of Theory/Basis Set used in the Experiment, in 3. Results and Discussion*).

The 3D-structure of mitragynine obtained with HF/6-31G** (HF631G2) in the most stable conformation allowed the construction of the training set of ten (10) opioids (Figure 3), whose structures were subjected to full optimization, to provide the lowest energy conformations, to be used in the next stage of the experiment. The numbering used in the opioids is shown in this figure, mitragynine (1).

The structure optimization calculations were developed with the Gaussian 98 software package [22]. The construction and visualization of the training set structures were performed with the HyperChem 8.06 [23] and Molden 6.3 [24] programs, respectively. The data matrix was explored with the Pirouette 3.10 program [25].

2.2 Molecular Descriptors

To provide information on the influence of steric, electronic, hydrophobic, and hydrophilic characteristics on opioid analgesia, several molecular properties were calculated and computed as molecular descriptors: steric properties (bond lengths, bond angles, and torsion angles); electronic properties (total energy (ET), HOMO energy, HOMO-1 energy, HOMO-2 energy, HOMO-3 energy,

LUMO energy, LUMO+1 energy, LUMO+2 energy, LUMO+3 energy, Mulliken's electronegativity (c = [HOMO energy – LUMO energy]/2), molecular hardness (h = I – AE/2), where I is ionization potential and AE electron affinity, molecular softness (1/h), defined as de inverse of molecular hardness; dipole moment (m), GAP energy = HOMO energy – LUMO energy), and atomic charge on an Nth atom (qN)); and physicochemical properties (molecular polarizability (POL), molecular refractivity (MR), hydration energy (HE), octanol-water partition coefficient (log P), molecular mass

(MM), total surface area (TSA), and molecular volume (VOL)). In addition, holistic properties were calculated and computed as descriptors to represent different sources of chemical information about size, symmetry, and distribution of atoms in the molecule.

The molecular properties sources of the descriptors were calculated with the programs

Gaussian 98 [22]/HF/6-31G** (HF631G2) [17, 18], HyperChem 8.06 [23], Dragon [26-28].

2.3 Data Preparation

In data preparation, the Pirouette 3.10 [28] and Minitab 16 [29] programs were used as follows: Pirouette [28] program (Organization and Visualization; and Pre-Treatment) Minitab [29] program (Variable Selection); Pirouette [28] program (Normalization and Standardization). In this procedure, the starting point was the construction of a data matrix with 10 rows (compounds) versus

216 columns (molecular descriptors) and the selection of variables was performed considering the Person coefficient r. Thus, when two descriptors presented a Person correlation coefficient $r < 6$ and, consequently, a lower correlation between them, one of them was excluded from the data matrix at random, considering that theoretically both describe the same property, and this formation produced a compression of the data.

2.4 Chemometrics

The compressed data were subjected to chemometric scrutiny as follows:

2.4.1 Exploratory Data Analysis

2.4.1. 1 Principal Components Analysis (PCA)

The PCA [10–13, 30] is widely used to simplify large data sets in a way that patterns and relationships can be readily recognized and understood. The underlying purpose of the technique is the dimension reduction.

The method generates a new set of variables called principal components, PCs, as linear combination of all the initial variables so that the first new variable, PC1, describes the largest variance in the data set, the second new

variable, PC2, must be chosen orthogonally (uncorrelated) to the first one and in the direction to describe as much variance left as possible and so on.

The initial data matrix, represented by **X**, is decomposed into two matrices, **T** and **P** where

$$\mathbf{X} = \mathbf{T}\,\mathbf{P}^{T} \quad (1)$$

and **T**, known as "scores" matrix, represents the position of the samples in the new coordinate system. The second matrix, **P**, is the "loadings" matrix and describes how the new axis, i.e. the PCs, are built from the original variables. The samples are mapped through scores and the variables by the loadings in the new low dimensional vector space defined by the principal components.

2.4.1.2 Hierarchical Cluster Analysis (HCA)

The HCA [10-13, 30] has become, together with principal components, another important tool in exploratory data analysis. Its primary purpose 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 in the form of dendograms. In the HCA, the distances between samples or variables are calculated and compared through the similarity index which ranges from zero, i.e. no similarity and large distance among samples, to one for identical samples.

2.4.2 Data Classification Method

2.4.2.1 The K-Nearest Neighbor (KNN) Method

The KNN method [10-13,30] classifies the objects based on distance comparison among them. The multivariate Euclidean distances between every pair of samples with known class membership is calculated. The closest K samples are used to build the model. The optimal K is determined by cross validation applied to the training set samples. The classification of a test samples is determined based on the multivariate distance of this sample with respect to the K samples in the training set.

2.4.2.3. Stepwise Discriminant Analysis (SDA)

The SDA is also a classification method that tries to maximize the probability of correct allocation [7, 10-13, 31]. This method has two main goals, which are to separate objects from distinct populations and to allocate new objects into previously defined populations. A stepwise procedure is used in this method that is, at each step, the most powerful variable is an input into the discriminant function. The criterion function for selecting the next variable depends on the number of specified groups. The SDA is an approximation based on the F test for significance of variables. At each stage, a variable is selected based on its significance and, after several stages, the most significant variables are extracted from the entire set subject to analysis [7, 10-13, 31].

3. RESULTS AND DISCUSSION

3.1 Selection of Theory/Basis Set used in the Experiment

Table 2 shows the geometric parameters used in the selection of the theory/basis set used in the experiment and the correlation matrix between them. These parameters, according to the numbering used for the opioids, mitragynine (1), Figure 3, are the dihedral angles C2N1C13C8, N2C3C2N1, C7C7C3N2, and N1C13C8C7, respectively. As one can see, the correlation between these parameters is less than 0.556. The separation was made in PC1, with three (3) principal components: PC1, PC2, and PC3 and explains 97% of the total information. Table 3 shows the contribution of the geometric parameters to the 3 PCs; and the corresponding variances are PC1 = 54.5, PC2 = 29.2, and PC3 = 13.5%, respectively.

Figure 4 shows the scores plot (PC1 versus PC2) for the theoretical and experimental methods used in the selection of the theory/basis set. According to this figure, the separation between semiempirical methods on the

right of the graph and other methods on the left of the same graph is made in PC1, with HF/6-31G* (HF631G1) and HF/6-31G** (HF631G2) being closer to the experimental value, therefore proving to be more adequate in reproducing the experimental data.

Figure 5 shows the dendrogram obtained with the HCA for the theoretical and experimental methods used in the selection of the theory/basis set. In this figure, it is possible to notice the presence of two clusters, one formed by the semi-empirical methods and the other constituted by the other methods, confirming the separation by the PCA. However, as can be seen by one, the HF/6-31G** method (HF631G2) presents greater similarity to the experimental data, which distinguishes it in relation to the HF/6-31G* method (HF631G1), leading to its selection to be used in the experiment.

Table 2. Geometric parameters responsible for selecting the theory/basis set used in the experiment and correlation matrix between them

Theory/Basis set	Geometric parameters[1]			
C2N1C13C8	N2C3C2N1	C7C2C3N2	N1C13C8C7	
HF/STO-3G	-1.4777	-0.24636	0.12623	0.74278
HF/3-21G	-1.4777	-0.24636	0.12623	0.74278
HF/6-31G	-0.0035	-0.26731	0.27170	0.60413
HF/6-31G*	1.7656	-0.25684	0.40100	-0.59752
HF/6-31G**	0.9547	-0.25422	0.06157	-1.2446
HF/CEP-31G	-0.0772	-0.20185	0.54648	-0.68996
BLYP/STO-3G	-0.0772	-0.20185	0.54648	-0.68996
BLYP/3-21G	-0.0772	-0.20185	0.40100	-0.68996
BLYP/6-31G	1.7656	-0.25946	0.04542	-0.59752
BLYP/6-31G*	0.21762	-0.58150	-0.05157	-0.89793
BLYP/6-31G**	-0.0035	-0.25160	0.74044	-0.99037
BLYP/CEP-31G	-0.0772	-0.19923	0.28786	-0.68996
B3LYP/STO-3G	-0.8881	-0.20447	0.28786	-0.43576
B3LYP/3-21G	-0.8881	-0.20447	0.20705	-0.43576
B3LYP/6-31G	0.8810	-0.25946	0.20705	-0.34333
B3LYP/6-31G*	0.0702	0.05473	-3.8661	2.6608
B3LYP/6-31G**	1.6919	4.3303	-1.5548	2.0137
B3LYPCEP-31G	-1.4777	-0.22804	0.06157	0.85831
AM1	-1.4777	-0.24636	0.12623	0.74278
PM3	-1.4777	-0.24636	0.12623	0.74278
Geometric parameter				
C2N1C13C8		0.096	0.415	0.555
N2C3C2N1			0.072	0.016
C7C2C3N2				0.000

1degree.

3.2 Applying Exploratory Analysis to Training Set Data

3.2.1 PCA

After data compression, the PCA was employed and the opioids were separated, with the help of three (03) variables, into two classes, according to their analgesia: MAO class (1, 2, 5, 6, 9, and 10) and LAO class (3, 4, 7, and 8). Table 4 shows the variables: LUMO+1 energy, CENT (Centralization), and EPS0 (Measure of electronegativity ETA) responsible for the separation of the opioids from the training set and the correlation matrix between these variables. As one can see in this table, the correlation between the variables is less than 0.603. The separation was made into PC1, with three (03) principal components, PC1, PC2, and PC3, and explains 100% of the total information, with PC1 and PC2 retaining 78.6 and 21.0%, respectively, of this information. According to Table 4, in general, MAO class opioids exhibit high LUMO+1 energy values combined with low CENT and EPS0 values.

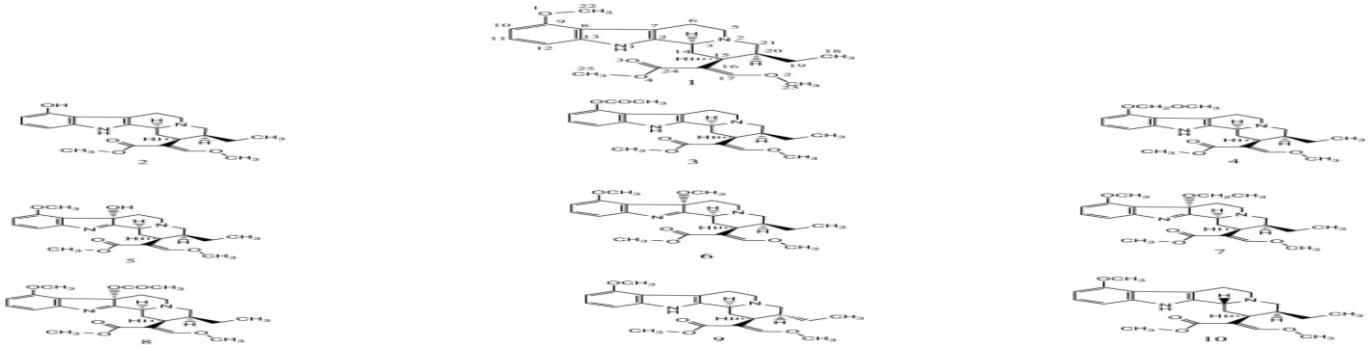


Figure 3. Opioids studied (training set), mitragynine (1) reports the atomic numbering used in the experiment

Table 3. Contribution of geometric parameters in the three (3) Principal Components

Variable	PC1	PC2	PC3
C2N1C13C8	0.184	0.850	0.343
N2C3C2N1	0.519	0.310	-0.734
C7C2C3N2	-0.590	0.144	-0.585
N1C13C8C7	0.590	-0.400	-0.050
Variance (%)	54.5	29.2	13.5
Cumulative (%)	54.5	83.7	97.2

Figure 6 shows the scores plot obtained with the PCA for the opioids in the training set. In this figure, the MAO class opioids are located on the left side due to the displacement in that direction produced by the LUMO+1 energy property (Figure 7), while the LAO class opioids are located on the right side of the same figure due to the action of the CENT and EPS0 properties, Figure 7.

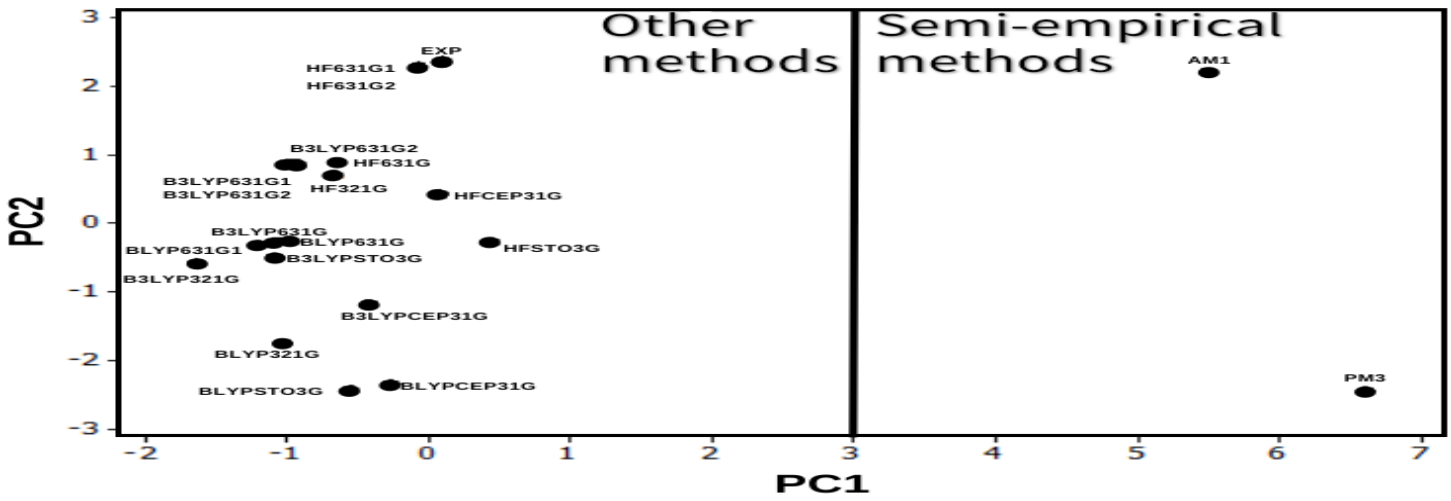


Figure 4. Score plot (PC1 versus PC2) for the theoretical and experimental methods used in theory/basis set selection

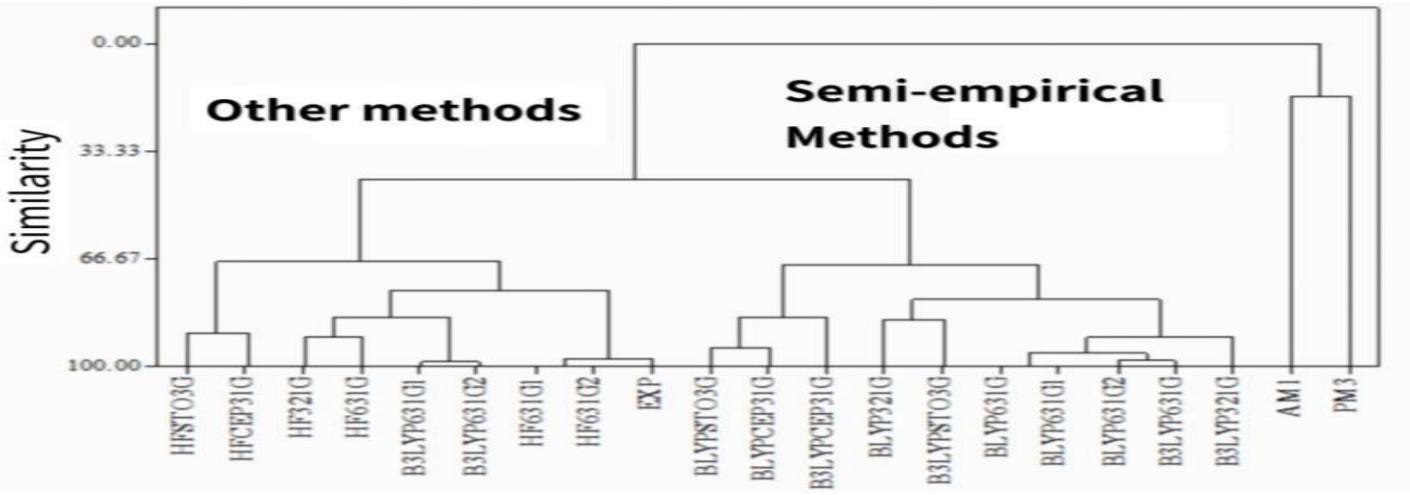


Figure 5. Dendrogram obtained with HCA for the theoretical and experimental methods used in the selection of the theory/basis set

Table 5 shows the contribution of the variables to the three (3) principal components: PC1, PC2, and PC3, with the respective variances, PC1 = 78.6, PC2 = 21.0, and PC3 = 0.40%. From this table, one can extract the following equation for PC1:

PC1 = -0.463 (LUMO+1 energy) + 0.627 (CENT) + 0.627 (EPS0) (2)

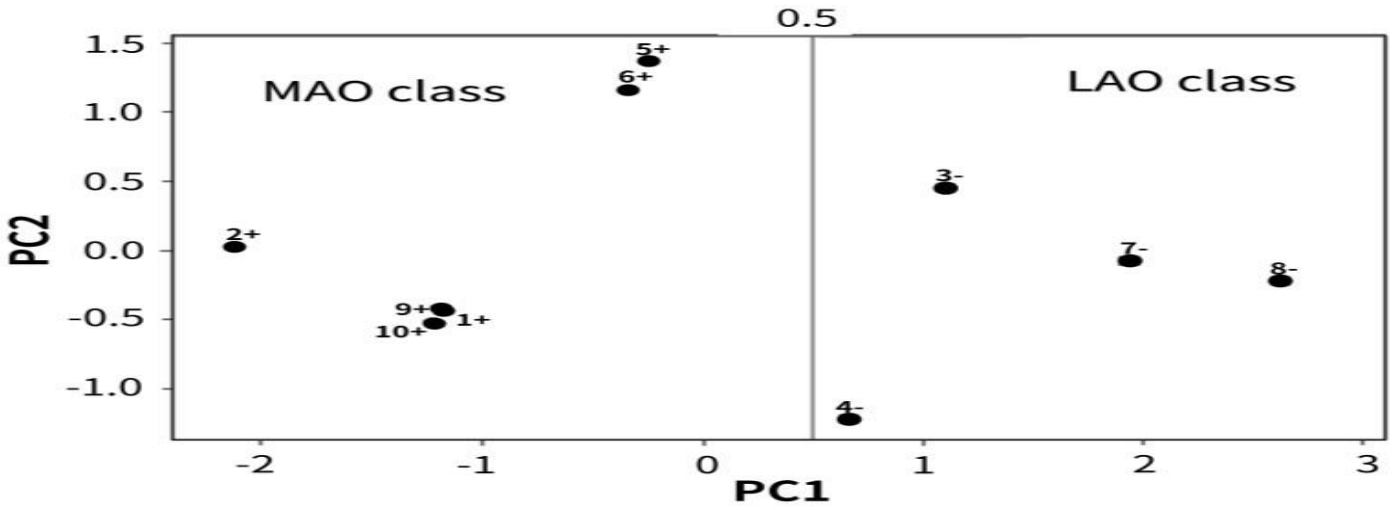


Figure 6. Score plot (PC1 versus PC2) for the separation of the opioids into two classes: MAO class and LAO class

Table 4. Variables responsible for separating of the opioids into two classes: most active opioids (MAO) and less active opioids (LAO), relative inhibitory activity (morphine) in % (RIA), and correlation matrix between descriptors

Opioid	LUMO+1 energy (kcal/mol)	CENT	EPS0	RIAa, b, c, d	Activity label
1+(mitragynine)	97.295	1230	20.1	95a, b	MAO
2+	97.40	1100	19.2	57a, b	MAO
3-	88.77	1510	21.1	38a, b	LAO
4-	96.58	1540	21.5	Inc	LAO
5+	87.89	1230	20.1	99a, b	MAO
6+	88.97	1230	20.1	70a, b	MAO
7-	89.13	1600	22.1	26a, b	LAO
8-	88.28	1740	22.4	15a, b	LAO
9+	97.74	1230	20.1	101a	MAO
10+	97.78	1230	20.1	85a	MAO
morphine				100d	
Descriptor					
EPS0	-0.501	-0.602			
CENT	-0.502				

a[3]. b[4]. cInactive. dRef. RIA □ 38 (MAO) and RIA □ 38 (LAO).

Table 5. Contribution of variables in the three (3) Principal Components

Variable	PC1	PC2	PC3
LUMO+1 energy	-0.463	0.886	0.0013
CENT	0.627	0.326	0.707
EPS0	0.627	0.328	-0.707
Variance (%)	78.6	21.0	0.40
Cumulative (%)	78.6	99.6	100.0

Equation (2) confirms the information previously extracted from Table 4, that is, new opioids derived from mitragynine can be classified as MAO by combining high energy values of LUMO+1 with low values of CENT and EPS0, respectively.

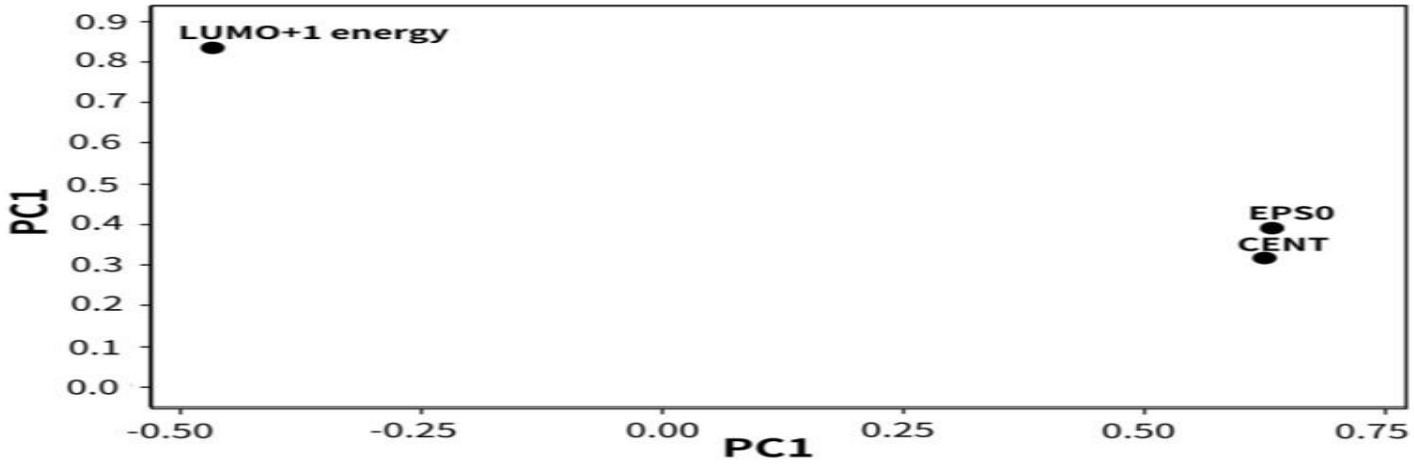


Figure 7. Loadings plot (PC1 versus PC2) for the variables responsible for separating of the opioids into two classes: MAO class and LAO class

3.2.2 HCA

Figure 8 shows the dendrogram obtained with the the HCA, incremental method, for the opioids of the training set. The variables in the HCA were the same as those selected in the PCA exploration, namely: LUMO+1 energy, CENT, and EPS0. One can see in this figure that the ten (10) opioids in the training set are also distributed into two classes, according to their analgesia: MAO class and LAO class. The MAO class consists of two clusters containing opioids 1, 2, 5, 6, 9, and 10; and the LAO class presents a cluster formed by opioids 3, 4, 7, and 8. For better interpretation of the dendrogram, the structures of the MAO and LAO classes will be detailed separately.

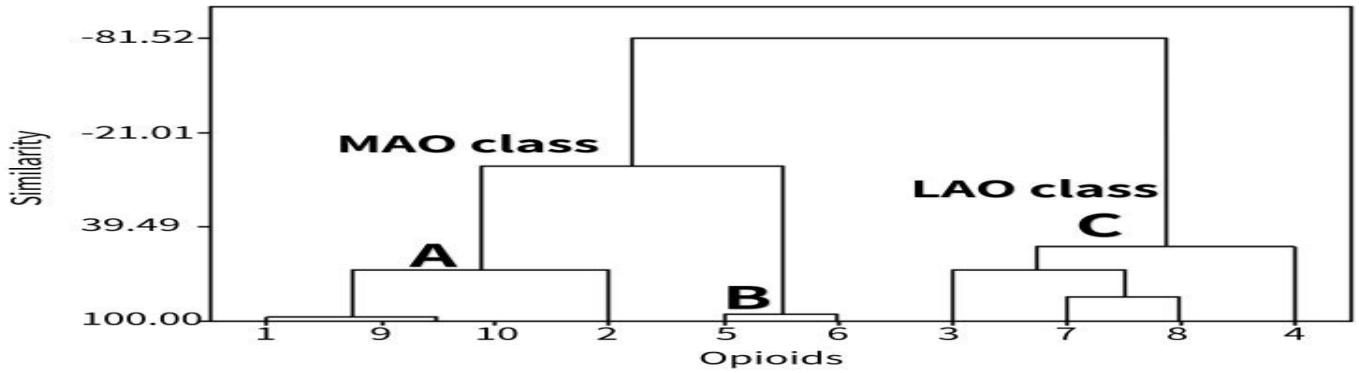


Figure 8. Dendrogram obtained by the HCA method for separating opioids into two classes: MAO class and LAO class

3.2.2.1 MAO class

Clusters A and B appear in this class. Figure 8a shows Cluster A. This cluster groups compounds 1, 2, 9, and 10. Opioids 1, 9, and 10 present the methoxyl radical (-OCH3) at C9; the difference between opioids 1 (mitragynine) and opioids 9 and 10 lies in the following aspect: opioid 1 (mitragynine) presents the H atom at C3 and the methyl radical (-CH3) at C20 behind and in front of the planes containing C3 and C20, respectively, whereas in opioid 9 the methyl radical (CH3) is oriented behind the plane of C20. In opioid 10, the H atom at C3, unlike in opioid 1 (mitragynine), is oriented behind the plane containing C3. Opioid 2 differs from mitragynine (1) by the replacement of the methoxyl radical (-OCH3) by the hydroxyl (-OH) at C9.

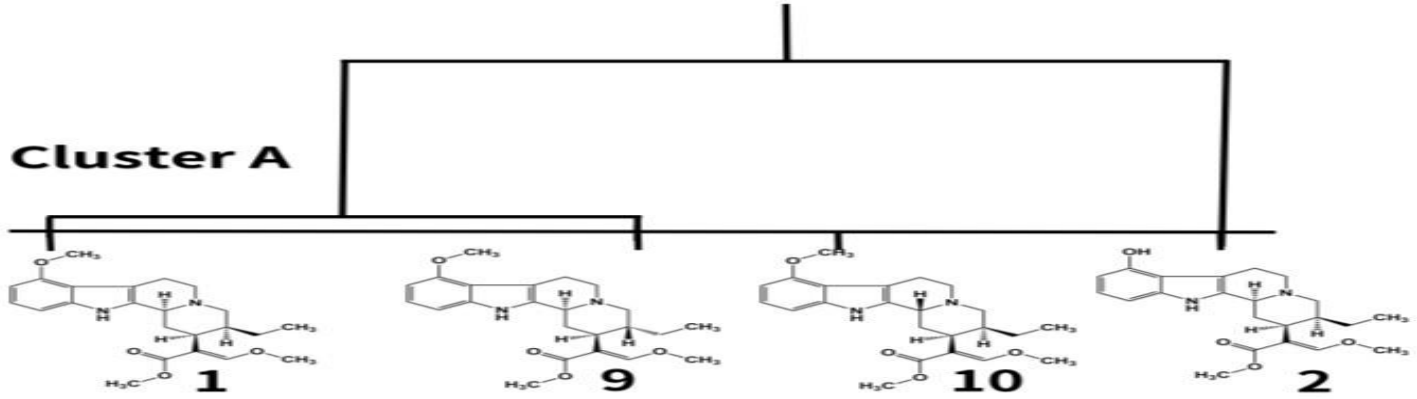


Figure 8a. Cluster A

Cluster B (Figure 8b) consists of opioids 5 and 6. These opioids differ from mitragynine (1) by the presence of hydroxyl (-OH) and methoxyl radical (-OCH3) at the C7 position, respectively.

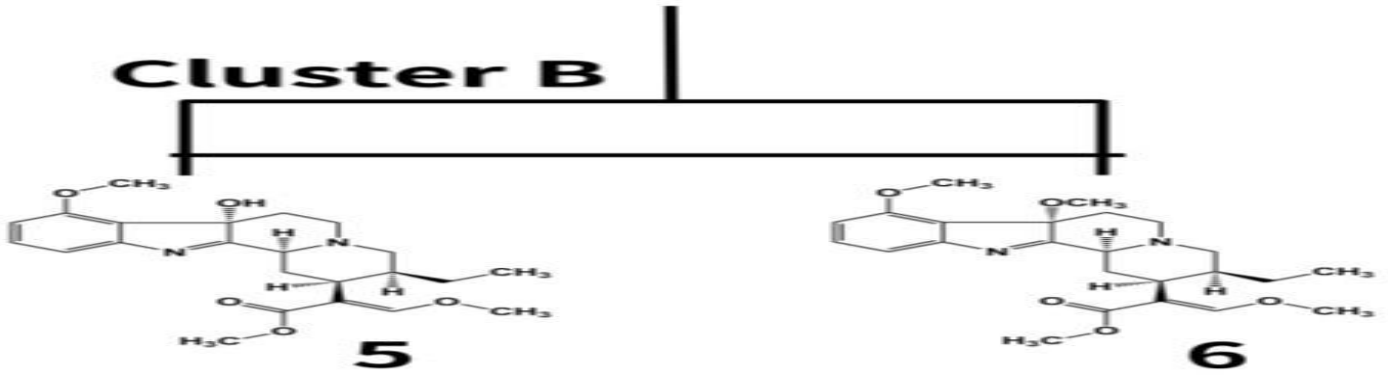


Figure 8b. Cluster B

3.2.2.2 LAO class

Cluster C is shown in Figure 8c and consists of opioids 3, 4, 7, and 8. When compared to mitragynine (1), opioids 3 and 4 have –OCOCH3 and –OCH2OCH3 radicals at the C9 position, respectively, whereas opioids 7 and 8 have the loss of the H atom at N1 and the introduction of – OCH2CH3 and –OCOCH3 radicals at the C7 position, respectively.

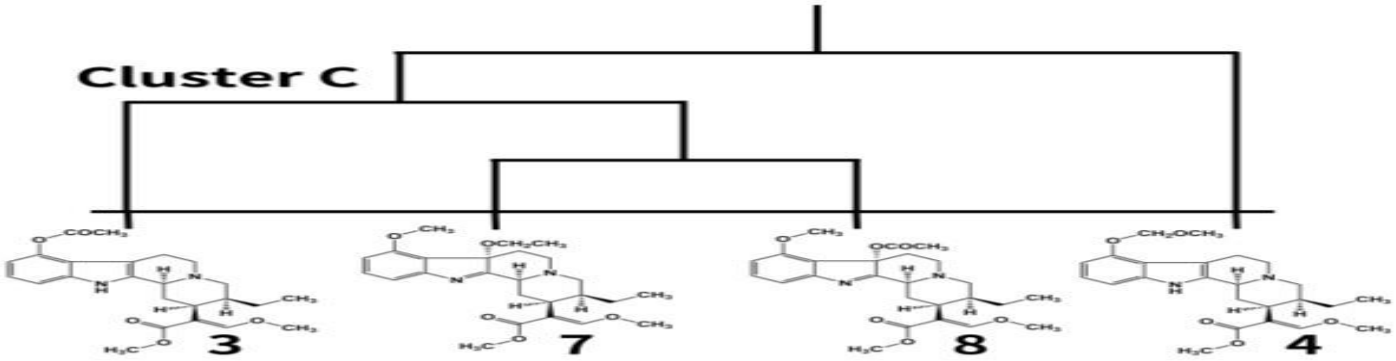


Figure 8c. Cluster C

As can be seen from the detailed class structures: MAO class and LAO class, the HCA investigation confirms the PCA for opioids.

3.3 Applying Classification Methods to Training Set Data

3.3.1 KNN Method

The selected variables are the same as those evidenced in the exploratory analysis of the opioid data from the training set, i.e.: LUMO+1 energy, CENT, and EPS0. Table 6 shows the results of the classification by the KNN method, with 100% correct information for 1KNN, 2KNN, and 4KNN, i.e.: all opioids were correctly classified into the predicted classes, indicating that under these conditions the classification obtained with the variables for the 4KNN model [10] provides good predictive capacity; therefore, the model was considered with four nearest neighbors.

3.3.2 SDA

By classifying the opioids in the training set with SDA, the discriminant functions generated with the variables LUMO+1 energy, CENT, and EPS0 are given by Equations (2a) and (2b) below:

MAO class = −0.461 (LUMO+1 energy) – 15.4 (CENT) + 7.71 (EPS0) – 2.85 (2a)
LAO class = 0.691 (LUMO+1 energy) + 23.1 (CENT) – 11.6 (EPS0) – 6.40 (2b)
Table 7 shows the classification matrix for the opioids in the training set obtained with the discriminant functions (2a) and (2b) with the correct percentage of information in the MAO and LAO classes of 100%.
Table 6. Classification of mitragynine and derivatives in MAO class and LAO class obtained by the KNN method

Class	Number of opioids	Incorrectly classified opioids				
		1KNN	2KNN	3KNN	4KNN	5KNN
MAO	6	0	0	0	0	0
LAO	4	0	0	1	0	1
Total	10	0	0	1	0	1
% Correct information		100	100	90	100	90

Table 7. Mitargynine and derivatives classification matrix in MAO class and LAO class obtained by the SDA method

Class	Number of opioids	Correct classification	
		MAO class	LAO class
MAO	6	6	0
LAO	4	0	4
TOTAL	10	6	4
% Correct information		100	100

The reliability of the SDA was assessed through the leave-one-out cross-validation test, which consist of omitting an opioid from de date set and constructing the discriminant functions with the other samples. Then, this omitted opioid was classified with the generated discriminant functions, and the procedure was repeat until the last opioid in the training set was omitted. Table 8 shows the SDA classification matrix using the cross-validation procedure.
Table 8. Matrix of classification of opioids in MAO class and LAO class obtained by the SDA method with cross validation

Class	Number of opioids	Correct classification	
		MAO class	LAO class
MAO	6	6	0
LAO	4	0	4

TOTAL	10	6	4
% Correct information	100	100	

With the model SDA constructed, the allocation rule for news opioids derivatives of mitragynine MOA e LAO investigated was established: (a) initially calculate, for the new opioids, the values of properties used in the construction of the model, (b) consider the auto scaled values of the properties in the discriminant functions, Equations (2a) e (2b), and (c) verify which discriminant function has the highest value. The new opioid derivative of mitragynine will be classified as MAO if the highest value obtained corresponds to the MAO class and vice-versa.

The chemometric experiment with the opioids from the training set showed that LUMO+1 energy, CENT, and EPS0 are the most important properties to describe their analgesic activities. For this reason, it becomes relevant to reflect on the behavior of MAO class opioids in a biological process:

(a) Molecular orbital energies are particularly useful descriptors when an ionic charge or charge transfer reaction is part of the ligand-receptor interaction [32]. Table 4 shows that MAO class opioids generally have higher values for the LUMO+1 energy property, while for LAO class opioids this value is generally lower. This may indicate that biological processes occur that allow electronic interaction between MAO class opioids and the biological receptor MOR [33], producing an increase in analgesic activity. For the hypothesis of ligand-MOR complex formation, one can verify a reaction with electron transfer from the highest occupied molecular orbital of the MOR protein to the LUMO+1 of MAO class opioids.

(b) the CENT property (centralization) is a tropological descriptor that derives from the hydrogendepleted molecular graph and summarizes in matrix form the tropological distance between all pairs of different hydrogen atoms in a molecule [32]. Furthermore, the CENT property quantifies the degree of molecular compaction to distinguish between molecular structures organized differently in relation to the centers. In Table 4, the opioids with lower values for the CENT property belong to the MAO class, when compared to the opioids of the LAO class. This may indicate that biological processes occur through the topological interaction between the opioids of the MAO class and the biological receptor - MOR, producing an increase in analgesic activity.

(c) the EPS0 property belongs to the class of descriptors called ETA electronegativity measures. According to Todeschin and Consonni [32], the ETA indices (Extended Topochemical Atom indices) are local and invariant vertices of graphs, defined in the theory of the mobile environment of valence electrons, in which the hydrogen-depleted molecular graph is considered to be composed of a nucleus and an electron valence environment. In Table 4, the MAO class opioids have lower EPS0 values than those of the LAO class. This may indicate that biological processes occur through the interaction between the MAO class opioids and the biological receptor-MOR with a decrease in the electronegativity of the opioids (ability to receive electrons from the MOR protein) and an increase in the analgesic activity.

3.4 Application of the Chemometric Models Obtained in Opioids from the Prediction Set

The insights gained from the chemometric experiment and chemical intuition allowed us to propose nine (9) new opioids derived from mitragynine (prediction set), which were evaluated to test the performance of the obtained models. Figure 9 shows the 2D structures of the opioids in the prediction set, and Table 9 shows the results of applying the PCA, HCA, KNN, and SDA models to this set. In this table, as one can see, the new opioids 13, 15-

19 are classified by the four (4) models as MAO, while opioid 11 is classified as MAO by the PCA, HCA, and KNN models and LAO by the SDA model. Opioid 12 is classified as MAO by the HCA and KNN models and LAO by the PCA and SDA models, while opioid 14 is classified as LAO by all models. According to Table 9, new opioids derived from mitragynine with analgesic potential can be proposed for synthesis and biological testing 13, 15-19. Table 10 shows the LUMO+1 energy, CENT and EPS0 variables for the opioids in the prediction set.

Table 9. Results of applying the PCA, HCA, KNN, and SDA models to the opioids of the prediction set

Opioids	PCA model	HCA model	KNN model	SDA model
11	MAO	MAO	MAO	LAO
12	LAO	MAO	MAO	LAO
13	MAO	MAO	MAO	MAO
14	LAO	LAO	LAO	LAO
15	MAO	MAO	MAO	MAO
16	MAO	MAO	MAO	MAO
17	MAO	MAO	MAO	MAO
18	MAO	MAO	MAO	MAO
19	MAO	MAO	MAO	MAO

Most active opioids (MAO) and less active opioids (LAO).

Table 10. Variables for the opioids of the prediction set

Opioids	LUMO+1 energy (kcal/mol)	CENT	EPS0
11	82.71	1104.0	19.25
12	98.68	1104.0	19.25
13	87.02	1332.0	20.51
14	87.85	1202.0	19.64
15	86.10	1202.0	19.64
16	74.24	1202.0	19.64
17	92.11	1378.0	20.83
18	85.96	1104.0	19.25
19	80.92	1233.0	20.12

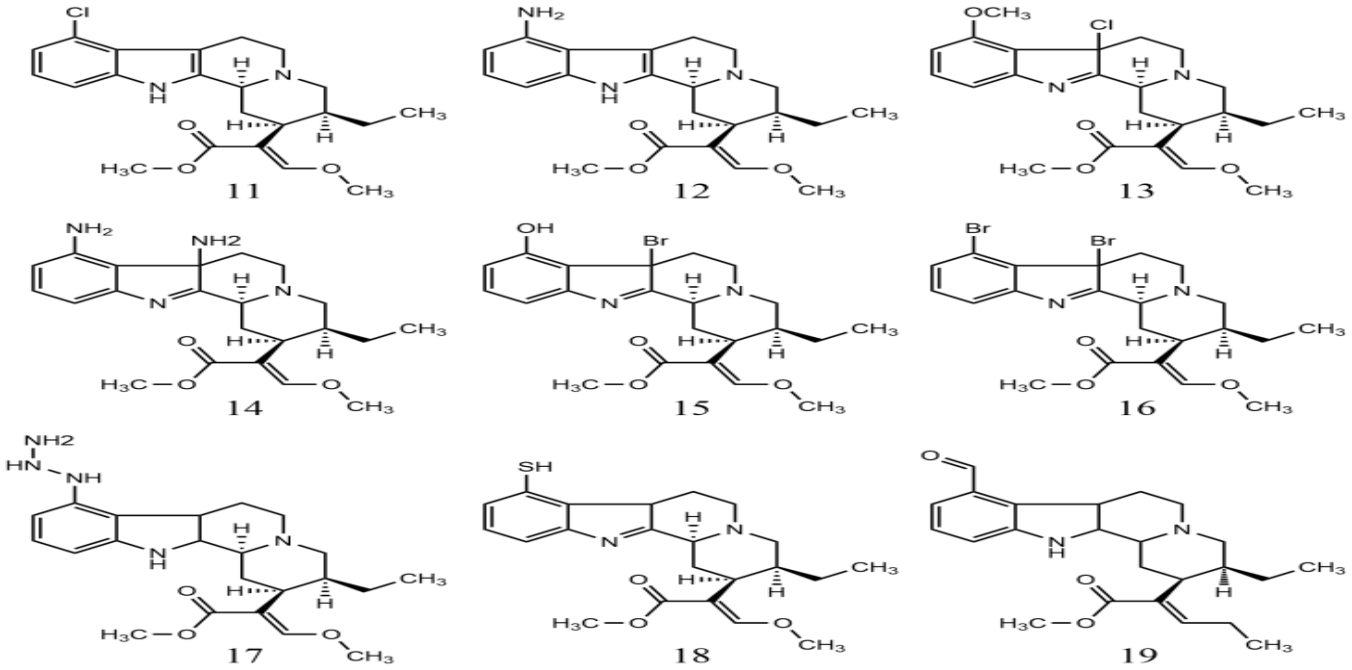


Figure 9. 2D structures for the opioids in the prediction set

4 CONCLUDING REMARKS

The chemometric experiment involving opioids with analgesic potential enabled the construction of models – PCA, HCA, KNN, and SDA – to explore and classify them into MAO and LAO classes, according to their analgesia. The properties LUMO+1 energy, CENT, and EPS0 are responsible for the exploration and classification, respectively, of the investigated opioids into the MAO and LAO classes.

It is interesting to note that these properties represent two distinct classes of interaction between mitragynine and derivatives and the biological receptor-MOR: electronic (LUMO+1 energy and EPS0) and topological (CENT). In addition, the insights accumulated in the chemometric experiment and chemical intuition allowed us to propose nine (9) new opioids derived from mitragynine; and the application of the models to these opioids suggested six (6) of them with analgesic potential for synthesis and biological testing.

Finally, the experiment can be reproduced at any level of Chemistry Education.

Acknowledgements

The authors gratefully acknowledge the LQTC-UFPA for computational support, the Swiss Center for Scientific Computing, the CAOS/CAMM center, The Netherlands, and the OCHEM platform for the use of MOLEKEL and MODEN programs, and for the generation of descriptors, respectively. The text of this manuscript was produced in collaboration with the IASA.

Conflicts of interest

The authors declare no conflict of interest.

REFFERENCES

Fain, A. Nociceptors and the perception of the pain. University of Connecticut Health Center.https://health.uconn.edu/cellbiology/wpcontent/uploads/sites/115/2023/05/Revised-Book2014.pdf.

Duarte, D. F. Uma breve história do ópio e dos opióides. *Revista Brasileira de Anestesiologia*, 5(1), 135-146, 2005. https://doi.org/10.1590/S0034-70942005000100015.

Takayama, H. Chemistry and Pharmacology of Analgesic Indole alkaloids from the Rubiaceous Plant, *Mitragyna speciose*. *Chemical and Pharmaceutical Bulletin*, 52, 916-928, 2004. https://doi.org/10.1248/cpb.52.916.

Raffa, R.B.; Beckett, J. R.; Brahmbhatt, V. N.; Ebinger, T. M.; Fabian, C. A.; Nixon, J. R.;

Orlando, S. T.; Rana, C.A.; Tejani, A. H.; Tomazic, R. J. Orally Active Opioid Compounds from Non-Poppy Source. *Journal of Medicinal Chemistry,* 56(12), 4840-4848, 2013. https://doi.org/10.1021/jm400143z.

Li, Y.; Chase, A. R.; Slivka, P. F.; Baggett, C. T.; Zhao, T. X.; Yin, H. Design, synthesis, and evaluation of biotinylated opioid derivatives as novel probes to study opioid pharmacology. *Bioconjugate Chemistry*, (19)12, 2585-2589, 2008. https://doi.org/10.1021/bc8003815 [6] Varmuza K. Pattern Recognition in Chemistry. Springer-Verlag, Berlin, 1980.

Johnson RA, Wichem DW. Applied Multivariate Statistical Analysis. Prentice-Hall, New Jersey, 1992.

Brown, S. D.; Sum, S. T.; Despagne, F.; Lavine, B. K. Chemometrics. *Analytical Chemistry*, 68(12), 21R-61R, 1996.

Brown, S. D. The chemometrics revolution re-examined. *Journal of Chemometrics*. 2017, 31:e2856. https://doi.org/10.1002/cem.2856

Ferreira MMC. Quimiometria: Conceiro, Métodos e Aplicações. Editora da Unicamp, Campinas, 2015.

Araújo, J. J. O.; Miranda, R. M.; Castro, J. S. O.; Figueiredo, A. F.; Pinheiro, A. C.; Morais, S.

S. S.; Santos, M. A. B.; Pinheiro, A. L. R.; Gil, F. S.; Bitencourt, H. R.; Alves, G. N. R.; Pinheiro,

J. C. Designing Artemisinins with Antimalarial Potential, Combining Molecular Electrostatic Potential, Ligand-Heme Interaction and Multivariate Models. *Computational Chemistry*, 11, 1-23, 2023, https://doi.org/10.4236/cc.2023.111001.

Castro, J. S. O.; Pinheiro, J. C.; Morais, S. S. S.; Bitencourt, H. R.; Figueiredo, A. F.; Santos,

M. A. B.; , Gil, F S.; Pinheiro, A. C. B. Design of N-11-Azaartemisinins Potentially Active against *Plasmodium falciparum* by Combined Molecular Electrostatic Potential, Ligand-Receptor Interaction and Models Built with Supervised Machine Learning Methods. *Journal of Biophysical Chemistry*, 14, 1-29, 2023. https://doi.org/ 10.4236/jbpc.2023.141001.

Silva, O. P. P.; Figueiredo, A. F.; Barbosa, J. P.; Santos, M. A. B.; Bitencourt, H. R.; Pinheiro, A. C. B.; Pinheiro, J. C. SAR and QSAR study of antimalarial artemisinins. *Contribuciones a las Ciencias Sociales*, 16(8), 2023, 9084–9104. https://doi.org/10.55905/revconv.16n.8-053

Carvalho, P.; Furr III, E. B.; McCurdy, C. (*E*)-Methyl 2-[(2*S*,3*S*,12b*R*)-3-ethyl-8-methoxy1,2,3,4,6,7,12,12b-octahydroindolo[2,3-*a*] quinolizin-2-yl]-3-methoxyacrylate ethanol solvate. *Acta Crystallographica, Section E*, 65 (Part 6), o1441-o1442, 2009. https://doi.org/10.1107/S1600536809017309.

Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and use of quantum mechanical molecular models. 76. AMI: a new general-purpose quantum mechanical molecular model. *Journal of American Chemistry Society*, 107 (13), 3902-3909, 1985. https://doi.org/10.1021/ja00299a024

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*, 9(6), 2672-2686, 2013. https://doi.org/10.1021/ct400224n

Roothaan, C. C. J. New Developments in Molecular Orbital Theory. *Reviews of Modern Physics*, 23(2), 69 – 89, 1951. https://doi.org/10.1103/RevModPhys.23.69

Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. Ab Initio Molecular Theory. Wiley, New York, 1986.

Stevens, W. J.; Basch, H.; Krauss, M. Compact effective potentials and efficient sharedexponent basis-sets for the 1st-row and 2nd-row atoms. *Journal of Chemical Physics*, 81(12), 6026-

33, 1984. https://doi.org/10.1063/1.447604

Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *Journal of Chemical Physics*, 98(7), 5648-5652, 1993. https://doi.org/10.1063/1.464913 [21] Lee, C.; Yang, W.; Parr, R. G. Development of the colic-salvetti correlation-energy formula into a functional of the electron density. *Physical Review B*, **37**, 785-789, 1988. https://doi.org/10.1103/PhysRevB.37.785 [22] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G.A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, O.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian 98, Revision A.6. Gaussian, Inc., Pittsburgh, 1998.

ChemPlus: Modular Extension to HyperChem Release 8.06. Molecular Modeling for Windows. Hyperchem, Inc; Gainesvill, 2008.

Schaftenaar, G.; Noordik, J H. Molden: a pre- and post-processing program for molecular and electronic structures. *Journal of Computer-Aided Molecular Design*, 14, 123-134, 2000. https://doi.org/10.1023/A:1008193805436

Pirouette 3.01, Informetrix, Inc. Woodinville, 2001.

Todeschinni, R.; Consonni, V.; Gramatica, P. Molecular Descriptor for Chemoinformatics. Willey-VCH, New York, 2009.

Todeschini, R.; Consonni, V.; Ballabio, D.; Grisoni, F. Chemometrics for QSAR Modeling. In: Brown, S.; Tauler, R.; Walczak, B. (Eds.), Comprehensiv Chemometrics (pp. 599-634). Elsevier, New York, 2020.

Sushko, I.; Novotarskyi, S.; Körner, R.; Pandey, A. K.; Rupp, M.; Teetz, W.; Brandmaier, S.; Abdelaziz, A.; Prokopenko, V. V.; Tanchuk, V. Y.; Todeschini, R.; Varnek, A.; Marcou, G.; Ertl,

P.; Potemkin, V.; Grishina, M.; Gasteiger, J.; Schwab, C.; Baskin, I. I.; Palyulin, V. A.; Radchenko,

D.; Cherrkasov, A.; Aires-de-Sousa, J.; Zhang, Q. Y.; Bender, A.; Nigsch, F.; Patin, L.; Williams, A.; Tkachenko, V.; Tetko, I. V. Online Chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information. *Journal of ComputerAided Molecular Design*, 25(6), 533-554, 2011. https://doi.org/10.1007/s10822-011-9440-2 [29] Minitab Statistical Software, Release 16 for Windows. Minitab Inc. State College, Pennsylvania, 2009.

Beebe, K. R.; Pell, R. J.; Seashottz, M. B. Chemometrics: A Practical Guide. Wiley, New York, 1998.

Mardia, K. V.; Kent, J. T.; Bibbly, J. M. Multivariate Analysis. Academic Press, New York, 1979.

Todeschini, R.; Consonni, V. Handbook of Molecular Descriptors**.** In: R. Mannhold, R.; Kubinyi, H.; Timmerman, H. (Eds.). Methods and Principles in Medicinal Chemistry. Wiley-VCH, New York, 2000.

Lohman, R.-J.; Tupally, K. R.; Kandale, A.; Cabot, P. J.; Parekh, H. S. Design and development of novel, short, stable dynorphin-based opioid agonists for safer analgesic therapy. *Frontier in Pharmacology*, 14, 1150313, 2023. https://doi.org/ 10.3389/fphar.2023.1150313