Available online at www.sciencedirect.com
ScienceDirect
Bioorganic & Medicinal Chemistry 15 (2007) 2927–2934
Bioorganic & Medicinal Chemistry

Quantitative structure–activity relationship of sesquiterpene lactones with cytotoxic activity

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

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

Received 9 January 2007; revised 1 February 2007; accepted 8 February 2007
Available online 11 February 2007

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). In this study, we investigated a set of 37 different sesquiterpene lactones, represented by 4 skeletons (14 germacranolides, 6 elemanolides, 9 guaianolides and nor-derivatives, and 8 pseudoguaianolides), in what it says respect of their cytotoxic properties. The experimental results were submitted to a QSAR study. A single model for the entire data set was described using 3D molecular descriptors and genetic algorithms establishing structure–activity relationships among the compounds. Important properties for the inhibition potency are discussed for the whole data set and for subsets of the different structural skeletons.

1. Introduction

Sesquiterpene lactones (SLs) are a large group of natural products that have been isolated from numerous genera of the Asteraceae family. They are described as the active compounds of various medicinal plants used in traditional medicine and are known to possess a wide variety of biological and pharmacological activities, such as antimicrobial, cytotoxic, and anti-inflammatory activities, effects on the central nervous and cardiovascular systems as well as allergenic potency.[1] The activities 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. These structural groups react with nucleophiles, especially cystein sulfhydryl groups, by a Michael-type addition.[2,3] Therefore, exposed thiol groups, such as cystein residues in proteins, appear to be the primary targets of sesquiterpene lactones, thereby inhibiting a variety of cellular functions which directs the cells into apoptosis.[4,5] The differences in activity among individual SLs may be explained by different numbers of alkylating structural elements.[6–9] However, other factors such as lipophilicity, molecular geometry, and the chemical environment or the target sulfhydryl may also influence the activity of these compounds.[6,9,10]

A number of significant cytotoxic sesquiterpene lactones have been isolated during the course of a continuous search for agents with this activity from plant sources.[11] However, little is known about the effects of different alkylant structure elements and of other structural factors on cytotoxicity in terms of quantitative structure–activity relationships (QSAR). This would be an important step in the direction of rational lead optimization.[12] The various chemometrical tools that are useful or used in QSAR were utilized by our group in other sources of natural products, such as in the pattern recognition of ^{13}C NMR,[13] in the prediction of the H NMR,[14] in the prediction of secondary metabolites in *Artemisia genus* (Asteraceae),[15] in the identification of chemical constituents isolated,[16] and in the classification of skeletons.[17]

The physical-chemical properties as well as the biological activity of organic compounds depend on their molecular structures. With the purpose of obtaining relationships between chemical structures and biological activities using computational approaches, it

Keywords: Cytotoxic activity; Sesquiterpene lactones; QSAR; Molecular descriptors.
* Corresponding author. Fax: +55 11 38155579; e-mail: vdpemere@iq.usp.br

0968-0896/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2007.02.005

is necessary to find appropriate representations of the molecular structure of the compounds.[18]

A molecular descriptor can be considered as a result of a logical and mathematical procedure, applied to chemical information codified through a molecule's representation.[18]

The encoded information of a molecular descriptor depends on the kind of molecular representation that is used and on the defined algorithm for its calculation. There are simple molecular descriptors derived from the counting of some atom-types or structural fragments in the molecule, others derived from algorithms applied to a topological representation (molecular graph) and usually called topological or 2D-descriptors, and there are molecular descriptors derived from a geometrical representation, which are called geometrical or 3D-descriptors.[18]

DRAGON program generates different groups of descriptors: 2D autocorrelations,[19–21] geometrical descriptors,[22–24] RDF descriptors,[25] 3D-MoRSE descriptors,[26,27] Weighted Holistic Invariant Molecular (WHIM),[28,29] GETAWAY (GEometry, Topology, and Atom-Weights AssemblY),[30,31] among others.

Some QSAR studies about SLs were developed.[7,12,32,33] Kupchan[7] obtained models using log *P*, the partition coefficient in octanol-H_2O solvent. They showed little significance for the 26 compounds studied. He suggested that electronic and steric characteristics can possibly explain the biological activity. Here, we report a more extensive study of structure–cytotoxicity relationships, in order to delineate further the structural requirements for this biological activity and to predict the cytotoxicity potency of unknown SLs.

2. Results and discussion

Regression analysis of training set generated model **I** that contains SPAN, G(O…O), Mor15u, Mor13m, and R8e$^+$ descriptors, which are able to explain 82.6% of variance in cytotoxic activity.

$$\begin{aligned}(\mathbf{I})\quad pED_{50} = {} & +0.484(\pm 0.232)\ \text{SPAN} \\ & -0.011(\pm 0.005)\text{G(O}\ldots\text{O)} \\ & +0.791(\pm 0.459)\ \text{Mor13m} \\ & +0.297(\pm 0.260)\ \text{Mor15u} \\ & -84.459(\pm 27.104)\text{R8e}^+ + 6.250(\pm 1.456)\end{aligned}$$

($n = 28$; $r^2 = 0.826$; $s = 0.258$; $F = 21.04$; $Q^2_{cv} = 0.743$; S-PRESS = 0.314; $n_{ext} = 9$; $r^2_{ext} = 0.800$; $Q^2_{ext} = 0.704$).

Analyzing Table 1 and Figure 1, it can be noticed that the latter depicts the adjustment regarding a straight line of the points used for the calibration of the model. This linear approach is validated analyzing Figure 2, which is a graph of the residues between the experimental and the predicted biological activity, presenting a random distribution of the points.

Table 1. Experimental values of pED_{50}, values calculated by the Eq. **I** for the training set and the respective errors

Compound	pED_{50} experimental	pED_{50} calculated	Errors (calculated – experimental)
2	4.58	4.44	−0.14
3	5.19	5.48	0.29
4	5.61	5.13	−0.48
5	5.29	5.25	−0.04
6	4.98	5.12	0.14
7	3.91	3.95	0.04
9	5.00	5.29	0.29
10	5.06	5.52	0.46
11	5.89	5.49	−0.40
12	5.74	5.61	−0.13
13	5.86	5.95	0.09
14	5.90	5.75	−0.15
16	6.29	6.34	0.05
17	5.08	5.17	0.09
18	5.07	5.12	0.05
19	5.12	5.42	0.30
23	6.57	6.36	−0.21
24	5.06	5.03	−0.03
25	5.39	5.01	−0.38
26	6.25	6.00	−0.25
27	5.15	5.20	0.05
29	5.18	5.44	0.26
30	6.28	6.07	−0.21
31	5.49	5.77	0.28
34	5.30	5.21	−0.09
35	5.39	5.26	−0.13
36	5.30	5.40	0.10
37	5.41	5.55	0.14

Table 2 and Figure 3 show a significant validation of the test set. There is a significant linear adjustment and the model **I** was able to predict and differentiate the most active compounds from the others.

The descriptors that appear in the most statistically significant equation (Eq. **I**) that best correlates the sesquiterpene lactone structures to their cytotoxic activity are 3D MoRSE (Mor15u e Mor13m), Geometrical (SPAN e G(O…O)), and GATEWAY (R8e$^+$). It is important to notice that a small number of descriptors

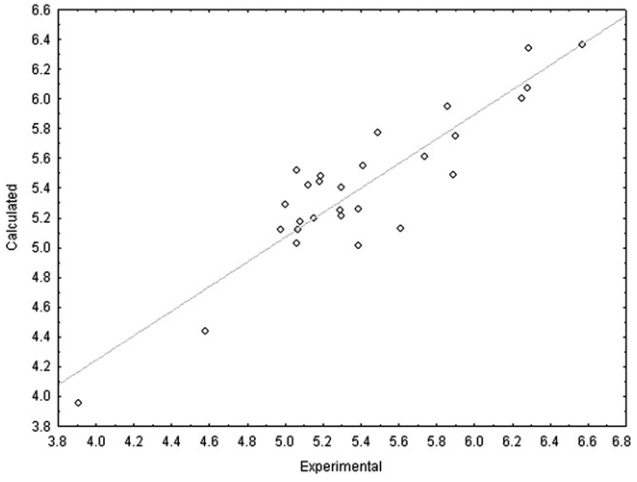

Figure 1. Graphic of experimental activity values (pED_{50}) versus calculated activity values for the training set.

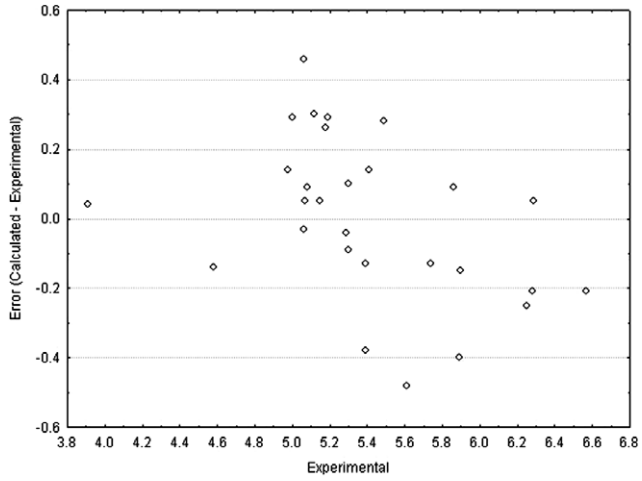

Figure 2. Graphic of experimental activity (pED_{50}) values versus respective errors for the training set.

Table 2. Experimental values of pED_{50}, values predicted by Eq. **I** for the test set, and the respective errors

Compound	pED_{50} experimental	pED_{50} predicted	Errors (predicted – experimental)
1	3.90	4.86	0.96
8	5.26	5.47	0.21
15	6.12	6.04	−0.08
20	6.46	5.86	−0.60
21	5.21	5.13	−0.08
22	5.68	5.79	0.11
28	5.98	6.02	0.04
32	5.60	5.49	−0.11
33	5.91	6.02	0.11

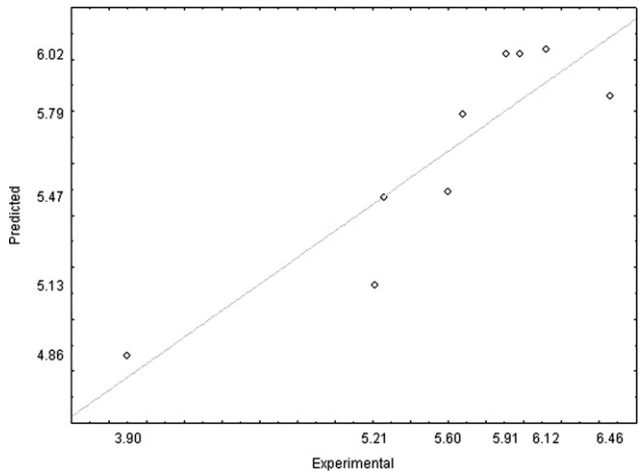

Figure 3. Graphic of experimental activity values (pED_{50}) versus predicted activity values for the test set.

were sufficient to predict cytotoxicity for the studied molecule set.

Analyzing Eq. **I**, it can be verified that the value of the coefficient of internal prediction Q^2_{cv} is significant (0.743), what indicates a robust model. The F value is highly significant, because for 95% of trust with 5 and 22° of freedom, the necessary minimum value is 2.66.

Biological activity is intimately tied to the three-dimensional structure of molecules and to electronic properties of specific sites of the molecule. The potential of the 3D MoRSE code simultaneously considers the 3D structure and atomic properties, such as partial atomic charges, making it particularly suited for studying biological data.[12] The Mor15u descriptor is strictly related to the stereochemistry of the compounds, as well as the descriptor Mor13m. However, the latter also uses the atoms' weight for the calculations.

The SPAN descriptor has its calculation based on the choice of the radius of the smallest sphere, centered on the center of mass, enclosing completely all atoms of a molecule. So, compounds that have a great number of ramifications and groups able to dislocate the center of mass have their radius increased.[18]

The G(O...O) descriptor represents the sum of the geometric distances between all pairs of oxygen atoms. Greater the number of these atoms in a molecule, and/or farther the distances between them, bigger will be this sum.[24]

The GETAWAY descriptors (Geometric Topology and Atom Weights Assembly) are related to the influence of the atoms in the determination of the molecular form (leverages), and to the distance between them. The $R8e^+$ descriptor is the maximum value of the calculation that uses the multiplication of the leverages between two atoms and the maximum value of the respective electronegativities of Sanderson, with topological distance equal to 8, both divided by the geometrical distance between them. Greater the influence of two atoms in the molecular form and their electronegativity and nearer the distance between them, greater will be the value of $R8e^+$.[20–22]

Regression analysis of the training set with all the 37 compounds generated model **II**, which contains SPAN, H0e, Mor15u, Mor13m, and $R8e^+$ descriptors, which are able to explain 80.3% of variance of cytotoxic activity.

$$\begin{aligned}(\mathbf{II}) \quad pED_{50} = {} & +0.214(\pm 0.121)\ \text{SPAN} \\ & - 3.288(\pm 1.450)\text{H0e} \\ & + 1.285(\pm 0.384)\text{Mor13m} \\ & + 0.872(\pm 0.527)\ \text{Mor15u} \\ & - 52.521(\pm 30.814)R8e^+ \\ & + 16.898(\pm 4.067)\end{aligned}$$

($n = 37$; $r^2 = 0.803$; $s = 0.288$; $F = 25.19$; $Q^2_{cv} = 0.742$; $S\text{-}_{PRESS} = 0.314$).

As in Eq. **I**, it can be verified that the value of the coefficient of internal prediction Q^2_{cv} is significant (0.742), what indicates a robust model. The F value is highly

significant, because for 95% of trust with 5 and 31 degrees of freedom, the necessary minimum value is 2.53.

All the descriptors already cited (except the G(O...O)) and explained above appear in this equation as well, but a new one can be noticed: H0e. This descriptor, a GETAWAY one, is related to the electronegativity, size, and location of the atom in the molecule. Greater the electronegativity, the size of the atom and further the distance between the atom and the center of the molecule, greater will be the descriptor's value.

The most statistically significant equation found (**I**) utilizes the descriptors previously cited. All of them are calculated by a 3D representation of the molecules. This reveals that the parameters related to conformation and stereochemistry are the most important in what it says respect to the cytotoxic activity of these sesquiterpene lactones.

Analyzing the structures of the compounds, separating them into sets that obtained the greatest and the smallest values of biological activity and comparing them by groups they own, which can be responsible for the high or low activity, some considerations could be made.

First, it is important to stand out that all the compounds, showing low or high activities, possess the α-methylene-γ-lactone structure.

The compounds that showed the highest biological activities (activity >6.0) were **15**, **16**, **20**, **23**, **26**, **28**, and **30**. The compounds **15**, **16**, **20**, **26**, and **30** have a double bond in the five-membered ring, at position 3 (Fig. 4). The compounds **16**, **20**, **26**, **28**, and **30** have a hydroxyl in the carbon that binds the cycloheptane and the cyclopentane (position 5). There is no distance between the cyclopentane and the lactone, that is, there is no carbon between them.

The compounds that have both structures (**16**, **20**, **26**, and **30**) possess values of activity which do not show a great variation, and the compound **20** has the highest value among them.

The highest active substances possess as skeletal the guaianolide (**16**, **20**, **26**, **28**, and **30**) and the pseudoguaianolide (**11**, **12**, **13**, **14**, and **15**) types (Fig. 4), and the first shows the greatest activities. This fact tells that the methyl group at position 4 can be more important to the activity than position 5. The exception is compound **28** (5.98), which has lower activity than **15** (6.12), probably because that it is the only one among the guaianolides that does not have the double bond at position 3. It can be noticed that the other guaianolide compounds possess a hydrophilic group at that same position.

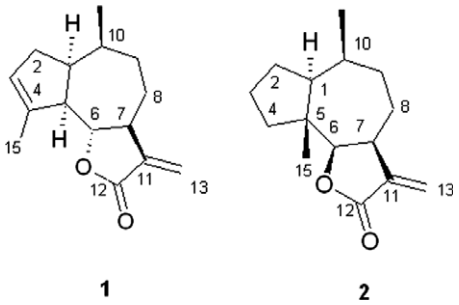

Figure 4. Guaianolide (**1**) and pseudoguaianolide (**2**) skeletons.

The substance **19**, which is a guaianolide, has lower activity, probably because there is not the double bond in the five-membered ring, as a double bond at position 10.

The compounds **17** and **28** show a very similar structure; however, the presence of a methylene group instead of a hydroxyl and a methyl chloride at position 10 contributes to a considerable increase in the activity in **28**, which shows that the double bond in the position is important to the activity. It is evident that the importance of the existence of a double bond is greater to the activity than the existence of a hydroxyl at position 5, since both compounds have this group.

This importance can also be visualized comparing the compounds **26**, **27**, **28**, **29**, and **30**. The first (**26**), which possesses a considerable activity, but lower than the compound **30**, has the double bond in the five-membered ring and the hydroxyl group at position 5. The third substance (**28**), with lower activity than compound **30**, shows the double bond as a methylene group at position 10 and the hydroxyl group at position 5. The compounds **27** and **29**, which have the lowest activities among this group, possess an epoxy group at position 10 in the ring, instead of a methylene group, a hydroxyl group at position 5, and do not have the double bond in the five-membered ring (position 3), as the compound **28**. The latter (**30**), which shows the highest activity, has the double bond at position 3 and an epoxy group at position 10 in the ring. It can be concluded that the double bond (as a methylene group at position 10 or at position 3) is important to the activity. It can be probably due to the alkylating property of the sesquiterpene lactones.

Another important feature in compounds of the guaianolide skeleton is the presence of a 8β-angeloyloxy and, consequently, the 6,12-lactonization in the structures. These compounds (**16**, **17**, **20**, **26**–**30**) exhibited the highest cytotoxic activities. From this observation one can suggest that these structural features are very relevant to the biological activity.

Among the compounds **9**–**15**, which are pseudoguaianolides of the ambrolide types, the presence of a double bond in the cyclopentanone (α,β-unsaturated carbonyl) increases the activity in **11**–**15** which shows, as already cited, the importance of these groups in the biological activity of the sesquiterpene lactones that possess a pseudoguaianolide skeleton.

As stereochemistry is an important factor to the biological activity and as the highest active compounds are

guaianolides and pseudoguaianolides, it can be suggested that these skeletal types probably have the most suitable stereochemistry to the cytotoxic activity than the other types.

The electronic features, which can also explain the descriptors used in this model, can be associated to the presence of double bonds. These structures increase the electron cloud of the molecules as well as they can influence the molecular form with conformations. The presence of oxygen atoms increases the electron cloud as well, but these structures seem to have secondary importance in activity, since there are compounds with a great number of oxygen atoms that do not show a

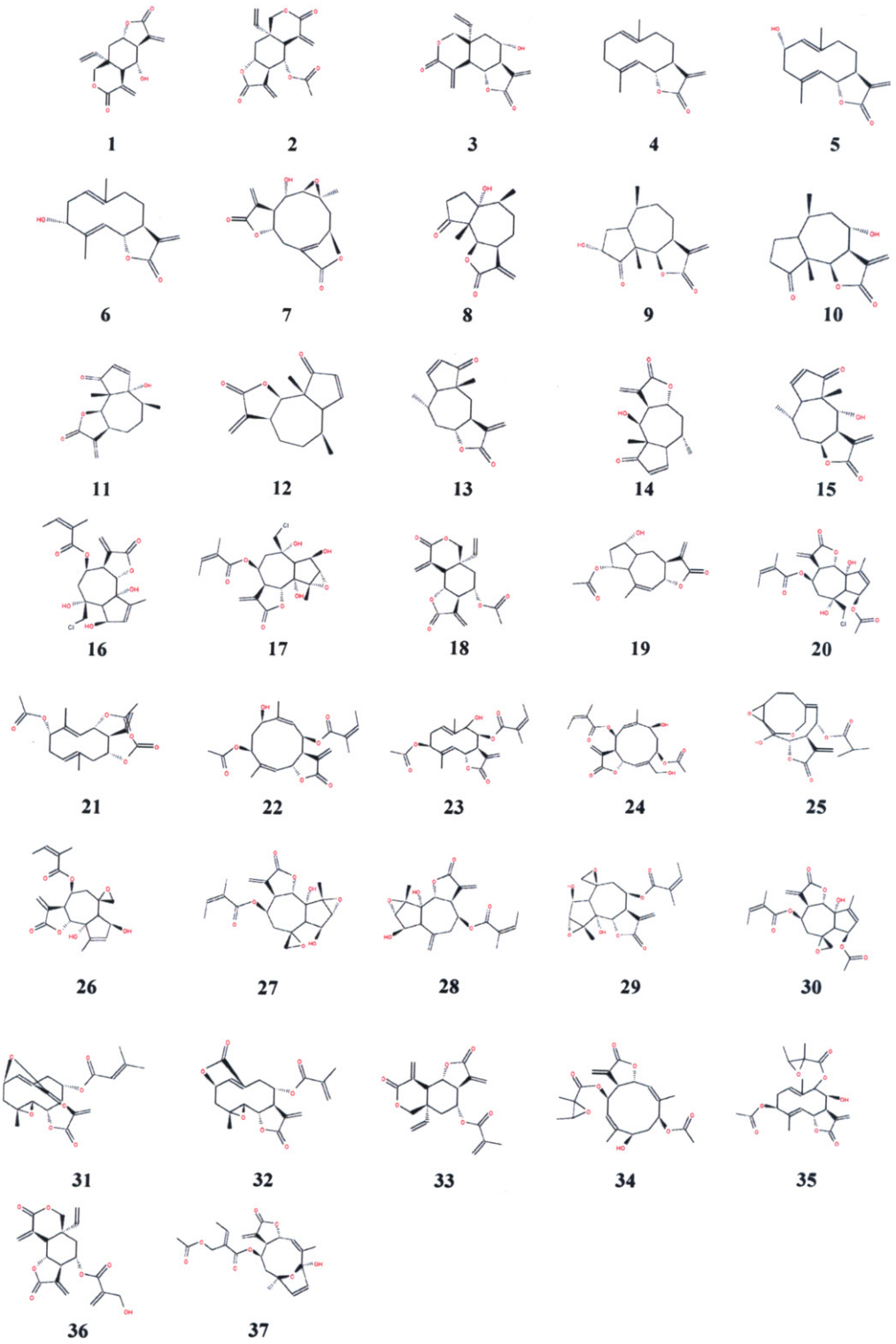

Figure 5. Structures of the investigated sesquiterpene lactones.

considerable activity such as, for example, the compounds **7**, **36**, and **37**. Thus, the type of descriptors shows an agreement with the features analyzed above.

3. Conclusions

First, it is important to focus on the type of information encoded in the QSAR descriptors in this study. Clearly, the descriptors appearing in the most statistically significant equation found represent a global description of steric properties and electronic features for each molecule, which was also found in previous studies.[8,9,12]

It could be concluded using the compounds present in this study, 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 this study it could also be concluded that the most active compounds are those which have the guaianolide and pseudoguaianolide skeletal types.

A more extensive study is needed to compare a greater number of compounds, including substances that do not possess the α-methylene-γ-lactone structure, to elucidate if it is important to the activity or not, as well as sesquiterpene lactones with a more varied skeletal types, to conclude if the guaianolide and pseudoguaianolide show greater activity than the other ones.

4. Materials and methods

Sesquiterpene lactones **1**–**37** (Fig. 5 and Table 3) were obtained from different Asteraceae species and were assayed for inhibitory activity in vitro against cells derived from human carcinoma of the nasopharynx (KB) carried in cell culture ($pED_{50} = -\log ED_{50}$) by Kupchan et al.[7,34–36] The parameter pED_{50} was calculated by ED_{50} values expressed as μmol/L.

Using the parameter ED_{50}, one can only conclude that the substances are cytotoxic. It would not be suitable if the word antitumoral were used, because this activity can only be measured when a specific action mechanism is proved and when a therapeutic index is evaluated.

Table 3. Substances, their respective skeletons, and experimental values of ED_{50} and pED_{50}

Number	Substance[a]	Skeleton	ED_{50} (μg/mL)	ED_{50} (μmol/L)	pED_{50}
1	Vernomenin (**12a**)[7]	Elemanolide	35	127.00	3.9
2	Vernomenin acetate (**12b**)[7]	Elemanolide	8	26.20	4.58
3	Vernolepin (**7a**)[7]	Elemanolide	1.8	6.52	5.19
4	Costunolide (**11a**)[7]	Germacranolide	0.57	2.46	5.61
5	Tamaulipin A (**11b**)[7]	Germacranolide	1.26	5.08	5.29
6	Tamaulipin B (**11c**)[7]	Germacranolide	2.6	10.50	4.98
7	Elephantol (**5a**)[7]	Germacranolide	36	123.20	3.91
8	Coronopilin (**20a**)[7]	Pseudoguaianolide	1.45	5.49	5.26
9	3-Hydroxydamsin (**20b**)[7]	Pseudoguaianolide	2.65	10.00	5
10	Desacetylconfertiflorin (**20c**)[7]	Pseudoguaianolide	2.3	8.71	5.06
11	Parthenin (**21a**)[7]	Pseudoguaianolide	0.34	1.30	5.89
12	Ambrosin (**21b**)[7]	Pseudoguaianolide	0.45	1.83	5.74
13	Aromaticin (**22a**)[7]	Pseudoguaianolide	0.34	1.38	5.86
14	Mexicanin I (**23**)[7]	Pseudoguaianolide	0.33	1.26	5.9
15	Helenalin (**22b**)[7]	Pseudoguaianolide	0.2	0.76	6.12
16	Eupachlorin (**6**)[35]	Guaianolide	0.21	0.51	6.29
17	Eupachloroxin (**13**)[35]	Guaianolide	3.6	8.39	5.08
18	Vernolepin acetate (**12b**)[7]	Elemanolide	2.7	8.49	5.07
19	Gaillardin (**13**)[7]	Guaianolide	2.3	7.52	5.12
20	Eupatundin (**14**)[7]	Guaianolide	0.47	1.25	6.46
21	Eupachlorin acetate (**15**)[7]	Germacranolide	0.16	0.35	5.21
22	Chammissonin diacetate (**17**)[7]	Germacranolide	2.13	6.12	5.68
23	Eupatocunin (**6**)[36]	Germacranolide	0.11	0.27	6.57
24	Eupacunolin (**19**)[36]	Germacranolide	3.7	8.80	5.06
25	Vernomygdin (**8**)[34]	Germacranolide	1.5	4.12	5.39
26	Euparotin (**2**)[35]	Guaianolide	0.21	0.56	6.25
27	Eupatoroxin (**7**)[35]	Guaianolide	2.8	7.14	5.15
28	Eupatundin (**14**)[7]	Guaianolide	0.47	1.04	5.98
29	10-Epieupatoroxin (**12**)[35]	Guaianolide	2.6	0.63	5.18
30	Euparotin acetate (**16**)[7]	Guaianolide	0.22	0.53	6.28
31	Elephantin (**1b**)[7]	Germacranolide	1.16	3.22	5.49
32	Elephantopin (**1a**)[7]	Germacranolide	0.94	2.51	5.6
33	Vernolepin methacrylate (**7c**)[7]	Elemanolide	0.42	1.22	5.91
34	Eupacunoxin (**2**)[36]	Germacranolide	2.1	5.00	5.3
35	Eupatocunoxin (**7**)[36]	Germacranolide	1.7	4.04	5.39
36	Vernodalin (**1**)[34]	Elemanolide	1.8	5.00	5.3
37	Liatrin (**18**)[7]	Germacranolide	1.62	3.93	5.41

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

All structure data were extracted from the SISTEMAT database.[37–39] An in-house program for data extraction was written in Java and subsequently used to select the SL compounds.

Molecular modeling computations were performed on SPARTAN for Windows v. 4.0 software (Wavefunction, Inc., Irvine, Calif.). The molecules were subjected to geometry optimization and conformational analysis (systematic analysis with dihedral angle rotationed at each 30°). The semi-empirical quantum chemical method used was AM1 (Austin Model 1)[40,41] and the root mean square (RMS) gradient value of 0.001 kcal/mol as termination condition. Molecules with their energies minimized were saved as MDL MolFiles for computing various molecular descriptors using DRAGON Professional version 5.4.[42]

The groups of descriptors calculated were: RDF (150 descriptors), 3D-MoRSE (160 descriptors), GETAWAY (197 descriptors), WHIM (99 descriptors), and Geometrical descriptors (74). All of them are 3D descriptors.

The variable selection was evaluated for each descriptor set as well as for the whole set of these molecular descriptors (680 descriptors), because their calculation is based on representation in three dimensions of the molecules.[18]

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 pairwise correlation ($r < 0.99$) analysis was performed to exclude those highly correlated.[43] Thus, the number of DRAGON descriptors used in our calculations was reduced to: ALL 3D: 396 descriptors; 3D-MORSE: 104 descriptors; GETAWAY: 128 descriptors; WHIM: 66 descriptors; Geometrical Descriptors: 22 descriptors; RDF: 71 descriptors.

We used MobyDigs[44] program for the calculation of regression models by using genetic algorithms. The compounds were first divided into two sets based on hierarchical clustering: one training set composed of 28 and one external test set composed of nine compounds. The latter is a suitable analysis of the predictive performance.[43] Models for cytotoxic activity were constructed based on the training set and the generated models were than validated internally and externally. Additionally, we calculated a regression model using all the compounds as the training set.

The search for the best models is performed using the ordinary least squares regression (OLS) under the Genetic Algorithm (GA) approach, that is, by the Variable Subset Selection-Genetic Algorithm (VSS-GA) method. In the GA terminology, a population is characterized by a set of candidate variables (the genetic heritage of the population) and is constituted by individuals, that is, models made of one or more population variables.[45]

Acknowledgments

This work was supported by grants from the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), (CAPES), and by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2007.02.005.

References and notes

1. Picman, A. K. *Biochem. Syst. Ecol.* **1986**, *14*, 255.
2. Kupchan, S. M.; Fessler, D. C.; Eakin, M. A.; Giacobbe, T. J. *Science* **1970**, *168*, 376.
3. Schmidt, T. J. *Bioorg. Med. Chem.* **1997**, *5*, 645.
4. Schmidt, T. J. *Curr. Org. Chem.* **1999**, *3*, 577.
5. Dirsch, V. M.; Stuppner, H.; Vollmar, A. M. *Cancer Res.* **2001**, *61*, 5817.
6. Heilmann, J.; Wasescha, M. R.; Schmidt, T. J. *Bioorg. Med. Chem.* **2001**, *9*, 2189.
7. Kupchan, S. M.; Eakin, M. A.; Thomas, A. M. *J. Med. Chem.* **1971**, *14*, 1147.
8. Karabunarliev, S.; Mekenyan, O. G.; Karcher, W.; Russom, C. L.; Bradbury, S. P. *Quant. Struct-At. Relat.* **1996**, *15*, 302.
9. Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959.
10. Kupchan, S. M. *Trans. N. Y. Acad. Sci.* **1970**, *32*, 85.
11. O'Hansch, C. Comprehensive Medicinal Chemistry: The Rational Design, Mechanistic Study and Therapeutic Application of Chemical Compounds. Oxford, 1990.
12. Heilmann, J.; Schmidt, T. J. *Quant. Struct-Act. Relat.* **2002**, *21*, 276.
13. Ferreira, M. J. P.; Oliveira, F. C.; Rodrigues, G. V.; Emerenciano, V. P. *Internet Electron. J. Mol. Des.* **2004**, *3*, 737–749, http://www.biochempress.com.
14. Costantin, M. B.; Ferreira, M. J. P.; Rodrigues, G. V.; Emerenciano, V. P. *Internet Electron. J. Mol. Des.* **2004**, *3*, 759–770, http://www.biochempress.com.
15. Schwabe, T.; Ferreira, M. J. P.; Alvarenga, S. A. V.; Emerenciano, V. P. *Internet Electron. J. Mol. Des.* **2005**, *4*, 9–16, http://www.biochempress.com.
16. Ramos, M. P. P.; Ferreira, M. J. P.; Lopes, L.; Emerenciano, V. P. *Internet Electron. J. Mol. Des.* **2005**, *4*, 173–180, http://www.biochempress.com.
17. Emereciano, V. P.; Scotti, M. T.; Stefani, R.; Alvarenga, S. A. V.; Nuzillard, J. M.; Rodrigues, G. V. *Internet Electron. J. Mol. Des.* **2006**, *5*, 213–223, http://www.biochempress.com.
18. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim (Germany), 2000.
19. Moran, P. A. P. *Biometrika* **1950**, *37*.
20. Geary, R. C. *The Incorpored Statistician* **1954**, *5*, 115.
21. Moreau, G.; Broto, P. *Nouveau Journal de Chimie—New Journal of Chemistry* **1980**, *4*, 359.
22. Diudea, M. V.; Horvath, D.; Graovac, A. *J. Chem. Inf. Comp. Sci.* **1995**, *35*, 129.
23. Balaban, A. T. *J. Chem. Inf. Comp. Sci.* **1997**, *37*, 645.
24. Randic, M.; Kleiner, A. F.; Dealba, L. M. *J. Chem. Inf. Comp. Sci.* **1994**, *34*, 277.

25. Hemmer, M. C.; Steinhauer, V.; Gasteiger, J. *Vibrat. Spectrosc.* **1999**, *19*, 151.
26. Gasteiger, J.; Sadowski, J.; Schuur, J.; Selzer, P.; Steinhauer, L.; Steinhauer, V. *J. Chem. Inf. Comp. Sci.* **1996**, *36*, 1030.
27. Schuur, J. H.; Selzer, P.; Gasteiger, J. *J. Chem. Inf. Comp. Sci.* **1996**, *36*, 334.
28. Consonni, V.; Todeschini, R.; Pavan, M. *J. Chem. Inf. Comp. Sci.* **2002**, *42*, 682.
29. Consonni, V.; Todeschini, R.; Pavan, M.; Gramatica, P. *J. Chem. Inf. Comp. Sci.* **2002**, *42*, 693.
30. Todeschini, R.; Gramatica, P. *Quant. Struct-Act. Relat.* **1997**, *16*, 113.
31. Todeschini, R.; Gramatica, P. *Quant. Struct-Act. Relat.* **1997**, *16*, 120.
32. Siedle, B.; García-Piñeres, A. J.; Murillo, R.; Schulte-Mönting, J.; Castro, V.; Rüngeler, P.; Klaas, C. A.; Da Costa, F. B.; Kisiel, W.; Merfort, I. *J. Med. Chem.* **2004**, *47*, 6042.
33. Wagner, S.; Hofmann, A.; Siedle, B.; Terfloth, L.; Merfort, I.; Gasteiger, J. *J. Med. Chem.* **2006**, *49*, 2241.
34. Kupchan, S. M.; Hemingway, R. J.; Karim, A.; Werner, D. *Del. J. Org. Chem.* **1969**, *12*, 3908.
35. Kupchan, S. M.; Kelsey, J. E.; Maruyama, M.; Cassady, J. M.; Hemingway, J. C.; Knox, J. R. *J. Org. Chem.* **1969**, *12*, 3876.
36. Kupchan, S. M.; Maruyama, M.; Hemingway, R. J.; Hemingway, J. C.; Shibuya, S.; Fujita, T. *J. Org. Chem.* **1973**, *12*, 2189.
37. Gastmans, J. P.; Furlan, M.; Lopes, M. N.; Borges, J. H. G.; Emerenciano, V. P. *Química Nova* **1990**, *13*, 10.
38. Gastmans, J. P.; Furlan, M.; Lopes, M. N.; Borges, J. H. G.; Emerenciano, V. P. *Química Nova* **1990**, *13*, 75.
39. Ferreira, M. J. P.; Rodrigues, G. V.; Emerenciano, V. P. *Can. J. Chem.—Revue Canadienne de Chimie* **2001**, *79*, 1915.
40. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
41. Dewar, M. J. S.; Healy, E. F.; Holder, A. J.; Yuan, Y. C. *J. Comput. Chem.* **1990**, *11*, 541.
42. Talete, S.R.L. DRAGON for Windows (Software for Molecular Descriptor Calculations). Version 5.4 – 2006 – http://www.talete.mi.it.
43. Livingstone, D. *Data Analysis for Chemists*; Oxford Science Publications.: New York, 1995, pp. 48–120.
44. Talete, S. R. L. Mobydigs Academic version - Version 1.0 – 2004 – http://www.talete.mi.it.
45. Leardi, R.; Boggia, R.; Terrile, M. *J. Chemom.* **1992**, *6*, 267.