Available online at www.sciencedirect.com
ScienceDirect
European Journal of Medicinal Chemistry 43 (2008) 2197–2205
EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY
http://www.elsevier.com/locate/ejmech

Original article

Use of self-organizing maps and molecular descriptors to predict the cytotoxic activity of sesquiterpene lactones

Mariane B. Fernandes, Marcus T. Scotti, Marcelo J.P. Ferreira, Vicente P. Emerenciano*

Instituto de Química, Universidade de São Paulo, Caixa Postal 26077, 05513-970, São Paulo, SP, Brazil

Received 25 April 2007; received in revised form 23 November 2007; accepted 9 January 2008
Available online 25 January 2008

Abstract

Some sesquiterpene lactones (SLs) are the active compounds of a great number of traditionally medicinal plants from the Asteraceae family and possess considerable cytotoxic activity. Several studies in vitro have shown the inhibitory activity against cells derived from human carcinoma of the nasopharynx (KB). Chemical studies showed that the cytotoxic activity is due to the reaction of α,β-unsaturated carbonyl structures of the SLs with thiols, such as cysteine. These studies support the view that SLs inhibit tumour growth by selective alkylation of growth-regulatory biological macromolecules, such as key enzymes, which control cell division, thereby inhibiting a variety of cellular functions, which directs the cells into apoptosis.

In this study we investigated a set of 55 different sesquiterpene lactones, represented by 5 skeletons (22 germacranolides, 6 elemanolides, 2 eudesmanolides, 16 guaianolides and nor-derivatives and 9 pseudoguaianolides), in respect to their cytotoxic properties. The experimental results and 3D molecular descriptors were submitted to Kohonen self-organizing map (SOM) to classify (training set) and predict (test set) the cytotoxic activity.

From the obtained results, it was concluded that only the geometrical descriptors showed satisfactory values. The Kohonen map obtained after training set using 25 geometrical descriptors shows a very significant match, mainly among the inactive compounds (~84%). Analyzing both groups, the percentage seen is high (83%). The test set shows the highest match, where 89% of the substances had their cytotoxic activity correctly predicted. From these results, important properties for the inhibition potency are discussed for the whole dataset and for subsets of the different structural skeletons.

Keywords: Cytotoxic activity; Sesquiterpene lactones; Self-organizing maps; Molecular descriptors

1. Introduction

Although sesquiterpene lactones (SLs) are terpenoid compounds characteristic of the Asteraceae (Compositae), they can be also found in other angiosperm families [1]. In Asteraceae, differences in skeletal types and quantities of SLs in different genera and species have been utilized in chemotaxonomic studies [2–8].

The classification of SLs according to their carbocyclic skeleton would place most of them into one of four major groups: germacranolides (with a 10-membered ring); elemanolides (with a 6-member ring); eudesmanolides (6/6 = bicyclic compounds); guaianolides and pseudoguaianolides (both 5/7-bicyclic compounds) [2]. However, SLs exhibit a variety of other skeletal arrangements [9].

An important usual feature of the SLs is the presence of a γ-lactone ring (closed towards either C-6 or C-8) containing, in many cases, an α-methylene group. Among other modifications, the incorporation of hydroxyls or esterified hydroxyls and epoxide rings are common. A few SLs occur in glycoside form [10–12] and some contain halogens or sulphur [13–15] atoms. The wide variety of chemical structures is correlated with a large diversity of biological activities [1].

During the search for anti-tumour compounds from plants, many SLs active against various types of organisms and

* Corresponding author. Tel.: +55 11 3091 2058; fax: +55 11 3815 5579.
E-mail address: vdpemere@iq.usp.br (V.P. Emerenciano).

0223-5234/$ - see front matter © 2008 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2008.01.003

systems have been discovered. They form one of the largest groups of cytotoxic and anti-tumour compounds of plant origin. Most of these active compounds are found in Asteraceae species, although some of them are found too in the Magnoliaceae and Apiaceae families and even Fungi [1].

A majority of known active SLs showed cytotoxic activity (KB and P388 leukemia in vitro) and activity against in vivo p388 leukemia [16–19]. Relationship between chemical structure and cytotoxic activity of sesquiterpene lactones was investigated in many papers on numerous tumour models [20–32].

Chemical studies showed that various cytotoxic SLs react with thiols, such as cysteine, by rapid Michael-type addition [22,23,29]. These reactions are mediated chemically by α,β-unsaturated carbonyl structures, such as an α-methylene-γ-lactone, an α,β-unsaturated cyclopentenone or a conjugated ester [22,30]. These studies support the view that SLs inhibit tumour growth by selective alkylation of growth-regulatory biological macromolecules, such as key enzymes, which control cell division, thereby inhibiting a variety of cellular functions, which directs the cells into apoptosis [31,32].

Differences in activity between individual SLs may be explained by different numbers of alkylating structural elements [23,33–35]. However, other factors such as lipophilicity, molecular geometry and the chemical environment or the target sulfhydryl may also influence the activity of sesquiterpene lactones [21,33,35].

Previously, we presented a QSAR study on the cytotoxic activity of a considerable variety of structurally different SLs [36]. Multiple linear regression analysis revealed that some structures are very important to the biological activity, such as the double bond in the cyclopentane ring and the double bond at position 10, as well as the hydroxyl group at position 5 and the angeloyloxy group at C-8. With the study it could be also concluded that the most active compounds can be those which have the guaianolide and pseudoguaianolide skeletal types.

Some artificial neural network studies about SLs have already been developed targeting other biological activities, such as the anti-inflammatory property [37]. Artificial neural networks (ANN) are defined as computational models having structures derived from the simplified concept of the brain in which a number of nodes, called neurons, are interconnected in a network-like structure [38]. ANNs are robust and able to detect groupings and patterns that may be unclear even by a trained human expert. Since ANNs are not restricted to linear correlations and can also take into account non-linear data correlations, they can be efficiently applied for modeling, prediction and classification. The most used ANN architecture for pattern recognition and classification is the self-organizing maps (SOM) [39]. This procedure can map multivariate data onto a two dimensional grid, grouping similar patterns near each other [40].

Kohonen called his ANN as "self-organizing network" [41], which projects objects from a multidimensional space into a space of lower-dimensionality, usually into a 2D plane. In this projection, the similarity relationship between objects is conserved. Thus, in principle, Kohonen networks can be used for clustering of objects. So, SOMs accomplish two things, they reduce dimensions and display similarities, which makes the SOM a powerful visualization tool. The training of these networks (Kohonen) is unsupervised, that is, the investigated property is not used during the training process. Each neuron in the grid is associated to a weight and similar patterns stimulate neurons with similar weight, so that similar patterns are mapped near each other [40], helping the QSAR analysis by simplifying the visualization and understanding.

For instance, SOM was successfully used in several applications using chemistry database, such as: in classification of photochemical reactions [42]; in chemotaxonomy of the Asteraceae family [43,44]; for a series of 103 sesquiterpene lactones which showed anti-inflammatory activity [37]; in drug design [45]; for comparing dataset compounds [46]; in classification of metabolites [47] and in the prediction of the diterpene skeletons [48].

The aim of this paper is to develop a SOM model to predict the cytotoxic potency of 55 SLs improving the prediction of the inhibitory potency of unknown compounds and comparing to the previous QSAR study [36].

2. Materials and methods

2.1. Dataset

This work uses the same dataset of 37 SLs as in the previous QSAR study [36], adding more 18 new SLs, which had their biological activities measured in vitro using KB human tumour cell tissue culture (nasopharynx carcinoma) with DMSO as solvent [49–51]. These SLs represent five different skeletal types (Fig. 1) with a wide structural diversity: 22 germacranolides, 18 guaianolides, 9 pseudoguaianolides, 5 elemanolides and 2 eudesmanolides (Table 1).

All structural data were extracted from the SISTEMAT database [52–54]. An in-house program for data extraction was written in Fortran/Java and subsequently used to select the SL compounds. The structures are shown in Fig. 2. Molecular modeling computations were performed on SPARTAN for Windows v. 4.0 software (Wavefunction, Inc., Irvine, Calif.). The molecules were subjected to geometrical optimization and conformational analysis (systematic analysis with dihedral angle rotated at each 30°). Semi-empirical quantum chemical method used was AM1 (Austin Model 1) [55,56] and the root mean square (RMS) gradient value of 0.001 kcal/mol as termination condition. Energy minimized molecules were saved as MDL MolFiles for computing various molecular descriptors using DRAGON Professional version 5.4 [57].

2.2. Molecular descriptors

The physico-chemical properties as well as the biological activity of organic compounds depend on their molecular structures. With the purpose of obtaining relationships among chemical structures and biological activities using computational approaches, it is necessary to find appropriate representations of the molecular structure of the compounds. A

Table 1
Substances, their skeleton, biological activity and activity class

Number	Substance[a]	Skeleton	ED_{50} (μmol/L)	Activity class
1	Vernomenin (**12a**) [7]	Elemanolide	127	Inactive
2	Vernomenin acetate (**12b**) [7]	Elemanolide	26.2	Inactive
3	Vernolepin (**7a**) [7]	Elemanolide	6.52	Inactive
4	Costunolide (**11a**) [7]	Germacranolide	2.46	Active
5	Tamaulipin A (**11b**) [7]	Germacranolide	5.08	Inactive
6	Tamaulipin B (**11c**) [7]	Germacranolide	10.5	Inactive
7	Elephantol (**5a**) [7]	Germacranolide	123.2	Inactive
8	Coronopilin (**20a**) [7]	Pseudoguaianolide	5.49	Inactive
9	3-Hydroxydamsin (**20b**) [7]	Pseudoguaianolide	10	Inactive
10	Desacetylconfertiflorin (**20c**) [7]	Pseudoguaianolide	8.71	Inactive
11	Parthenin (**21a**) [7]	Pseudoguaianolide	1.3	Active
12	Ambrosin (**21b**) [7]	Pseudoguaianolide	1.83	Active
13	Aromaticin (**22a**) [7]	Pseudoguaianolide	1.38	Active
14	Mexicanin I (**23**) [7]	Pseudoguaianolide	1.26	Active
15	Helenalin (**22b**) [7]	Pseudoguaianolide	0.76	Active
16	Eupachlorin (**6**) [23]	Guaianolide	0.51	Active
17	Eupachloroxin (**13**) [23]	Guaianolide	8.39	Inactive
18	Vernolepin acetate (**12b**) [7]	Elemanolide	8.49	Inactive
19	Gaillardin (**13**) [7]	Guaianolide	7.52	Inactive
20	Eupatundin (**14**) [7]	Guaianolide	1.25	Active
21	Eupachlorin acetate (**15**) [7]	Germacranolide	0.35	Active
22	Chammissonin diacetate (**17**) [7]	Germacranolide	6.12	Inactive
23	Eupatocunin (**6**) [34]	Germacranolide	0.27	Active
24	Eupacunolin (**19**) [34]	Germacranolide	8.8	Inactive
25	Vernomygdin (**8**) [33]	Germacranolide	4.12	Active
26	Euparotin (**2**) [23]	Guaianolide	0.56	Active
27	Eupatoroxin (**7**) [23]	Guaianolide	7.14	Inactive
28	Eupatundin (**14**) [7]	Guaianolide	1.04	Active
29	10-Epieupatoroxin (**12**) [23]	Guaianolide	0.63	Active
30	Euparotin acetate (**16**) [7]	Guaianolide	0.53	Active
31	Elephantin (**1b**) [7]	Germacranolide	3.22	Active
32	Elephantopin (**1a**) [7]	Germacranolide	2.51	Active
33	Vernolepin methacrylate (**7c**) [7]	Elemanolide	1.22	Active
34	Eupacunoxin (**2**) [34]	Germacranolide	5	Inactive
35	Eupatocunoxin (**7**) [34]	Germacranolide	4.04	Active
36	Vernodalin (**1**) [33]	Elemanolide	5	Inactive
37	Liatrin (**18**) [7]	Germacranolide	3.93	Active
38	Xerantolide (**12**) [49]	Guaianolide	6.10	Inactive
39	Erivanin (**16**) [49]	Eudesmanolide	1.10	Active
40	Lactucin (**7**) [49]	Guaianolide	260.00	Inactive
41	Eupatolide (**I**) [48]	Germacranolide	5.00	Inactive
42	Deoxylactucin (**8**) [49]	Guaianolide	3.10	Active
43	Eupatoriopicrin (**II**) [48]	Germacranolide	3.30	Active
44	Vernoflexuoside (**9**) [49]	Guaianolide	19.00	Inactive
45	Vernoflexuoside tetraacetyl (**10**) [49]	Guaianolide	8.75	Inactive
46	Arctolide (**13**) [49]	Guaianolide	29.00	Inactive
47	Alatolide (**III**) [50]	Germacranolide	4.00	Active
48	Vernoflexin (**11**) [49]	Guaianolide	4.30	Active
49	Epoxynobilin (**4**) [49]	Eudesmanolide	0.99	Active
50	Hydroxyisonobilin (**2**) [49]	Germacranolide	3.70	Active
51	Hirsutolide (**6**) [49]	Germacranolide	7.00	Inactive
52	Salonitenolide (**3**) [49]	Germacranolide	9.10	Inactive
53	Linifolin A (**15**) [49]	Pseudoguaianolide	2.00	Active
54	Dehydronobilin (**1**) [49]	Germacranolide	2.70	Active
55	Parthenolide (**5**) [49]	Germacranolide	5.30	Inactive

[a] Original numbers, used in the references (between square brackets), are between parentheses.

molecular descriptor can be considered as a result of a logical and mathematical procedure, applied to chemical information codified through a molecule's representation [58].

DRAGON computer software was employed to calculate the molecular descriptors that were used for performing a Kohonen Neural Network. The groups of descriptors calculated were: RDF [59] (150 descriptors); 3D-MoRSE [60,61] (160 descriptors); GETAWAY [62,63] (197 descriptors); WHIM [64,65] (99 descriptors); geometrical [66–68] (74 descriptors); all 3D (680 descriptors); 2D autocorrelations [69–71] (96 descriptors); functional group counts [58] and atom-centered fragments [72,73] (274 descriptors).

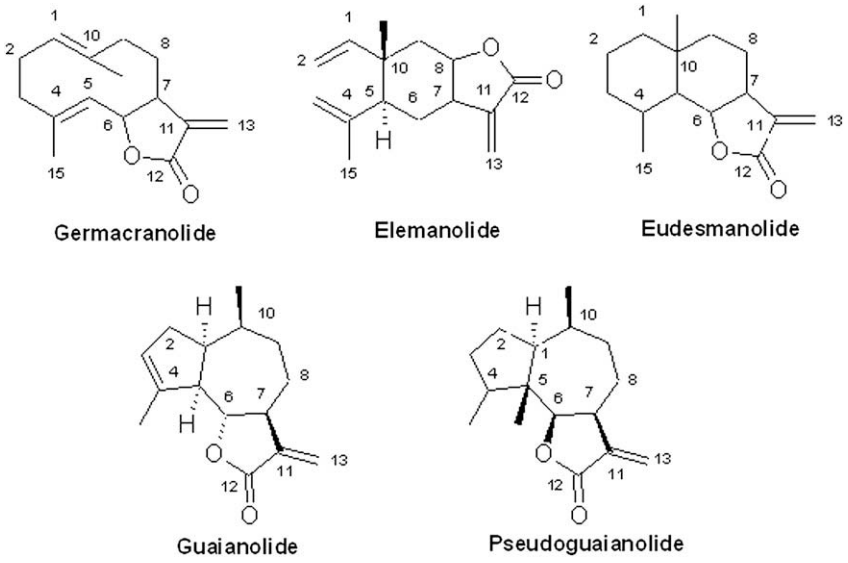

Fig. 1. Five skeletal types of sesquiterpene lactones used in this study: germacranolides (with a 10-membered ring), elemanolides (with a 6-member ring), eudesmanolides (6/6 = bicyclic compounds), guaianolides and pseudoguaianolides (both 5/7-bicyclic compounds).

For each block of descriptors, the constant variables were excluded, as well as those that presented only a different value of the series (near constant variable). For the remaining descriptors pair wise correlation ($r < 0.99$) analysis was performed to exclude those which are highly correlated [74]. Thus, the number of *DRAGON* descriptors used in our calculations was reduced to: RDF (71 descriptors); 3D-MoRSE (100 descriptors); GETAWAY (136 descriptors); WHIM (67 descriptors); geometrical (25 descriptors); all 3D (391 descriptors); 2D autocorrelations (34 descriptors); functional group counts and atom-centered fragments (33 descriptors).

2.3. Generation of Kohonen neural networks

A Kohonen-ANN was trained using Matlab 6.5 computing environment by Mathworks and SOM Toolbox 2.0 [75]. Matlab is a powerful and easy program to use scientific computing language and is the choice for most scientific simulation and data analysis. SOM toolbox is a set of Matlab functions that can be used to develop and implement SOM neural networks, which contains functions for creation, visualization and analysis of self-organizing maps. A SOM grid with square geometry and 9×3 dimension size was created and trained [76].

The training was conducted through the Batch-training algorithm. In this algorithm, the whole dataset is presented to the network before any adjustment is made. In each training step, the dataset is partitioned according to the regions of the map weight vectors. The dataset was divided according to the data that intends to classify. This kind of classification was made in order to reduce the intrinsic error in the dependent variable, dividing the compounds into two sets: active and inactive substances.

During the process of a self-organized map, two phases are necessary: the training and the test phases. This last one is extremely important because after several cycles of training, there must be a trend between the learning of the network and its ability of prediction that is verified in the results. The size of the test set was around 17% of the whole set. It is important to choose sets that contain representative samples of the training group, including all the ranges of the biological activity. When the test set contains the major characteristics present in the training set, the results offer a more robust assessment. Following this logic, six series of compounds were chosen and listed in Table 2.

Training and test performance are evaluated by computing the ratio of the number of samples correctly classified by SOM. A summary of the number of training and test sets is shown in Table 3.

3. Results and discussion

The set of descriptors earlier cited (2D autocorrelations, 3D-MoRSE, all the 3D, functional group counts and atom-centered fragments, geometrical, GETAWAY, RDF and WHIM) were used for the training set. Analyzing all the obtained results, it was concluded that only the geometrical descriptors showed satisfactory values, while the others presented a matching accuracy of less than 70%.

A summary of the obtained results using the geometrical descriptors, both as absolute values and as percentages, is shown in Table 4. Fig. 3 shows the obtained Kohonen map after training set 1 using 25 geometrical descriptors.

Table 4 presents a very significant match in the training set, mainly among the inactive compounds (~84%). Analyzing both groups, the percentage seen is high (83%). The test set 1 shows the highest match, where 89% of the substances in this set had their cytotoxic activity correctly predicted. Table 4 also shows that the inactive compounds were so correctly predicted (75%) as the active ones (73%). This could reveal that the features of the substances – encoded as descriptors – that

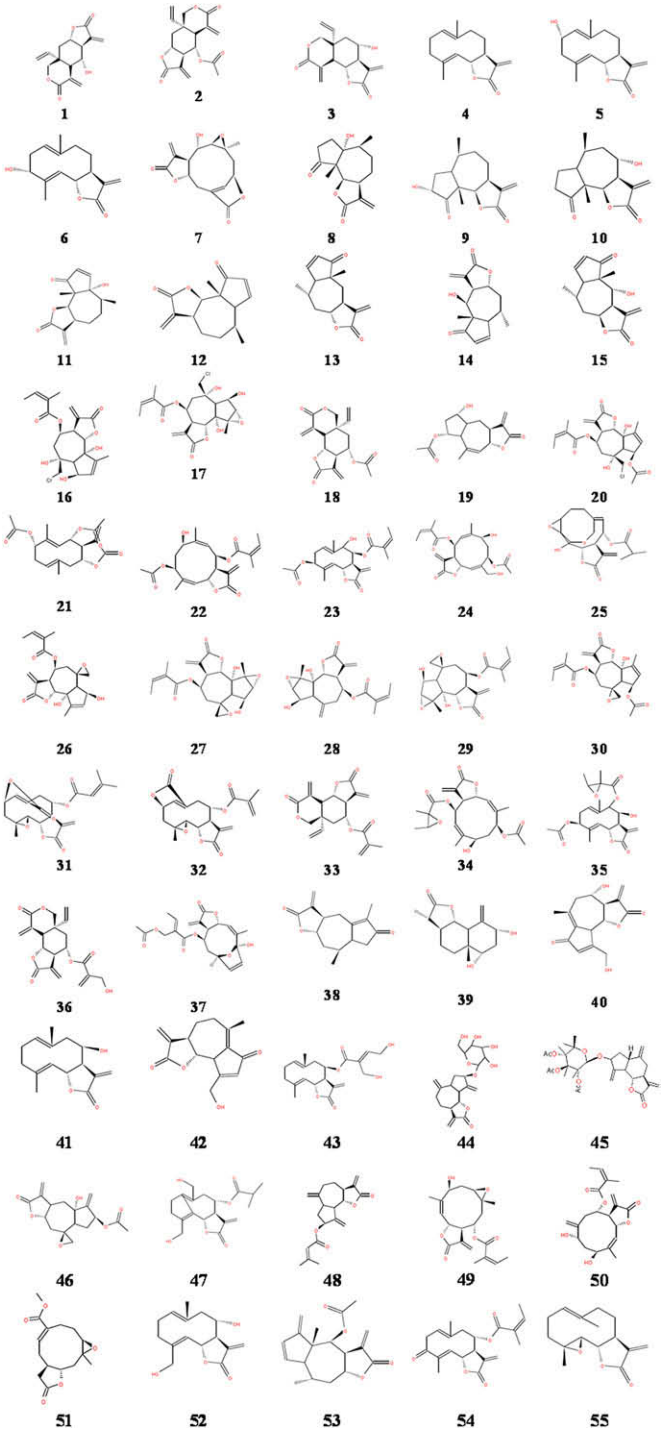

Fig. 2. The chemical structure of the 55 sesquiterpene lactones used for the training and the test sets in self-organizing maps.

form both groups can place them with the same efficiency. In other words, the components that show greatest correlation among clusters, therefore, the most influent in the Kohonen map, do not belong specifically to either group (e.g. actives, inactives).

A way to interpret the SOM map is to think that the training phase weights of the whole neighborhood are moved in the same direction, so similar items tend to excite adjacent neurons. Therefore, SOM forms a semantic map where similar samples are mapped close together and dissimilar apart.

Table 2
Compounds chosen for the six different test sets

	Test set
Set 1	1, 6, 9, 11, 20, 23, 26, 36, 54
Set 2	2, 8, 15, 25, 28, 35, 41, 44, 53
Set 3	7, 13, 17, 29, 33, 34, 43, 49, 50
Set 4	10, 14, 21, 38, 46, 47, 51, 52, 55
Set 5	3, 4, 16, 18, 30, 31, 40, 45, 48
Set 6	5, 12, 19, 22, 24, 27, 32, 37, 39, 42

Table 3
Summary of the six different training and test sets

		Training set		Test set		Total	
		Train	% Total	Test	% Total	Total	% Total
Set 1	Active	24	83	5	17	29	100
	Inactive	22	85	4	15	26	100
	Total	46	84	9	16	55	100
Set 2	Active	24	83	5	17	29	100
	Inactive	22	85	4	15	26	100
	Total	46	84	9	16	55	100
Set 3	Active	23	79	6	21	29	100
	Inactive	23	88	3	12	26	100
	Total	46	84	9	16	55	100
Set 4	Active	26	90	3	10	29	100
	Inactive	20	77	6	23	26	100
	Total	46	84	9	16	55	100
Set 5	Active	24	83	5	17	29	100
	Inactive	22	85	4	15	26	100
	Total	46	84	9	16	55	100
Set 6	Active	24	83	5	17	29	100
	Inactive	21	81	5	19	26	100
	Total	45	82	10	18	55	100

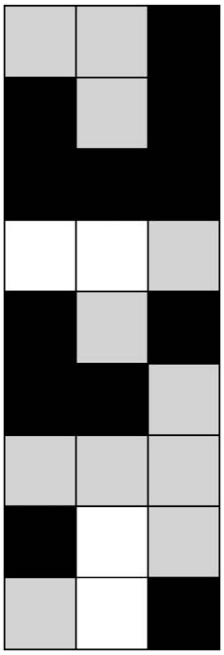

Fig. 3. Kohonen map obtained after the training phase of set 1 using 25 geometrical descriptors. The grey color represents the active substances, while the black represents the inactive ones.

In order to analyze a Kohonen map, set 1 was chosen since it represented our best results. Fig. 3, which shows the Kohonen map of set 1, depicts the analysis of the two groups of the training set. The black and grey squares represent the inactive and active compounds, respectively. It can be visualized that the top of the map shows a predominance of the inactive set and the bottom, a majority of active compounds. This distribution shows a satisfactory separation between both groups.

In Fig. 4, it can be seen the U-matrix of the set 1, which visualizes distances between neighboring map units, and thus shows the cluster structure of the map: high values of the U-matrix indicate a cluster border, uniform areas of low values indicate clusters themselves. Each component plane shows the values of one variable in each map unit [77]. It can also be seen the weight that each descriptor has in this Kohonen map. In a general way, the most representative descriptors are those that have two main characteristics:

a) The greatest weights in the inactive predominant region or in the active one.
b) A considerable difference between the highest and lowest descriptors' values.

The following descriptors appear to be most important for active substances: SPAN, G(O…O), H3D, AGDD, G1, RGyr, QXXm, QXXv and QZZv; as well as to the inactive ones: DISPm and DISPe [58]. An important fact to be highlighted is the presence of the two first descriptors related to the active compounds (SPAN e G(O…O)). These are included among the five descriptors that appear in the equation calculated in our last QSAR study. The importance of these descriptors is emphasized regarding the cytotoxic activity of these SLs. Besides, all kinds of descriptors in this Kohonen map, especially the geometrical descriptors, certify the fact that the sterical properties and electronic features, as concluded in latter studies and so in our QSAR study, have a great importance when they are related to this biological activity.

The descriptors cited previously were useful in the perception of certain characteristics present in the substances, which were able to differentiate them between active and inactive compounds with good efficiency. So, one can point out that the obtained map could be used for screening the respective biological activity in a virtual database, i.e. the map could

Table 4
Summary of the training and test match results (%) using 25 geometrical descriptors

	Train set 1	Train set 2	Train set 3	Train set 4	Train set 5	Train set 6	Average	Standard deviation
Map size	9 × 3	9 × 3	7 × 6	10 × 2	10 × 2	9 × 2	–	–
Active match	79	79	91	81	79	79	81	5
Inactive match	91	91	78	80	91	71	84	9
Total	85	85	85	80	85	76	83	4
	Test set 1	Test set 2	Test set 3	Test set 4	Test set 5	Test set 6	Average	Standard deviation
Active match	80	60	83	67	80	80	75	9
Inactive match	100	100	33	67	75	60	73	26
Total	89	78	67	67	78	70	75	9

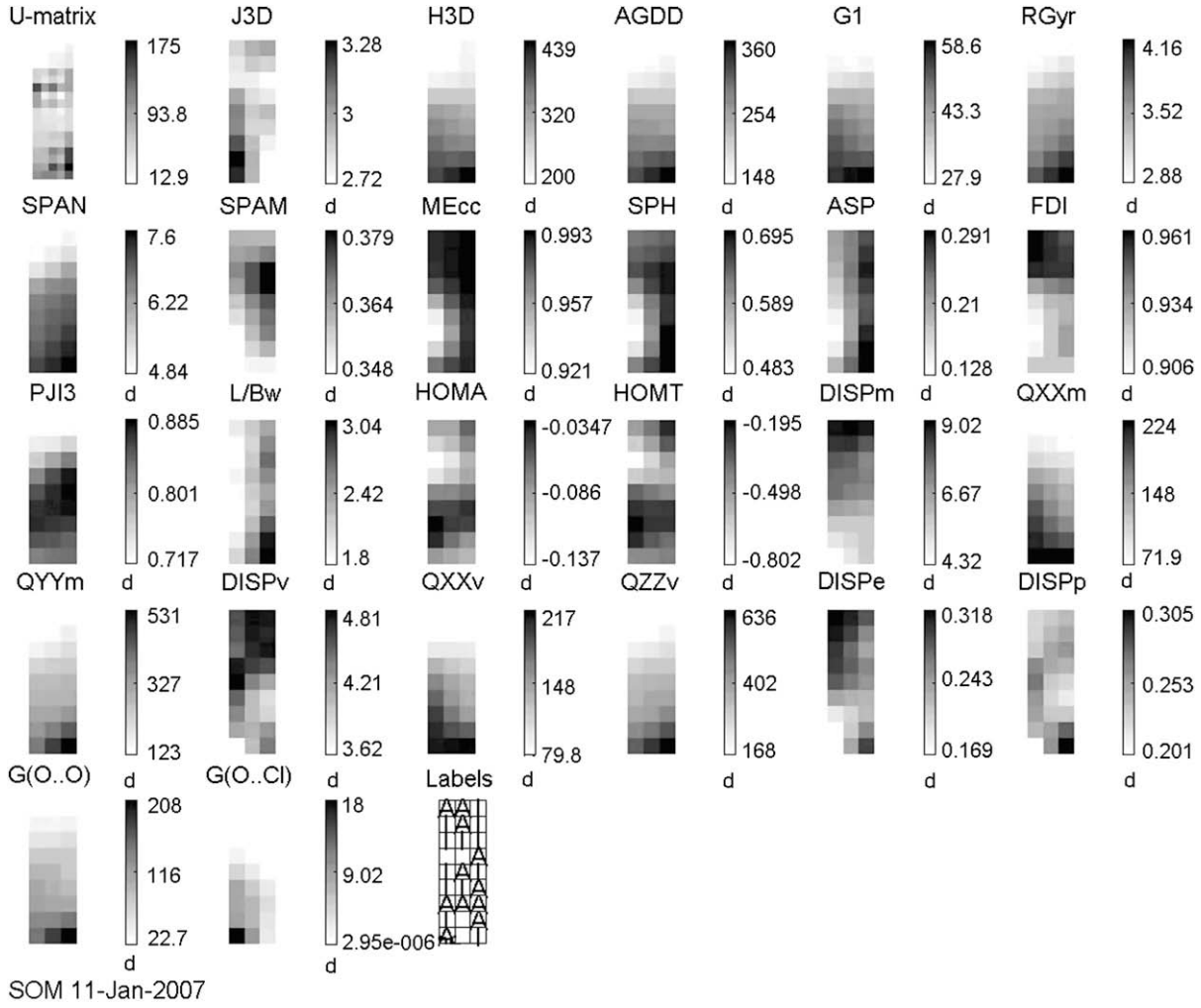

Fig. 4. The 25 descriptors' weight map of the Kohonen map obtained with set 1.

be used to select the cytotoxic activity for unknown compounds, without the necessity of a biological assay. A Kohonen map was created for each of the sets using only these 11 descriptors. The results are shown in Table 5.

Table 5, as Table 4, presents a very significant match in the training set, mainly among the inactive compounds (~88%). Analyzing both groups, the percentage seen is high (83%). The test set 1 shows the highest match, where 78% of the substances in this set had their cytotoxic activity correctly predicted, thus supporting the results of Table 4, although the match in the test set. Besides, the match in the test set was 69%, which is still a significant result. Table 5 also shows that the inactive compounds were less correctly predicted (58%) than the active ones (72%). This could be explained by the fact that these eleven descriptors can place the compounds as active ones with more efficiency than the inactive ones. In other words, the characteristics that show the greatest importance, that is, the most influent in the Kohonen map, using these specific descriptors, belong to the active compounds.

Table 5
Summary of training and test match results (%) using 11 geometrical descriptors

	Train set 1	Train set 2	Train set 3	Train set 4	Train set 5	Train set 6	Average	Standard deviation
Map size	11 × 3	13 × 3	6 × 6	10 × 2	10 × 2	11 × 2		
Active match	88	96	87	85	83	92	89	5
Inactive match	82	64	91	80	77	67	77	10
Total	85	80	89	83	80	80	83	4
	Test set 1	Test set 2	Test set 3	Test set 4	Test set 5	Test set 6	Average	Standard deviation
Active match	60	80	100	33	60	100	72	26
Inactive match	100	50	0	83	75	40	58	36
Total	78	67	67	67	67	70	69	4

There are two compounds in the dataset that do not have the α-methylene-γ-lactone structure. One is in the active group, while the other is in the inactive group. It is not possible to make conclusions about the importance of this structure to the cytotoxic activity, since the sample's size of compounds owning this structure is small. However, a few observations can be made:

- 13 of the 15 guaianolide/pseudoguaianolide active compounds have a double bond in the cyclopentane ring or at position 10, while only 3 of the 9 guaianolide inactive compounds and none of the pseudoguaianolide have these structures.
- 6 of the 9 guaianolide active substances contain the angeloyloxy group at position 8 and the hydroxyl group at position 5, while only 2 of the 9 guaianolide inactive substances have these structures.

The majority of the compounds that show the highest activity (<2.0 μmol/L) are guaianolides (**16**, **20**, **21**, **26**, **28**, **29** and **30**) and pseudoguaianolides (**11**, **12**, **13**, **14** and **15**) while there are two germacranolides (**21** and **23**), one elemanolide (**33**) and two eudesmanolide (**39** and **49**) in this group. All these guaianolide have the angeloyloxy group at position 8 and the hydroxyl group at 5. The germacranolide **23** shows the angeloyloxy group at position 8 (Fig. 1) and the germacranolide **21**, an acetate group at the same position; the elemanolide **33** has the metacrilate group, which seems the angeloyloxy, at position 6 and one of the eudesmanolides (**49**) has an angeloyloxy group at position 8. So, the importance of this group in the cytotoxic activity is evident.

All pseudoguaianolides possess a double bond in the cyclopentane ring. As none of the compounds that have this skeleton (pseudoguaianolide), among the inactive substances, show the double bond, it can be concluded that this structure is important for activity.

Among the inactive compounds, six show the angeloyloxy structure or similar, for instance the metacrilate group. These six substances have activities between 5.0 and 8.8 μmol/L. The values are relatively close to the active compounds' values, revealing these structures are outstanding to the biological activity, but alone they are not able to provide satisfactory values to the same, i.e. other characteristics are necessary in the molecule to the cytotoxic activity.

4. Conclusions

SOM classifications produce acceptable separations of the active and inactive groups in our test set 1, and support the use of SOMs as filters for virtual screening of compound databases. Noteworthy in this study is the result that two of the eleven weight descriptors (SPAN and GO(O…O)) appeared in our previous QSAR study. The results reported here provide additional support for these structures as important for cytotoxic activity in these SLs. Our results also find that the greatest prediction accuracy was obtained using descriptors from the geometrical group. Together these results provide support for the importance of 3D analysis in the study of structural features important to biological activity.

A more extensive study using resonance database, such as ^{13}C NMR, and neural networks would improve the results and the knowledge of the necessary structures for cytotoxic activity. A pharmacokinetic study using calculated molecular properties from 3D molecular fields of interaction energies is an urgent necessity for strategies to modify rationally compound properties in order to achieve a desired ADME (absorption, distribution, metabolism and excretion) profile. This kind of correlation is still poorly explored and an insufficient pharmacokinetic profile is the major cause for the failure of a drug discovery project. Since pharmacokinetics is closely linked with physico-chemical properties, experimental design and quantitative structure–property modeling are key factors to systematically explore physico-chemical property space and to establish stable, predictive models for lead optimization. Usually large-scale measurements of physico-chemical properties or in vitro ADME data are either too time-consuming or too much material is needed. Furthermore, experimental approaches require the synthesis of compounds and cannot be used for the prioritization of synthetic targets. Therefore, computational approaches, using the Volsurf [78] as a tool, for instance, for the correlation of molecular structures with pharmacokinetic properties are of great interest.

This global model provides additional information, established from previous QSAR studies, about structural influences on biological activity. Structural models can contribute to searching strategies and improvements in lead structures for the development of anti-tumour drugs based on sesquiterpene lactones.

Appendix. Supplementary data

Supplementary data associated with this article can be found in the online version, at doi:10.1016/j.ejmech.2008.01.003.

References

[1] A.K. Picman, Biochem. Syst. Ecol. 14 (1986) 255–281.
[2] H. Yoshioka, T.J. Mabry, B.N. Timmermann, Sesquiterpene Lactones: Chemistry, NMR and Plant Distribuition, University of Tokyo Press, 1973.
[3] V.H. Heywood, J.B. Harborne, B.L. Turner, The Biology and Chemistry of the Compositae, Vols. I and II, Academic Press, New York, 1977.
[4] M.J.P. Ferreira, A.J.C. Brant, A.R. Rufino, S.A.V. Alvarenga, F.M.M. Magri, V.P. Emerenciano, Phytochem. Anal. 15 (2004) 389–396.
[5] S.A.V. Alvarenga, M.J.P. Ferreira, G.V. Rodrigues, V.P. Emerenciano, Bot. J. Linn. Soc. 147 (2005) 291–308.
[6] M.J.P. Ferreira, A.J.C. Brant, S.A.V. Alvarenga, V.P. Emerenciano, Chem. Biodivers. 2 (2005) 633–644.
[7] V.P. Emerenciano, D. Cabrol-Bass, M.J.P. Ferreira, S.A.V. Alvarenga, A.J.C. Brant, M.T. Scotti, K.O. Barbosa, Nat. Prod. Commun. 1 (2006) 495–507.
[8] L.M. Calabria, V.P. Emerenciano, M.J.P. Ferreira, M.T. Scotti, T.J. Mabry, Nat. Prod. Commun. 2 (2007) 277–285.
[9] F.C. Seaman, Bot. Rev. 48 (1982) 121–595.

[10] W. Herz, K. Aota, M. Holub, Z. Samek, J. Org. Chem. 35 (1970) 2611–2624.
[11] W. Herz, G. Ramakrishnan, R. Murari, Phytochemistry 17 (1978) 1953–1955.
[12] W. Kisiel, Pol. J. Pharmacol. Pharm. 27 (1975) 461–467.
[13] J.F. Siuda, J.F. DeBernardis, Lloydia 35 (1973) 107–143.
[14] A.G. Gonzalez, J. Bermejo, J.L. Breton, G.M. Massanet, J. Triana, Phytochemistry 13 (1974) 1193–1197.
[15] Y. Kondo, F. Yoshizaki, F. Hamada, J. Imai, G. Kusano, Tetrahedron Lett. 25 (1977) 2155–2158.
[16] J.L. Hartwell, Cancer Treat. Rep. 60 (1976) 1031–1067.
[17] G.R. Pettit, Biosynthetic Products for Cancer Chemotherapy, Plenum Press, New York, 1977.
[18] G.A. Cordell, in: H. Wagner, P. Wolff (Eds.), New Natural Products and Plant Drugs with Pharmacological, Biological or Therapeutical Activity, Springer, Berlin, 1977.
[19] O. Sticher, in: H. Wagner, P. Wolff (Eds.), New Natural Products and Plant Drugs with Pharmacological, Biological or Therapeutical Activity, Springer, Berlin, 1977.
[20] J.L. Hartwell, B.J. Abbott, Adv. Pharmacol. Chemother. 7 (1969) 117–209.
[21] S.M. Kupchan, Trans. N. Y. Acad. Sci. 32 (1970) 85–106.
[22] S.M. Kupchan, D.C. Fessler, M.A. Eakin, T.J. Giacobbe, Science 168 (1970) 376–378.
[23] S.M. Kupchan, M.A. Eakin, A.M. Thomas, J. Med. Chem. 14 (1971) 1147–1152.
[24] B. Hladon, B. Drozdz, M. Holub, T. Bobkiewicz, Arch. Immunol. Ther. Exp. 23 (1975) 845–855.
[25] K.H. Lee, E.S. Huang, C. Piandosi, J. Pagano, T.A. Geissman, Cancer Res. 31 (1971) 1649–1654.
[26] K.H. Lee, R. Meck, C. Piandosi, E.S. Huang, J. Med. Chem. 16 (1973) 299–301.
[27] K.H. Lee, I.H. Hall, E.C. Mar, C.O. Starnes, S.A. ElGebaly, T.G. Waddel, R.I. Hadgraft, C.G. Ruffner, I. Weidner, Science 196 (1977) 533–536.
[28] I. Ganjian, I. Kubo, P. Fludzinski, Phytochemistry 22 (1983) 2525–2526.
[29] S.M. Kupchan, T.J. Giacobbe, I.S. Krull, A.M. Thomas, M.A. Eakin, D.C. Fessler, J. Org. Chem. 35 (1970) 3539–3543.
[30] T.J. Schmidt, Bioorg. Med. Chem. 5 (1997) 645–653.
[31] T.J. Schmidt, Curr. Org. Chem. 3 (1999) 577–608.
[32] V.M. Dirsch, H. Stuppner, A.M. Vollmar, Cancer Res. 61 (2001) 5817–5823.
[33] J. Heilmann, M.R. Wasescha, T.J. Schmidt, Bioorg. Med. Chem. 9 (2001) 2189–2194.
[34] S. Karabunarliev, O.G. Mekenyan, W. Karcher, C.L. Russom, S.P. Bradbury, Quant. Struct.-Act. Relat. 15 (1996) 302–310.
[35] R.D. Cramer III, D.E. Patterson, J.D. Bunce, J. Am. Chem. Soc. 110 (1988) 5959–5967.
[36] M.T. Scotti, M.B. Fernandes, M.J.P. Ferreira, V.P. Emereciano, Bioorg. Med. Chem. 15 (2007) 2927–2934.
[37] S. Wagner, A. Hofmann, B. Siedle, L. Terfloth, I. Merfort, J. Gasteiger, J. Med. Chem. 49 (2006) 2241–2252.
[38] J. Zupan, J. Gasteiger, Neural Networks in Chemistry and Drug Design, Wiley-VCH, Weinheim, Germany, 1999.
[39] T. Kohonen, Self-Organization and Associative Memory, third ed.). Springer Verlag, Berlin, 1989.
[40] L.A. Fraser, D.A. Mulholland, D.D. Fraser, Phytochem. Anal. 8 (1997) 301–311.
[41] T. Kohonen, Biol. Cybern. 43 (1982) 59–69.
[42] Q.-Y. Zhang, J. Aires-de-Souza, J. Chem. Inform. Model. 45 (2005) 1775–1783.
[43] F.B. Da Costa, L. Terfloth, J. Gasteiger, Phytochemistry 66 (2005) 345–353.
[44] D. Hristozov, F.B. Da Costa, J. Gasteiger, J. Chem. Inform. Model. 47 (2007) 9–19.
[45] J. Gasteiger, A. Teckentrup, L. Terfloth, S. Spycher, J. Phys. Org. Chem. 16 (2003) 232–245.
[46] P. Bernard, A. Golbraikh, D. Kireev, J.R. Chrétien, N. Rozhkova, Analysis 26 (1998) 333–341.
[47] S. Gupta, J. Aires-de-Sousa, Mol. Divers. 11 (2007) 23–36.
[48] V.P. Emerenciano, M.T. Scotti, R. Stefani, S.A.V. Alvarenga, J.M. Nuzillard, G.V. Rodrigues, Internet Electron. J. Mol. Des. 5 (2006) 213–223.
[49] B. Hladon, B. Drozdz, H. Grabarczyk, T. Bobkiewicz, J. Olszewski, Pol. J. Pharmacol. Pharm. 27 (1975) 429–438.
[50] B. Hladon, B. Drozdz, M. Holub, P. Szafarek, O. Klimaszewska, Pol. J. Pharmacol. Pharm 30 (1978) 611–620.
[51] M. Bialecki, B. Hladon, B. Drozdz, E. Bloszyk, S. Szwemin, T. Bobkiewicz, Pol. J. Pharmacol. Pharm. 26 (1974) 511–518.
[52] J.P. Gastmans, M. Furlan, M.N. Lopes, J.H.G. Borges, V.P. Emerenciano, Quim. Nova 13 (1990) 10–15.
[53] V.P. Emerenciano, G.V. Rodrigues, J.P. Gastmans, Quim. Nova 16 (1993) 431–434.
[54] M.J.P. Ferreira, G.V. Rodrigues, V.P. Emerenciano, Can. J. Chem. 79 (2001) 1915–1925.
[55] M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, J.J.P. Stewart, J. Am. Chem. Soc. 107 (1985) 3902–3909.
[56] M.J.S. Dewar, E.F. Healy, A.J. Holder, Y.C. Yuan, J. Comput. Chem. 11 (1990) 541–542.
[57] S.R.L. Talete, DRAGON for Windows (Software for Molecular Descriptor Calculations). Version 5.4 (2006), http://www.talete.mi.it.
[58] R. Todeschini, V. Consonni, Handbook of Molecular Descriptors, Wiley-VCH, Weinheim, Germany, 2000.
[59] M.C. Hemmer, V. Steinhauer, J. Gasteiger, Vib. Spectrosc. 19 (1999) 151–164.
[60] J. Gasteiger, J. Sadowski, J. Schuur, P. Selzer, L. Steinhauer, V. Steinhauer, J. Chem. Inf. Comput. Sci. 36 (1996) 1030–1037.
[61] J.H. Schuur, P. Selzer, J. Gasteiger, J. Chem. Inf. Comput. Sci. 36 (1996) 334–344.
[62] R. Todeschini, P. Gramatica, Quant. Struct.-Act. Relat. 16 (1997) 113–119.
[63] R. Todeschini, P. Gramatica, Quant. Struct.-Act. Relat. 16 (1997) 120–125.
[64] V. Consonni, R. Todeschini, M. Pavan, J. Chem. Inf. Comput. Sci. 42 (2002) 682–692.
[65] V. Consonni, R. Todeschini, M. Pavan, P. Gramatica, J. Chem. Inf. Comput. Sci. 42 (2002) 693–705.
[66] M.V. Diudea, D. Horvath, A. Graovac, J. Chem. Inf. Comput. Sci. 35 (1995) 129–135.
[67] A.T. Balaban, J. Chem. Inf. Comput. Sci. 37 (1997) 645–650.
[68] M. Randic, A.F. Kleiner, L.M. Dealba, J. Chem. Inf. Comput. Sci. 34 (1994) 277–286.
[69] P.A.P. Moran, Biometrika 37 (1950) 17–23.
[70] R.C. Geary, Incorporated Statistician 5 (1954) 115–145.
[71] G. Moreau, P. Broto, New J. Chem. 4 (1980) 359–360.
[72] V.N. Viswanadhan, A.K. Ghose, G.R. Revankar, R.K. Robins, J. Chem. Inf. Comput. Sci. 29 (1989) 163–172.
[73] V.N. Viswanadhan, M.R. Reddy, R.J. Bacquet, M.D. Erion, J. Comput. Chem. 14 (1993) 1019–1026.
[74] D. Livingstone, Data Analysis for Chemists, Oxford Science Publications, New York, 1995.
[75] J. Vesanto, J. Himberg, E. Alhoniemi, J. Parhankangas, SOM Toolbox for Matlab 5 (2005), http://www.cis.hut.fi/projects/somtoolbox.
[76] F. Marini, J. Zupan, A.M. Magrì, Anal. Chim. Acta 544 (2005) 306–314.
[77] J.Vesanto, J. Himberg, E. Alhoniemi, J. Parhankangas, Proceedings of the Matlab DSP Conference, Espoo, Finland (1999) 35–40.
[78] P. Crivori, G. Cruciani, P.A. Carrupt, B. Testa, J. Med. Chem. 43 (2000) 2204–2216.