Struct Chem (2009) 20:577–585
DOI 10.1007/s11224-009-9437-9

ORIGINAL RESEARCH

A chemometric study on the analgesic activity of cannabinoid compounds using SDA, KNN and SIMCA methods

A. Arroio · E. F. Lima · K. M. Honório · A. B. F. da Silva

Received: 31 July 2008 / Accepted: 9 February 2009 / Published online: 7 April 2009
© Springer Science+Business Media, LLC 2009

Abstract The supervised pattern recognition methods K-Nearest Neighbors (KNN), stepwise discriminant analysis (SDA), and soft independent modelling of class analogy (SIMCA) were employed in this work with the aim to investigate the relationship between the molecular structure of 27 cannabinoid compounds and their analgesic activity. Previous analyses using two unsupervised pattern recognition methods (PCA—principal component analysis and HCA—hierarchical cluster analysis) were performed and five descriptors were selected as the most relevants for the analgesic activity of the compounds studied: R_3 (charge density on substituent at position C_3), Q_1 (charge on atom C_1), A (surface area), log P (logarithm of the partition coefficient) and MR (molecular refractivity). The supervised pattern recognition methods (SDA, KNN, and SIMCA) were employed in order to construct a reliable model that can be able to predict the analgesic activity of new cannabinoid compounds and to validate our previous study. The results obtained using the SDA, KNN, and SIMCA methods agree perfectly with our previous model. Comparing the SDA, KNN, and SIMCA results with the PCA and HCA ones we could notice that all multivariate statistical methods classified the cannabinoid compounds studied in three groups exactly in the same way: active, moderately active, and inactive.

Keywords Cannabinoid compounds · Analgesic activity · Supervised pattern recognition methods

A. Arroio
Faculdade de Educação, Departamento de Metodologia do Ensino e Educação Comparada, Universidade de São Paulo, Av. da Universidade 308, bloco B, sala 1, 05508-900 São Paulo, SP, Brazil

E. F. Lima · A. B. F. da Silva (✉)
Instituto de Química de São Carlos, Universidade de São Paulo, C. P. 780, 13560-970 São Carlos, SP, Brazil
e-mail: alberico@iqsc.usp.br

K. M. Honório
Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, Av. Arlindo Bettio, 1000, 03828-000 São Paulo, SP, Brazil

K. M. Honório
Universidade Federal do ABC (UFABC), Rua Santa Adélia, 166, Bairro Bangu, 09210-170 Santo André, SP, Brazil

Introduction

Cannabis sativa is one of the first plants to have been used as medicine, for religious ceremonies and recreations. The first accounts of its use for these purposes stretch back 5,000 years [1]. This plant is the unique source of a set of more than 60 oxygen-containing aromatic hydrocarbon compounds collectively known as cannabinoids. It also contains a number of other compounds of potential interest including at least 120 different terpenes and 21 flavonoids [2]. From these constituents, only two of them have many discoveries about their pharmacology: (1) Δ^9-tetrahydrocannabinol (Δ^9-THC), which has psychoactivity and (2) cannabidiol, which is not psychoactive [2].

About 41 years ago, Gaoni and Mechoulam identified Δ^9-tetrahydrocannabinol (THC) as the main psychoactive molecule present in C*annabis sativa* [3]. The pharmacological effects of cannabinoids are mediated through at least two receptors, named as CB_1 and CB_2, and these effects include tachycardia, hypothermia, analgesia, and the appetite-enhancing effects [3, 4]. In this work, we have studied cannabinoid compounds with analgesic activity and this is

significant since the control of pain requires a knowledge of the pain mechanisms and an understanding of the drugs used for this purpose. Therefore, analgesic drugs have a very important role in the therapy against pain as, for instance, the therapy of patients in the terminal stage of cancer.

The rational search for new analgesic drugs is a very efficient strategy to obtain more specific and potent compounds without side effects. Some methods used for this strategy include studies based on structure-activity relationships (SAR) and quantitative structure-activity relationships (QSAR). The main goal of applying these methods is to transform the chemical structure of a compound into a set of numbers (parameters, properties, or variables) and correlate them with the biological activity establishing a qualitative/quantitative relationship between calculated molecular properties and biological activity [5]. In this context, we can cite the multivariate statistical methods or pattern recognition methods, which are very helpful to extract meaningful information of the system studied and construct mathematical models that can be able to predict some properties of interest. There are two kinds of pattern recognition methods: the unsupervised and the supervised ones. In the first category (the unsupervised methods), the main goal is to reduce the data complexity so that inherent clusters in the samples can be visualized. In the supervised methods, a training set of samples with known class is used to produce a mathematical model that can predict the class of unknown samples.

In a previous work [6], we used two unsupervised pattern recognition methods known as Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA) with the aim to investigate which molecular properties (variables or descriptors) would be more efficient in classifying cannabinoid compounds according to their degree of analgesic activity. According to previous results [6], three categories of compounds were observed from PCA and HCA: active, moderately active, and inactive compounds. The PCA and HCA results indicated that five descriptors were the most important for the discrimination of the compounds: R_3 (charge density on substituent at position C_3), Q_1 (charge on atom C_1), A (surface area), log P (logarithm of the partition coefficient) and MR (molecular refractivity) [6].

Now, the main goal of this work is to investigate the therapeutic aspects of cannabinoid compounds according to their analgesic potency by using supervised pattern recognition methods (or classification methods) in order to predict the class of new cannabinoid compounds and validate our previous results employing PCA and HCA [6]. Since the unsupervised pattern recognition PCA and HCA are not appropriate for the prediction of unknown compounds, we have decided to use more robust classification methods: SDA [7], KNN [8], and SIMCA [9], which will help us to construct a more reliable model for future prediction studies [10].

Methodology

Molecular structure of cannabinoid compounds

The general structure of cannabinoid compounds and the numbering system used in this work are shown in Fig. 1. Figure 2 shows the chemical structure and the biological activity of the 27 cannabinoid compounds studied in this work, which were tested using the same biological assays [11–15]. The 27 analgesic cannabinoid compounds studied here were classified into three classes: actives, moderately actives, and inactives, based on the qualitative effects caused on rhesus monkeys when the compounds were intravenously injected [11–15].

Following the previous PCA and HCA analyses [6], the compounds of the training set were classified into three classes: active, moderately active, and inactive regarding their analgesic activity. The molecular descriptors were calculated based on the most stable conformation for each cannabinoid structure.

The geometry optimization and all of the quantum chemical calculations were performed by using the quantum chemical semi-empirical method AM1 (Austin model 1) [16]. The calculated properties were selected so that they could represent electronic, stereochemical, and hydrophobic characteristics of the compounds studied. The molecular descriptors were calculated based on the most stable conformation for each cannabinoid structure and the most important descriptors selected by PCA and HCA analyses [6], as mentioned before, were: R_3 (charge density on substituent at position C_3), Q_1 (charge on atom C_1), A (surface area), log P (logarithm of the partition coefficient), and MR (molecular refractivity) [6]. The numerical values of the five selected variables (R_3, Q_1, A, log P, and MR) are displayed in Table 1. It is important to say that the five variables were autoscaled before all analyses.

The three supervised pattern recognition methods used here (KNN, SDA, and SIMCA) made use of the PCA and HCA results obtained previously [6] in order to obtain a reliable model that can be able to predict the class of new samples [17]. The classification rules achieved were validated by means of a leave-one-out cross-validation procedure due to the small number of cannabinoid compounds tested with the same biological assays [11–15]. In order to perform the statistical analyses by using the KNN and SIMCA methods we employed the PIROUETTE 3.11 program [18], and to perform the SDA analysis we made use of the STATISTICA program [19].

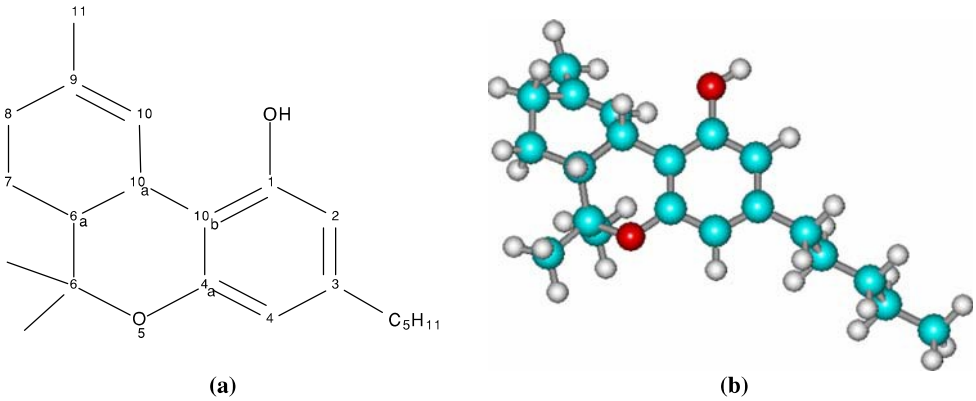
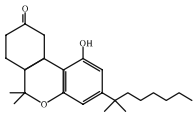
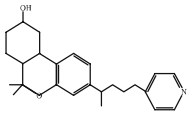


Fig. 1 General structure **a** bidimensional with the numbering system used for the cannabinoid compounds studied and **b** tridimensional structure



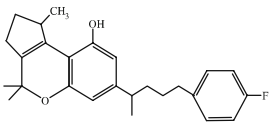
1

Nabilone (active)



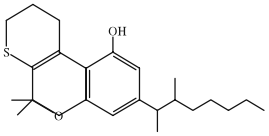
2

DHHP (active)



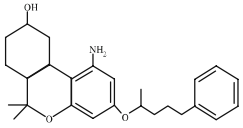
3

Homolog of CP41988 (active)



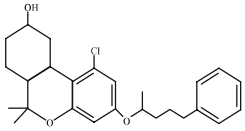
4

Homolog of DMHP (active)



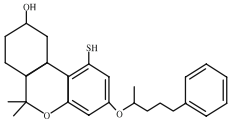
5

Homolog of 9-Nor-9-β-hydroxy-HHC (moderately active)



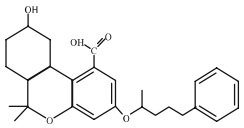
6

Homolog of 9-Nor-9-β-hydroxy-HHC (moderately active)



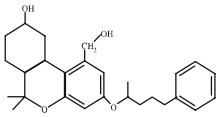
7

Homolog of 9-Nor-9-β-hydroxy-HHC (moderately active)



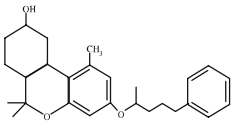
8

Homolog of 9-Nor-9-β-hydroxy-HHC (moderately active)



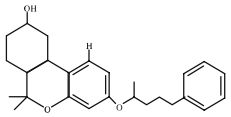
9

Homolog of 9-Nor-9-β-hydroxy-HHC (moderately active)



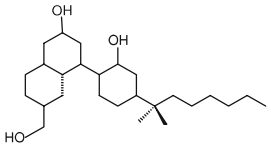
10

Homolog of 9-Nor-9-β-hydroxy-HHC (moderately active)



11

Homolog of 9-Nor-9-β-hydroxy-HHC (moderately active)



12

CP5524 (active)

Fig. 2 Chemical structure of the 27 cannabinoid compounds studied

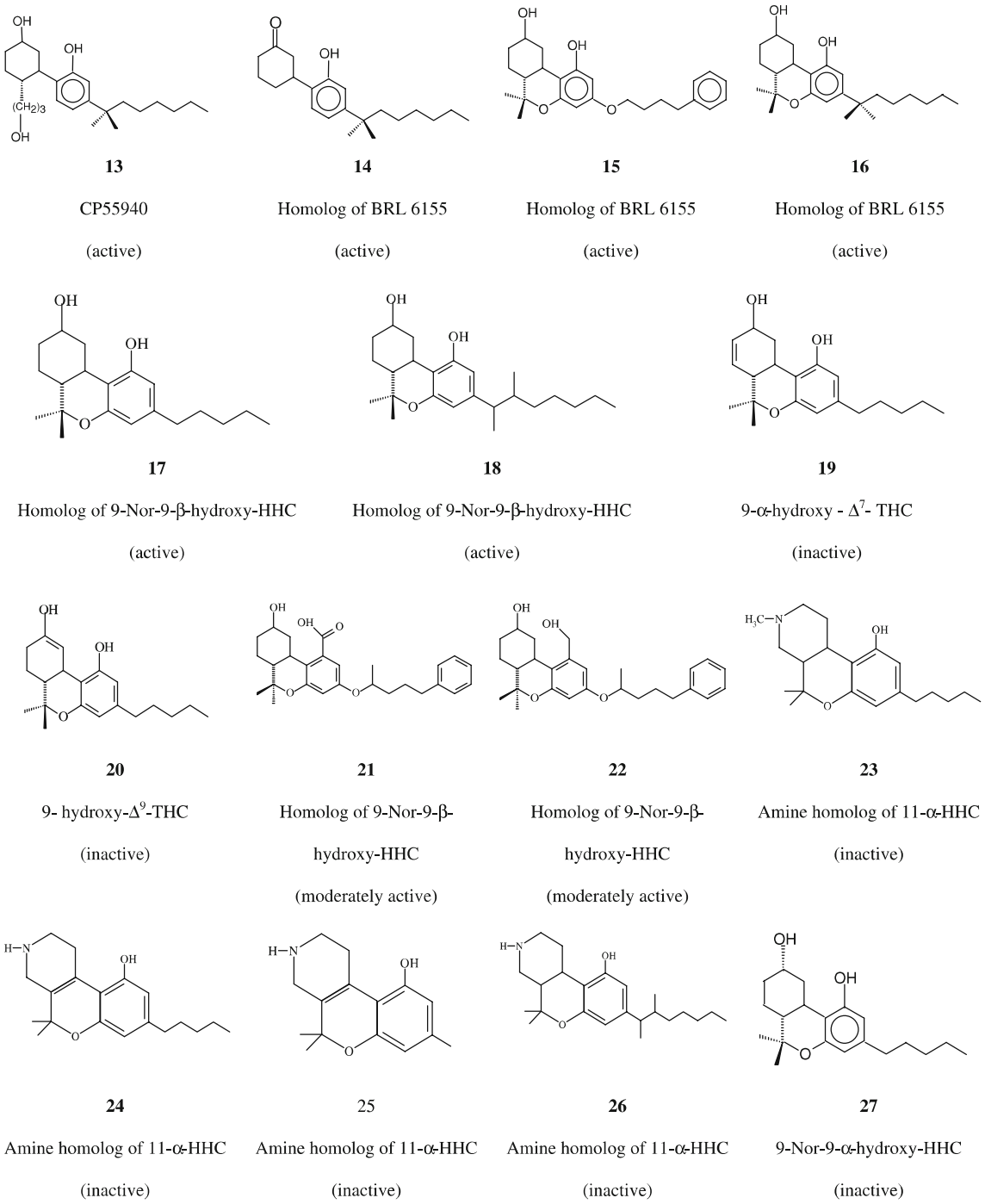


Fig. 2 continued

SAR analyses

Structure-activity relationship (SAR) studies with bioactive molecules can help in the developing of more effective compounds (by selecting the molecular properties that can be responsible for the biological activity) with the goal of understanding the interaction mechanism between ligands and their biological receptors [20]. For this purpose, several multivariate statistical methods can be used.

Stepwise discriminant analysis (SDA)

Stepwise discriminant analysis (SDA) can be used for discrimination (recognition or classification) and

Table 1 Value of the five most relevant descriptors selected by the PCA and HCA methods [6]

Compound	R_3	Q_1	A (Å^2)	log P	MR (Å^3)
1	0.053	0.438	665.297	3.601	113.116
2	0.043	0.021	625.808	2.638	119.493
3	0.075	0.681	638.244	2.470	123.613
4	0.066	0.475	650.838	2.625	117.060
5	−0.135	0.405	650.214	0.883	128.677
6	−0.144	−0.021	611.168	2.378	129.843
7	−0.339	0.535	660.107	1.817	133.146
8	−0.144	−0.426	613.764	1.985	131.126
9	−0.175	−0.222	696.602	1.752	131.183
10	−0.173	0.081	683.156	2.754	129.409
11	−0.159	−0.061	603.943	2.601	125.127
12	0.051	0.305	641.516	4.580	123.471
13	0.040	0.341	648.036	4.350	116.253
14	0.040	0.380	562.287	4.374	100.574
15	0.049	0.360	607.592	4.582	99.547
16	0.049	0.387	592.857	4.374	100.574
17	0.540	0.391	606.116	3.393	114.143
18	0.062	0.431	616.797	3.291	114.217
19	0.065	0.446	512.916	1.721	96.881
20	0.065	0.440	526.331	1.465	97.125
21	−0.147	−0.328	579.533	1.614	132.333
22	−0.155	−0.139	574.953	1.380	132.390
23	0.050	0.555	563.101	1.441	98.674
24	0.056	0.535	553.243	1.079	93.379
25	0.036	0.556	445.743	−0.506	74.975
26	0.062	0.447	649.390	2.533	111.678
27	0.077	0.418	516.640	1.837	95.918

prediction of samples. Its main objective is to determine discriminant functions, which represent linear combinations of the calculated variables [20]. The procedure used in the SDA method is to build discriminant functions (one function for the active compounds and another one for the inactive ones) adding one variable at a time until obtaining the final discriminant functions based on the set of the best variables that discriminate the groups of compounds [21]. This method is useful for selecting variables with the highest relevance to separate the compounds into different groups (often referred to as discriminant power of the variables), since it builds the discriminant functions using one variable at a time until the best discriminant function is obtained. After the statistical validation of the model through this procedure, the discriminant functions may be used to make predictions with unknown compounds. [22].

K-nearest neighbours (KNN)

The KNN method classifies the compounds studied based on the distance among them. In general, Euclidean distance is used to measure the nearness between samples. The predicted class of a test compound is determined according to its distance regarding the closest K compounds in the training set [7, 10, 17]. Leave-one-out cross-validation is used to select the optimal number of the nearest neighbors (K). In this procedure, each sample in the training set is excluded and then classified using the remaining training set compounds. This is repeated for different values of K and each one of the K nearest samples "votes" once for its class. The class receiving the highest number of "votes" is assigned to that sample [23].

Soft independent modeling of class analogy (SIMCA)

SIMCA is a classification technique that minimizes assumptions on the linearity of relationships between descriptors and classes [24]. This technique uses PCA to model the shape and position of the object formed by the samples [8, 17]. The number of optimal principal components (PCs) is determined for each class and the model is finished by defining boundary regions for each PCA model. A multidimensional box is constructed for each class and the classification of future samples (prediction) is performed by determining within which box, if any, the samples lies. This is in contrast to KNN where only the physical closeness of samples is used for classification. The main advantage of SIMCA over other classification methods is its ability to detect outlier samples [17].

Results and discussion

According to a previous study [6], the most important descriptors found in PCA and HCA analyses were: R_3, Q_1, A, log P, and MR, which are responsible for the discrimination of the cannabinoid compounds under study here into three groups (active, moderately active, and inactive cannabinoid compounds). As can be seen from Fig. 3, the dendrogram showed a good discrimination among the three classes of compounds. The similarity value between the classes of active (including the moderately active compounds) and inactive compounds was 0.0, indicating there was a very good distinction between these classes of compounds. After the use of the unsupervised pattern recognition methods (PCA and HCA), we employed the classification methods (KNN, SDA, and SIMCA) in order to validate and construct a model that can be used to predict the class of new analgesic cannabinoid molecules with similar molecular structure.

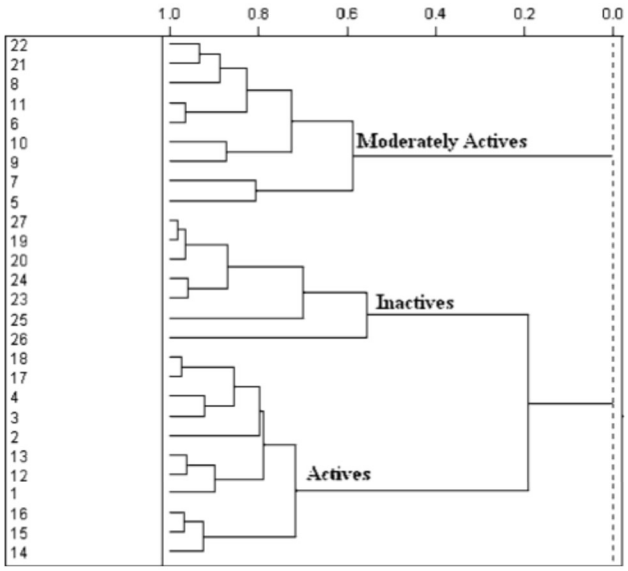


Fig. 3 Dendrogram obtained for the set of selected variables

K-nearest neighbor (KNN)

Table 2 presents the results obtained using the five selected descriptors (R_3, Q_1, A, log P, and MR) taking into account 1–9 nearest neighbors (9NN). For the cases of 1 until 7 nearest neighbors, the accuracy of classification was 100% (i.e., all compounds were classified correctly), where 11 samples belong to the group of active compounds (class 1), 7 belong to the group of inactive compounds (class 2) and 9 to the group of moderately active compounds (class 3). For the case of 9NN, one compound was incorrectly classified (compound 23), decreasing the accuracy of classification (see Table 2). So, in order to construct our KNN model we considered 7NN as the limit of confidence.

Furthermore, as the outcomes obtained by the KNN method agree perfectly with those from PCA and HCA, we can conclude that these multivariate statistical models obtained by the three techniques (PCA, HCA, and KNN) are reliable to be used in the prediction of new cannabinoid compounds with analgesic activity.

Stepwise discriminant analyses (SDA)

The SDA analyses were performed using the variables R_3, log P, Q_1, MR, and A. The discriminant functions obtained by SDA for each group of cannabinoid compounds, written as a linear combination of the five original variables, are:

Group 1 = −171.88 + 109.66 R_3 + 5.79 Log P – 0.05 Q_1 + 0.39 MR + 0.43 A (*actives*)

Group 2 = −186.46 – 138.47 R_3 + 0.78 Log P – 34.61 Q_1 + 0.39 MR + 0.47 A (*moderately actives*)

Table 2 Classification obtained with the KNN method

Compound	Observed class	1NN	3NN	5NN	7NN	9NN
1	1	Class 1	Class 1	Class 1	Class 1	Class 1
2	1	Class 1	Class 1	Class 1	Class 1	Class 1
3	1	Class 1	Class 1	Class 1	Class 1	Class 1
4	1	Class 1	Class 1	Class 1	Class 1	Class 1
5	3	Class 3	Class 3	Class 3	Class 3	Class 3
6	3	Class 3	Class 3	Class 3	Class 3	Class 3
7	3	Class 3	Class 3	Class 3	Class 3	Class 3
8	3	Class 3	Class 3	Class 3	Class 3	Class 3
9	3	Class 3	Class 3	Class 3	Class 3	Class 3
10	3	Class 3	Class 3	Class 3	Class 3	Class 3
11	3	Class 3	Class 3	Class 3	Class 3	Class 3
12	1	Class 1	Class 1	Class 1	Class 1	Class 1
13	1	Class 1	Class 1	Class 1	Class 1	Class 1
14	1	Class 1	Class 1	Class 1	Class 1	Class 1
15	1	Class 1	Class 1	Class 1	Class 1	Class 1
16	1	Class 1	Class 1	Class 1	Class 1	Class 1
17	1	Class 1	Class 1	Class 1	Class 1	Class 1
18	1	Class 1	Class 1	Class 1	Class 1	Class 1
19	2	Class 2	Class 2	Class 2	Class 2	Class 2
20	2	Class 2	Class 2	Class 2	Class 2	Class 2
21	3	Class 3	Class 3	Class 3	Class 3	Class 3
22	3	Class 3	Class 3	Class 3	Class 3	Class 3
23	2	Class 2	Class 2	Class 2	Class 2	Class 1
24	2	Class 2	Class 2	Class 2	Class 2	Class 2
25	2	Class 2	Class 2	Class 2	Class 2	Class 2
26	2	Class 2	Class 2	Class 2	Class 2	Class 2
27	2	Class 2	Class 2	Class 2	Class 2	Class 2

Group 3 = −114.47 + 109.07 R_3 + 1.47 Log P − 2.21 Q_1 + 0.23 MR + 0.38 A (*inactives*)

Comparing the SDA results with the PCA and HCA analyses [6], we can notice that the three methods classified the cannabinoid compounds studied in three groups exactly in the same way, i.e., active, moderately active, and inactive compounds.

Another way to perform the classification by using the discriminant functions is to calculate the classification matrix by using the coefficients shown by the discriminant functions. According to this matrix, the accuracy of classification obtained using the SDA method was 100%, indicating a good separation of the three groups of cannabinoid compounds (active, moderately active, and inactive).

The allocation rule derived from the SDA results, when the analgesic activity of a new cannabinoid compound is investigated, is: (a) initially one calculates, for the new

analgesic cannabinoid compound, the values of the most important variables obtained with the SDA methodology; (b) substitute these values in the three discriminant functions obtained in this work; (c) see which discriminant function (Group 1: active; Group 2: moderately active and Group 3: inactive compounds) presents the highest value. The new cannabinoid compound is active if the highest value is related to the discriminant function of Group 1 and so on.

Soft independent modeling of class analogy (SIMCA)

We performed the SIMCA analyses using the autoscaled values of the five selected variables (R_3, log P, Q_1, MR, and A). The best SIMCA model found was the one built with the same variables used in the SDA method. Figure 4 shows the three-dimensional projection of the compounds obtained with three PCs. The hyperboxes for the three classes of cannabinoid compounds studied are represented in Fig. 4 by the points around each class.

From Fig. 4 we can see the splitting of the set of compounds into three well distinct classes, corresponding to class 1 (active compounds), class 2 (moderately active compounds), and class 3 (inactive compounds). The coordinates of the hyperboxes that determine the limits of the classes are based on the standard deviations of the sample scores in the direction of each PC and state a confidence limit of 95% for the distribution of the classes (represented by dotted surfaces in Fig. 4). The rotation of Fig. 4 shows that no compound is allocated out of the confidence limits and that there is no superposition between the three classes.

Figure 5 displays the class distances calculated according to the residuals of the samples when they are adjusted to the

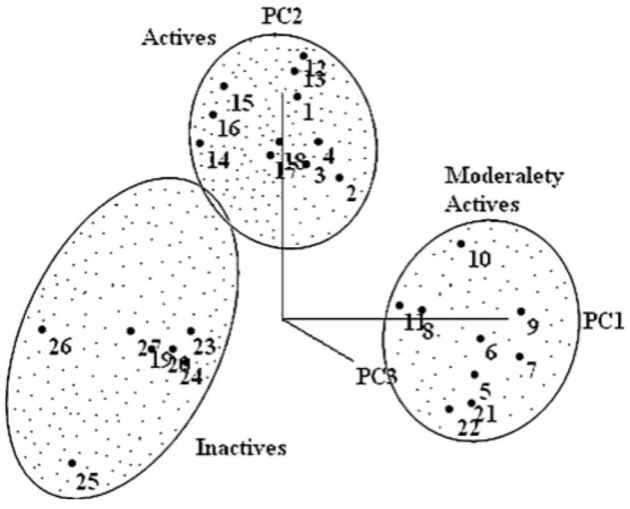


Fig. 4 Three-dimensional projection of the hyperboxes for the three classes of compounds obtained by SIMCA using the five selected variables (R_3, log P, Q_1, MR, and A)

classes. The compounds lying in the north–west quadrant (NW) belong solely to the x-axis class. Analogously, the compounds in the south–east quadrant (SE) are members of the y-axis class only. Compounds positioned in the south–west quadrant (SW) may belong to both classes. However, the compounds in the north–east quadrant (NE) belong to a third class of compounds (the moderately active ones).

From the SIMCA method, we are able to obtain the following data: (i) the distance between classes, which is a measure of how separated are the classes in a model; (ii) the modeling power (MP) of the variables used in the classification model, indicating the influence of each variable in the model; (iii) the discriminant power (DP) of the variables, which is indicative of the importance of each variable in the discrimination of the compounds into different classes [25]. Table 3 shows the values of the modeling and discriminant powers for the five selected variables. The most important variable in the two models is R_3 (charge density on substituent at position C_3), since it has the highest value for MP and DP, indicating that electronic effects can have an important role in the interaction between the cannabinoid compounds and the biological receptor.

Table 4 shows the calculated distances between classes 1, 2, and 3. The distance between the classes 1 and 1, 1 and 2, 1 and 3, 2 and 3 are: 0.000, 4.080, 8.279, and 18.933 (all above 3), respectively. In chemometrics, the distances above 3 are considered suitable for a good distinction between classes [26].

From our results, we can see that for a cannabinoid compound to become an analgesic active molecule it must have some important characteristics: (1) high log P values (the analgesic compounds studied have a higher lipophylic character than the inactive ones, indicating a high capacity of crossing the biological membrane and, consequently, they are able to reach the biological receptor more easily); (2) a suitable surface area (A) value (compounds with small surface area cannot interact with the biological receptor); (3) positive MR values, which indicate that some substituents in the active compounds can interact by two ways: (a) through polar groups of the biological receptor (due to polarizability effects) and (b) through steric effects (due to the size of the substituents), since some modifications on the receptor can occur avoiding the interaction between the compounds and the biological receptor; (4) negative values for R_3 and Q_1 (this indicates that the cannabinoid compounds studied need to have electron donor substituents at C_1 and C_3 positions in order to present analgesic activity).

The results obtained in this work using the SDA, KNN, and SIMCA methods confirm the importance of the five descriptors employed in all statistical analyses (R_3, Q_1, A, log P and MR), which can represent the main interactions between the analgesic cannabinoid compounds and the

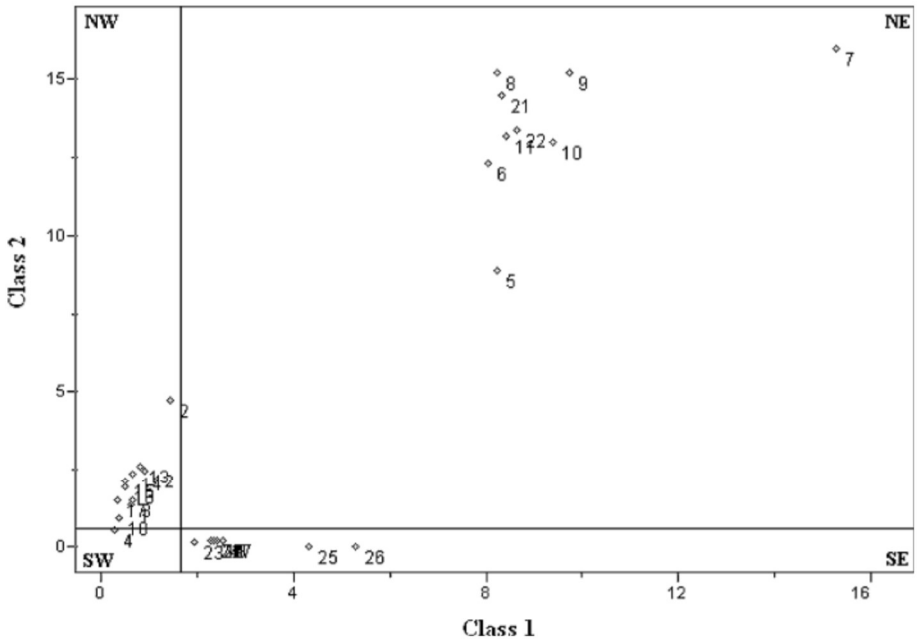


Fig. 5 Class distances obtained by SIMCA for the 27 cannabinoid compounds studied

Table 3 Modeling and discriminant powers for the five selected variables according to SIMCA method

Variable	Discriminant power (DP)	Modeling power (MP)
R_3	478.926	0.724
Q_1	86.393	0.610
A	21.123	0.607
Log P	29.874	0.557
MR	199.805	0.581

Table 4 Distance between classes for the selected variables according to the SIMCA method

	Class 1	Class 2	Class 3
Class 1	0.000	4.080	8.279
Class 2	4.080	0.000	18.933
Class 3	8.280	18.933	0.000

biological receptor. It is important to notice that all multivariate statistical methods classified the analgesic cannabinoid compounds studied here in three groups exactly in the same way: active, moderately active, and inactive.

Conclusions

The results obtained in this work, using three supervised pattern recognition methods (KNN, SDA, and SIMCA), agree perfectly with our previous model, in which we used two unsupervised pattern recognition methods (PCA and HCA). This is a strong indication that our previous selection of variables was suitable, since it provided 100% of accuracy in the classification and this indicates that the models obtained with all of the five methodologies (PCA, HCA, KNN, SDA, and SIMCA) are reliable.

Comparing the SDA, KNN, and SIMCA results with the PCA and HCA ones, we concluded that all methods classified the analgesic cannabinoid compounds studied in three groups exactly in the same way: active, moderately active, and inactive. Therefore, from our KNN, SDA, and SIMCA results we can see that our models are very reliable and they can help in the design of new analgesic cannabinoid compounds.

Acknowledgments The authors would like to thank FAPESP and CNPq (Brazilian agencies) for the financial support.

References

1. Roger GP (2006) Br J Pharmacol 147:S163. doi:10.1038/sj.bjp.0706406
2. Roger GP (2004) Euphytica 140:73. doi:10.1007/s10681-004-4756-9
3. Tiziano T, Píer LF, Clementina M, Gabriella O, Giuseppe S, Adriano M (2006) J Med Chem 49:984. doi:10.1021/jm050875u
4. Carl RL, Arthur CR, Alexander FH (2004) Br J Pharmacol 143:227. doi:10.1038/sj.bjp.0705931
5. Honório KM, da Silva ABF (2005) J Mol Model 11:200. doi:10.1007/s00894-005-0243-z
6. Arroio A, Honório KM, da Silva ABF (2004) J Mol Struct THEOCHEM 709:223. doi:10.1016/j.theochem.2003.10.079
7. Romeder JM (1973) Methode Programme d'analyse Discriminante. Dunod, Paris
8. Kowalski BR, Bender CF (1972) J Am Chem Soc 94:5632. doi:10.1021/ja00771a016

9. Wold S (1976) Pattern Recognit 8:127. doi:10.1016/0031-3203(76)90014-5
10. Molfetta FA, Honório KM, Alves CN, da Silva ABF (2004) J Mol Struct THEOCHEM 674:191. doi:10.1016/j.theochem.2003.12.007
11. Mechoulam R (1973) Marijuana: chemistry, pharmacology, metabolism and clinical effects. Academic Press, New York
12. Petersen RC (1980) Marijuana research findings.. Department of Health and Human Services, Maryland
13. Mechoulam R (1970) Science 168:1166. doi:10.1126/science.168.3936.1159
14. da Silva ABF, Trsic M (1995) J Mol Struct 356:256. doi:10.1016/0022-2860(95)08959-Y
15. Razdan RK (1986) Pharmacol Rev 38:49
16. Dewar MJ, Zoebisch EG, Healy EF, Stewart JJP (1985) J Am Chem Soc 107:3902. doi:10.1021/ja00299a024
17. Beebe KR, Pell RJ, Seasholtz MB (1998) Chemometrics: a practical guide. Wiley-Interscience, New York
18. INFOMETRIX. Pirouettte Multivariate Data analysis form IBM PC Systems, Version 2.02. Seattle, WA, 1990-1996
19. STATSOFT INC. Statistica 4.3 Tulsa, 1993
20. Souza J Jr, Molfetta FA, Honório KM, Santos RHA, da Silva ABF (2004) J Chem Inf Comput Sci 44:1153. doi:10.1021/ci030384n
21. Gardiner WP (1997) Statistical Analysis Methods for Chemists. Royal Society Chemistry, Cambridge
22. Weber KC, Honório KM, Bruni AT, da Silva ABF (2006) J Mol Model 12:915. doi:10.1007/s00894-005-0083-x
23. Ferreira MMC (2002) J Braz Chem Soc 13:742
24. Clark RD (2004) J Mol Graph Model 22:487. doi:10.1016/j.jmgm.2004.03.009
25. Sharaf MA, Illman DL, Kowalski BR (1986) Chemometrics. Wiley, New York
26. Brereton RG (1992) Multivariate pattern recognition in chemometrics, illustrated by case studies. Elsevier, Amsterdam