J Mol Model (2005) 11: 200–209
DOI 10.1007/s00894-005-0243-z

ORIGINAL PAPER

Káthia M. Honório · Albérico B. F. da Silva

A study on the influence of molecular properties in the psychoactivity of cannabinoid compounds

Received: 18 March 2004 / Accepted: 27 January 2005 / Published online: 3 May 2005

Abstract Several molecular properties are calculated for a set of 26 cannabinoid compounds with the goal of connecting the psychoactivity of the compounds with an appropriate set of calculated properties. For this purpose we used quantum chemical (the AM1 semi-empirical method) and chemometric methods. The AM1 method was employed to calculate the set of quantum chemical molecular properties and the chemometric methods were employed with the aim of selecting the most relevant properties to be correlated with psychoactivity. The chemometric methods used were Principal Component Analysis (PCA), Hierarchical Cluster Analysis (HCA) and the K-Nearest Neighbor (KNN) method. The chemometric analysis showed that an electronic property (energy of LUMO), a hydrophobic property (log *P*), a steric property (volume of the substituent at the C4 position) and a topological property (Lovasz–Pelikan index) were the most important variables for the separation between the psychoactive and psychoinactive compounds. In order to validate our PCA, HCA and KNN results, eight new cannabinoid compounds (with known psychoactivity) were used in a prediction study and were classified correctly by the methods used in this work, indicating that our PCA, HCA and KNN models are able to predict reliable psychoactivity of cannabinoid compounds.

Keywords Cannabinoid compounds · Psychoactivity · Quantum chemical and chemometric methods

Introduction

The plant *Cannabis sativa L.* has been used by man for centuries for eating, medicinal practices and religious rituals [1]. The Assyrians considered the *C. sativa* as the major drug of their pharmacopoeia and named it according to its use, e.g.: *qunnabu*, when it was used in religious rituals; *azallu*, a medicinal term as well as hemp; *gan-zi-gun-nu* which is translated as "the drug that takes the mind away" [1]. Nowadays, there are many valuable discoveries regarding *Cannabis*, but many myths and uncertainties still persist. Due to the great interest in the effects caused by the compounds extracted from the *Cannabis*, several studies have been carried out with the aim to better understand the relationship between the chemical structure and the biological activity of cannabinoid compounds [2–4].

By the mid-1970s, most cannabinoids had been isolated, synthesized and their metabolic pathways elucidated [5]. Some industries and academic laboratories initiated projects to develop cannabinoid-based drugs. However, separation between the psychotropic effects and the medically useful ones was not achieved, except for Nabilone (a potent THC-type drug) [5]. Nabilone has been used in some countries, UK for example, as an antiemetic agent. Δ^9 -THC is also used for this purpose as well as for enhancement of appetite. Δ^8 -THC, which is considerably less expensive to prepare and as active as Δ^9 -THC in antiemetic studies is not marketed, apparently for purely commercial reasons [5].

The rational search for new drugs is a very efficient strategy to obtain more specific, 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) [6]. The main goal of applying these methods in this work is to transform the chemical structure of a compound into a set of numbers (parameters, properties or variables) that correlate with the biological activity, establishing a qualitative/quantitative relationship between calculated molecular properties and biological activity [7].

Here, the quantum chemical AM1 semi-empirical method [8] is employed to calculate the molecular parameters (properties) that are possibly correlated with

K. M. Honório · A. B. F. da Silva (✉)
Departamento de Química e Física Molecular,
Instituto de Química de São Carlos, Universidade de São Paulo,
C. P. 780, 13560-970 São Carlos, SP, Brasil
E-mail: alberico@iqsc.usp.br

the psychoactivity and psychoinactivity within a set of 26 cannabinoid compounds. In sequence, pattern recognition methods [9, 10] are employed to correlate the calculated parameters and the psychoactivity presented by the cannabinoid compounds. The pattern recognition methods used were Principal Component Analysis (PCA) [11–13], Hierarchical Cluster Analysis (HCA) [14] and K-Nearest Neighbor (KNN) [15]. The PCA, HCA and KNN methods were employed in this work with the aim of extracting the most relevant information from the data set and classifying the compounds studied into two different groups (psychoactive or psychoinactive).

Cannabinoid molecules have a typical structure as shown in Fig. 1 (the numbering system used in this work is also shown). The chemical structure of each cannabinoid molecule studied, along with its activity indication, is displayed in Fig. 2. The 26 cannabinoid compounds were classified into two classes: actives and inactives, based on the effects caused on rhesus monkeys when the compounds were intravenously injected [16–20]. Other than our 26 cannabinoid compounds studied (the training set), eight new cannabinoid compounds with known psychoactivity [16–20] were used in a prediction study with the aim of assessing our PCA, HCA and KNN results obtained with the training set. The chemical structures of these eight new compounds are presented in Fig. 3.

Methodology

Geometry optimization

In order to obtain the most stable conformation for each compound under study, we used a combination of molecular mechanics and quantum chemical semi-empirical calculations. A pre-optimization of the geometries was carried out by using the MM+ molecular mechanics method [21], as implemented in the Hyperchem program [22]. Afterwards, we carried out a conformational search with the goal of obtaining the lowest energy conformation of the compounds by using the Chemplus program [23]. After that we performed a full geometry optimization using the AM1 semi-empirical method [8] of the AMPAC 6.5 program [24] with the EF and PRECISE keywords. In this way, we tried to ensure that the most stable geometry for each cannabinoid compound was found. In fact there is no guarantee that the lowest energy conformations found here by optimization are exactly the bioactive ones, but we believe that the calculated molecular descriptors associated to the lowest energy conformers will not change drastically for the biologically active ones.

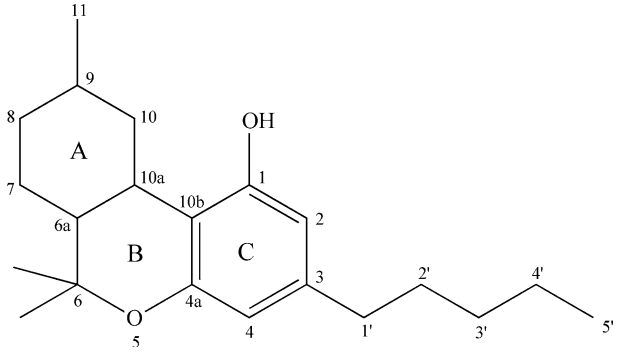

Fig. 1 General structure and numbering used for the 26 cannabinoid molecules studied

Molecular descriptors

In our structure-activity relationships (SAR) study, some molecular properties (variables or descriptors) are evaluated in order to determine which among them could be the most important variables in explaining the psychoactivity presented by the cannabinoid compounds. Four kinds of properties were considered in this work: electronic, steric, hydrophobic and topological, as the interaction between a compound and the biological receptor can very often occur due to electronic, hydrophobic, steric and topological features of the compounds.

The molecular descriptors were calculated making use of the AM1 semi-empirical method [8]. This method was employed to determine structural and electronic parameters to be correlated with the psychoactivity. These parameters include bond distances, torsion angles, bond orders, ionization potential, energies of the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) frontier orbitals, etc. Besides the AM1 method, we made use of the program Chemplus [23], to calculate hydration energy, molar refractivity, molecular volume and molecular surface area, and Spartan 5.0 [25], to calculate Log *P* (the *n*-octanol/water partition coefficient).

For the calculation of topological properties, we made use of the WHIM/3D-QSAR program [26] that evaluates a large variety of these descriptors that represent different sources of chemical information. The WHIM descriptors contain information on the whole 3D molecular structure in terms of size, shape, symmetry and atom distribution [27].

SAR analyses

In this work, we also employed the chemometric methods PCA, HCA and KNN in order to identify which of the calculated molecular properties would be responsible for the psychoactivity presented by the cannabinoid compounds.

The PCA method was used with the aim of reducing the set of molecular properties and exploring the main information contained in the data set. The PCA is a multivariate technique that is designed to reduce the dimensionality of a data set (or training set), that presents a large number of interrelated variables, while

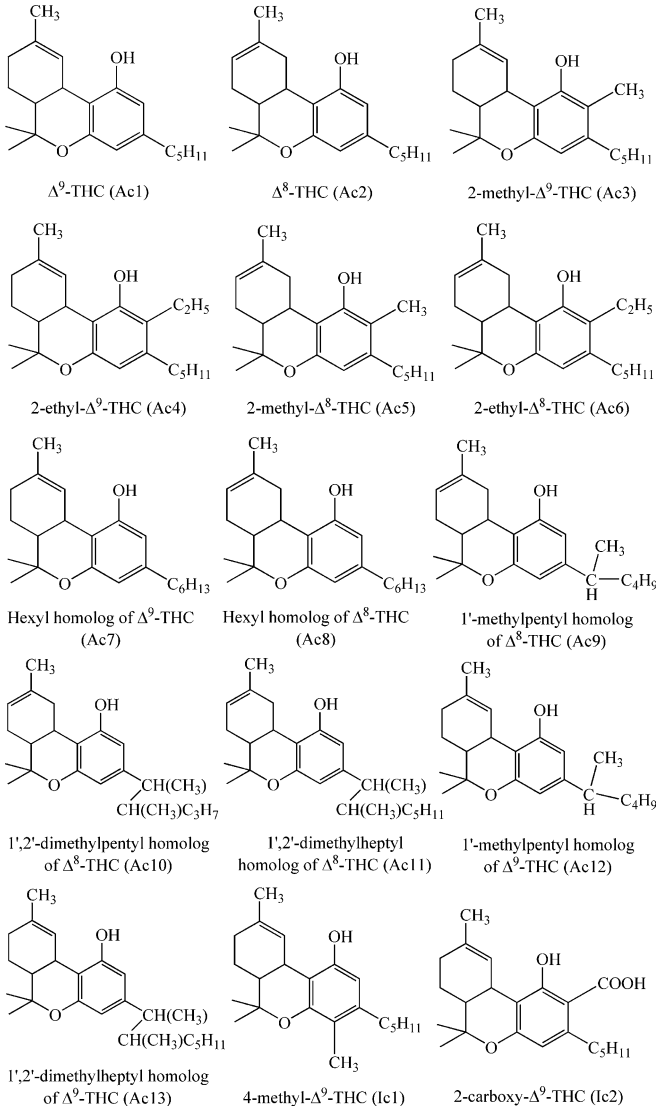

Fig. 2 Chemical structure of the 26 cannabinoid molecules studied

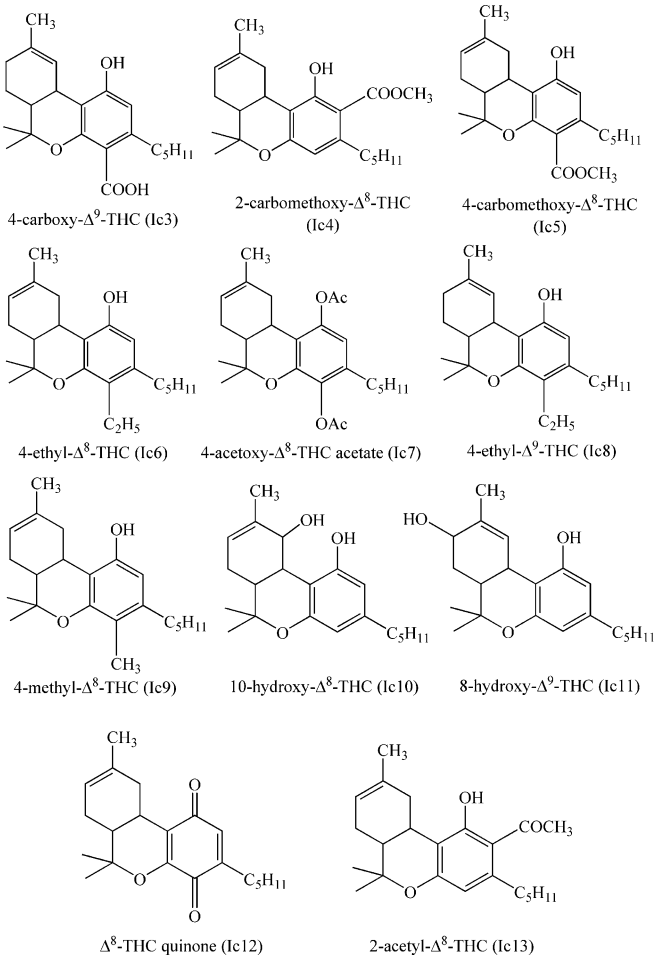

Fig. 2 (Contd.)

retaining as much as possible the variation of the data set. This is achieved by transforming the original variables into a new set of variables called principal components (PCs). The PCs are uncorrelated and ordered so that the first PC retains most of the variance present in the data [28].

The HCA method was used as this technique examines the distances between the samples in a data set and represents this information as a two-dimensional plot called a dendrogram. The HCA method is an excellent tool for preliminary data analysis and is informative when comparing the resulting dendrogram in conjunction with PCA, as they provide complementary information. In HCA, each element is used to generate a similarity matrix. The two most similar elements are fused to form a cluster and the process of generating and analyzing a new similarity matrix is repeated. Again the two most similar elements are fused. The process is then repeated iteratively until all the elements belong to a single group, after the final fusion [9, 14, 29].

The KNN method was used to validate our initial data set, as it classifies a new object (compound) according to its distance to an object in the training set. The closest neighbors of the training set are found and an object is assigned into a given class that has the largest number of nearest neighbors. This method is self-validating because in the training set each sample (object) is compared with all of the objects in the set but not with itself. The best value of K can be chosen based on the results from the training set alone [30, 31].

Results and discussion

Reduction of the number of descriptors

After the calculation of the relevant variables, they were autoscaled so that each had equal importance in each analysis. This kind of treatment ensures that the relative influences of different variables on the calculation are independent of their respective units. Afterwards, in

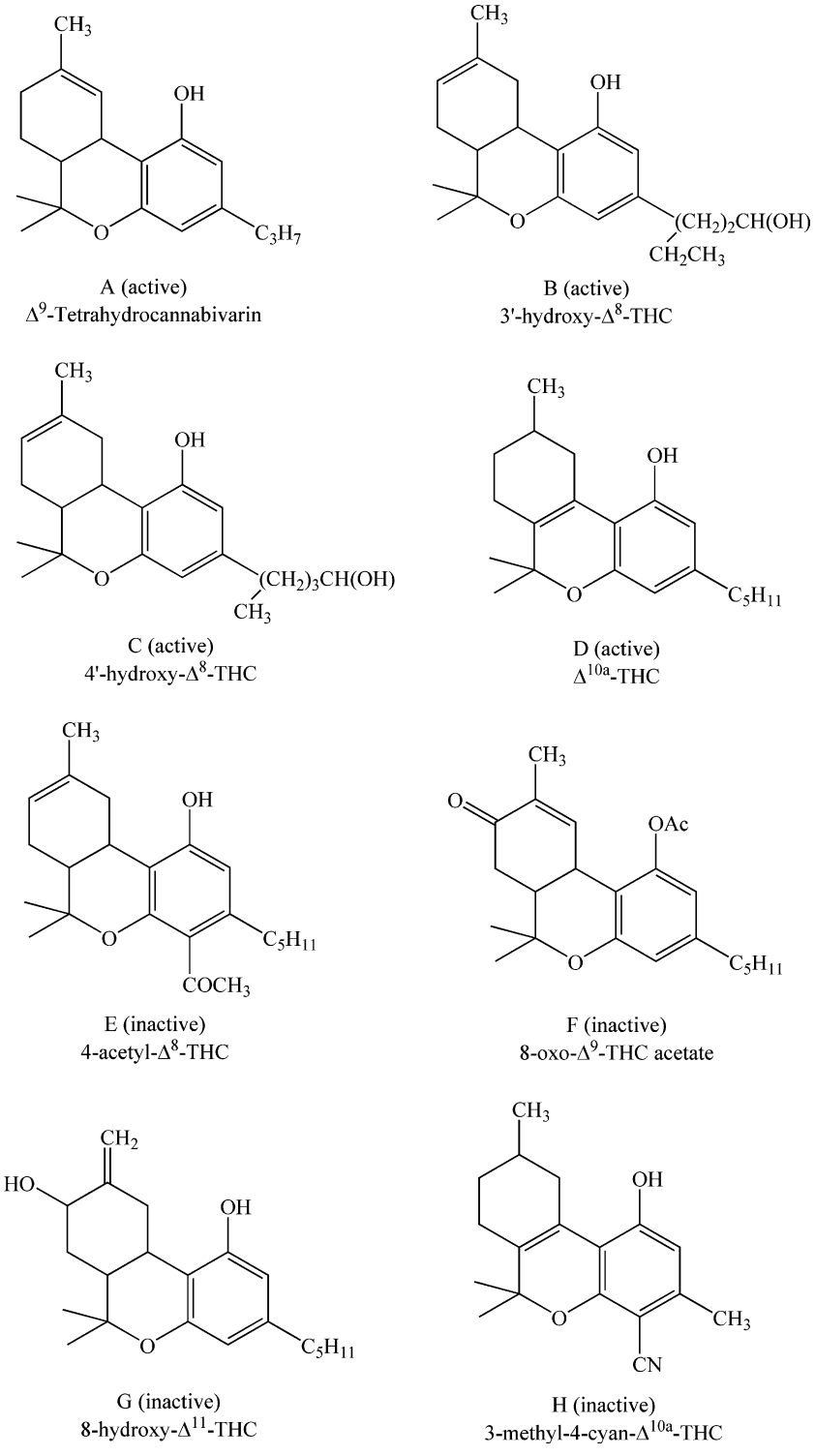

Fig. 3 Chemical structure of the eight cannabinoid molecules used in the prediction study

order to perform our PCA analysis with the calculated molecular descriptors, it was necessary to reduce their number by choosing the most relevant ones, from an initial 116 (see Table 1). Such a large number of variables can cause difficulties with PCA analysis. The reduction of the number of descriptors was carried out using the Fisher's weight [10, 14] and a correlation matrix between the calculated variables. PCA can be used to reduce the number of descriptors, but we decided to use, preliminarily, the Fisher's weight because this statistic takes into account the variance of the calculated descriptors into each class of compounds (active and inactive) giving an indication of the discriminatory power (i.e., the relative importance) of each descriptor. The values of the Fisher's weight for the calculated variables are presented in Table 1.

After reducing our initial set of 116 descriptors (Table 1), we selected only 20 variables that showed significant weight values, i.e. gave a Fisher's weight above 0.60. These variables were those that possessed a higher ability in the discrimination (separation) between psychoactive and psychoinactive molecules.

From the twenty selected variables obtained by using the Fisher's weight, we also attained the correlation matrix among them (see Table 2). From Table 2 we can see that some variables are correlated to each other (we considered correlated variables as only those that possess correlation coefficients above 0.70), and according to the results showed in Tables 1 and 2 (Fisher's weight and correlation matrix) only a remaining eight variables were considered important for the separation between active and inactive compounds. These variables were: μ (dipole moment), E_{LUMO} (energy of the lowest unoccupied molecular orbital), Log P (logarithm of the partition coefficient), $E_{\mathrm{L+1}}$ (energy of the lowest unoccupied molecular orbital above LUMO), Q4′ (total charge of the substituent at the C4 position), VC4 (volume of the substituent at the C4 position), E1m (variable regarding the atomic masses) and Lovasz–Pelikan index (LPI)—molecular branching index.

PCA results

Using the eight variables selected by the Fisher's weight and the correlation matrix, we carried out several PCA analyses with the aim of selecting the most important descriptors for the discrimination between active and inactive cannabinoid compounds. After that, the best discrimination was obtained by using four out of the 116 variables calculated initially: E_{LUMO}, Log P, VC4 and LP1. The values for E_{LUMO}, Log P, VC4 and LP1 in the PCA analysis are presented in Table 3.

In Fig. 4 we can see the score plot which presents the coordinates (scores) of each cannabinoid molecule in the new coordinate system based on PCs. From Fig. 4 one can see the good separation obtained with PCA between the two groups of compounds (psychoactive and psychoinactive). The PCA results show that the three first PCs describe 95.14% of the overall information (variance) of the data set (see Table 4). The first PC is responsible for 55.39% of the total information and it is mainly responsible for the separation between the active and inactive cannabinoid compounds.

Figure 5 shows the eigenvalues (loadings) obtained to each one of the variables used in the PCA analysis. The results obtained showed that the variables could be grouped in four distinct classes of variables: electronic (E_{LUMO}), hydrophobic (Log P), steric (VC4) and topological (LP1). This can be interpreted as an indication that each class of molecular property contributes for the biological activity presented by the cannabinoid compounds. In fact, we can say that the electronic property (E_{LUMO}) would be responsible for short-range electrostatic interactions that are established between the drug and the biological receptor. The hydrophobic property (Log P) would be responsible for the transport of the drug through membrane and contribute to the hydrophobic interactions in the

Table 1 Values of the Fisher's weight (W_{Fish}) for all calculated variables

Variable	W_{Fish}	Variable	W_{Fish}	Variable	W_{Fish}	Variable	W_{Fish}
ΔH_{f}	0.992	qC_8	0.086	NHO	0.050	CIC	0.103
E_{e}	0.249	qC_9	0.138	NHD	0.050	BIC	0.096
E_{T}	0.541	qS_1	0.095	NHA	1.366	1 K	0.147
IP	0.567	qS_2	0.018	ISIZ	0.211	2 K	0.070
μ	0.675	qX_1	0.478	MIC	1.355	PHI	0.055
E_{HOMO}	0.567	qX_2	0.161	IAC	0.289	Sv	0.000
E_{LUMO}	0.636	L_1	0.502	AAC	1.654	Se	0.000
H–it L	0.346	L_2	0.011	ZM1	0.249	Sp	0.005
χ	0.673	L_3	0.166	ZM2	0.331	Ss	0.961
η	0.347	L_4	0.017	CHL0	0.231	MDN	1.039
S	0.320	L_5	1.294	CHL1	0.207	MAXDP	0.372
A	0.009	A_1	0.020	CHL2	0.191	DELS	1.472
V	0.008	A_2	0.070	CHL0A	0.390	HYF	0.162
E_{Hid}	0.323	A_3	0.102	CHL1A	0.079	L1u	0.038
Log P	0.903	A_4	0.162	CHL2A	0.367	L2u	0.370
MR	0.008	A_5	0.062	ROUV	0.075	P1u	0.081
α	0.001	VC4	1.011	WIT	0.075	P2u	0.288
$E_{\mathrm{H\text{-}1}}$	0.416	VC2	0.000	WIA	0.090	L1m	0.138
$E_{\mathrm{L+1}}$	0.831	VC1	0.062	FP	0.274	L2m	0.554
Q_1	0.085	θ_1	0.057	GSI	0.274	P1m	0.218
Q_2	0.006	θ_2	0.531	BAL	1.345	P2m	0.477
Q_3	0.039	θ_3	0.315	BAC	0.382	E1m	0.864
$Q_{2'}$	0.143	NAT	0.012	ICEN	0.144	E2m	0.381
$Q_{4'}$	0.770	SKC	0.212	IDE	0.466	E1v	0.466
A_{D}	0.343	NBO	0.212	IDM	0.272	E1e	0.385
qC_1	0.075	NH	0.240	IDDE	0.025	E1p	0.378
qC_2	0.027	NC	0.008	IDDM	0.217	L1s	0.148
qC_3	0.003	NO	1.366	IC	0.330	L2s	0.679
qC_4	0.582	NCO	0.849	SIC	0.145	LP1	2.623

Table 2 Correlation matrix between the selected variables obtained by using the Fisher's weight

	ΔH_f	μ	E_{LUMO}	χ	Log P	E_{L+1}	$Q_{4'}$	L_5	VC4	NO
ΔH_f	1.00	−0.77	0.48	−0.58	0.44	−0.40	0.02	−0.29	−0.43	−0.96
μ		1.00	−0.47	0.55	−0.59	0.33	0.01	0.08	0.17	0.80
E_{LUMO}			1.00	−0.98	0.65	−0.40	0.43	−0.15	−0.16	−0.67
χ				1.00	−0.70	0.44	−0.36	0.13	0.17	0.75
Log P					1.00	−0.39	0.18	0.00	−0.01	−0.63
E_{L+1}						1.00	−0.23	0.26	0.28	0.44
$Q_{4'}$							1.00	0.95	0.30	0.30
L_5								1.00	0.95	0.30
VC4									1.00	0.41
NO										1.00
	NCO	NHA	MIC	AAC	BAL	Ss	MDN	DELS	E1m	LP1
ΔH_f	−0.69	−0.96	−0.97	−0.89	−0.73	−0.95	−0.89	−0.92	0.35	−0.74
μ	0.67	0.80	0.80	0.81	0.61	0.77	0.79	0.83	−0.30	0.60
E_{LUMO}	−0.92	−0.67	−0.66	−0.72	−0.47	−0.63	−0.56	−0.75	0.49	−0.49
χ	0.94	0.75	0.75	0.80	0.48	0.71	0.65	0.83	−0.47	0.49
Log P	−0.61	−0.63	−0.61	−0.78	−0.34	−0.42	−0.55	−0.64	0.60	−0.32
E_{L+1}	0.43	0.44	0.45	0.46	0.34	0.45	0.40	0.48	−0.35	0.39
$Q_{4'}$	0.30	0.30	0.30	0.25	0.57	0.30	0.21	0.26	−0.31	0.67
L_5	0.30	0.30	0.30	0.25	0.57	0.30	0.21	0.26	−0.31	0.67
VC4	0.36	0.41	0.42	0.33	0.66	0.43	0.26	0.36	−0.34	0.72
NO	0.83	1.00	1.00	0.97	0.76	0.95	0.88	0.98	−0.48	0.75
NCO	1.00	0.83	0.84	0.83	0.68	0.83	0.66	0.88	−0.51	0.67
NHA		1.00	1.00	0.97	0.76	0.95	0.88	0.98	−0.48	0.75
MIC			1.00	0.96	0.76	0.96	0.87	0.98	−0.47	0.76
AAC				1.00	0.69	0.87	0.90	0.96	−0.54	0.69
BAL					1.00	0.76	0.61	0.74	−0.61	0.94
Ss						1.00	0.83	0.96	−0.34	0.76
MDN							1.00	0.90	−0.34	0.62
DELS								1.00	−0.46	0.75
E1m									1.00	−0.57
LP1										1.00

Table 3 Calculated values for the most important variables used in the chemometric analyses

Compound		E_{LUMO} (eV)	Log P	VC4 (Å^3)	LP1
Ac1	Active	0.412	5.346	71.70	2.551
Ac2	Active	0.403	5.346	71.70	2.551
Ac3	Active	0.399	5.814	71.70	2.564
Ac4	Active	0.417	6.210	71.70	2.568
Ac5	Active	0.399	5.814	71.70	2.564
Ac6	Active	0.407	6.210	71.70	2.568
Ac7	Active	0.407	5.743	71.70	2.551
Ac8	Active	0.402	5.743	71.70	2.551
Ac9	Active	0.415	5.677	71.70	2.554
Ac10	Active	0.413	6.007	71.70	2.556
Ac11	Active	0.412	6.800	71.70	2.556
Ac12	Active	0.428	5.667	71.70	2.554
Ac13	Active	0.428	6.800	71.70	2.556
Ic1	Inactive	0.427	5.814	151.80	2.569
Ic2	Inactive	−0.328	5.045	71.70	2.573
Ic3	Inactive	0.008	5.045	180.35	2.580
Ic4	Inactive	−0.244	5.076	71.70	2.574
Ic5	Inactive	0.044	5.076	240.07	2.582
Ic6	Inactive	0.423	6.210	211.55	2.574
Ic7	Inactive	−0.522	5.091	240.54	2.594
Ic8	Inactive	0.433	6.210	211.59	2.574
Ic9	Inactive	0.413	5.814	152.12	2.569
Ic10	Inactive	0.224	4.505	71.70	2.571
Ic11	Inactive	0.337	4.433	71.70	2.560
Ic12	Inactive	−1.431	4.105	92.73	2.569
Ic13	Inactive	0.059	4.655	71.70	2.573

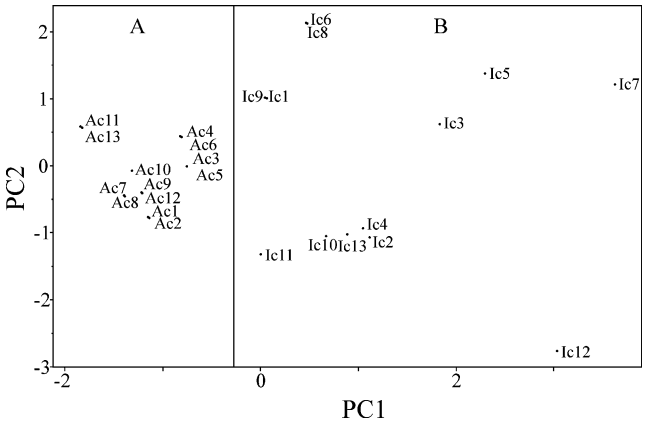

Fig. 4 Score plot for the separation between the psychoactive (*group A*) and psychoinactive (*group B*) cannabinoid compounds studied

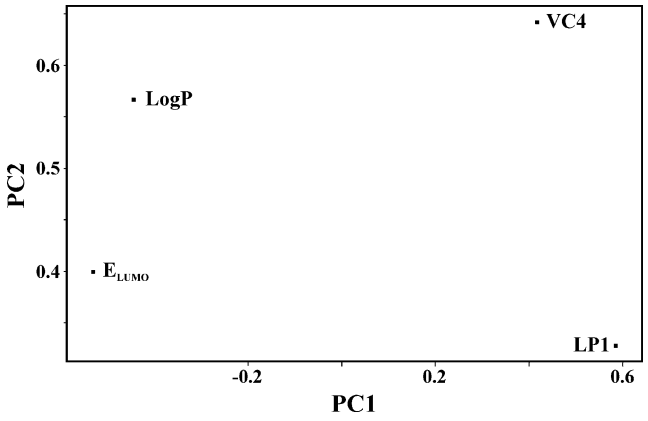

Fig. 5 Loading plot using only the variables responsible for separation between the psychoactive (*group A*) and psychoinactive (*group B*) cannabinoid compounds studied: E_{LUMO} (energy of the lowest unoccupied molecular orbital), Log *P* (logarithm of the partition coefficient), VC4 (volume of the substituent at the C4 position) and LP1 (Lovasz–Pelikan index)

formation of the drug-receptor complex. The steric and topological properties (VC4 and LP1) can be related to the three-dimensional complementarity between the drug and its receptor-binding site [32]. The property LP1 is defined by Lovasz and Pelikan [33] as a molecular branching index.

In Table 5 we can see the numerical values of the loadings (coefficients) of the four selected variables for the construction of the two first PCs (PC1 and PC2). From Table 5 we can write PC1 as follows:

$$\mathrm{PC1} = -0.530\,E_{\mathrm{LUMO}} - 0.444\,\mathrm{Log}\,P + 0.424\,\mathrm{VC4} + 0.585\,\mathrm{LP1} \quad (1)$$

From the PC1 values in Table 6 we can see that the active compounds have PC1 $<$ 0 and the inactive compounds have PC1 $>$ 0. From Eq. 1 we can also see that E_{LUMO} and Log *P* contribute for PC1 negatively whereas VC4 and LP1 contribute for PC1 positively. In this way, it is possible to say that the contribution of all four variables (multiplied by their coefficients in Eq. 1) determines the presence or absence of psychoactivity in cannabinoid compounds. Thus, for a cannabinoid compound to present psychoinactivity, VC4 and LP1 need to present high values (i.e. the compounds need to present a bulky substituent at the C4 position and several branchings in the overall structure) whereas the other variables (E_{LUMO} and Log *P*) need to have low values, as PC1 needs to present a positive value (PC1 $>$ 0). For the active compounds, E_{LUMO} and Log *P* need to have high values and VC4 and LP1 need to have low values, as PC1 needs to present a negative value (PC1 $<$ 0). Therefore, in order to design new psychoinactive cannabinoid compounds it is necessary to consider that the candidate molecule presents high values for VC4 and LP1 with low values for E_{LUMO} and Log *P*.

HCA results

In our HCA study, the incremental technique [34] was used for the calculation of distances among the compounds and the dendrogram obtained with the selected descriptors (E_{LUMO}, Log *P*, VC4 and LP1) is presented in Fig. 6. The vertical lines in Fig. 6 correspond to the cannabinoid compounds and the horizontal lines correspond to the similarity values between pair of compounds, a compound and a class of compounds and between the classes of compounds.

From Fig. 6 we can see that the better separation between the two groups of cannabinoid compounds, i.e. psychoactive (group A) and psychoinactive (group B), was obtained only using the descriptors E_{LUMO}, Log *P*, VC4 and LP1, exactly the same descriptors obtained in our PCA study. The similarity value between the two classes of compounds was 0.0 and this means these two classes are distinct. From Fig. 6 we can also see that the HCA results are very similar to those obtained in our PCA study, i.e. the compounds studied were grouped

Table 4 Variance of the data set for the four PCs

Component	Percentage	Cumulative
PC1	55.39	55.39
PC2	31.40	86.79
PC3	8.35	95.14
PC4	4.86	100.00

Table 5 Loadings of the most important variables in the first two PCs

Variable	PC1	PC2
E_{LUMO}	−0.530	0.400
Log *P*	−0.444	0.567
VC4	0.424	0.641
LP1	0.585	0.328

Table 6 PC1 score values for each cannabinoid compound studied

Compound		PC1
Ac1	Active	−1.147
Ac2	Active	−1.136
Ac3	Active	−0.750
Ac4	Active	−0.818
Ac5	Active	−0.750
Ac6	Active	−0.805
Ac7	Active	−1.396
Ac8	Active	−1.390
Ac9	Active	−1.206
Ac10	Active	−1.311
Ac11	Active	−1.819
Ac12	Active	−1.216
Ac13	Active	−1.839
Ic1	Inactive	0.043
Ic2	Inactive	1.119
Ic3	Inactive	1.837
Ic4	Inactive	1.047
Ic5	Inactive	2.301
Ic6	Inactive	0.480
Ic7	Inactive	3.628
Ic8	Inactive	0.467
Ic9	Inactive	0.063
Ic10	Inactive	0.674
Ic11	Inactive	0.003
Ic12	Inactive	3.035
Ic13	Inactive	0.888

into two categories: psychoactive compounds (labeled Ac1–Ac13 in Fig. 2) and psychoinactive compounds (labeled Ic1–Ic13 in Fig. 2). Thus, we can consider the results obtained with HCA as a confirmation of those obtained with PCA.

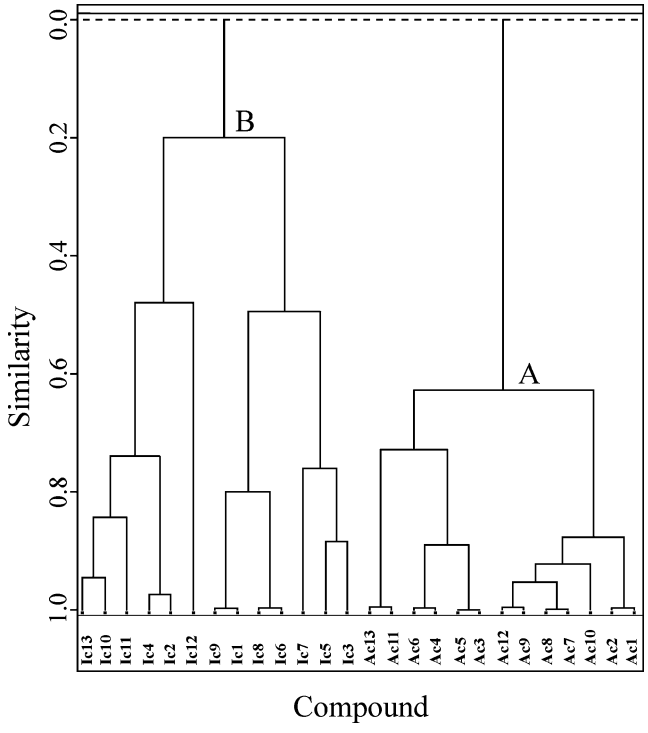

Fig. 6 Dendrogram obtained for the separation between the psychoactive (*group A*) and psychoinactive (*group B*) cannabinoid compounds studied

KNN results

This method was used for the validation of the initial data set (the 26 cannabinoid compounds) and we used the same variables employed in the PCA and HCA analyses (E_{LUMO}, Log *P*, VC4 and LP1). Table 7 presents the KNN results obtained with 1, 3 and 5 nearest neighbors. For the cases of 1 and 3 nearest neighbors (1NN and 3NN, respectively), the percentage of correct information is 100.0%, whereas with five nearest neighbor (5NN) this percentage is lower (96.2%). We decided to use three nearest neighbor (3NN) instead of 5NN because the percentage of correct information is higher (100.0%), and we used 3NN instead of 1NN because the greater the number of nearest neighbors, the better the reliability of the KNN method.

The results obtained with the KNN method were similar to those obtained with PCA and HCA showing that the outcomes obtained with the three chemometric methods (PCA, HCA and KNN) were informative, as the three methodologies classified the cannabinoid compounds into two classes (psychoactive and psychoinactive) exactly of the same way. From these results we can say that one is able to have a high percentage of success in classifying a new cannabinoid compound as psychoactive or psychoinactive by using our PCA, HCA and KNN models.

Remarks on the variables selected with the PCA, HCA and KNN methods

The main goal of this work was to find the properties responsible for the psychoinactivity of cannabinoid compounds in order to help in the design of new cannabinoid molecules without psychoactivity that can be used as therapeutic agents. It is important, therefore, to make some comments on the variables that were found responsible for the separation between the psychoactive and psychoinactive compounds and to draw some conclusions from our chemometric study:

(a) From our psychoinactive PC1 equation (PC1 needs to be positive for this kind of compounds, i.e. PC1

Table 7 Classification obtained with the KNN method

Category	Number of compounds	Compounds incorrectly classified		
		1NN	3NN	5NN
Active	13	0	0	0
Inactive	13	0	0	1
Total	26	0	0	1
Percentage of correct information		100.0	100.0	96.2

> 0), we could see that Log P needs to present low values. From the definition of partition coefficient (partition of the compound between organic and water phases) we can say that the psychoinactive compounds studied present a high hydrophilic character (i.e. with low lipophilicity) than the psychoactive ones. This indicates a low capacity of crossing biological membranes and, consequently, not able to reach the biological receptor since the biological membrane is constituted mainly by lipid containing cells.

(b) The energies of the frontier orbitals are important properties in several chemical and pharmacological processes. The reason for this is the fact that these properties give information on the electron-donating and electron-accepting character of a compound and, consequently, on the formation of a charge transfer complex (CTC). So, the energy of the highest occupied molecular orbital (E_{HOMO}) measures the electron-donating character of a compound, and the higher the E_{HOMO}, the higher the electron-donating capability of the compound [35]. Also, the energies of the lowest unoccupied molecular orbital (E_{LUMO}) measure the electron-accepting character of a compound and the lower the E_{LUMO}, the higher the electron-accepting capability of the compound [35]. Analyzing the E_{LUMO} for the psychoinactive compounds, we observed that the E_{LUMO} values need to have low values since PC1 needs to present a positive value. In other words, the psychoinactive compounds will have a pronounced electron-accepting character (low E_{LUMO} values) and, consequently, a charge transfer reaction is likely to occur between the psychoinactive compounds and the biological receptor-binding site.

(c) From our PC1 equation we can notice, as pointed out previously, that the psychoinactive cannabinoid compounds need to have PC1 > 0. Since VC4 and LP1 are the variables in the PC1 equation that have positive coefficients, we can see that VC4 and LP1 are the key variables that make PC1 > 0. Thus, we can conclude that our PCA model shows that steric and topological properties (VC4 and LP1, respectively) have a dominant role in the psychoinactivity mechanism of the cannabinoid compounds studied in this work, as the compounds that present a bulky substituent at the C4 position (high VC4) and several branchings in the overall structure (high LP1) will probably be psychoinactive.

(d) Perhaps one of the most interesting outcomes of this work is the fact that VC4 is a key variable responsible for the separation between the psychoactive and psychoinactive cannabinoid compounds under study. In previous works [36, 37], we have already paid attention to the fact that VC4 could have an important role in the inhibition of the psychoactivity mechanism of cannabinoid compounds. The main effect observed by Honorio et al. [36] and Honorio and da Silva [37] was the influence of the size of substituents at the C4 position, as this variable could give support to the idea that substitutions at this region could cause loss of activity in cannabinoid molecules due to steric hindrances.

It is interesting to see that the importance of VC4 in the psychoactivity of cannabinoid compounds (described previously just using quantum chemical calculations [36, 37]) is also observed in this work through chemometric methods (PCA, HCA and KNN) by using a large number of molecular variables (116 variables). In fact, we have found in this work mathematical evidences (the PC1 equation and the HCA and KNN results) that show VC4 is an important variable to be considered when one is trying to study (understand) the psychoactivity of cannabinoid compounds.

In order to validate our PCA, HCA and KNN results obtained with our training set (the 26 cannabinoid compounds) we decided to perform a prediction study with the aim to test our PCA, HCA and KNN models in classifying a set of eight new cannabinoid compounds with known psychoactivity (see Fig. 3) [16–20]. The prediction results obtained with this new set of cannabinoid compounds are summarized in Table 8. It is interesting to see that all of the eight new cannabinoid compounds were correctly classified according to our PCA, HCA and KNN models. However, it is important to notice that the compounds B and C were incorrectly classified by HCA; but, as two of the three models (PCA and KNN) classified compounds B and C correctly and only HCA classified them incorrectly, it is reasonable to consider compounds B and C as psychoactive molecules.

Conclusions

According to our PCA, HCA and KNN studies, we have shown that four variables can be considered important to discriminate psychoactive and psychoinactive cannabinoid compounds: E_{LUMO} (energy of the lowest unoccupied molecular orbital), Log P (logarithm of the partition coefficient), VC4 (volume of the substituent at the C4 position) and LP1 (Lovasz–Pelikan index—a molecular branching index).

Table 8 The psychoactivity prediction results obtained with the PCA, HCA and KNN methods for a set of eight new cannabinoid compounds: active (+) and inactive (−)

Compound	PCA	HCA	KNN
A (active)	+	+	+
B (active)	+	−	+
C (active)	+	−	+
D (active)	+	+	+
E (inactive)	−	−	−
F (inactive)	−	−	−
G (inactive)	−	−	−
H (inactive)	−	−	−

According to our PCA study, a cannabinoid compound will be psychoinactive when it presents high values for VC4 and LP1 and low values for E_{LUMO} and Log *P*. Therefore, in order to design new psychoinactive cannabinoids it is necessary that the candidate molecule presents high values for VC4 and LP1 and low values for E_{LUMO} and Log *P*.

Our PCA model also showed that steric and topological properties (VC4 and LP1, respectively) have an important role in the psychoinactivity mechanism of the cannabinoid compounds studied in this work, as the molecules that present a bulky substituent at the C4 position (high VC4) and several branchings in the overall structure (high LP1) will probably be psychoinactive.

The four variables (E_{LUMO}, Log *P*, VC4 and LP1) responsible for the discrimination between psychoactive and psychoinactive cannabinoid compounds belong to four different classes: electronic (E_{LUMO}), hydrophobic (Log *P*), steric (VC4) and topological (LP1), indicating that the interaction between the cannabinoid compounds studied and the biological receptor can occur due to electronic, hydrophobic, steric and topological features of the compounds.

From the prediction study performed with our PCA, HCA and KNN models, we can conclude that these models are able to give reliable information on the psychoactivity of cannabinoid compounds, as the eight new cannabinoid molecules (with known psychoactivity) were correctly classified by our PCA, HCA and KNN models.

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

References

1. Mechoulam R, Ben-Shabat S (1999) Nat Prod Rep 16:131–143
2. Joy JE, Watson SJ, Benson JA (1999) Marijuana and medicine: assessing the science base. National Academy Press, New York
3. Reggio PH, Greer KV, Cox SM (1989) J Med Chem 32:1630–1635
4. Reggio PH, Panu AM, Miles S (1993) J Med Chem 36:1761–1771
5. Mechoulam R, Hanus L (2000) Chem Phys Lipids 108:1–13
6. Montanari CA (1995) Quím Nova 18:56–64
7. Martin YC (1981) J Med Chem 24:229–237
8. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) J Am Chem Soc 107:3902–3909
9. Kowalski BR, Bender CF (1972) J Am Chem Soc 94:5632–5639
10. Bruns RE, Faigle JF (1985) Quim Nova 8:84–99
11. Massart DL, Vandeginste BGM, Dening SN, Michotte Y, Kaufman L (1988) Chemometrics: a textbook. Elsevier, Amsterdam
12. Malinowski ER (1991) Factor analysis in chemistry, 3rd edn. Wiley, New York
13. Kowalski BR (1984) Chemometrics: mathematics and statistics in chemistry. D Reidel Publishing Company, Cosenza
14. Sharaf MA, Illman DL, Kowalski BR (1986) Chemometrics. Willey, New York
15. Beeb KR, Pell RJ, Seasholtz MB (1998) Chemometrics: a practical guide. Willey, New York
16. Mechoulam R (1973) Marijuana: chemistry, pharmacology, metabolism and clinical effects. Academic Press, New York
17. Petersen RC (1980) Marijuana research findings. Department of Health and Human Services, Maryland
18. Mechoulam R (1970) Science 168:1159–1166
19. Da Silva ABF, Trsic M (1995) J Mol Struct 356:247–256
20. Razdan RK (1986) Pharmacol Rev 38:75–149
21. Allinger NL, Yuh YH, Lin JH (1989) J Am Chem Soc 111:8551–8566
22. Ostlund NS (1995) HyperChem: program for molecular visualization and simulation. University of Waterloo, Canada
23. Ostlund NS (1994) Hypercube. ChemPlus: extensions for hyperchem. Ontario, Canada
24. Dewar MJS (1994) AMPAC 6.5: program for semi-empirical calculations. University of Texas, USA
25. Hehre WJ, Huang WW, Klunzinger PE, Deppmeier BJ, Driessen AJ (1996) Spartan 5.0, user's guide. Irvine, CA
26. Todeschini R (1997) WHIM/3D-QSAR. Talete srl, Milano
27. Todeschini R, Gramatica P (1998) Persp Drug Discov Des 9:355–380
28. Jolliffe IT (1986) Principal component analysis. Springer-Verlag, New York
29. Johnson RA, Wichern DW (1992) Applied multivariate statistical analysis. Prentice-Hall, New Jersey
30. Lindon JC, Holmes E, Nicholson JK (2001) Prog Nucl Magn Reson Spectrosc 39:1–40
31. Dillon WR, Goldstein M (1984) Multivariate analysis. Wiley, New York
32. Tavares LC (1993) D Sc Thesis. Universidade de São Paulo, São Paulo
33. Lovasz L, Pelikan J (1973) Period Math Hung 3:175–182
34. Manly BFJ (1986) Multivariate statistical methods. Chapman& Hall, New York
35. Honorio KM, da Silva ABF (2003) Int J Quant Chem 95:126–132
36. Honorio KM, Freitas LG, Trsic M, da Silva ABF (2001) J Mol Struct (THEOCHEM) 538:99–106
37. Honorio KM, da Silva ABF (2002) J Mol Struct (THEOCHEM) 578:111–117