Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Phytochemistry 66 (2005) 345–353
PHYTOCHEMISTRY
www.elsevier.com/locate/phytochem

# Sesquiterpene lactone-based classification of three Asteraceae tribes: a study based on self-organizing neural networks applied to chemosystematics

Fernando B. Da Costa [a,b,*], Lothar Terfloth [b], Johann Gasteiger [b]

[a] Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo, Av. do Café s/n, 14040-903 Ribeirão Preto, SP, Brazil
[b] Computer-Chemie-Centrum und Institut für Organische Chemie, Universität Erlangen-Nürnberg, Nägelsbachstraße 25, D-91052 Erlangen, Germany

Received 8 July 2004; received in revised form 7 December 2004

## Abstract

This work describes an application of artificial neural networks on a small data set of sesquiterpene lactones (STLs) of three tribes of the family Asteraceae. Structurally different types of representative STLs from seven subtribes of the tribes Eupatorieae, Heliantheae and Vernonieae were selected as input data for self-organizing neural networks. Encoding the 3D molecular structures of STLs and their projection onto Kohonen maps allowed the classification of Asteraceae into tribes and subtribes. This approach allowed the evaluation of structural similarities among different sets of 3D structures of sesquiterpene lactones and their correlation with the current taxonomic classification of the family. Predictions of the occurrence of STLs from a plant species according to the taxa they belong to were also performed by the networks. The methodology used in this work can be applied to chemosystematic or chemotaxonomic studies of Asteraceae.


*Keywords:* Asteraceae; Chemoinformatics; Chemosystematics; Chemotaxonomy; Kohonen networks; Sesquiterpene lactones; 3D structures; Taxonomic markers

## 1. Introduction

Recently, there has been great interest in the study of chemosystematics and chemotaxonomy based on secondary metabolites from higher plants (Alvarenga et al., 2001; Gottlieb, 1972, 1982; Harborne and Williams, 1994), including studies of the family Asteraceae (Bohlmann, 1990; Emerenciano et al., 2001; Harborne and Turner, 1984; Seaman, 1982; Spring and Buschmann, 1996; Zdero and Bohlmann, 1990).

The family Asteraceae is one of the largest angiosperm family and comprises about 1300 genera and 25,000 species distributed over three subfamilies and 17 tribes (Bremer, 1994). Basically, species of Asteraceae biosynthesize polyacetylenes, flavonoids and terpenoids. Nevertheless, the sesquiterpene lactones (STLs) are the most studied class of secondary metabolites, which are used as taxonomic markers (Seaman, 1982; Spring and Buschmann, 1996).

More than 4000 structures of STLs with around 30 different skeletal types have so far been reported from several tribes of Asteraceae (Emerenciano et al., 1998; Seaman, 1982). However, these compounds are accumulated in a few of these tribes (Seaman, 1982; Spring and Buschmann, 1996; Zdero and Bohlmann, 1990). Although some special skeletal types are unique to certain tribes or subtribes, others are more widespread and have special trends of accumulation. Due to their

* Corresponding author. Tel.: +49 9131 85 26579; fax: +49 9131 85 26566.
*E-mail address:* fernando.dacosta@chemie.uni-erlangen.de (F.B. Da Costa).

0031-9422/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.phytochem.2004.12.006

chemodiversity, the STLs are the most suitable class of natural products for chemosystematic studies within the family.

Comparisons of structural differences between sets of STLs from different taxa can yield interesting results concerning systematics. Thus, these micromolecules have great utility for ascertaining tribal, subtribal, generic and specific boundaries within the family as well as providing biogenetic information regarding evolutionary relationships (Seaman, 1982). The underlying assumption in using STLs for a systematic study of Asteraceae is that different taxa may have evolved slightly different metabolisms. Thus, a successful chemosystematic study can also provide insights into these different metabolic pathways, although only hypothetical biosynthetic routes which generate most of the different STL types are currently known.

However, these metabolic pathways in some tribes are not different enough to have completely resulted in STLs of different skeletal types in a way to allow a clear separation. Rather, the same skeletal types can be found in different tribes and, in some cases, sets composed of the same STLs are found in different tribes, resulting in a region of overlap. Thus, in most cases, the observation of a specific STL skeletal type cannot be used for its classification into one tribe. In such cases, a method to distinguish one tribe from the rest is through comparison of sets of shared skeletons or substitutional diversification of such skeletons. The same criterion may be applied to the subtribal level when one subtribe of a certain tribe needs to be differentiated.

The hypotheses investigated in this study were: (1) when there is not a single skeletal type to distinguish tribes or subtribes it may be that the set of metabolites with their substitutional features can be used for this purpose. The rationale behind this is that different metabolic pathways are characterized by a set of compounds along such a pathway; (2) in addition, when characterizing metabolites, their 3D structures should be used during the process of differentiation.

Two reasons can be put forth for that choice: first, metabolites are formed in the pockets of enzymes which are inherently 3D in nature and slight changes in the pockets of these enzymes might have produced metabolites of slightly different 3D structure. Furthermore, a representation of the 3D structure of a metabolite can account both for different skeletal types along with substitutional features as well as different stereochemistry. Such features are influenced by the shape of the binding pocket of the enzyme.

Among the different computer-based methods that are able to treat and analyze chemical data, artificial neural networks have particular importance (Gasteiger and Engel, 2003; Zupan and Gasteiger, 1999). The Kohonen network (self-organizing neural network) is a method that allows the projection of $n$-dimensional input objects into a 2D-plane resulting in a so-called self-organizing map (Kohonen, 1995; Zupan and Gasteiger, 1999) (Figs. 1(a) and (b)). In this way, similar $n$-dimensional objects are located close to each other in the 2D array of neurons of a map. Thus, the topological relationships of the objects in the high-dimensional space are, after projection, preserved or at least kept as closely as possible to their original aspect. This approach has been used in different studies concerning chemistry (Kohonen, 1995; Zupan and Gasteiger, 1999).

## 2. Results and discussion

In the present study, self-organizing neural networks were used for the study of a data set of three STL-containing tribes of Asteraceae (Table 1, Scheme 1). The 3D atomic coordinates of the structures of STLs coded with appropriate descriptors (see Section 4) were used as input variables ($n$-dimensional object) and presented to Kohonen networks (Figs. 1(a) and (b)). The network learns in an unsupervised manner to recognize clusters and to display similar cases close to each other. Thus, the Kohonen map obtained with small a data set of structures can be used to visualize the distribution of chemical characters in STL-producing taxa. Furthermore, the analysis of a map can also help a phytochemist to easily locate those taxa which contain specific types of STLs. For example, this is a useful feature when there is interest in the collection of new plant material with the aim to find special types of compounds for dif-

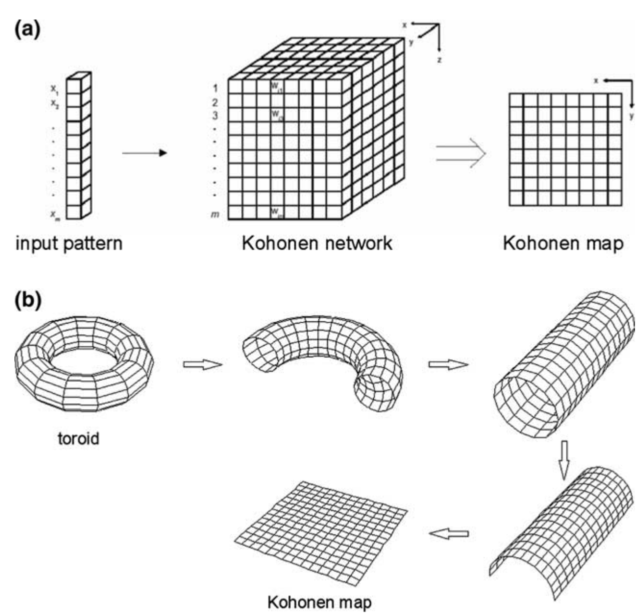



Fig. 1. (a) Architecture of a Kohonen network: the vector (left) with 128 input variables ($m = 128$) representing a coded 3D structure is presented to the Kohonen network (centre); the Kohonen map (right) is displayed in the $x, y$-plane while the neuron weights ($m = 128$) are in the $z$-plane. (b) Wrapping the 2D Kohonen map into a toroid, a plane without beginning and without end.

Table 1
Sesquiterpene lactone-containing tribes and subtribes of Asteraceae selected for this study, their codification, the number and skeleton subtypes of selected structures of sesquiterpene lactones (STLs)[a]

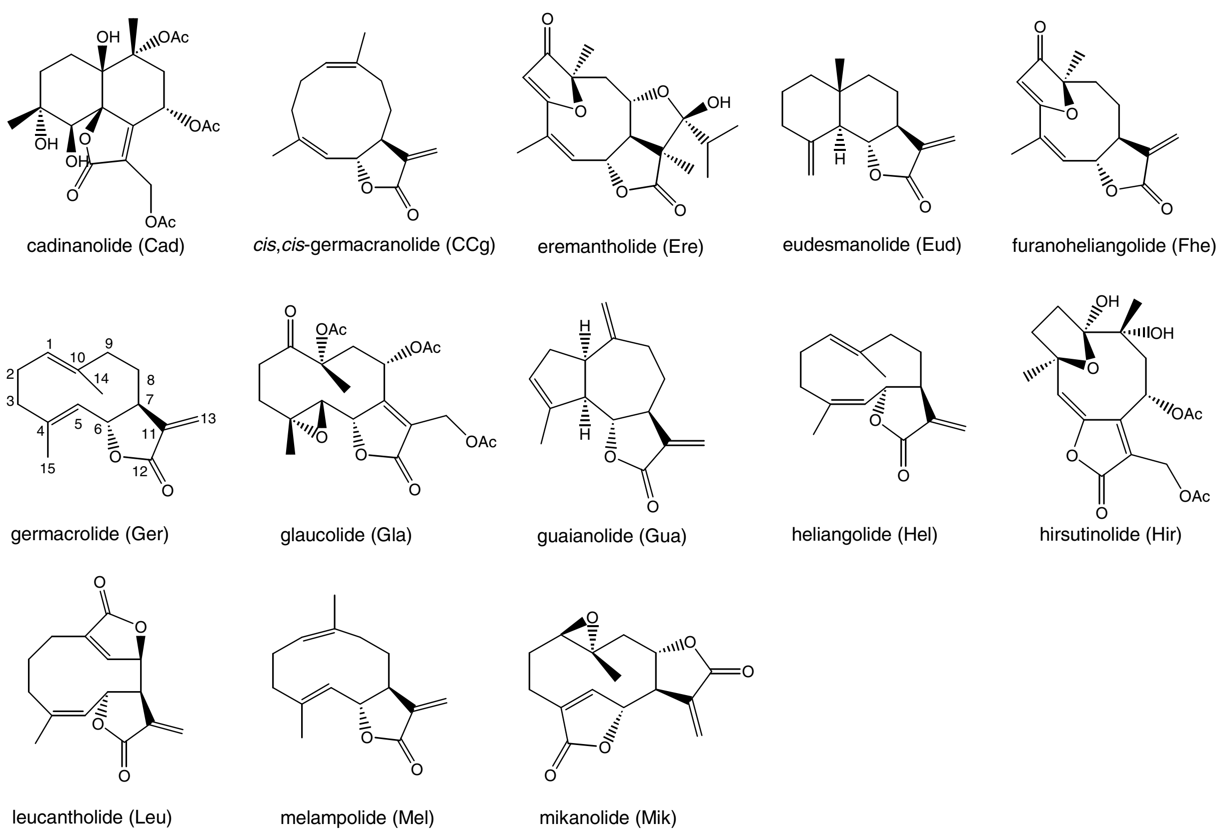
| Taxonomic classification (tribes and subtribes) | Codes | No. of STLs | Skeleton subtypes of STLs | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Cad | Ccg | Ere | Eud | Fhe[b] | Ger | Gla | Gua | Hel | Hir | Leu | Mel | Mik |
| Eupatorieae | Eup | | | | | | | | | | | | | | |
| Eupatoriinae | Eup/eup | 22 | – | – | – | – | – | 7 | – | 8 | 7 | – | – | – | – |
| Mikaniinae | Eup/mik | 20 | – | – | – | – | – | 13[c] | – | – | – | – | – | – | 7 |
| | Total | 42 | – | – | – | – | – | 20 | – | 8 | 7 | – | – | – | 7 |
| Heliantheae | Hlt | | | | | | | | | | | | | | |
| Helianthinae | Hlt/hlt | 21 | – | – | – | – | 7 | 7 | – | – | 7 | – | – | – | – |
| Melampodiinae | Hlt/mel | 21 | – | 2 | – | – | 3 | 5 | – | – | – | – | 2 | 9 | – |
| Verbesininae | Hlt/ver | 20 | – | – | – | 9 | 1 | 7 | – | – | 1 | – | – | 2 | – |
| | Total | 62 | – | 2 | – | 9 | 11 | 19 | – | – | 8 | – | 2 | 11 | – |
| Vernonieae | Ver | | | | | | | | | | | | | | |
| Lychnophorinae | Ver/lyc | 20 | – | – | 5 | – | 8 | 5 | – | 2 | – | – | – | – | – |
| Vernoniinae | Ver/ver | 20 | 2 | – | – | – | – | – | 9 | – | – | 9 | – | – | – |
| | Total | 40 | 2 | – | 5 | – | 8 | 5 | 9 | 2 | – | 9 | – | – | – |
| | Sum | 144 | 2 | 2 | 5 | 9 | 19 | 44 | 9 | 10 | 15 | 9 | 2 | 11 | 7 |

Cad, cadinanolide; Ccg, *cis,cis*-germacranolide; Ere, eremantholide; Eud, eudesmanolide; Fhe, furanoheliangolide; Ger, germacrolide; Gla, glaucolide; Gua, guaianolide; Hel, heliangolide; Hir, hirsutinolide; Leu, leucantholide; Mel, melampolide; Mik, mikanolide.

[a] Except when mentioned, all STLs are 6α,12-*trans*-olides.

[b] 3,10-Epoxy-heliangolides.

[c] Six 8α,12-*trans*-germacranolides.




Scheme 1. Examples of skeleton types and germacranolide derivatives of the selected structures of sesquiterpene lactones.

ferent purposes. Another feature of this approach is that it can be used to predict the classification of entire sets of 3D structures at different hierarchical levels taking into account skeletal types, substitutional features as well as different stereochemistry. Thus, this tool is helpful for chemosystematic or chemotaxonomic analyses.

### 2.1. The chemical data set

The taxonomic division used in this work is based on the Bremer classification system (Bremer, 1994). The most representative STLs that occur in seven subtribes of the tribes Eupatorieae, Heliantheae (subfamily Asteroideae), and Vernonieae (subfamily Cichorioideae) were selected as input variables to the networks (Table 1, Scheme 1). The subfamily Barnadesioideae is not represented here since its genera do not produce STLs.

Examples of selected skeletal types of STLs are the germacranolides (comprising the four isomers germacrolides, heliangolides, melampolides and *cis*,*cis*-germacranolides), guaianolides and eudesmanolides (Scheme 1), which together comprise the basic tripartite chemistry of the STL-producing tribes of the family (Seaman, 1982). Minor STL subclasses, which are restricted to some of the seven subtribes, were also taken into account, like glaucolides, hirsutinolides, eremantholides and leucantholides, among others (Table 1, Scheme 1).

A data set of 144 STL structures was elaborated and substitutional features were taken into account in the selection procedure, such as the presence, types, and stereochemistry of functionalities at different positions in the sesquiterpene rings.

In order to investigate the assignment of sets of STLs to their tribes and subtribes using Kohonen networks, four different studies were performed. In the first three sets of experiments, each of the three tribes was studied individually and the analyses were focused onto some of their subtribes. The fourth experiment was based on the analysis of the three tribes together in one map. The assignment of an entire set of STLs of a single plant species into taxa was performed by a process called majority decision.

### 2.2. The tribe Eupatorieae

The 42 patterns of the STLs of the tribe Eupatorieae (Table 1) were used as input data in a network with $6 \times 5$ neurons. The separation between the patterns of the subtribes Eupatoriinae and Mikaniinae is quite evident and can be visualized in Figs. 2(a) and (b). A single Kohonen map (Fig. 2(a)) shows the 30 neurons and some clusters of STLs according to their skeletons. In Fig. 2(b), four identical maps are displayed together in order to reflect the toroidal topology of the network, resulting in a plane without beginning and without end. It means that the four edges of the rectangles are connected (Fig. 1(b)).

The subtribe Eupatoriinae comprises four genera (Bremer, 1994). Nevertheless, species of the genus *Eupatorium* are the only STL-containing taxa in this subtribe (Herz, 2001). The subtribe Mikaniinae comprises only the genus *Mikania* (Bremer, 1994). Then, in this particular example, the mapping of the structures of these two subtribes allows the evaluation of the STL chemistry of two genera.

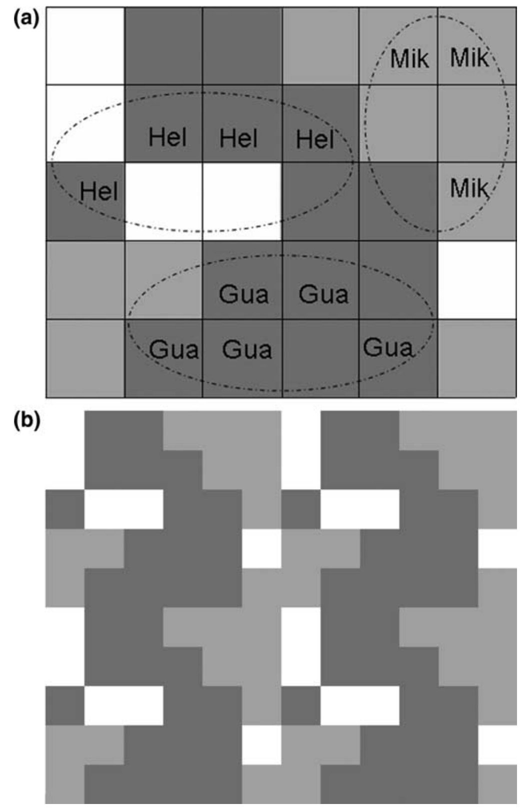



Fig. 2. Kohonen map with toroidal topology and dimension of $6 \times 5$ neurons of the 42 patterns from the tribe Eupatorieae, (eup/eup = 22) + (eup/mik = 20): (a) single map showing clusters of structures; (b) four folded map illustrating the separation of the two subtribes. Color scheme: dark gray = eup/eup; light gray = eup/mik; white = empty neurons. Legend: Gua, guaianolide; Hel, heliangolide; Mik, mikanolide.

The germacranolides and their derivatives of both subtribes comprise the largest set of structures (Table 1) and appeared as the largest group in the map, which was separated from the guaianolides (Gua) of Eupatoriinae (Fig. 2(a)). The heliangolides (Hel) of Eupatoriinae occur together in one cluster and the mikanolides (Mik) of Mikaniinae are in another one (Fig. 2(a)). The 8α,12-*trans*-germacranolides of the subtribe Mikaniinae formed a third group, close to the mikanolides (not shown). Thus, this approach allowed the separation of the skeletal types and subtypes of STLs into the two subtribes.

In this experiment, it was possible to differentiate and easily separate the STLs from both subtribes (or both genera) according to their skeletons. Even when the same skeletal type was present in the two subtribes, like the germacrolides (Table 1), the separation was easily achieved. No conflicts (or collisions) were observed in the map. A conflict means that patterns of different classes appear together within a single neuron. In other words, a conflict occurs when the STLs from different subtribes are not differentiated by the network and are grouped together, because of their very close chemical similarity.

Different clusters contain structures of the same skeletal type. In such clusters, those STLs with quite similar substitutional features in their rings were grouped together into one single neuron. This shows the capacity of the network for this purpose. One example which illustrates this observation involves two pairs of 8α,12-*trans*-germacranolides of Mikaniinae, one consisting of two 1,10-epoxy derivatives with an acetoxyl moiety at C-15, and the other consisting of two compounds with a 1(10)(*Z*) double bond, an aldehyde at C-14 and a side chain ester at C-15 (Scheme 2). Each of these pairs appeared in a different neuron close to each other.

### 2.3. The tribe Heliantheae

Heliantheae is the tribe from which the largest number of STLs has been isolated so far. Therefore, a larger number of structures have been selected to perform this study. The 62 STLs of Heliantheae (Table 1) were input into a network with 7 × 6 neurons (Fig. 3(a)) and the structures were grouped into their subtribes.

Typical STLs of the subtribe Helianthinae are the germacrolides, heliangolides and furanoheliangolides (Scheme 1). The subtribe Melampodiinae is basically characterized by the presence of germacrolides, melampolides and *cis,cis*-germacranolides, whereas other minor groups are also present (Scheme 1). In the subtribe Verbesininae, the typical compounds are the germacrolides and eudesmanolides, but heliangolides, furanoheliangolides and melampolides also occur (Scheme 1) (Bohlmann, 1990; Spring and Buschmann, 1996; Zdero and Bohlmann, 1990). This non-uniform pattern distribution of STLs indicates that some skeletal types or subtypes can not easily be separated and then assigned into to their subtribes. In addition, regions of overlap of STLs occur due to the presence of quite similar structures within these three subtribes. Thus, in such special cases the assignment of STL skeletons into taxa should be evaluated with care.

As shown in the map (Fig. 3(a)), groups of different STL skeletons appeared in different clusters within the subtribes. As observed in the previous experiment,

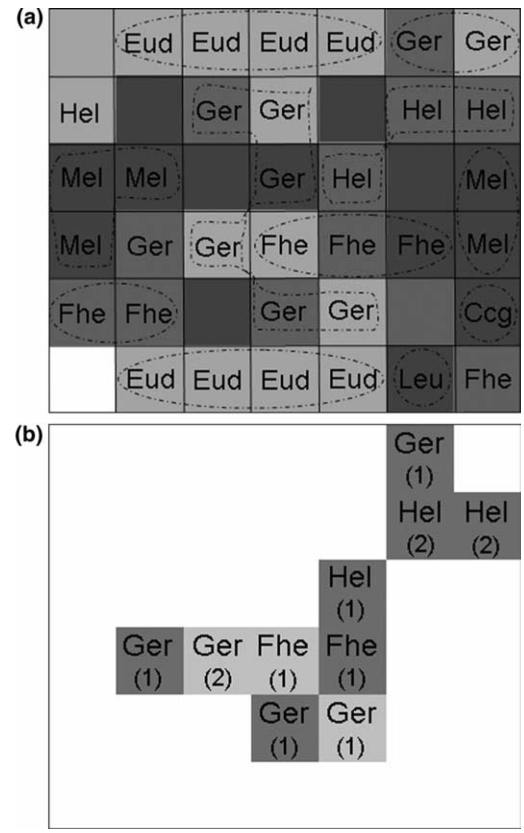
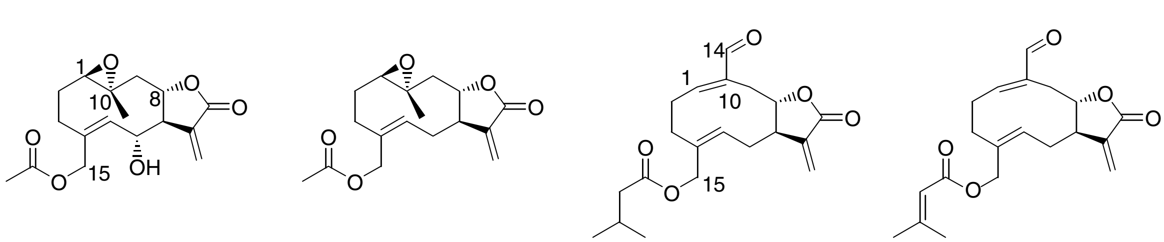



Fig. 3. (a) Kohonen map with toroidal topology and dimension of 7 × 6 neurons of the 62 patterns from the tribe Heliantheae showing clusters of structures, (hlt/hlt = 21) + (hlt/mel = 21) + (hlt/ver = 20). Color scheme: gray = hlt/hlt; dark gray = hlt/mel; light gray = hlt/ver; white = empty neuron. Legend: Ccg, *cis,cis*-germacranolide; Eud, eudesmanolide; Fhe, furanoheliangolide; Ger, germacrolide; Hel, heliangolide; Leu, leucantholide; Mel, melampolide. (b) Kohonen map showing the prediction of 13 STLs of a species from the subtribe Helianthinae. The skeletal types and number of structures (in parentheses) are indicated in the neurons. Color scheme: dark gray = hlt/hlt; light gray = hlt/ver; white = empty neurons.

groups of structures with the same skeletal type formed clusters and those containing similar substitutional features were projected into individual neurons. In this study, some clusters included more than one subtribe. Due to the overlap of STLs within the three subtribes, some collisions were observed and they occurred in five neurons, comprising 13 structures. One example is a neuron containing four furanoheliangolides with different side chain esters at C-8, one from the subtribe



Scheme 2. Two pairs of 8α,12-*trans*-germacranolides of the subtribe Mikaniinae grouped into two different neurons.

Melampodiinae and three from Helianthinae (not shown).

An important consideration is that entire sets of structures from those taxa which biosynthesize heterogeneous sets of compounds can be split up in the Kohonen map. It is because the map shows the areas with special trend of accumulation of similar types of STLs. Furthermore, the map also reflects the metabolic profile of such taxa since similar 3D structures which are derived from a certain metabolic route are more concentrated in a particular area. Such features of the Kohonen maps are useful for chemotaxonomic analysis of plants species, including those with non-uniform production of STLs types.

In order to exemplify the features discussed above as well as the aspect regarding regions of overlap of STLs, an experiment to evaluate the prediction ability of the Kohonen network was performed as follows. An entire, heterogeneous set of STLs from a single species of the subtribe Helianthinae (*Viguiera radula*) was selected from the literature (Spring et al., 2003). The set consists of 13 STLs comprising six germacrolides (Ger), five heliangolides (Hel) and two furanoheliangolides (Fhe). From the 13 STLs, five of them are from the initial data set of the 62 structures from Heliantheae (Table 1). All the 3D structures were encoded and then projected into the previously trained Kohonen network (Fig. 3(a)). Then, a prediction map was obtained (Fig. 3(b)) where a majority decision process took place. The STLs were separated according to their skeletal types. Structures with similar substitutional features were grouped into single neurons (Fig. 3(b)). As can be seen in Fig. 3(b), nine structures were projected into the area which corresponds to the subtribe Helianthinae. The remaining four structures were projected into three neurons of the subtribe Verbesininae. Nevertheless, the analysis of these four structures (one furanoheliangolide and three germacrolides) reveals that they were correctly assigned to the appropriate STL class. This is because their corresponding STL skeletons as well as substitutional features also occur in taxa of the subtribe Verbesininae and are present in the data set of this subtribe. This experiment assigned *V. radula* correctly to the subtribe Helianthinae (Figs. 3(a) and (b)) and thus, confirms the ability of the networks to assign a plant species to the subtribe it belongs to by majority decision based on a set of STLs.

This result emphasizes that a chemotaxonomic classification of Asteraceae involving tribes or subtribes should not be based on a single metabolite or one skeleton but on an entire set of metabolites.

### 2.4. The tribe Vernonieae

The 40 patterns of the tribe Vernonieae were analyzed in a map with $6 \times 5$ neurons. The patterns were easily separated into the subtribes Vernoniinae and Lychnophorinae, showing a good clustering among the STLs (Fig. 4). No collisions were observed.

The typical STLs of the subtribe Lychnophorinae are special types of germacranolides along with furanoheliangolides and eremantholides (Herz, 1996) (Scheme 1). Some guaianolides also occur to a minor extent. The subtribe Vernoniinae is basically characterized by the presence of glaucolides and hirsutinolides (Scheme 1), but cadinanolides were also reported (Herz, 1996; Zdero and Bohlmann, 1990). All these STL types have been selected and the most important STL-producing genera are represented.

The larger groups comprising the glaucolides (Gla), hirsutinolides (Hir), and furanoheliangolides (Fhe) were clustered onto different areas in the Kohonen map (Fig. 4). It shows the separation of the skeleton types into their subtribes. As occurred in the two previous experiments, clusters composed by closely related structures were observed and compounds with similar substitutional patterns were grouped together into single neurons.

This experiment shows that if the STL skeleton types of the subtribes are very different among each other and there are no regions of overlap, it results in a good separation by the network.

### 2.5. Analysis of the three tribes

The final approach in this study was the projection of the 144 patterns of STLs altogether into a single map with $10 \times 9$ neurons. In this step, the STLs were separated into their tribes (Eupatorieae, Heliantheae and Vernonieae), as can be observed in the map in Fig. 5(a). Due to the structural similarity of some groups of STL skeletons, a complete separation of the patterns

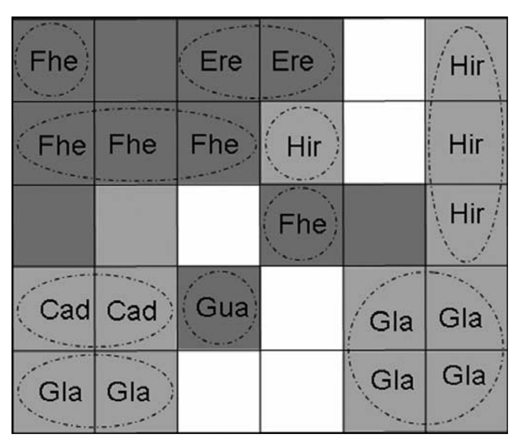



Fig. 4. Kohonen map with toroidal topology and dimension of $6 \times 5$ neurons of the 40 patterns from the tribe Vernonieae showing clusters of structures, (ver/lyc = 20) + (ver/ver = 20). Color scheme: dark gray = ver/lyc; light gray = ver/ver; white = empty neurons. Legend: Cad, cadinanolide; Ere, eremantholide; Fhe, furanoheliangolide; Gla, glaucolide; Gua, guaianolide; Hir, hirsutinolide.

into their corresponding tribes could not be achieved. It indicates that some groups of taxa from the three tribes share the same STL skeletons or similar compounds. Due to this reason, regions of overlap are present and collisions were observed. They occurred in five neurons and comprised 14 STLs.

In general, STLs with different skeletal types were well separated into large clusters, like eudesmanolides and germacranolides. Those having the same skeleton or belonging to particular subtypes were clustered, no matter to which tribe they belong, like the germacrolides (Ger), heliangolides (Hel) and furanoheliangolides (Fhe) (Fig. 5(a)). Within the different clusters composed by STLs of the same skeletal type, grouping of structures according to their substitutional patterns as well as to their 3D structures also took place. Although regions of overlap and collisions were present, the separation was achieved in such a way to allow the prediction of patterns by majority decision, as shown in the experiment with the tribe Heliantheae.

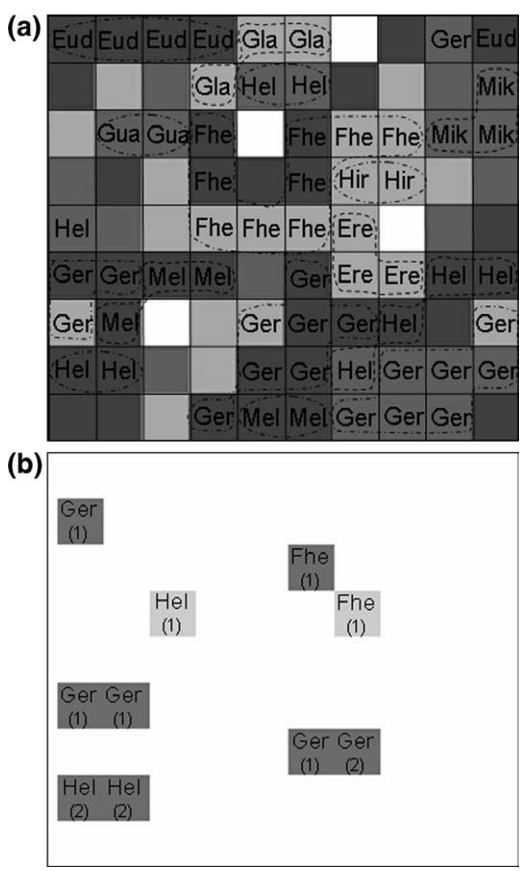



Fig. 5. (a) Kohonen map with toroidal topology and dimension of $10 \times 9$ neurons of the 144 patterns of tribes Eupatorieae, Heliantheae and Vernonieae showing clusters of structures (abbreviations according to Table 1): (eup = 42) + (hlt = 62) + (ver = 40). Color scheme: gray = eup; dark gray = hlt; light gray = ver; white = empty neurons. (b) Kohonen map showing the prediction of 13 STLs of a species from the tribe Heliantheae. The skeletal types and number of structures (in parentheses) are indicated in the neurons. Color scheme: dark gray = hlt; light gray = ver; white = empty neurons.

The same data set of STLs from *V. radula* (as analyzed in Section 2.3) was used to evaluate the prediction ability of the Kohonen network by the majority decision process. The task here was to assign this plant species to the tribe it belongs to. From the 13 STLs, 11 appeared in the area of the tribe Heliantheae and were placed into their corresponding clusters (Fig. 5(b)). Two STLs (one heliangolide and one furanoheliangolide) appeared in the area of the tribe Vernonieae (Fig. 5(b)). Based on this mapping, *V. radula* was assigned to the tribe Heliantheae. This experiment reinforces the good prediction ability of the networks through the majority decision process even if areas of overlap are present or if a heterogeneous set of compounds is selected.

## 3. Conclusions

In this work, a novel view on the distribution of STLs within tribes and subtribes of the family Asteraceae has been obtained. It is based on the projection of the 3D structures of STLs encoded by their RDF representations into Kohonen maps. Due to the fact that the data set of chemical structures was linked to taxonomic information, the separation of compounds of different taxa in the maps was achieved according to the different 3D structures of the metabolites. The 3D structures were well classified according to their different skeletons. The structures in the clusters contained the same skeleton as well as the same or similar substitutional features.

The results show that small and representative data sets of structures can be used to help the phytochemist in the analysis of the distribution of chemical characters in the STL-producing taxa by means of self-organizing maps. The prediction ability of the Kohonen networks was evaluated for an entire set of STLs from a single species. This showed that a plant can be assigned to a tribe or subtribe on the basis of their set of STLs.

When different STL skeletons were present in a data set, their separation was easy to perform and they were correctly assigned to their taxa. Even if regions of overlap involving quite similar STLs were observed, the network was able to separate and assign different structures into different taxa. In such cases, the prediction ability of the network through majority decision achieved good results.

In summary, the Kohonen maps are an important tool for the phytochemists who are interested in the selection of plant species for further studies involving special types of STLs or to help them in chemotaxonomic analyses. This shows that the set of STLs might successfully be used to assign a plant to a subtribe or to a tribe even when there is not a single type of STL that can be used as a marker for such assignment. More

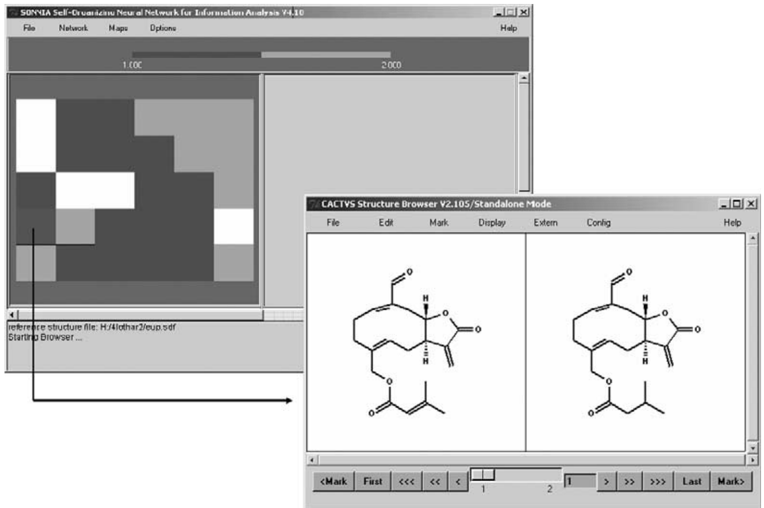



Fig. 6. Graphical user interface (GUI) of SONNIA (left side) depicting a Kohonen map (see Figs. 1 and 2 for more detailed information). The two closely related compounds contained in one neuron in the map are shown in the window on the right.

generally, this shows that self-organizing networks applied to chemical data can contribute to the understanding of living organisms.

## 4. Experimental

### 4.1. Data set of sesquiterpene lactones

The structures of STLs and additional information were taken from published data (Bohlmann, 1990; Herz, 1996, 2001; SciFinder Scholar, 2004; Seaman, 1982; Spring and Buschmann, 1996; Zamorano et al., 1995; Zdero and Bohlmann, 1990). The data set comprising 144 STLs was elaborated using their 2D structures.

### 4.2. Structure representation

The 3D coordinates of the STLs were generated from the constitution (2D structures) of the molecules by the 3D structure generator CORINA (Computer-Chemie-Centrum, 2004; Sadowski and Gasteiger, 1993). CORINA provides a single low energy conformation of a molecular structure. The 3D coordinates of the atoms of the molecules were transformed into a structure code with a fixed number of descriptors using the radial distribution function (RDF) code (Hemmer et al., 1999). This code represents a molecular structure by a radial distribution function and this function can be interpreted as a probability distribution to find an atom in a spherical volume of radius $r$. The basic equation of the RDF code of an ensemble of $N$ atoms is shown in the equation

$$g(r) = \sum_{i=1}^{N-1} \sum_{j>i}^{N} A_i A_j \mathrm{e}^{-B(r-r_{ij})^2},$$

where $N$ is the number of atoms in a molecule, $A_i$ and $A_j$ are properties associated with the atoms $i$ and $j$, respectively, $r_{\mathrm{ij}}$ represents the distance between atoms $i$ and $j$ and $B$ is a smoothing factor.

The RDF representation is uniform and invariant under translations and rotation of molecules. In this study, the atomic property $A$ used was the atomic number. The distribution function is given as a vector with equidistant values of $r$ whose dimension was set to 128 and the function $g(r)$ was defined in the interval 0.0–12.8 Å. Thus, the RDF was sampled at every 0.1 Å.

### 4.3. Kohonen maps

The software Self-Organizing Neural Network for Information Analysis (SONNIA) was used to generate the Kohonen maps with toroidal topology (Fig. 1(b)) (Molecular Networks GmbH, 2002). The input vector with 128 dimensions (Fig. 1(a)) containing the RDF code of the STLs was presented to the networks. The size of the networks as well as the training parameters varied according to the number of patterns analyzed in each of the four experiments. SONNIA provides many visualization techniques that make it particularly suitable for the study of chemical data sets. For example, the structures which are mapped into the neurons can be displayed in a window when a neuron is selected, as depicted in Fig. 6.

## Acknowledgement

FBC is grateful to the Alexander von Humboldt Foundation (Germany) for a Research Fellowship at the Computer-Chemie-Centrum.

## References


Alvarenga, S.A.V., Gastmans, J.P., Rodrigues, G.V., Moreno, P.R.H., Emerenciano, V.P., 2001. A computer-assisted approach for chemotaxonomic studies – diterpenes in Lamiaceae. Phytochemistry 56, 583–595.

Bohlmann, F., 1990. Chemistry of the Heliantheae. (Compositae) Plant Syst. Evol. (Suppl. 4), 67–75.

Bremer, K., 1994. Asteraceae: Cladistics and Classification. Timber Press, Portland.

Computer-Chemie-Centrum, Universität Erlangen-Nürnberg, 2004. Available from: <http://www2.chemie.uni-erlangen.de/software/corina/index.html>.

Emerenciano, V.P., Ferreira, M.J.P., Branco, M.D., Dubois, J.E., 1998. The application of Bayes' theorem in natural products as a guide for skeletons identification. Chemometr. Intell. Lab. Sys. 40, 83–92.

Emerenciano, V.P., Militao, J.S.L.T., Campos, C.C., Romoff, P., Kaplan, M.A.C., Zambon, M., Brant, A.J.C., 2001. Flavonoids as chemotaxonomic markers for Asteraceae. Biochem. Syst. Ecol. 29, 947–957.

Gasteiger, J., Engel, T., 2003. Chemoinformatics: A Textbook. Wiley-VCH, Weinheim.

Gottlieb, O.R., 1972. Chemosystematics of the Lauraceae. Phytochemistry 11, 1537–1570.

Gottlieb, O.R., 1982. Micromolecular Evolution, Systematics and Ecology. Springer Verlag, Berlin, Heidelberg, NY.

Harborne, J.B., Turner, B.L., 1984. Plant Chemosystematics. Academic Press, London.

Harborne, J.B., Williams, C.A., 1994. Recent advances in the chemosystematics of the monocotyledons. Phytochemistry 37, 3–18.

Hemmer, M.C., Steinhauer, V., Gasteiger, J., 1999. The prediction of the 3D structure of organic molecules from their infrared spectra. Vib. Spectrosc. 19, 151–154.

Herz, W., 1996. A review of the terpenoid chemistry of the Vernonieae. In: Hind, D.J.N., Beentje, H.J. (Eds.), Compositae: Systematics, Proceedings of the International Compositae Conference, 1994, vol. 1. Kew Scientist, Kew, pp. 229–251.

Herz, W., 2001. Chemistry of the Eupatoriinae. Biochem. Syst. Ecol. 29, 1115–1137.

Kohonen, T., 1995. Self-Organizing Maps. Springer-Verlag, Berlin, Heidelberg, NY.

Molecular Networks GmbH, 2002. Available from: <http://www.mol-net.de/software/sonnia/index.html>.

Sadowski, J., Gasteiger, J., 1993. From atoms and bonds to three-dimensional atomic coordinates: automatic model builders. Chem. Rev. 93, 2567–2581.

SciFinder Scholar, 2004. Available from: <http://www.cas.org/SCIFINDER/SCHOLAR2004> (access only with license agreement).

Seaman, F.C., 1982. Sesquiterpene lactones as taxonomic characters in the Asteraceae. Bot. Rev. 48, 123–551.

Spring, O., Buschmann, H., 1996. A chemosystematic survey of sesquiterpene lactones in the Helianthinae (Compositae). In: Hind, D.J.N., Beentje, H.J. (Eds.), Compositae: Systematics, Proceedings of the International Compositae Conference, 1994, vol. 1. Kew Scientist, Kew, pp. 307–311.

Spring, O., Zipper, R., Conrad, J., Vogler, B., Klaiber, I., Da Costa, F.B., 2003. Sesquiterpene lactones from glandular trichomes of *Viguiera radula* (Heliantheae; Asteraceae). Phytochemistry 62, 1185–1189.

Zamorano, G., Catalán, C.A.N., Díaz, J.G., Herz, W., 1995. Sesquiterpene dilactones from *Mikania ypacarayensis*. Phytochemistry 38, 1257–1260.

Zdero, C., Bohlmann, F., 1990. Systematics and evolution within the Compositae, seen with the eyes of a chemist. Plant Syst. Evol. 171, 1–14.

Zupan, J., Gasteiger, J., 1999. Neural Networks in Chemistry and Drug Design, second ed. Wiley-VCH, Weinheim.