UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE QUÍMICA
Programa de Pós-Graduação em Química

MARCUS TULLIUS SCOTTI

Emprego de Redes Neurais e de Descritores Moleculares em Quimiotaxonomia da Família Asteraceae

São Paulo

Data do Depósito na SPG:
30/05/2008

MARCUS TULLIUS SCOTTI

Emprego de Redes Neurais e de Descritores Moleculares em Quimiotaxonomia da Família Asteraceae

Tese apresentada ao Instituto de Química da Universidade de São Paulo para obtenção do Título de Doutor em Química (Química Orgânica)

Orientador(a): Prof(a). Dr(a). Nome do(a) Orientador(a)

São Paulo
2008

Aos meus pais, Tullio Scotti e Lélia de Medeiros Scotti pelo todo o apoio e companheirismo, principalmente nos momentos mais difíceis.

À minha irmã, Luciana Scotti, por ser a principal fonte de perseverança na minha vida.

AGRADECIMENTOS

À minha noiva, Kátia Fernandes Babesco, por ter sempre me apoiado.

Ao Professor Dr. Vicente de Paulo Emerenciano, do Instituto de Química da USP, pela sua paciência, orientação, pelas conversas agradáveis e pelo seu altruísmo em compartilhar idéias.

Aos Professores do Instituto de Química da USP, por suas aulas e pela atenção fornecida.

Ao Mauro Vicentini, por suas contribuições nesta tese e pelas conversas mais que agradáveis.

A Mariane Ballerini Fernandes por sua contribuição nesta tese.

Aos meus colegas de laboratório, Michelle Rossini e Harold Fokoue, por suas conversas agradáveis e pelo esforço em cooperar.

À CAPES pela bolsa de doutorado fornecida.

“Se esta ciência que traz grandes benefícios ao homem, não servir para entendê-lo, terminará voltando-se contra ele.”

Giordano Bruno

RESUMO

Scotti, M.T. Emprego de Redes Neurais e de Descritores Moleculares em Quimiotaxonomia da Família Asteraceae. 2008. 165p. Tese (Doutorado) - Programa de Pós-Graduação em Química. Instituto de Química, Universidade de São Paulo, São Paulo.

Esse trabalho descreve o desenvolvimento de uma nova ferramenta quimioinformática designada de SISTEMATX que possibilitou a análise quimiotaxonômica da família Asteraceae, empregando novos parâmetros moleculares, bem como o estudo da relação quantitativa estrutura química – atividade biológica de substâncias provenientes desse grupo vegetal.

A família Asteraceae, uma das maiores entre as angiospermas, caracteriza-se quimicamente pela produção de sesquiterpenos lactonizados (SLs). Um total de 1111 (SLs), extraídos de 658 espécies, 161 gêneros, 63 subtribos e 15 tribos da família Asteraceae foram representados e cadastrados em duas dimensões no SISTEMATX e associados à respectiva origem botânica. A partir dessa codificação, o grau de oxidação e as estruturas em três dimensões de cada SL foram obtidos pelo sistema. Essas informações, associadas aos dados botânicos, foram exportadas para um arquivo texto, o qual permitiu a obtenção de vários tipos de descritores moleculares. Esses parâmetros moleculares foram correlacionados com o grau de oxidação médio por tribo e tiveram sua seleção realizada por regressão linear múltipla utilizando algoritmo genético. Equações com coeficientes estatísticos variando entre $0,725 \leq r2 \leq 0,981$ e $0,647 \leq Qcv2 \leq 0,725$ foram obtidas com apenas um descritor, possibilitando a identificação de algumas características estruturais relacionadas ao grau de oxidação. Não foi obtida nenhuma relação entre o grau de oxidação dos SL e a evolução das tribos da família Asteraceae.

Os descritores moleculares também foram usados como dados de entrada para separar as ocorrências botânicas através de mapas auto-organizáveis (rede não supervisionada Kohonen). Os mapas gerados, com cada bloco de descritor, separaram as tribos da família Asteraceae com valores de índices de acerto total entre 66,7% e 83,6%. A análise desses resultados evidencia semelhanças entre as tribos Heliantheae, Helenieae, e Eupatorieae e, também, entre as tribos Anthemideae e Inuleae. Tais observações são coincidentes com as classificações sistemáticas propostas por Bremer, que utilizam principalmente dados morfológicos e, também, moleculares. A mesma abordagem foi utilizada para separar os ramos da tribo Heliantheae, segundo a classificação proposta por Stuessy, cuja separação é baseada no número de cromossomos das subtribos. Os mapas auto-organizáveis obtidos separam em duas regiões distintas os ramos A e C, com elevados índices de acerto total que variam entre 81,79% a 92,48%.

Ambos os estudos demonstram que os descritores moleculares podem ser utilizados como uma ferramenta para classificação de táxons em níveis hierárquicos baixos, tais como tribos e subtribos. Adicionalmente, foi demonstrado que os marcadores químicos corroboram parcialmente com as classificações que empregam dados morfológicos e moleculares. Os descritores obtidos por fragmentos ou pela representação da estrutura dos SLs em duas dimensões foram suficientes para

obtenção de resultados significativos, não sendo obtida melhora nos resultados com os descritores que utilizam a representação em três dimensões das estruturas.

Paralelamente, um estudo adicional foi realizado relacionando a estrutura química, representada pelos mesmos descritores moleculares anteriormente mencionados, com a atividade citotóxica de 37 SLs frente às células tumorais da nasofaringe KB. Uma equação com índices estatísticos significativos (r2=0,826 e Qcv2=0,743) foi obtida. Os cinco descritores, selecionados a partir de uma equação estatisticamente mais significativa, representam uma descrição global de propriedades estéricas e características eletrônicas de cada molécula que auxiliaram na determinação de fragmentos estruturais importantes para a atividade citotóxica. Tal modelo permitiu verificar que os esqueletos carbônicos dos tipos guaianolídeo e pseudoguaianolídeo são encontrados nos SLs que apresentam maior atividade citotóxica.

Palavras-chave: Descritores Moleculares, Redes Neurais, Asteraceae, Quimiotaxonomia, Mapas Auto-Organizáveis, Kohonen.

ABSTRACT

Scotti, M.T. **Use of Neural Networks and Molecular Descriptors in Chemotaxonomy of the Asteraceae Family.** 208. 165p. PhD Thesis - Graduate Program in Chemistry. Instituto de Química, Universidade de São Paulo, São Paulo.

This work describes the development of a new chemoinformatic tool named SISTEMATX that allowed the chemotaxonomic analysis of the Asteraceae family employing new molecular parameters, as well as the quantitative structure activity relationship study of compounds produced by this botanical group.

The Asteraceae, one of the largest families among of angiosperms, is chemically characterized by the production of sesquiterpene lactones (SLs). A total of 1111 (SLs), extracted from 658 species, 161 genera, 63 subtribes and 15 tribes of the Asteraceae, were represented and registered in two dimensions in the SISTEMATX and associated with their botanical source. From this codification, the degree of oxidation and the structures in three dimensions of each SL were obtained by the system. These data linked with botanical origin were exported for a text file which allow the generation of several types of molecular descriptors. These molecular parameters were correlated with the average oxidation degree by tribe and were selected by multiple linear regressions using genetic algorithms. Equations with statistical coefficients varying between $0{,}725 \leq r2 \leq 0{,}981$ and $0{,}647 \leq Qcv2 \leq 0{,}725$ were obtained with only one descriptor, making possible the identification of some structural characteristics related to the oxidation level. Any relationship between the degree of oxidation of SL and the tribes evolution of the family Asteraceae was not obtained.

The molecular descriptors were also used as input data to separate the botanical occurrences through the self organizing-maps (unsupervised net Kohonen). The generated maps with each block descriptor, divide the Asteraceae tribes with total indexes values between 66,7% and 83,6%. The analysis of these results shows evident similarities among the Heliantheae, Helenieae and Eupatorieae tribes and, also, between the Anthemideae and Inuleae tribes. Those observations are in agreement with the systematic classifications proposed by Bremer, that use mainly morphologic and, also, molecular data. The same approach was utilized to separate the branches of the Heliantheae tribe, according to the Stuessy's classification, whose division is based on the chromosome numbers of the subtribes. From the obtained self-organizing maps, two different areas (branches A and C) were separated with high hit indexes varying among 81,79% to 92,48%.

Both studies demonstrate that the molecular descriptors can be used as a tool for taxon classification in low hierarchical levels such as tribes and subtribes. Additionally, was demonstrated that the chemical markers partially corroborate with the classifications that use morphologic and molecular data. Descriptors obtained by fragments or by the representation of the SL structures in two dimensions were sufficient to obtain significant results, and were not obtained better results with descriptors that utilize the structure representation in three dimensions.

An additional study was accomplished relating the chemical structure, represented by the same molecular descriptors previously mentioned, with the cytotoxic activity of 37 SLs against tumoral cells derived from human carcinoma of the nasopharynx (KB). An equation with significant statistical indexes was obtained. The five descriptors, selected from the more statistical significant equation, shows a global description of sterical properties and electronic characteristics of each molecule that aid in the determination of important structural fragments for the cytotoxic activity. From the model can be verified that the carbon skeletons of the guaianolide and pseudoguaianolide types are encountered in the SLs that show the higher cytotoxic activity.

Keywords: Molecular Descriptors, Neural Networks, Asteraceae, Chemotaxonomy, Self-Organizing Maps, Kohonen.

ÍNDICE DE FIGURAS

1. INTRODUÇÃO

ÍNDICE DE TABELAS

1. INTRODUÇÃO

3. Metodologia

4. Resultados

LISTA DE ABREVIATURAS E SIGLAS

r - coeficiente de correlação

s - desvio padrão

F- fator de confiabilidade

Q^2 - coeficiente de predição

ED_{50} – concentração que para obter 50% do efeito

pED_{50} – logaritmo negativo de ED_{50}

PCA – análise de componentes principais

PLS – método dos mínimos quadrados parciais

MLR – regressão linear múltipla

NN – Redes Neurais

SOM – Mapas Auto-Organizáveis

SLs – Sesquiterpenos Lactonizados

1. INTRODUÇÃO

1.1. Quimiossistemática

A quimiossistemática consiste na classificação dos organismos através de caracteres químicos, fornecendo algumas respostas e/ou propostas para uma compreensão maior sobre evolução. A quimiossistemática se restringe a análise onde se emprega substâncias como caracteres.

As substâncias produzidas pelos vegetais são chamadas de metabólitos primários e secundários. As plantas utilizam a energia do sol para produzir compostos orgânicos a partir do dióxido de carbono, em um processo chamado fotossíntese. Os produtos iniciais da fotossíntese são os carboidratos, posteriores alterações metabólicas geram uma diversidade de compostos orgânicos de estruturas simples e com baixo peso molecular, entre estes estão açúcares, ácidos carboxílicos e amino ácidos, sendo encontrados em todos os seres vivos. Estes compostos são formados nas tranformações denominadas de processo metabólico primário. Os metabólitos secundários apresentam uma distribuição restrita, e de fontes botânicas específicas (Geissman & Crout 1969). Os metabólitos secundários são divididos em classes como os mostrados na figura 1.1.1.

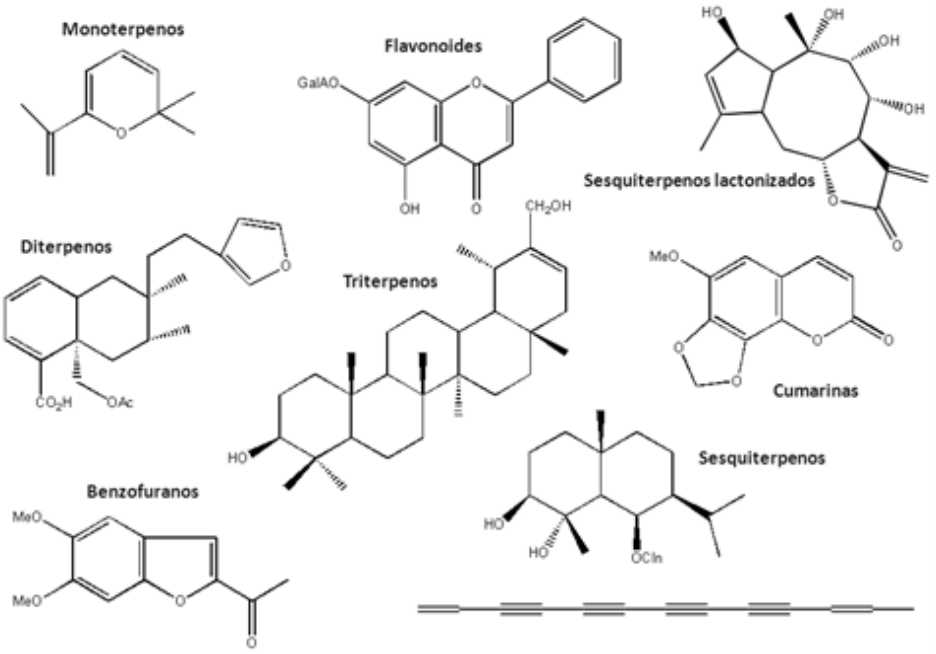


Figura 1.1.1. Exemplos de algumas classes de metabólitos secundários

Na Química de Produtos Naturais, os metabólitos secundários são importantes marcadores quimiotaxonômicos (Harborne 1988). Técnicas como espectroscopia de ressonância magnética nuclear 2D aliada a programas computacionais específicos podem diminuir o tempo de identificação estrutural de compostos quando as estruturas forem altamente complexas. A cada ano há uma explosão de quantidade de dados relativos a diversas estruturas de compostos orgânicos. Estes dados fornecem uma riqueza de informação disponível em bancos de dados químicos que são de interesse inestimável por elevar o conhecimento da composição química de plantas. Também são úteis na proposição de esqueletos das estruturas através de comparação com padrões de compostos já identificados. Tais bancos de dados podem ser utilizados para propósitos quimiotaxonômicos (Emerenciano *et al.* 1998a; Emerenciano *et al.* 1998b) .

Na família Asteraceae, uma das maiores entre as Angiospermas, são isoladas compostos pertencentes às classes mostradas na figura 1.1.1.

1.2. Sistemática da Família Asteraceae

A família Asteraceae é uma das maiores famílias de angiospermas no mundo. Cerca de 23.000 espécies dessa família já foram descritas botanicamente, e diversas revisões com relação a sua química e biologia foram publicadas (Heywood 1977; Bremer 1992; Hind & Beentje 1994). Esta família foi classificada por vários botânicos (Cassini 1816; Bentham 1873; Hoffman 1890; Carlquist 1876; Wagenitz 1876; Cronquist 1988; Bremer 1996).

O botânico francês Henry Cassini (Cassini 1816) foi o primeiro classificador da família Asteraceae, e através de seus estudos identificou numerosos gêneros e tribos que atualmente aindasão reconhecidos. Em 1816, Cassini publicou um diagrama mostrando as inter-relações de 19 tribos (figura 1.2.1). Em 1873, Bentham apresentou uma nova classificação (figura 1.2.2), onde a família é dividida em 13 tribos com algumas modificações em relação ao esquema apresentado por Cassini. Em 1890, Hoffman (Hoffman 1890) repete a classificação de Cassini, com um número pequeno de alterações.

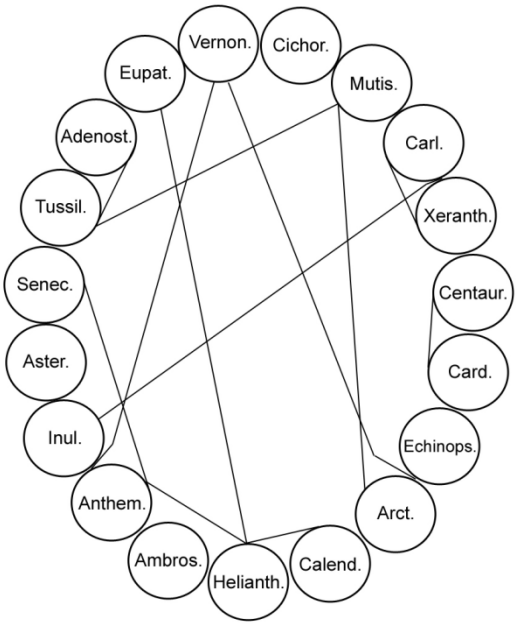


Figura 1.2.1. Diagrama, segundo Cassini (Cassini 1816), mostrando as inter-relações de 19 tribos de Asteraceae.

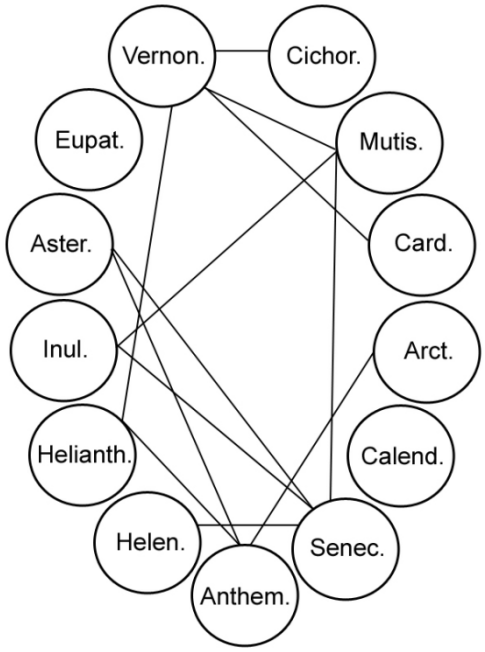


Figura 1.2.2. Diagrama, segundo Bentham (Benthan 1873), reduzindo o número de tribos de 19 (Cassini 1816) para 13, e suas inter-relações.

Em 1876, Carlquist (Carlquist 1876) dividiu a Asteraceae em 2 subfamílias em base de estudos morfológicos, Asteroideae e Cichorioideae (figura 1.2.3). No

mesmo ano, Wagenitz (Wagenitz 1876) também propôs uma divisão em 2 subfamílias que diferenciou da Carlquist, colocando a tribo Eupatorieae na subfamília Asteroideae ao invés na Cichorioideae (tabela 1.2.1). Essa visão bifilética da família foi o maior passo no entendimento das relações entre as tribos da Asteraceae.

Wagenitz (1976)	**Carlquist (1976)**
I.	Subfamily Cichorioideae
Vernonieae	Mutisieae
Liabeae	Vernonieae
Mutisieae	Cardueae
Cardueae	Arctoteae
Echinopeae	Cichorieae
Arctoteae	Eupatorieae
II.	
Eupatorieae	Subfamily Asteroideae
Heliantheae	Heliantheae
Helenieae	Astereae
Senecioneae	Inuleae
Calenduleae	Calenduleae
Astereae	Senecioneae
Inuleae	Anthemideae
Anthemideae	

Figura 1.2.3. As classificações de Carlquist (Carlquist 1876) e Wagenitz (Wagenitz 1876) baseados em caracteres morfológicos.

Em 1987, Bremer apresentou um cladograma da Asteraceae baseado em 81 caracteres, 10 dos quais químicos (Bremer 1987). Os caracteres restantes foram na sua maioria características morfológicas e de DNA. Este estudo é um exemplo de classificação incorporando caracteres químicos combinados com morfológicos e moleculares.

Oito anos depois, Bremer apresentou uma nova apresenteu uma nova classificação da Asteraceae baseado principalmente na morfologia, propondo 4 sub-famílias: Asteroideae (As)t, Cichorioideae (Cic), Carduoideae (Car) e Barnadesioideae (Bar) (Bremer 1996). Bremer colocou a tribo Mutiseae em um ramo

(clado) não bem posicionado entre Barnadesioideae e Carduoideae. Figura 1.2.4 mostra o diagrama de Bremer, o qual foi modificado do diagrama original apresentado um ano antes.

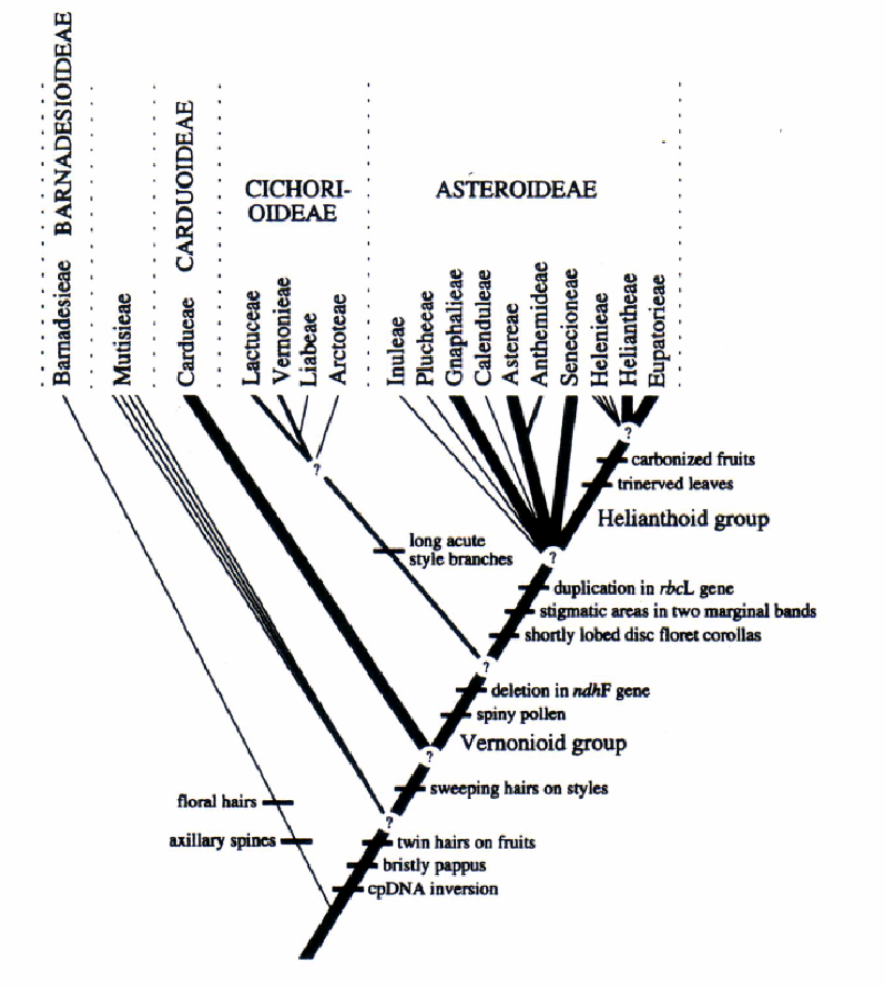


Figura 1.2.4. Diagrama Filogenético de tribos da Asteraceae de acordo com Bremer (Bremer 1996).

Kim e Jansen (Kin & Jansen 1995; Kin & Jansen 1996) apresentaram uma análise filogenética de 94 seqüência do gene ndhF do cloroplasto que representam todo os ramos principais de Asteraceae (Figura 1.2.5). O gene ndhF do cloroplasto provou ser mais filogeneticamente informativo para Asteraceae que os gene previamente usado, como rbcL.

Tabela 1.2.1 - Acrônimos de 3 letras das tribos apresentadas na figura 4 e utilizadas no estudo e o respectivo número de espécies conhecidas. A nomenclatura das tribos são as fornecidas pelo estudo de Bremer (Bremer 1996), exceto onde há um asterisco (*), os quais indicam a nomenclatura de Kim e Jansen (Kin & Jansen 1996).

Taxon	Three-Letter Acronyms	Nº of Species
Anthemideae	ANT	1737
Arctoteae	ARC	139
Astereae	AST	2846
Athroisma group*	ATH	26
Barnadesieae	BAR	92
Calenduleae	CAL	113
Cardueae	CAR	2513
Calyceraceae*	CAY	50
Eupatorieae	EUP	2396
Gnaphalieae	GNA	1728
Gochnatieae	GOC	68
Goodeniaceae*	GOD	380
Helenieae	HEL	835
Heliantheae	HLT	2449
Inuleae	INU	480
Lactuceae	LAC	2486
Liabeae	LIA	159
Mutisieae	MUT	321
Nassauvieae	NAS	318
Plucheeae	PLU	220
Senecioneae	SEN	3247
Tageteae*	TAG	216
Tarchonantheae*	TAR	2
Vernonieae	VER	1346

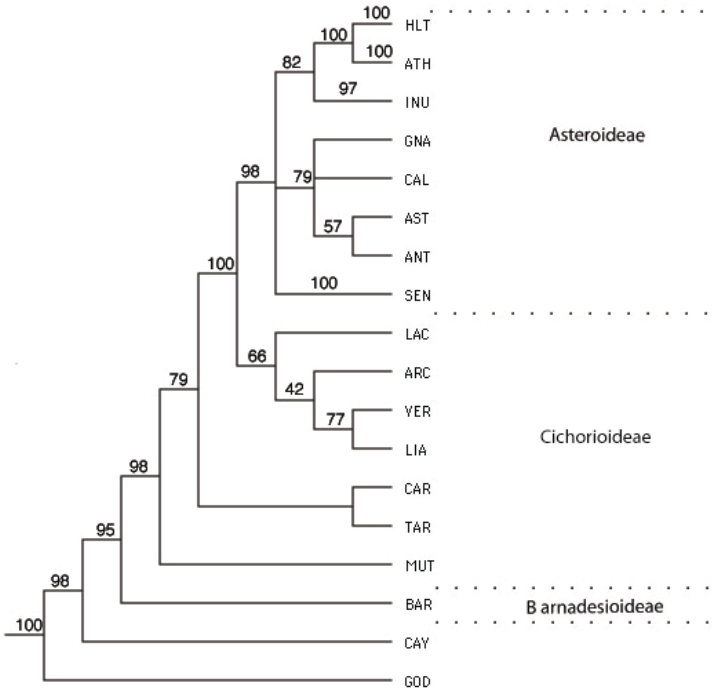


Figura 1.2.5. Árvore gerada por Kin e Jansen para as tribos da família Asteraceae utilizando dados moleculasres (Kin & Jansen 1995; Kin & Jansen 1996).

Funk e colaboradores produziram uma “supertree” (figura 1.2.6) mostrando a filogenia da família Asteraceae (Funk *et al.* 2005), utilizando os trabalhos mais recentes publicados e também dados ainda não publicados, mas que foram fornecidos por autores que contribuíram na época. Portanto o trabalho é o resultado de uma compilação de árvores utilizando diversos dados principalmente dados moleculares em conjunto com dados morfológicos.

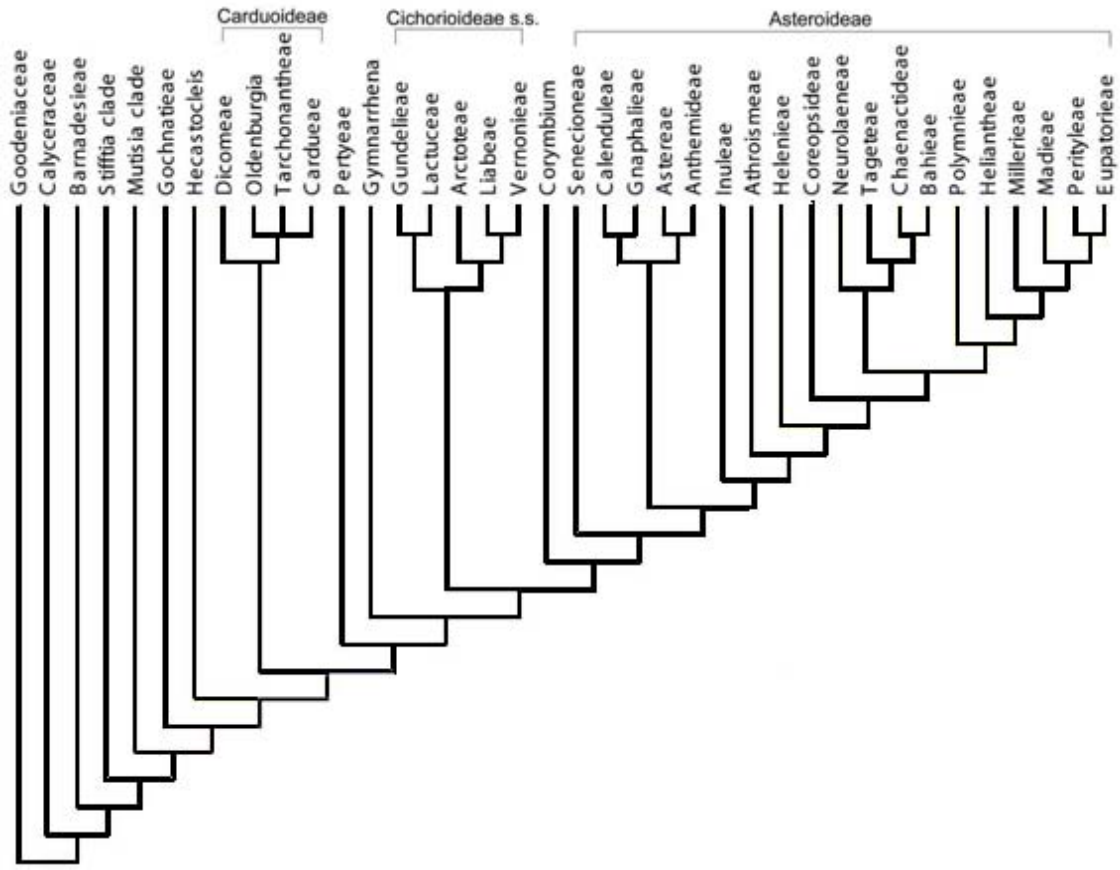


Figura 1.2.6. Cladograma da super-árvore de Funk (Funk *et al.* 2005) e colaboradores mostrando as relações da tribos da família Asteraceae.

Diversos autores atribuíram diferentes classificações para grupos onde as classificações não estão muito claras. Entre estas estão as tribos Mutisae e Heleniae. Para alguns autores, certos gêneros de Helenieae, pertencem a subtribos de Heliantheae.

Em 1977 Stuessy (Stuessy 1977) estabeleceu uma filogenética relação entre as subtribos de Heliantheae utilizando dados de morfologia e número de cromossomos (x). Entre as 15 subtribos reconhecidas pelo autor (figura 1.2.7), foram enfatizadas 3 linhas evolucionárias:

1. A primeira com a subtribo Verbesiniae no centro com o número de cromossomos baseados principalmente em x=15, x=16 e 17 e seus derivados aneuplóides.
2. A segunda com tendo como centro a subtribo Coreopsidinae com x = 12.

3. A terceira tendo a subtribo Galisonginae no centro com número de cromossomos variando de x=8 a x=18.

As subtribos Gaillardinae e Bahiinae, classificadas por Hoffmann (Hoffmann 1890) como pertencentes a tribo Heleniae, foram transferidas para Heliantheae por Stuessy (Stuessy 1977).

Considerando n=8 and n=9 o menor número de cromossomos encontrados em Heliantheae, e que a maioria de seus representantes tenha característica morfológicas herbáceas, Stuessy sugeriu que o complexo ancestral da tribo possui atributos das 3 linhas evolucionárias, especialmente das duas maiores. As características morfológicas pertencentes à linha evolucionária da Galinsoginae refletem a condição ancestral.

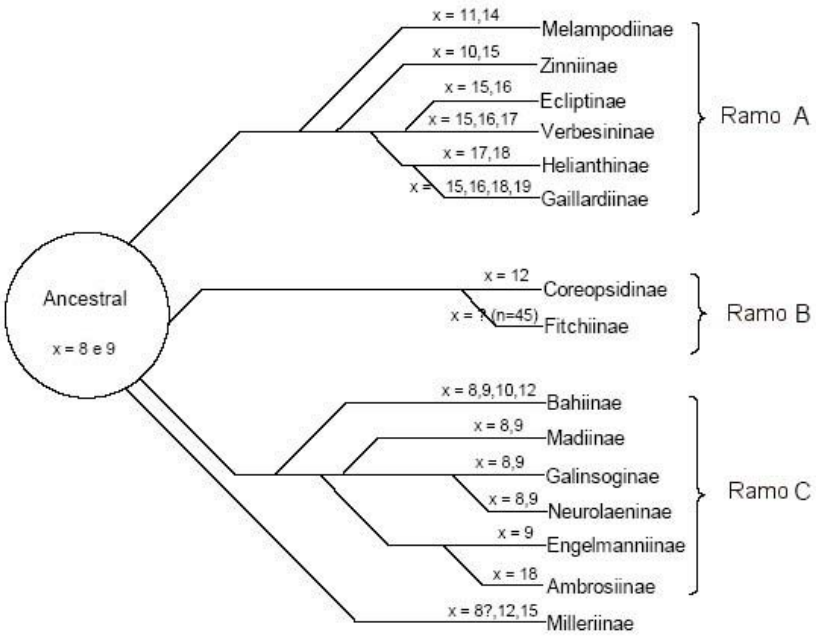


Figura 1.2.7. Similaridade entre as subtribos da tribo Heliantheae segundo Stuessy (Stuessy 1977).

1.2.1. Quimiossistemática da Família Asteraceae

Basicamente, nas espécies de Asteraceae, os principais metabólitos secundários isolados são monoterpenos, sesquiterpenos, sesquiterpenos lactonizados - SLs (Seaman *et al.* 1982), poliacetilenos (Zdero *et al.* 1990), flavonóides (Harborne *et al.* 1975; Bohm *et al.* 2001), benzofuranos e benzopiranos (Proksch *et al.* 1983), cumarinas (Murray *et al.* 1982), diterpenóides (Seaman *et al.* 1990) e triterpenóides (Macari *et al.* 1994). Atualmente todas as classes têm números de representantes (no mínimo algumas centenas) que são considerados satisfatórios para o estudo quimiotaxonômico.

A quimiossistemática apresenta limitações como as descritas a seguir:

1. Muitas espécies ainda não foram estudadas.
2. Nos estudos publicados na literatura, algumas vezes não são mencionadas as partes das plantas os quais os compostos são extraídos, ou partes diferentes de espécies diferentes são estudadas.
3. Existe uma grande diferença quantitativa como qualitativa entre as partes das plantas estudadas.
4. O estudo fitoquímico é dirigido muitas vezes para novos compostos ou compostos pouco usuais.
5. Existem resultados com falsos positivos ou negativos, de compostos nas espécies estudadas.

Dados químicos podem ser aplicados na comparação de árvores filogenéticas utilizando dados morfológicos e/ou macromoleculares, independente das limitações

de se usar metabólitos secundários, secundários. Esta metodologia é ainda de grande valia na para entender as diferenças das classificações utilizando os outros dois tipos de marcadores taxonômicos (Calabria *et al.* 2007).

1.2.2. Evolução Química de Sesquiterpenos Lactonizados em Asteraceae

A seminal contribuição de Gottlieb e colaboradores para a quimiotaxonomia resultou em diversos postulados sobre a evolução de metabólitos secundários nas plantas uma das quais sugerem que “A evolução de micromoléculas ocorre por oxidação. Os compostos mais oxidados caracterizam novas rotas químicas. Dentro de cada linha a evolução ocorre por desoxigenação”. “Os compostos relativamente altamente oxidados caracterizam novas linhagens químicas” (Gottlieb 1989). Uma correlação entre o número de espécies e a quantidade de oxigênio atmosférico desde a época geológica, em que ouve a diversificação deste táxon, foi proposta entre o crescimento do principal táxon da planta e sua variação na quantidade de oxigênio atmosférico.

Cronquist sugeriu que na Asteraceae a grande produção e variabilidade de metabólitos secundários é a maior causa do seu sucesso evolucionário (Cronquist 1988). Mais recentemente, foi sugerido que as rotas oxidativas nas plantas ocorrem paralelamente aos mecanismos de proteção contra a degradação oxidativa (Gottlieb & Kaplan 1993).

O oxigênio é essencial para a vida animal, vegetal e para processos de combustão. Os metabólitos secundários são evidentemente formados também por reações de óxido-redução do metabolismo de plantas (Gottlieb & Kaplan 1993). Em qualquer ciclo metabólico das plantas, as reações de redução-oxidação ocorrem e

podem envolver mecanismos químicos complexos. Um grande número de reações é catalisado eficientemente por enzimas, mas poucas dessas reações (óxido-redução e outras) utilizam luz para ocorrerem (Gottlieb & Kaplan 1993). Recentemente, os avanços da quimiotaxonomia de plantas levaram a desenvolver teorias e postulados que não foram extensivamente testados pelo uso de métodos matemáticos aplicados nos extensos bancos de dados de metabólitos secundários.

Os estudos realizados por Emerenciano e colaboradores (Emerenciano *et al.* 1986), com relação aos sesquiterpenos lactonizados em Asteraceae compararam as tribos com relação aos graus de especialização e oxidação.

Para se calcular o grau de oxidação para cada composto, o número de oxidação (NOX), foi calculado de acordo com as regras de Hendrickson (Hendrickson *et al.* 1970). Essas regras podem ser sumarizadas pela equação 1.2.2.1:

$$NOX = \sum n_i C - B \qquad \textbf{Equação 1.2.2.1}$$

Onde: n_i é o número de ligações entre os átomos de carbono e *B*.

Na equação 1.2.2.1, *B* pode ser átomo de H, C, ou um heteroátomo X, portanto as ligações resultantes são: C-H, C-C, e C-X respectivamente. O estado de oxidação de um átomo de carbono com um desses átomos é obtido pela adição dos seguintes valores:

1. -1 para ligações com átomo de hidrogênio;
2. 0 para ligações com átomo de carbono;
3. +1 para ligações com heteroátomos.

A somatória inclui todas as ligações entre C-B de um composto orgânico em consideração, fornecendo o total do estado de oxidação como um número de oxidação em relação aos seus átomos de carbono. Obviamente uma ligação dupla entre dois átomos de carbono tem um valor igual a zero. Por fim o grau de oxidação é calculado dividindo-se o número de oxidação pelo número de átomos de carbono presente na molécula (equação 1.2.2.2) (Gottlieb *et al.* 1996).

$$O = \frac{NOX}{n}$$ **Equação 1.2.2.2**

O grau de especialização E (equação 1.2.2.3), é calculado a partir do número de ligações formadas (**f**), quebradas (**q**), número de sistemas cíclicos formados que envolvem heteroátomos (**c**), número de carbonos adicionais com relação ao precursor e número de átomos de carbono da molécula (**n**) (Gottlieb *et al.* 1996).

$$E = \frac{q + f + c + u}{n}$$ **Equação 1.2.2.3**

Para cada tribo foi calculado a média dos graus de oxidação (EAo) e de especialização (EAe) com relação ao número de ocorrências dos sesquiterpenos lactonizados.

1.3. Os Sesquiterpenos e sua Biossíntese

Os sesquiterpenos são um grupo dos compostos terpênicos contendo 15 carbonos. É a subclasse de terpenos mais diversificada. Esta sua diversificação

estrutural aliada a sua atividade biológica resultaram em um grande interesse de pesquisa neste grupo de terpenos. (Cordell 1976).

Os sesquiterpenos, assim como todos os compostos classificados como terpenóides, são formados de unidades chamadas de isoprenos , , que são considerados como os terpenóide mais simples, possuindo 5 carbonos. OS compostos tepenóides, também chamados de isoprenóides são classificados em:

1. Hemiterpenos (C5)
2. Monoterpenos (C10)
3. Sesquiterpenos (C15)
4. Diterpenos (C20)
5. Triterpenos (C30)
6. Tetraterpenos (C40)
7. Politerpenos

A biossíntese destes compostos consiste na condensação do precursor básico: o pirofosfato de isopentenila (IPP). São conhecidas duas rotas para a biossíntese do pirofosfato de isopentenila. A primeira rota é conhecida desde o fim da década de 1950 (Liechtenthaler 1999), e envolve a biossíntese do ácido mevalônico (MVA), a qual requer 3 moléculas de acel-CoA para produzir o composto 3-hidróxi-3-metilglutaril-SCoA. A condensação de Claisen entre duas unidades de acetil-CoA pode ocorrer de forma linear, porém o segundo ataque ocorre na carbonila cetônica. As próximas etapas são de redução, que são de grande importância nos animais pois agem como limitadores da biossíntese de colesterol (Goldstein & Brown 1990). O mevalonato resultante é fosforilado, formando o 5-

pirofosfomevalonato. A formação do IPP é catalisada pela pirofosfomevalonato decarboxilase (descarboxilação), e seguida de desidratação (figura 1.3.1).

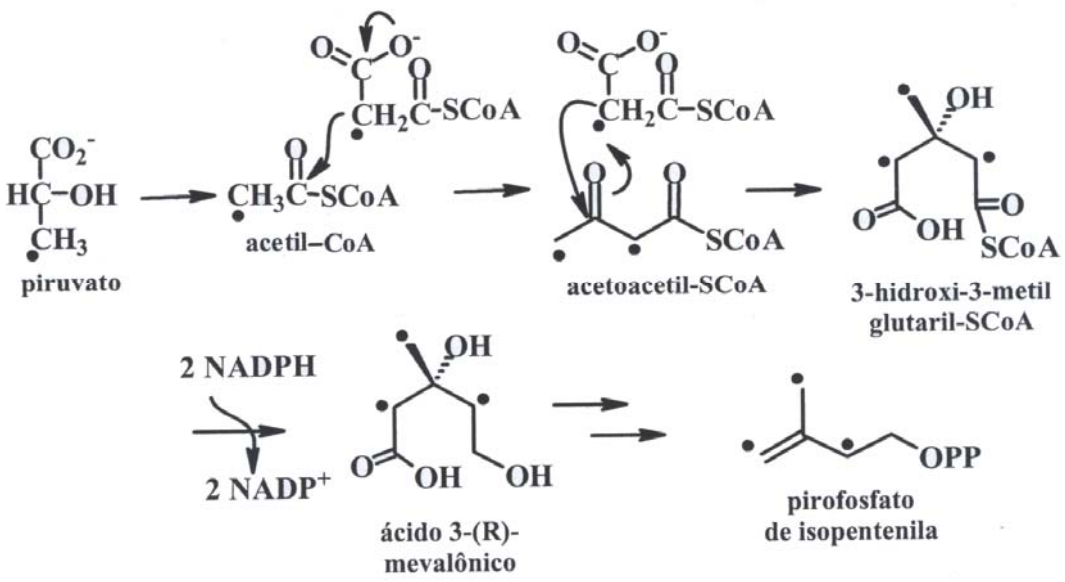


Figura 1.3.1. Biossíntese do IPP: rota do ácido mevalônico.

Uma rota que não dependia da via do ácido mevalônico para a síntese do IPP, via do 1-desoxi-D-xilose-5-fosfato (DOXP), foi descoberta em uma eubactéria (Rohmer 1993)(Lichtentaler 1999), e verificou-se que esta rota está presente em algas verdes (chlorophyta), plantas superiores e outros grupos de algas (Lichtentaler 1999). Nesta rota representada na figura 1.3.2, o piruvato após reagir com pirofosfato de tiamina-enzima (TPP-E), reage com gliceraldeido-3-fosfato (GA-3-P) formando 1-desoxi-D-xilose-5-fosfato, que a seguir produz o pirofosfato de isopentenila (IPP).

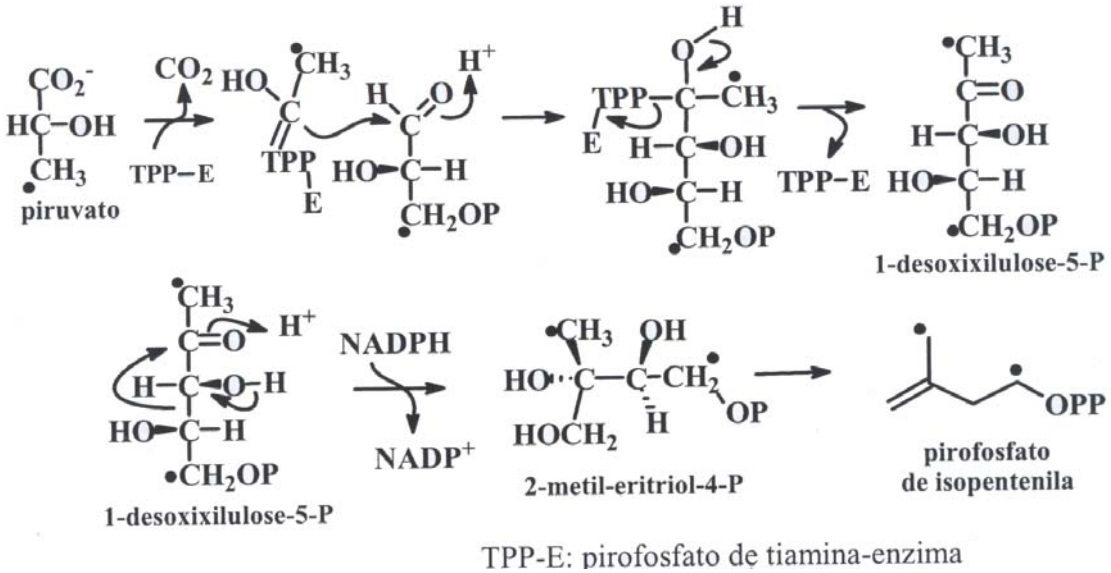


Figura 1.3.2. Biossíntese do IPP: rota do 1-desoxi-D-xilose-5-fosfato.

Após a formação do IPP, este pode ser convertido através da ação de uma isomerase em pirofosfato de 3,3-dimetilalila (DMPA). O isopreno, o mais simples terpenóide, é sintetizado diretamente do DMPA, sendo responsável pela síntese dos hemiterpenos (Lange & Croteau 1999).

Uma reação nucleofílica entre o IPP e o DMPA (cauda-cabeça), mediadas por enzimas denominadas preniltrasferases, gera o pirofosfato de geranila (figura 1.3.3), unidade precursora dos monoterpenos. Por sua vez a reação do pirofosfato de geranila com uma unidade de pirofosfato de isopentenila gera o pirofosfato de farnesila (**2**). Este último é precursor imediato dos sesquiterpenos (C15).

A reação do pirofosfato de farnesila com o fosfato de isopentenila (figura 1.3.3) forma o pirofosfato de pirofosfato de geranilgeranila, precursor dos diterpenos (C20). Tanto o pirofosfato de farnesila como o pirofosfato de geranilgeranila podem reagir (cabeça-cabeça), formando respectivamente o esqualeno, precursor dos triterpenos (C30), e o fitoeno, precursor dos tetraterpenos (C40)

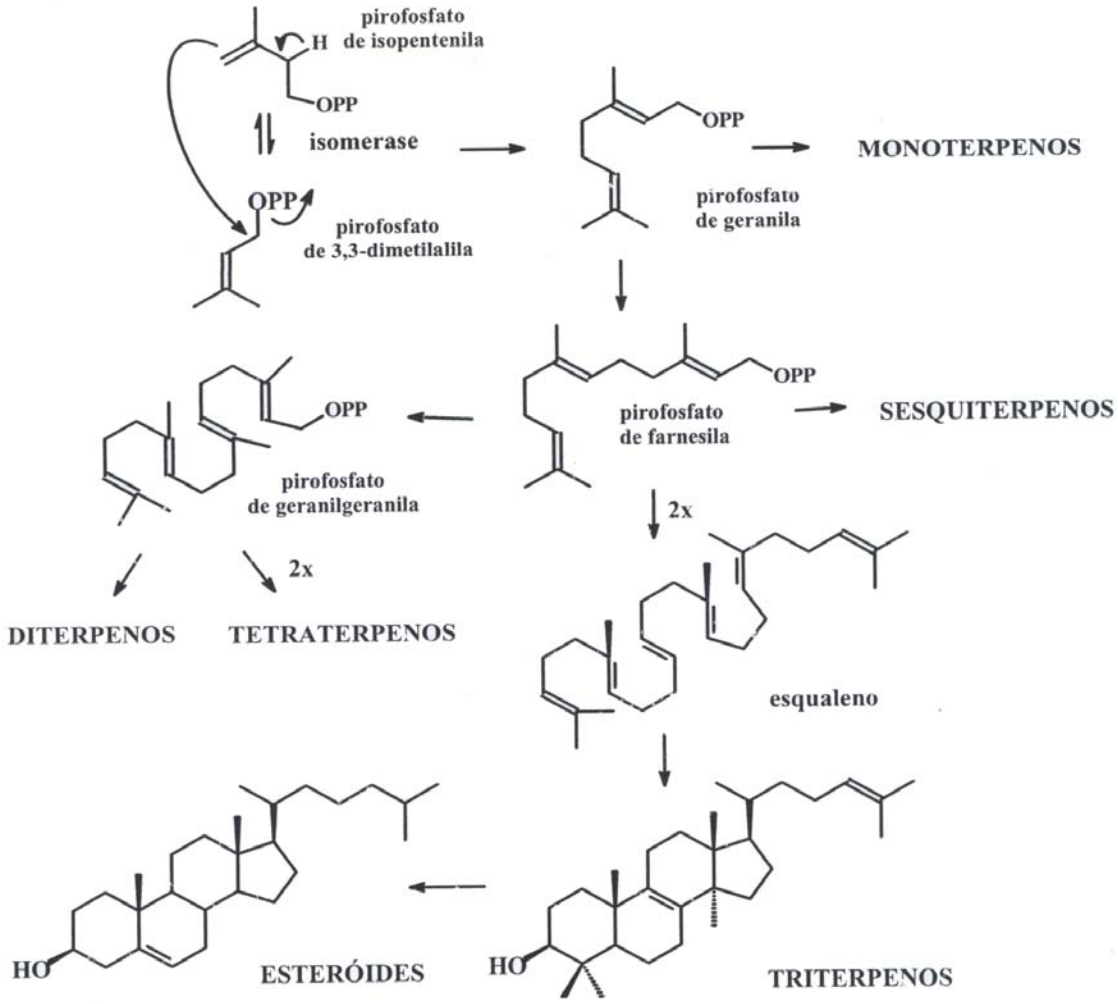


Figura 1.3.3. Esquema da rota biossintética dos terpenos a partir do pirofosfato de isopentenila e do pirofosfato de 3,3-dimetila.

1.3.1. Biossíntese de Sesquiterpenos Lactonizados

Os sesquiterpenos lactonizados são de grande interesse na pesquisa de produtos naturais, pois são usados com sucesso nos estudos quimiotaxonômicos (Seaman 1982), são responsáveis por diversas atividades biológicas derivados de sua estrutura (Picman 1986).

O germacrano é o intermediário dos sesquiterpenos lactonizados, após a formação deste intermediário, um carbono metílico da cadeia isopropílica é oxidado a um grupo carboxílico, enquanto a ligação dupla é introduzida entre o carbono 11 e o carbono 13 (carbono metílico da cadeia lateral). A incorporação do grupo hidroxila

no carbono 6 ou 8 no anel de 10 carbonos permite que a ligação éster entre este grupo hidroxila e o grupo funcional carboxila da cadeia lateral. Em função destes grupos estarem na mesma molécula, essa esterificação intramolecular é denominada de lactonização (figura 1.3.1.1) (Seaman 1982).

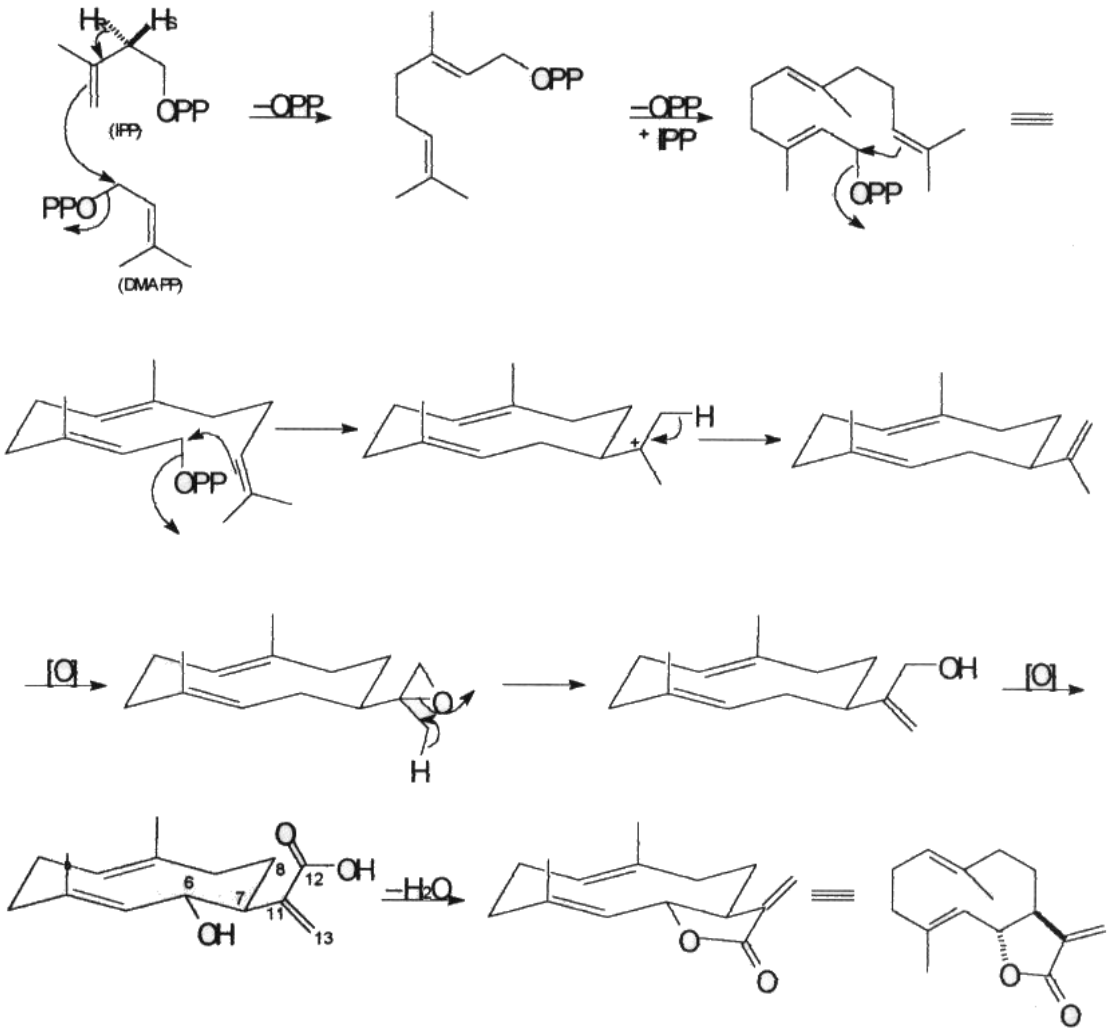


Figura 1.3.1.1. Biogênese de sesquiterpenos lactonizados a partir do isopreno.

1.4. A Atividade Citotóxica dos Sesquiterpenos Lactonizados

A extensa variedade de estruturas químicas descobertas ao longo dos anos é combinada com uma diversidade de atividades biológicas e farmacológicas. (Picman

1986) Os sesquiterpenos lactonizados são descritos como os princípios ativos de várias plantas medicinais usadas na medicina tradicional e são conhecidos por atuarem no sistema nervoso central e no sistema cardiovascular; possuírem atividades antimicrobianas, antitumorais, inflamatórias, além de potencial alergênico.

Embora os sesquiterpenos lactonizados sejam compostos terpênicos (três unidades isoprênicas ligadas covalentemente), são chamados sesquiterpenos, característicos de Asteraceae (Compositae) ques também podem ser encontrados em outras famílias de Angiospermas (Picman 1986). Em Asteraceae, diferenças nos tipos de esqueleto e quantidades de sesquiterpenos lactonizados encontrados em diferentes gêneros e espécies têm sido utilizados nos estudos taxonômicos. (Yoshioka *et al.* 1973; Heywood *et al.* 1977; Kelsey & Shafizadeh 1979)

A classificação dos sesquiterpenos lactonizados de acordo com o seu esqueleto carbônico divide a maioria destes em quatro grupos principais: germacranolídeos (com um anel de 10 membros); eudesmanolídeos (compostos 6/6-bicíclicos); guaianolídeos e pseudoguaianolídeos (ambos compostos 5/7-bicíclicos) (figura 1.4.1). (Yoshioka *et al.* 1973) Entretanto, os sesquiterpenos lactonizados exibem uma variedade de outros arranjos de esqueleto (Seaman 1982).

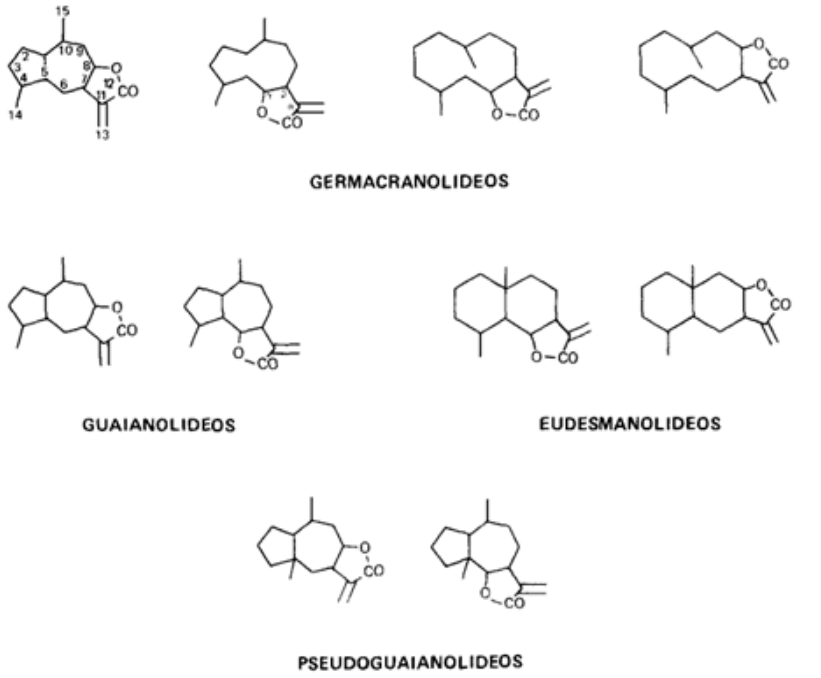


Figura 1.4.1. Esqueletos carbocíclicos das principais classes de sesquiterpenos lactonizados.

Durante a busca por compostos com atividade citotóxica presentes em plantas, muitos sesquiterpenos lactonizados ativos contra vários tipos de organismos e sistemas têm sido descobertos. Eles formam um dos maiores grupos de substâncias citotóxicas de origem vegetal. A maioria destes sesquiterpenos lactonizados ativos são encontrados em espécies de Asteraceae, embora alguns originem-se de Magnoliaceae, Apiaceae e até mesmo de fungos (Picman 1986).

As atividades são mediadas quimicamente por estruturas carbonílicas α,β-insaturadas, como uma α-metileno-γ-lactona, uma ciclopentanona α,β-insaturada ou um éster conjugado. Estes grupos reagem com nucleófilos, especialmente grupos sulfidrila de cisteínas, por uma adição de Michael (figura 1.4.2). (Kupchan *et al.* 1970a; Schmidt 1997). Grupos tióis expostos, como resíduos de cisteína em proteínas, parecem ser os primeiros alvos dos sesquiterpenos lactonizados, levando à inibição de uma variedade de funções celulares a qual direciona as células à

apoptose. (Schmidt 1999; Dirsch, 2001) As diferenças na atividade entre as várias estruturas de sesquiterpenos lactonizados pode ser explicada por diferentes números de elementos estruturais alquilantes. (Kupchan *et al.* 1971; Heilmann *et al.* 2001) Entretanto, outros fatores como lipofilicidade, geometria molecular, e o ambiente químico ou o alvo sulfidrila podem também influenciar a atividade dos sesquiterpenos lactonizados (Kupchan *et al.* 1970b; Heilmann *et al.* 2001).

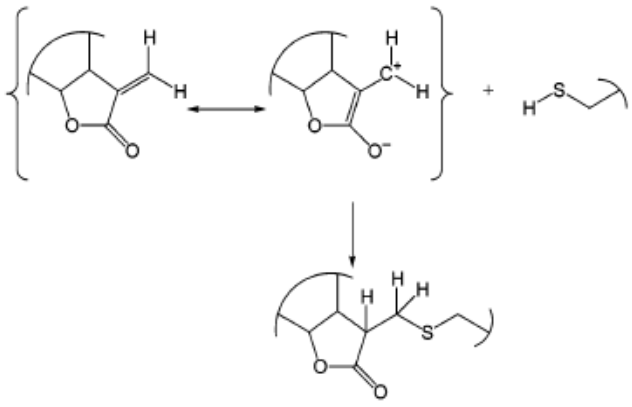


Figura 1.4.2. Reação entre lactona com grupo sulfidrila de cisteína, por uma adição de Michael.

1.5 Armazenamento de dados - O sistema especialista SISTEMATX

Um dos objetivos desta Tese foi o desenvolvimento do SISTEMAT X, cujo significado é SISTEMAT eXtended, é um novo sistema que vem sendo desenvolvido pelo nosso grupo, em conjunto com o Dr. Ricardo Stefani do Instituto de Ciências Exatas e Biológicas, Universidade Federal de Mato Grosso .Preto. Esforços foram feitos no sentido de dotar o SISTEMAT (Gastmans *et al.* 1990a;Gastmans *et al.* 1990b) de uma nova interface, um gerador de estruturas e outras novas funções, incluindo a capacidade de gerenciar bancos maiores e novos tipos de exportação de

dados (tabela 1.5.1). O objetivo seguinte é disponibilizar o trabalho para toda a comunidade científica, através de uma interface WEB. A diferença fundamental entre o SISTEMAT padrão e o SISTEMAT X é a forma de gerenciamento dos dados, pois o segundo utiliza um banco padrão SQL e o primeiro um banco proprietário. Muitos dos problemas encontrados ao executar o SISTEMAT X se mostraram relacionados com a configuração do SQL.

Tabela 1.5.1. Comparação das características do SISTEMAT e SISTEMATX

Característica	**SISTEMAT padrão**	**SISTEMAT X**
Associação RMN e biogenética	Obrigatória	Opcional
Banco de Dados	proprietário	servidor SQL (MySQL 5.0)
Editor de moléculas	Embutido	Embutido
Linguagem de programação	FORTRAN PASCAL	JAVA, C/C++
Máximo de átomos por molécula	60	999
Máximo de usuários simultâneos	1	10 (Entrada) 50 (consulta)
Sistema Operacional	MS-DOS / Windows	Windows / Linux / Mac OS X
Tamanho máximo do banco	50 MB	4 TB
Tipo de interface	Console	Gráfica

O SISTEMATX apresenta a possibilidade de se adicionar diversas propriedades como ponto de fusão, tempo de retenção de cromatografia, dados de espectroscopia de massa, de ressonância magnética, dados de atividade biológica e ocorrência botânica (figura 1.5.1).

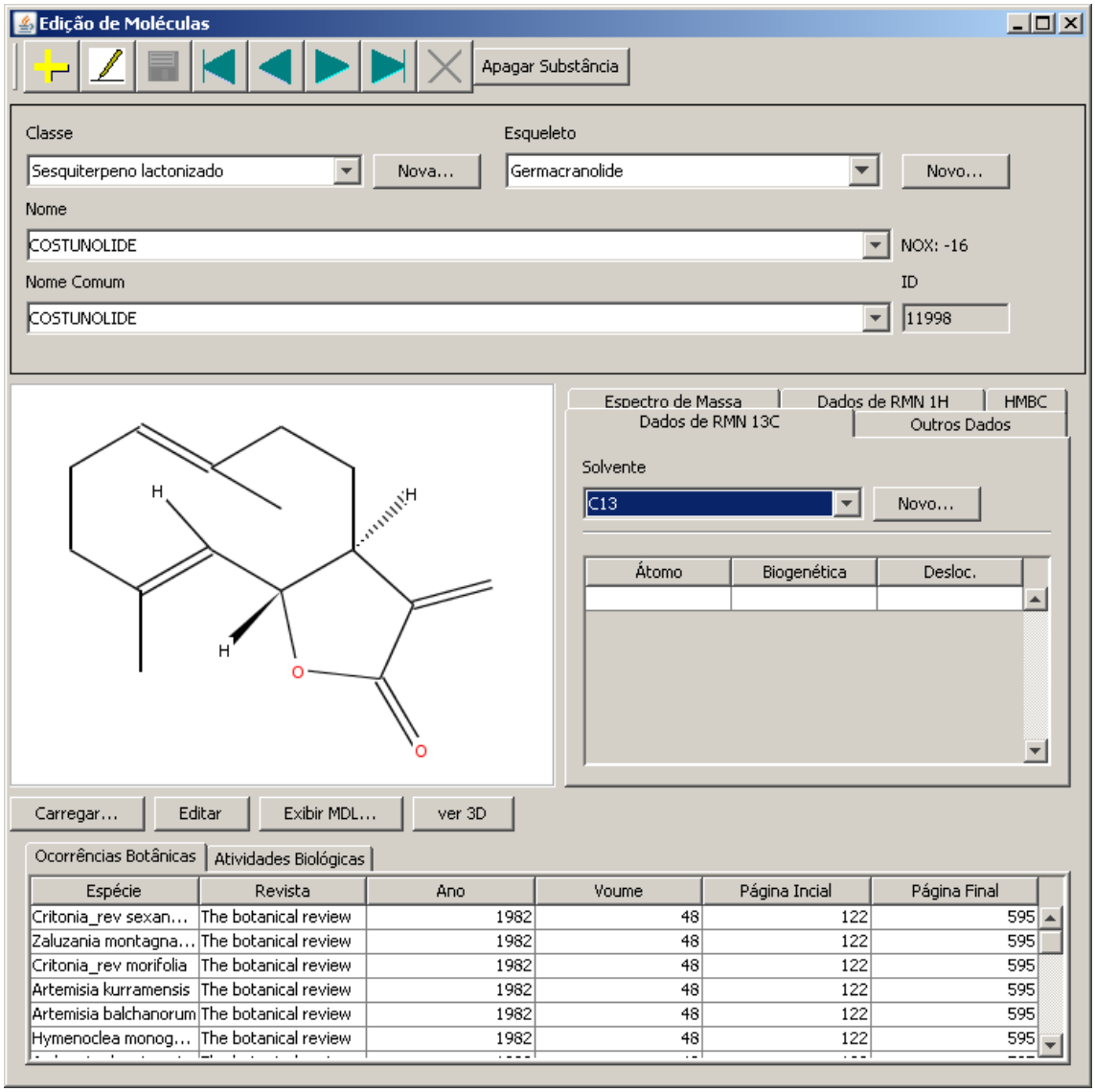


Figura 1.5.1. Tela de Edição de Moléculas do SISTEMATX

No SISTEMATX os compostos desenhados são em 2 dimensões, com as devidas informações estereoquímicas, em seguida, estruturas em 3D podem ser geradas automaticamente utilizando um software como o CORINA (Sadowski & Gasteiger 1993) ou CONCORD e salvas em um arquivo .mol ou .hin. Possibilitando que estas possam ser utilizadas como dados de entrada para gerar diversos descritores moleculares.

1.6. Descritores Moleculares

As propriedades físico-químicas como também a atividade biológica de compostos orgânicos dependem de suas estruturas moleculares. Com a finalidade de se obter relações entre as estruturas químicas e a atividades biológicas utilizando abordagens computacionais, é necessário encontrar representações apropriadas da estrutura molecular dos compostos (Hansch *et al.* 1990).

Um descritor molecular pode ser considerado como sendo o resultado obtido de procedimento lógico e matemático, aplicado às informações químicas codificadas através da representação de uma molécula (Consonni *et al.* 2002a). Este procedimento transforma estas informações em um valor numérico associado a uma determinada propriedade molecular importante para posterior análise, correlacionado com uma propriedade molecular, como por exemplo, ponto de fusão, ou a uma atividade biológica. Porém, estas correlações são raramente obtidas, pois os sistemas estudados são freqüentemente complexos e uma relação entre uma propriedade molecular com os descritores moleculares não é, em geral, claramente entendida e, conseqüentemente definido ambiguamente. O mais importante para ser considerado

e limitante é o fato dos sistemas, em muitos casos, não serem completamente conhecidos (Kubinyi 1993ª; Kubinyi 1993b).

Os métodos que podem ser aplicados para se obter relações entre as estruturas moleculares dos ligantes e as afinidades relativas destes com o receptor dependem se a estrutura do receptor é conhecida. Se a estrutura do receptor não for conhecida, as variações da atividade biológica, em uma determinada série de moléculas, podem ser relacionadas com as relativas diferenças dos descritores

moleculares. Alguns destes descritores necessitam de um alinhamento estrutural (superposição) das moléculas, e assim, um descritor pode diferenciar uma molécula de outra (Klebe *et al.* 1994).

O estudo das propriedades estéricas envolvidas nas interações entre os ligantes e os receptores biológicos é freqüentemente decisivo no entendimento das características estruturais dos ligantes para a atividade biológica. Os efeitos estéricos ocorrem de diversas maneiras. Sugere-se na literatura (Hansch *et al.* 1990) que este pode aparecer como resultado da repulsão entre os átomos não ligados. Tais repulsões podem determinar não apenas a influência intramolecular estérica dos substituintes nas propriedades moleculares, mas também a influência intermolecular específica da afinidade do ligante pelo o receptor. E, em particular, nos métodos de *QSAR* (Relação Quantitativa entre Estrutura Química e Atividade Biológica) clássico, consideram-se ainda insatisfatórios (Hansch *et al.* 1990), os métodos disponíveis para quantificar as características topológicas de um composto e a comparação com os outros descritores de propriedades físico-químicas. Apenas propriedades estéricas de substituintes ou, de certas subestruturas, podem ser adequadamente descritas, fornecendo informações precisas, necessária para análises precisas dos efeitos estéricos das interações dos ligantes com o sítio ativo dos receptores (Hansch *et al.* 1990).

Neste contexto, encontram-se na literatura (Carbo *et al.* 1980; Hodgkin & Richards 1987; Reynolds *et al.* 1992; Good 1992; Serilevy *et al.* 1994) vários trabalhos envolvendo cálculos de similaridade com o objetivo de serem utilizados como um método de gerar parâmetros para as análises de QSAR. Em geral, os cálculos de similaridade comparam os compostos da série estudada considerando algumas propriedades, como por exemplo, densidade eletrostática, potencial

eletrostático, e, formato (Serilevy *et al.* 1994; Good *et al.* 1993). Considerando-se as relações observadas entre similaridades moleculares e as correspondentes variações nos valores de atividade biológica, diferentes expressões de similaridade química têm sido investigadas (Kubinyi *et al.* 1998).

Adicionalmente, decorrente do enorme desenvolvimento dos sistemas de modelagem molecular, encontram-se na literatura (Sadowski & Gasteiger 1993; Sadowski *et al.* 1994) muitos bancos de dados, baseados em cristalografias de raio-X, e estes estão disponíveis para fornecer dados de diferentes tipos de estruturas em 3 dimensões. O desenvolvimento computacional possibilitou realizar mais rapidamente cálculos que geram as estruturas em 3 dimensões (Sadowski & Gasteiger 1993). Conseqüentemente, encontram-se na literatura inúmeros descritores moleculares, como por exemplo, índices topológicos como também índices que codificam as informações geométricas em 3D da molécula (Consonni *et al.* 2002a).

Também se observa na literatura, uma procura crescente (Todeschini & Gramatica 1997a; Consonni *et al.* 2002a) de descritores moleculares que sejam validados e de métodos de seleção (Baroni *et al.* 1993; Kubinyi 1994; Golbraikh & Tropsha 2002; Gasteiger *et al.* 2003) visando representar significativamente as informações relacionadas às propriedades físico-químicas e/ou à atividade biológica contidas nas séries de compostos estudadas.

Entre os programas existentes para cálculos de descritores moleculares (Xtsar, AMPAC, Molconnz, CODESSA). Todos estes descritores são facilmente e, rapidamente calculados, apropriados para análise de QSAR e de similaridade/diversidade de extensos bancos de dados (Consonni *et al.* 2002a). A grande maioria dos descritores no programa DRAGON 5.4 (Talete, 2006) usado

nesta tese (Topológicos, Geométricos, BCUT, Autocorrelação 2D, WHIM, GETAWAY, RDF, 3D-MoRSE entre outros} são holísticos (Guha *et al.* 2004)e utilizados para classificar séries de dados em termos de características globais.

1.6.1. Descritores GETAWAY

É uma sigla utilizada para *Geometric Topology and Atom Weights Assembly.* Estes descritores são calculados a partir de uma matriz de influência molecular *MIM* (H) (equação 1.6.1.1), que é calculada utilizando a matriz de coordenadas dos átomos (M) em relação ao centro da molécula com geometria em 3 dimensões, como definida no item 1.6.1.2. Na matriz de influência molecular (H), as linhas representam os átomos (inclusive o Hidrogênio) e as colunas as coordenadas x, y e z de cada átomo de uma estrutura molecular em 3 dimensões. A matriz de influência molecular é simétrica *A* X *A*, onde *A* representa o número de átomos.

Os elementos diagonais (hii) da matriz de influência molecular, denominados leverages, representam cada átomo na determinação da forma molecular. O valor da somatória dos elementos diagonais pode ser 1, 2 ou 3, para moléculas lineares, planares e em 3 dimensões, respectivamente. Os átomos presentes na periferia da molécula, grandes átomos e moléculas esféricas apresentam maiores valores de leverage que os localizados no centro. Átomos maiores também apresentam maiores valores de levarage que átomos menores. Moléculas esféricas apresentam átomos com menores valores de levarage que moléculas lineares. Para série de moléculas com aproximadamente a mesma conformação, o maior valor de leverage decresce com o aumento do número de átomos na molécula. Os valores de leverage

dependem da geometria da molécula e são sensíveis à mudança conformacional e ao comprimento das ligações e, portanto ao tipo de ligação.

Os elementos (hij) fora da diagonal representam os graus de acessibilidade do átomo j para interagir com o átomo i, e valor da somatória destes elementos é sempre 0. Valores negativos destes elementos significam que os átomos ocupam posições opostas em relação ao centro da molécula.

Os descritores calculados a partir da matriz de influência molecular (H), denominados descritores H-GETAWAY, podem ser ponderados pelas propriedades atômicas como massa atômica, polarizabilidade, volume de van der Waals e eletronegatividade, respectivamente.

$$H = M \bullet (M^T \bullet M)^{-1} \bullet M^T$$ **Equação 1.6.1.1**

Os descritores $H_k(w)$ (equação 1.6.1.2) estão entre os descritores obtidos através da matriz de influência molecular (H). Nesta equação *k* é a distância topológica fixada, w_i e w_j são as propriedades atômicas respectivamente dos átomos *i* e *j*, d_{ij} é a distância topológica entre os átomos *i* e *j*, h_{ij} são os elementos fora da diagonal da matriz de influência molecular e representam o grau de acessibilidade entre os átomos *i* e *j*. $\delta(k;d_{ij};h_{ij})$ é a função delta de Dirac definida na equação 1.6.1.3.

$$H_k(w) = \sum_{i=1}^{A-1} \sum_{j>i} h_{ij} w_i w_j \delta(k; d_{ij}; h_{ij})$$ **Equação 1.6.1.2**

$$\delta(k; d_{ij}; h_{ij}) = \begin{cases} 1 \text{ se } \mathrm{d_{ij}} = \mathrm{k} \text{ e } \mathrm{h_{ij}} > 0 \\ 0 \text{ se } \mathrm{d_{ij}} = \mathrm{k} \text{ ou } \mathrm{h_{ij}} \leq 0 \end{cases}$$ **Equação 1.6.1.3**

Os descritores $H_k(w)$ são descritores de autocorrelação, onde são considerados apenas os valores das propriedades dos átomos que estejam numa distância topológica igual a determinada (k) e apresentem valores de acessibilidade positivos (h_{ij}), pois este valor positivo significa que há uma chance de interagir entre estes átomos. Como todos os descritores de autocorrelação, os descritores $H_k(w)$ são utilizados para verificar similaridade/dissimilaridade numa série de compostos (Consonni *et al.* 2002a; Consonni *et al.* 2002b).

A partir da matriz de influência molecular (H), criou-se uma nova matriz R denominada matriz de influência/distância. A matriz R (equação 1.6.1.4) utiliza os valores de leverages hii, hjj (elementos diagonais da matriz de influência molecular - H) de dois átomos i e j quaisquer da molécula e a distância geométrica entre estes rij. Os elementos diagonais da matriz R apresentam valor 0 (zero) e aqueles que não estão na diagonal são resultantes da média geométrica dos elementos diagonais da matriz H com a distância geométrica entre os dois átomos (Consonni *et al.* 2002a).

$$[R]_{ij} = \left[\frac{\sqrt{h_{ii} \bullet h_{jj}}}{r_{ij}} \right]_{ij} \quad i \neq j \qquad \textbf{Equação 1.6.1.4}$$

Os maiores valores dos elementos da matriz R derivam dos átomos mais externos (mais altos *levarages*) e simultaneamente próximos um do outro no espaço molécula (uma pequena distância interatômica).

A somatória das linhas da matrix de influência/distância codifica alguma informação útil que poderia ser relacionada à presença de substituintes ou de fragmentos na molécula. Os autores (Consonni *et al.* 2002a) observaram que valores altos das somatórias das linhas correspondem a átomos terminais que estão

localizados a outros átomos terminais como aqueles presentes nos substituintes de uma molécula.

Os descritores calculados a partir da matriz de influência/distância (R), denominados descritores R-GETAWAY, podem ser ponderados pelas propriedades atômicas como massa atômica, polarizabilidade, volume de van der Waals e eletronegatividade.

Os descritores $R_k(w)$ (equação 1.6.1.5) estão entre os descritores obtidos através da matriz de influência/distância (R). Nesta equação k é a distância topológica fixada, w_i e w_j são as propriedades atômicas respectivamente dos átomos i e j, d_{ij} é a distância topológica entre os átomos i e j, h_{ii} e h_{jj} elementos da diagonal da matriz de influência molecular, representam a influência do átomo na forma da molécula, r_{ij} distância geométrica entre os átomos i e j, e $\delta(k;d_{ij})$ é a função delta de Dirac definida na equação 1.6.1.6.

$$R_k(w) = \sum_{i=1}^{A-1} \sum_{j>i} \frac{\sqrt{h_{ii} h_{jj}}}{r_{ij}} w_i w_j \delta(k; d_{ij}) \qquad \textbf{Equação 1.6.1.5}$$

$$\delta(k, dij) = \begin{Bmatrix} 1 \text{ se dij} = \text{k} \\ 0 \text{ se dij} \neq k \end{Bmatrix} \qquad \textbf{Equação 1.6.1.6}$$

Os descritores $R_k(w)$ são descritores de autocorrelação, onde é considerado apenas os valores das propriedades dos átomos que estejam numa distância topológica igual a determinada (k). Como todos os descritores de autocorrelação, os descritores $R_k(w)$ são utilizados para verificar similaridade/dissimilaridade numa série de compostos (Consonni *et al.* 2002a; Consonni *et al.* 2002b).

A classe GETAWAY apresenta um total de 197 descritores (Consonni *et al.* 2002b; Consonni *et al.* 2002a).

1.6.2. Descritores WHIM

Esta sigla é utilizada para *Weighted Holistic Invariant Molecular*. São descritores baseados na análise de componentes principais (PCA) (Wold *et al.* 1987) aplicadas a uma matriz de coordenadas dos átomos de uma molécula em relação ao seu centro com geometria em 3 dimensões (matriz molecular). Nesta matriz as linhas representam os átomos, portanto uma molécula (com **n** átomos) gera uma matriz com n linhas e três colunas representando as coordenadas x, y, z. Além da matriz molecular, é definida uma matriz diagonal n x n, onde os elementos da diagonal principal contêm os valores de uma propriedade atômica (sem nenhuma propriedade – valores unitários, massa atômica, volume de van der Walls, eletronegatividade, polarizabilidade, ou estado eletrotopológico (Kier *et al.* 1991).

A matriz de covariância ponderada (equação 1.6.2.1) (3X3 - invariância com relação à translação e rotação) é obtida através dos dados das duas matrizes (matriz molecular e a matriz com os valores de propriedade atômica), semelhante ao cálculo do momento de dipolo. Na equação, n é o número de átomos, wi é a propriedade atômica do átomo i, qij e qik são respectivamente os valores das coordenadas j (j = 1, 2 e 3) e k do átomo i, $\overline{q}_j$ e $\overline{q}_k$ são respectivamente os valores das médias dos valores da coordenada j e k.

$$s_{jk} = \frac{\sum_{i=1}^{n} w_i (q_{ij} - \overline{q}_j)(q_{ik} - \overline{q}_k)}{\sum_{i=1}^{n} w_i}$$

Equação 1.6.2.1

A análise de componentes principais (PCA) é executada sobre a matriz de covariância, obtendo 3 autovalores (λ_1, λ_2 e λ_3) e a matriz de autovetores. As coordenadas dos átomos são projetadas em cada componente principal t_m (m=1,2 e 3), gerando uma nova matriz de coordenadas (matriz T - invariância com relação à translação e rotação). Finalmente os descritores são calculados a partir dos dados desta matriz (Belvisi *et al.* 1994).

Os descritores WHIM são construídos de forma que tentem capturar as informações relevantes em 3 dimensões com relação, respectivamente ao tamanho, forma, simetria e distribuição dos átomos numa molécula independente da referência de coordenadas. Portanto, a abordagem WHIM pode ser definida como uma procura generalizada dos eixos principais com respeito a uma propriedade molecular definida.

Os descritores WHIM são divididos em dois tipos de descritores: direcionais e não direcionais.

Os descritores direcionais são divididos em 4 tipos relacionados, respectivamente ao tamanho, ao formato, à simetria da molécula e à distribuição dos átomos (acessibilidade entre os mesmos).

Os descritores relacionados ao tamanho da molécula são definidos diretamente pelos autovalores λ_1, λ_2 e λ_3. Os descritores relacionados ao formato da molécula são obtidos pela equação 1.6.2.2, onde ϑ_m (m = 1, 2 e 3) são os autovalores proporcionais calculados a partir dos valores dos autovalores (λ_1, λ_2 e λ_3). Como $\vartheta_1 + \vartheta_2 + \vartheta_3 = 1$, Só dois descritores são independentes.

$$\vartheta_m = \frac{\lambda_m}{\sum_m \lambda_m} \qquad \textbf{Equação 1.6.2.2}$$

Os descritores relacionados à simetria (γ_1, γ_2 e γ_3) são obtidos através das equações 1.6.2.3 e 1.6.2.4. Nestas, n_s é a soma de todos os grupos de átomos que apresentem os mesmos autovalores, com sinais opostos, presentes no mesmo componente *m*, na é o número de átomos os quais seus apresentem autovalores opostos simétricos presentes no mesmo componente. $0 < \gamma \leq 1$.

$$\gamma'_m = -\left[\frac{n_s}{n}\log_2\frac{n_s}{n} + n_a\left(\frac{1}{n}\log_2\frac{1}{n}\right)\right] \qquad \textbf{Equação 1.6.2.3}$$

$$\gamma_m = \frac{1}{1+\gamma'_m} \qquad 0 < \gamma \leq 1 \qquad \textbf{Equação 1.6.2.4}$$

O quarto tipo de descritor (η_m) relacionado à acessibilidade dos átomos, é calculado a partir da inversa da kurtosis k_m (equações 1.6.2.5. e 1.6.2.6.). Onde t_{im} é o valor da projeção do átomo i no eixo principal t_m.

$$k_m = \frac{\sum_i t_{im}^{\,4}}{\lambda_m^{\,2} n} \qquad \textbf{Equação 1.6.2.5}$$

$$\eta_m = \frac{1}{k_m} \qquad \textbf{Equação 1.6.2.6}$$

O grupo de descritores ηm, pode ser interpretado como a quantidade de espaço não preenchido por átomo projetado. Quanto menor for o valor da kurtosis, maior será o valor de ηm, portanto maior o espaço projetado não preenchido.

Os descritores não direcionais WHIM são diretamente derivados dos descritores direcionais, não dependendo dos eixos principais tm. Os descritores T, A

e V representam respectivamente às contribuições linear, quadrática e completa para o tamanho da molécula (equações 1.6.2.7 a 1.6.2.9). O formato da molécula, a simetria da molécula e sua densidade são representadas respectivamente por K, G, D (equação 1.6.2.10 a 1.6.2.12).

$$T = \lambda_1 + \lambda_2 + \lambda_3$$ **Equação 1.6.2.7**

$$A = \lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3$$ **Equação 1.6.2.8**

$$V = T + A + \lambda_1\lambda_2\lambda_3$$ **Equação 1.6.2.9**

$$K = \frac{\sum_m \left| \frac{\lambda_m}{\sum_m \lambda_m} - \frac{1}{3} \right|}{\frac{4}{3}} \qquad 0 \le K \le 1$$ **Equação 1.6.2.10**

$$G = (\gamma_1\gamma_2\gamma_3)^{1/3}$$ **Equação 1.6.2.11**

$$D = \eta_1 + \eta_2 + \eta_3$$ **Equação 1.6.2.12**

Esta classe apresenta 99 descritores (Todeschini & Gramatica 1997a) (Todeschini & Gramatica 1997b);

1.6.3. Descritores RDF

Esta sigla é utilizada para *Radial Function Distribution*. São descritores obtidos através da função (equação 1.6.3.1) de distribuição radial calculada sobre as distâncias interatômicas de uma molécula. A função pode ser interpretada como

sendo a distribuição de probabilidade para encontrar um átomo em um volume esférico de raio de valor r. (Hemmer *et al.* 1999).

Na equação 1.6.3.1, **N** é o número de átomos da molécula, *f* é um fator de escalonamento, A_i e A_j são propriedades dos átomos (massa atômica, eletronegatividade, volume de van der Waals e pela polarizabilidade) *i* e *j* respectivamente. No termo exponencial da equação, r_{ij} é a distância entre os átomos *i* e *j*, *B* é um parâmetro de aplainamento (que define a distribuição de probabilidade das distâncias individuais), e *r* é o raio pré-definido. Quanto maior o valor de *B*, maior é a influência da diferença das distâncias nos valores de *g(r)*.

Esta classe de descritores apresenta algumas características em comum com a classe de descritores 3D MoRSE desenvolvida pelo mesmo grupo de pesquisa (Schuur *et al.* 1996) (descrita no item 1.6.4. *Descritores 3D-MoRSE*). Estas características são:

1. independência da quantidade dos valores do número de átomos, ou seja, do tamanho da molécula;
2. exatidão relativa ao arranjo em 3 dimensões dos átomos;
3. invariância com relação à translação e rotação da molécula inteira;

$$g(r) = f \sum_{i}^{N-1} \sum_{j>i}^{N} A_i A_j e^{-B(r-r_{ij})^2} \qquad \textbf{Equação 1.6.3.1}$$

Esta classe apresenta 150 descritores (Hemmer *et al.* 1999);

1.6.4. Descritores 3D-MoRSE

Esta sigla é utilizada para Molecule Representation of Structure based on Electron diffraction. Estes descritores refletem a distribuição em três dimensões de diferentes propriedades moleculares e expressam informações sobre a ramificação das moléculas.

São obtidos através da somatória dos produtos de cada uma das propriedades atômicas, a saber: massa, eletronegatividade, volume de van der Waals e, polarizabilidade. A função de cálculo (equação 1.6.4.1), deriva daquela utilizada determinação da estrutura molecular através das medidas de difração eletrônica. Devido a característica desta função, o número de valores obtidos independe do tamanho da molécula. Nesta função, Ai e Aj são os valores das diferentes propriedades dos átomos i e j, rij é a distância interatômica entre os respectivos átomos, e s é um fator que divide a função em 32 valores. Por exemplo, para o cálculo do descritor Mor07m a propriedade utilizada para Ai e Aj é a massa atômica e o valor de s é 7 Å-1. Esta classe apresenta 160 descritores.(Schuur *et al.* 1996; Gasteiger *et al.* 1996)

$$I(s) = \sum_{i=2}^{N} \sum_{j=1}^{i-1} A_i A_j \frac{sin(sr_{ij})}{sr_{ij}}$$ **Equação 1.6.4.1**

Onde: s = 0,2,...31 $Å^{-1}$

1.6.5. Descritores de Auto-correlação 2D

Os descritores de autocorrelação 2D podem ser definidos como relação entre valores de uma única variável entre os átomos (considerando a distância topológica entre estes) de uma molécula representada em 2 dimensões.

Os descritores de autocorrelação derivam de funções matemáticas que foram utilizadas principalmente para estudos estatísticos geográficos (Moran 1950) (Geary 1954). Os descritores de autocorrelação gerados pelo programa DRAGON 5.4 são: ATS (descritor de autocorrelação de uma estrutura topológica Broto-Moreau), MATS (Moran autocorrelation), GATS (Geary autocorrelation).

Os descritores ATS, são derivados da função matemática (equação 1.6.5.1**)**, onde *f(x)* é a medida de uma propriedade associada a cada ponto do segmento AB, *f*(*x*+*t*) é a medida da mesma propriedade em um ponto diferente de *f(x)*. Portanto a função *F(t)* é a descrição da mesma propriedade, porém com uma precisão menor. Contudo a *F(t)* tem uma vantagem de independer de um referencial externo, já que t é uma variável interna e permanece inalterada quando é a função f(x) é transladada ao longo do eixo x. A autocorrelação também é utilizada no tratamento de sinais elétricos como a eletroencefalografia (Moreau & Broto 1980).

$$F(t) = \int_{AB} f(x) f(x+t) dx \qquad \textbf{Equação 1.6.5.1}$$

A função representada na equação 1.6.5.1 é adaptada para a forma vetorial considerando as distâncias topológicas entre os átomos (*i* e *j*) de uma molécula representada em 2 (equação 1.6.5.2).

$$S^2 = \sum_i f^2(i) + \sum_{i \neq j} 2 f(i) f(j)$$ **Equação 1.6.5.2**

O primeiro termo da equação 1.6.5.2 é o primeiro componente do vetor de autocorrelação, o qual é associado a um valor de distância topológica igual a 0. O segundo termo pode ser dividido em diversas somatórias parciais contendo pares de átomos separados com o mesmo valor de distância topológica. Estas somas parciais são os outros componentes do vetor de autocorrelação (Broto *et al.* 1984).

Os descritores ATS obtidos pelo programa DRAGON utilizam o segundo termo da equação 1.6.5.2, portanto são obtidos para os átomos com distâncias topológicas maiores ou iguais a 1 (equação 1.6.5.3). Nesta equação k é um valor de distância topológica pré-determinada, N é o número de átomos na molécula, *Ai* e *Aj* são propriedades atômicas (massa atômica, o volume de van der Waals, a polarizabilidade ou a eletronegatividade) dos átomos *i* e *j* que estejam a uma distância topológica k, e δ é a função delta de Dirac definida na equação 1.6.5.4. (Broto *et al.* 1984; Consonni *et al.* 2002a).

$$ATS_k = \sum_{i=1}^{N-1} \sum_{j>i} A_i A_j \delta(k, d_{ij})$$ **Equação 1.6.5.3**

$$\delta(k, dij) = \begin{cases} 1 \text{ se dij} = \text{k} \\ 0 \text{ se dij} \neq k \end{cases}$$ **Equação 1.6.5.4**

O descritor de autocorrelação Moran (MATS) (equação 1.6.5.5), um dos mais antigos descritores de autocorrelação, compara o valor de uma variável de um vértice (átomo), com todos os outros, que estejam numa separados por um valor de distância topológica k. Na equação 1.6.5.5 xi e xj são as propriedades dos átomos i e j respectivamente e $\overline{x}$ é a média dos valores da atômicas. Valores altos deste

descritor indicam uma autocorrelação positiva, valores negativos indicam uma autocorrelação negativa (Moran 1950).

$$MATS_k = \frac{\sum_i \sum_j (x_i - \bar{x})(x_j - \bar{x})\delta(k, d_{ij})}{\sum_i (x_i - \bar{x})^2}$$ **Equação 1.6.5.5**

O descritor de autocorrelação Geary (GATS) (equação 1.6.5.6), é semelhante ao descritor MATS, porém a interação não é calculada pelo produto dos desvios da média, mas pelos desvios dos valores da propriedade atômica de um vértice (átomo) com o de outro vértice. Valores maiores deste descritor indicam uma autocorrelação inversa, valores próximos de 0 indicam uma autocorrelação positiva (Geary 1954). Os descritores MATS fornecem valores mais representativos globalmente, enquanto o GATS é mais sensível a diferença de valores de propriedades de átomos vizinhos.

$$GATS_k = \frac{\sum_i \sum_j (x_i - x_j)\delta(k, d_{ij})}{\sum_i (x_i - \bar{x})^2}$$ **Equação 1.6.5.6**

Esta classe apresenta 96 descritores.

1.6.6. Descritores Geométricos Calculados pelo Programa DRAGON

São diversos descritores baseados na distância geométrica entre os átomos. Alguns destes descritores calculam a soma geométrica entre os átomos de nitrogênio, átomos de oxigênio, entre os átomos de enxofre, dentre outros.

Os descritores deste tipo são baseados na *layer distance matrix* (LM3D) (equação 1.6.6.2), a qual é obtida através da matriz de distância geométrica (equação 1.6.6.1). (Diudea *et al.* 1995). Esta classe apresenta 70 descritores

$$m_i = \sum_{i=1}^{N} d_{ij}$$ **Equação 1.6.6.1**

$$lm_{ik} = \sum_{u=1}^{N} m_i \delta(k, d_{ij})$$ **Equação 1.6.6.2**

$$\delta(k, dij) = \begin{Bmatrix} 1 \text{ se dij} = \text{k} \\ 0 \text{ se dij} \neq k \end{Bmatrix}$$ **Equação 1.6.6.3**

1.6.7. Descritores Topológicos Calculados pelo Programa DRAGON

A necessidade de usar descritores topológicos originou-se do fato de que propriedades físico-químicas podem ser expressas em números e, portanto têm uma possibilidade numérica de se fazer comparações e correlações. As estruturas químicas são entidades discretas, portanto é preciso que se traduzam estas estruturas em números com o objetivo de avaliar o grau de similaridade/dissimilaridade e fazer correlações com diversas propriedades físico-químicas. As estruturas em 3 dimensões das moléculas dependem da sua topologia, ou seja, das posições individuais dos átomos e das ligações entre eles (Hansch *et al.* 1990).

Os descritores topológicos, comumente (Balaban & Devillers 1999) chamados de índices topológicos (*TI*), são calculados baseados na matriz de adjacência e/ou na matriz de distância topológica de uma molécula representada em 2 dimensões. Nas representações das moléculas em 2 dimensões, os átomos e as ligações

correspondentes são representados como vértices e arestas, respectivamente (Balaban & Devillers 1999).

Quando dois átomos (vértices) estão ligados (vizinhos) por uma ligação covalente (aresta), sua distância topológica é definida como 1 (Balaban & Devillers 1999) e estes átomos são adjacentes. As distâncias e as adjacências entre dois átomos numa molécula representada em 2 dimensões são as menores possíveis. Exemplificando, para a representação da molécula do 1-metil-2-propil-ciclobutano (figura 1.6.7.1), as matrizes de adjacência e de distância são apresentadas nas figuras, respectivamente, 1.6.7.2 e 1.6.7.3.

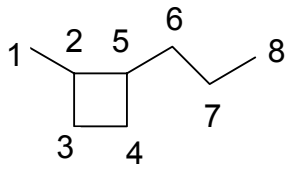


Figura 1.6.7.1. Representação em 2 dimensões da estrutura molecular do 1-metil-2-propil-ciclobutano.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	0	1	0	0	0	0	0	0
2	1	0	1	0	1	0	0	0
3	0	1	0	1	0	0	0	0
4	0	0	1	0	1	0	0	0
5	0	1	0	1	0	1	0	0
6	0	0	0	0	1	0	1	0
7	0	0	0	0	0	1	0	1
8	0	0	0	0	0	0	1	0

Figura 1.6.7.2. Matriz de adjacência da molécula do 1-metil-2-propil-ciclobutano. Os átomos foram numerados como atribuído na figura 1.6.7.1.

	1	2	3	4	5	6	7	8
1	0	1	2	3	2	3	4	5
2	1	0	1	2	1	2	3	4
3	2	1	0	1	2	3	4	5
4	3	2	1	0	1	2	3	4
5	2	1	2	1	0	1	2	3
6	3	2	3	2	1	0	1	2
7	4	3	4	3	2	1	0	1
8	5	4	5	4	3	2	1	0

Figura 1.6.7.3. Matriz de distâncias topológicas da molécula do 1-metil-2-propil-ciclobutano. Os átomos foram numerados como atribuído na figura 1.6.7.1.

Os descritores moleculares são regularmente criticados na literatura sobre *QSAR*. Algumas das principais críticas (Balaban & Devillers 1999) dos descritores topológicos são:

1. Têm um significado físico-químico pouco claro;
2. Existe uma probabilidade de correlação ao usar um grande número de descritores altamente intercorrelacionados como, por exemplo, conectividade normal e conectividade de valência;
3. O índice de degeneração de certos descritores topológicos pode ser alto;

Algumas das vantagens dos descritores topológicos que os fazem ser largamente utilizados nos estudos de QSAR e QSPR (Relação Quantitativa entre Estrutura Química e Propriedade Molecular) são:

1. Os descritores topológicos podem ser calculados para todas as moléculas existentes;

2. A obtenção dos valores dos descritores topológicos é relativamente rápida utilizando os computadores hoje existentes;
3. O cálculo de diferentes descritores de uma mesma molécula e e considerando-se cada descritor como uma variável permite uma abordagem usando estatística multivariada (Balaban & Devillers 1999);

Há uma extensa quantidade de diferentes descritores topológicos presentes no programa DRAGON v. 5.4 (119 descritores), como por exemplo:

a) O índice topológico CICk (equação 1.6.7.2) que como o índice ICk (índice de informação das moléculas) (equação 1.6.7.1), considera os átomos de hidrogênio nas moléculas. Os valores iguais a ordem zero representa grupos de átomos isolados em classes equivalentes e a ordem 1 denota pares de átomos ligados covalentemente, agrupados em ordem de equivalência (de acordo com a natureza dos átomos, e a multiplicidade da ligação). Para uma série de n vértices, estes são considerados equivalentes se representam o mesmo elemento químico, e possuem as mesmas características estruturais com os seus vizinhos de ordem k. Se há diferentes classes diferentes classificados na ordem k, estes elementos são numerados sucessivamente pi (i = 1, 2, 3, ..., r), onde r é o número total de diferentes elementos classificados na mesma ordem (Balaban & Devillers 1999).

$$IC_k = -\sum_{i=1}^{r} p_i \log_2 pi$$

Equação 1.6.7.1

$$CIC_k = \log_2 n - IC_k$$

Equação 1.6.7.2

$$p_i = \frac{n_i}{n}$$

Equação 1.6.7.3

Onde: ni - número de átomos de mesmo elemento com a mesma vizinhança de ordem k;
n – número total de átomos;

Pela equação 1.6.7.1, verifica-se que quanto maior a diversidade entre os vértices de mesma ordem k, maior será o valor do índice de informação das moléculas (ICk). Através da equação 1.6.7.2 verifica-se que quanto maior o ICk, menor será CICk, portanto quanto maior a diversidade entre os vértices de mesma ordem k, menor será o CICk.

b) O descritor PJI2 é calculado a partir do raio (R) e do diâmetro D generalizados (equação 1.6.7.4). O raio e o diâmetro são calculados a partir dos pontos extremos e centrais de uma molécula em duas dimensões. Todas as distâncias topológicas dos átomos (vértices) de uma molécula representada em 2 dimensões são calculadas com relação a todos os outros átomos (vértices) desta. O átomo que apresentar o maior valor de distância topológica com o átomo mais distante será considerado como ponto extremo e seu valor de distância topológica será o diâmetro generalizado (D). Conseqüentemente o átomo que apresentar o menor valor de distância topológica com o átomo mais distante será considerado como ponto central e seu valor de distância topológica será o raio generalizado (R). O ponto extremo e o centro da molécula não precisam ser únicos (Petitjean 1992).

O descritor PJI2 pode ser interpretado com uma medida de balanço entre uma molécula cíclica e uma acíclica. Um valor de PIJ2 igual a 0, indica uma

molécula estritamente cíclica, quanto maior o valor de PJI2, maior será o caráter acíclico do formato da molécula (Petitjean 1992).

$$PJI_2 = \frac{(D - R)}{R}$$ **Equação 1.6.7.4**

1.6.8. Descritores BCUT

Sigla utilizada para os descritores propostos por Burden (B), validados pelo *Chemical Abstracts Service* (CAS) *Registry* e ampliados na Universidade do Texas (UT). Os descritores BCUT são calculados através dos autovalores obtidos da matriz de adjacência (exemplo: figura 1.6.7.2) com elementos nulos da diagonal substituídos por alguma propriedade atômica (massa atômica, volume de van de Waals, eletronegatividade, e polarizabilidade) (Burden 1997).

A essência da obtenção dos descritores é resolver a equação de autovalor (equação 1.6.8.1).

$$[\mathrm{B}][\mathrm{V}] = [\mathrm{V}][\mathrm{e}]$$ **Equação 1.6.8.1**

Na equação 1.6.8.1, [V] é a matriz de autovetores, [e] é uma matriz diagonal de autovalores, e [B] é uma matriz de conectividade com as seguintes características (BURDEN 1989):

1. Os elementos diagonais dos átomos são valores de alguma propriedade atômica (massa atômica, volume de van de Waals, eletronegatividade, e polarizabilidade);

2. Os valores dos elementos não diagonais dependem da ligação existente entre os átomos i e j. O valor é $\sqrt{1}$ para uma ligação simples, $\sqrt{2}$ para uma ligação dupla, $\sqrt{3}$ para uma ligação tripla e $\sqrt{1,5}$ para uma ligação aromática;
3. Todos os outros elementos não diagonais recebem valor 0,001.

Considerando-se que esta classe de descritores depende das propriedades atômicas, pode-se aplicar esta em estudos de QSAR e QSPR (Pearlman & Smith 1999; Burden 1997) inclusive para moléculas isotopológicas, ou seja, com a mesma conectividade. Esta classe apresenta 64 descritores ;

1.6.9. Grupos Funcionais do Programa DRAGON

Coletânea de fragmentos moleculares, contendo poucos átomos. Como por exemplo, números de carbonos primário, secundário, terciário, quaternário; de anéis aromáticos substituídos ou não-substituídos; de cetonas alifáticas ou aromáticas. Esta classe apresenta 121 descritores (Todeschini & Consonni 2000);

1.6.10. Descritores de Átomo Centrado

Os descritores de átomo centrado, identificam diversas seqüências de átomos como fragmentos e, verificado que estes fragmentos (ou seja sua estrutura química) se correlacionam com a atividade biológica (Ghose *et al.* 1988). Estes fragmentos estão classificados a partir de um átomo central, portanto estes fragmentos classificam o átomo de acordo com sua vizinhança (dependem dos átomos aos quais o átomo central está ligado e, dos tipos de ligações envolvidas: simples, dupla, tripla, aromática). Como por exemplo: CRn, número de carbonos (sp^3) ligados

respectivamente à uma, a duas, a três ou, a quatro cadeias alifáticas; CXn, número de carbonos (sp^3) ligados a um, a dois, a três ou a quatro halogênios e, =CXn, número de carbonos (sp^2) ligados a um, a dois, a três ou a quatro halogênios. Esta classe apresenta 120 descritores (Viswanadhan *et al.* 1989);

1.6.11. Descritores Constitucionais Calculados pelo Programa DRAGON

São descritores independentes da conectividade e conformação moleculares. Alguns exemplos desta classe de descritores são: tipos de átomos e de ligações, peso molecular, e somatória do volume atômico de van der Waals. Esta classe de descritores não consegue distinguir a maioria dos isômeros moleculares e as moléculas similares. Esta classe apresenta 47 descritores (Todeschini & Consonni 2000);

1.7. As bases teóricas das Redes Neurais

Surgiu há algum tempo o interesse em criar programas de computador cujo mecanismo simulasse neurônios humanos. Em 1943 McCulloch publicou um trabalho denominado “A logical calculus of the ideas immanent in nervous actvity” (McCulloch *et al.* 1943). A pesquisa nesta área não se desenvolveu muito até os anos 70, quando houve um ressurgimento do interesse nas redes neurais (RN) devido a várias razões: a fabricação de computadores mais rápidos (onde se pode trabalhar com programas maiores), a descoberta de novas arquiteturas de redes neurais e de novos algoritmos de aprendizagem e o interesse em construir computadores com modelo de processamento paralelo. Para uma visão geral da teoria e das aplicações das RN, existem várias revisões bastante abrangentes

descritas na literatura (Zupan *et al.* 1993) (Minsk *et al.* 1969) (Smith *et al.* 1993) (Fraser *et al.* 1997).

Em geral os programas de computador feitos até o momento e denominados de redes neurais imitam o mecanismo de transmissão sináptica dos neurônios biológicos onde as transmissões são simples impulsos (entrada de dados). Nos neurônios biológicos as conexões entre um neurônio e outro, chamadas sinapses, são diferentes em termos de intensidade do sinal, no caso dos computadores pode-se dar pesos às entradas (*inputs*) e com isto obtém-se uma saída ponderada que é o resultado final da rede (Zupan *et al.* 1993).

Um neurônio artificial é um simbolismo computacional que supomos imitar um neurônio biológico, isto é, ele aceita muitos diferentes sinais x_j vindos de neurônios vizinhos e os processa de uma maneira predefinida figura (1.7.1) (Zupan *et al.* 1993).

Dependendo da saída deste processo, o neurônio j decide se dispara um sinal yj ou não. O sinal disparado pode ser 1, 0, ou pode ser um valor real entre 1 e 0; dependendo se estamos trabalhando com valores binários ou reais (Zupan *et al.* 1993).

A função que calcula a saída de um vetor multidimensional de entrada X, f(X), é composta de duas partes. A primeira avalia o que chamamos de “entrada da rede” (Net), enquanto a segunda transfere a entrada da rede, de uma maneira não linear para um valor de saída Y. A primeira função é uma combinação linear das variáveis x_1, x_2,...., x_m, multiplicados pelos coeficientes W_{ji}, chamados pesos, enquanto a segunda serve como uma função de transferência, passando o sinal (s) através do axônio para outros dendritos de neurônios (Zupan *et al.* 1993).

A saída y_j do ultimo neurônio pode ser calculada de acordo com a seguinte equação 1.7.1:

$$Netj = \Sigma\, w_{ji}\, x_i \qquad \textbf{Equação 1.7.1}$$

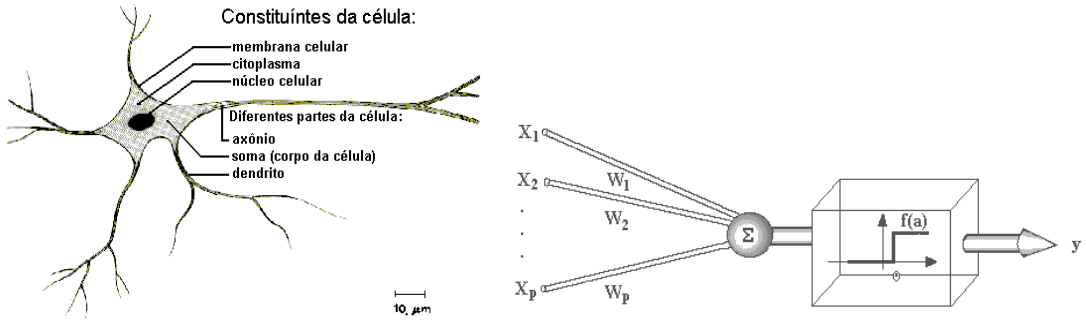


Figura 1.7.1. Comparação entre um neurônio artificial e outro biológico. O círculo que mimetiza o corpo celular do neurônio representa procedimentos matemáticos simples que fazem um sinal de saída (*output*) y, a partir do conjunto de sinais de entrada (*input*), serem representados pelo vetor multi-variado X.

Depois Netj é colocado como argumento em uma função de transferência. São utilizadas diversas funções de transferência como as listadas abaixo e mostradas na figura 1.7.2.

1. Identidade: $f(x) = a \times x$
2. Degrau: $f(x) = \begin{cases} \gamma\ se\ x \geq \theta \\ -\gamma\ se\ x < \theta \end{cases}$
3. Rampa: $f(x) = \begin{cases} \gamma\ se\ x \geq \theta_2 \\ x\ se\ x \geq \theta_1 \\ -\gamma\ se\ x < \theta_1 \end{cases}$
4. Sigmoidal: $f(x) = \frac{1}{1+e^{-x/T}}$

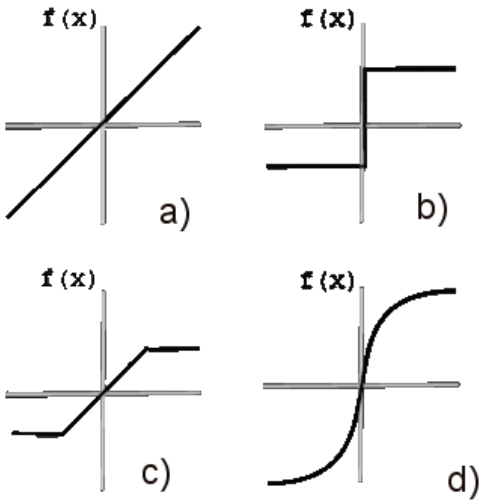


Figura 1.7.2. Funções de ativação utilizadas em redes neurais: a) função identidade; b) função degrau; c) função rampa; d) função sigmóide.

Os pesos Wi nos neurônios artificiais são análogos as forças das sinapses dos neurônios reais entre os axônios que disparam os sinais e os dendritos que recebem estes sinais (figura 1.7.1) (Zupan *et al.* 1993).

Acredita-se que o “conhecimento” no cérebro é conseguido pela adaptação das sinapses a diferentes entradas de sinais, causando melhores ou piores sinais de saída. Os resultados são constantemente mandados de volta como novos sinais de entrada (*inputs*). De maneira análoga ao cérebro humano, os neurônios artificiais tentam imitar o processo de adaptação da força das sinapses por uma adaptação interativa dos pesos *Wji* nos neurônios, observando as diferenças entre uma determinada saída *yj* e a saída desejada *Tj* (Zupan *et al.* 1993).

Redes neurais artificiais (RN) podem ser compostas de diferentes números de neurônios, nas aplicações em Química varia de 10 até milhares. Os neurônios nas RNs podem ser colocados em uma, duas, três ou várias camadas (figura 1.7.3).

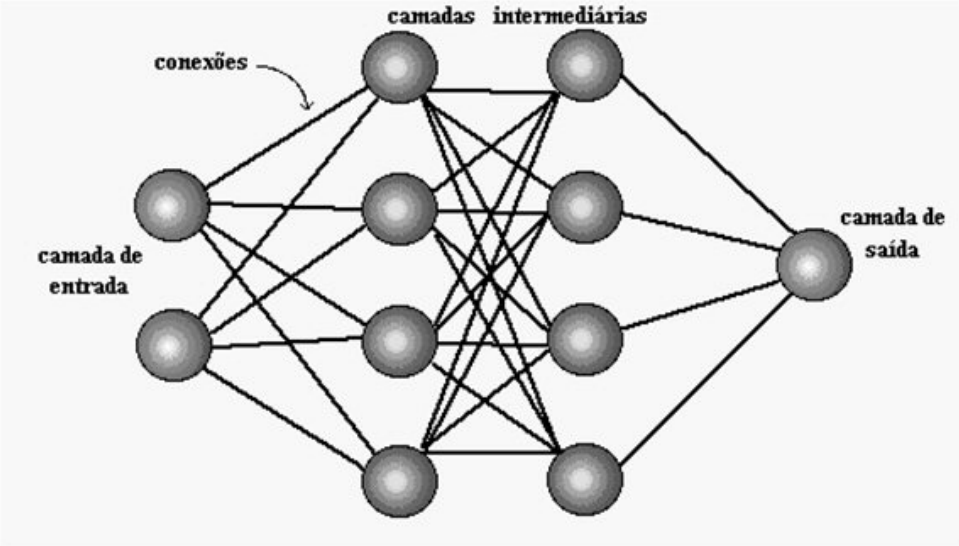


Figura 1.7.3. Rede neura artificial (RNA) de uma (esquerda) e de duas camadas (direita).

A seleção de um conjunto dados para treino é o primeiro passo quando se quer aplicar um método de aprendizagem para modelagem clássica, para reconhecimento de padrões, para sistemas especialistas e para redes neurais. O procedimento padrão consiste em dividir os dados em três partes: a primeira para treinamento, a segunda para controle e a terceira para testar a rede quando ela já estiver "treinada" (Zupan *et al.* 1993).

As aplicações de RN em vários campos da ciência são diversas. Como exemplos: comércio e administração (estimação de custos), engenharia (configuração de equipamentos), indústria (controle de qualidade), medicina (diagnóstico médico) e outras como reconhecimento de caracteres e processamento de linguagem, além da previsão do tempo (Zupan *et al.* 1993).

Em Química, as aplicações das RN são inúmeras e estão extensivamente discutidas (Zupan *et al.* 1993). Alguns dos trabalhos mais específicos a análise de limonóides em Meliaceae (Fraser *et al.* 1997), análise de espectros no infravermelho (Cleva *et al.* 1999), em RMN ^{13}C (Doucet *et al.* 1993) e em espectrometria de

massas (Lohninger *et al.* 1992) previsão de esqueletos terpênicos (Emerenciano *et al.* 2006).

1.7.1. Aprendizado Supervisionado em Redes Neurais Artificiais

O aprendizado supervisionado (figura 1.7.1.1) precisa ter uma série de entradas e saída (Xs,Ts). Para treinar a rede nós devemos ter uma série de variáveis *m* como entrada *Xs* (por exemplo, dados espectrais) e a cada *Xs* é associado uma resposta *Ts* (por exemplo, fragmentos em determinação estrutural). Os pesos dos neurônios são primeiramente corrigidos na camada de saída, depois na segunda e posteriormente na primeira, ou seja, na qual obtém os sinais diretamente da camada de entrada. Depois que a camada *n* de neurônios dispara suas saídas *Yi*, que pode ser vista como um vetor *Y* (y1, y2,...yi...Yn), estas repostas são comparadas com os valores do objetivo *tj* do vetor *Ts* que acompanha o vetor de entrada *Xs*. (Zupan *et al.* 1993) Deste modo, o erro δ i em cada nódulo de saída pode ser definido na equação 1.7.1.1):

$$\Sigma i = y\ i - t\ l \qquad \textbf{Equação 1.7.1.1}$$

O aprendizado é feito em ciclos ou épocas chamados *"epochs*", ou seja, define-se uma época como a apresentação completa do conjunto de padrões à rede. Cada ciclo corresponde a um período mínimo no qual todos os pares de entradas e saídas são apresentados uma vez para a rede. Em geral, depois de cada ciclo calcula-se o RMS (*root-mean-square*) segundo a equação 1.7.1.2:

$$RMS = \left(\left[\Sigma_{s=1} \Sigma_{j=1} \left(t_{si} - Y_{si} \right)^2 \right] / rn \right)^{1/2} \qquad \textbf{Equação 1.7.1.2}$$

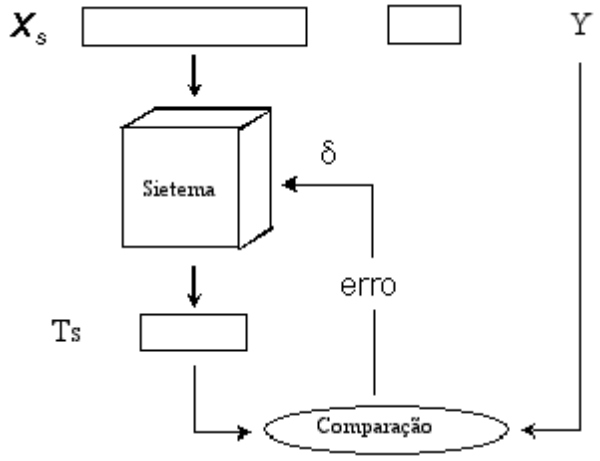


Figura 1.7.1.1. Esquema de uma rede supervisionada. Resultados da diferença entre os valores desejados e obtidos são utilizados no ajuste dos valores de pesos da rede.

1.7.2. Aprendizado Não Supervisionado em Redes Neurais Artificiais

No aprendizado não supervisionado nenhum “professor” é envolvido, ao invés disto, a rede é exposta a um número de entradas e se organiza de modo a fazer suas próprias classificações com base nestes dados. A aprendizagem não-supervisionada pode ser usada como módulo de “descoberta de características” que precede a aprendizagem supervisionada. O modelo de RN não supervisionado mais utilizado é o modelo de Kohonen (Kohonen, 2001). Com base em um conjunto de dados de entrada, a rede começa a analisá-los e tenta descobrir relações entre partes diferentes do conjunto. Os principais objetivos nas análises de ensino não supervisionado é diminuir a dimensionalidade dos dados para uma melhor visualização e verificar a relação entre estes, como mostrado na figura 1.7.2.1.

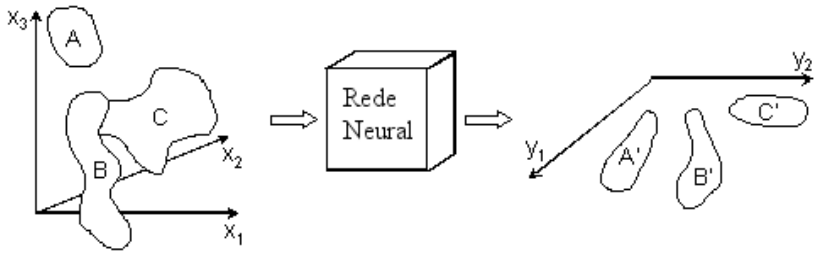


Figura 1.7.2.1. Esquema de uma rede neural não supervisionada. Neste exemplo as 3 variáveis originais foram combinadas gerando apenas 2 variáveis, facilitando a visualização da distribuição dos dados.

1.7.3. Mapas Auto-Organizáveis

A rede neural Kohonen pode ser vista como uma metodologia que permite projetar objetos de um espaço hiper-dimensional em um plano de duas dimensões resultando nos chamados mapas auto-organizáveis (SOM – *"Self Organization Maps"*) (Kohonen, 2001). A rede neural de Kohonen é tipicamente constituída de duas redes de neurônio que são conectados por uma conexão ponderada para cada entrada (pode se utilizar ordens maiores, porém para uma melhor visualização dos dados a bidimensionalidade é mais recomendável). No fim do treinamento os dados de amostra são associados com a rede de neurônios de acordo com sua similaridade baseada na distância Euclidiana no original hiper-espaço.

O uso da rede neural de Kohonen como uma técnica de aprendizagem não supervisionada, apresenta um baixo risco de *"overfitting"* ou *"overtraining"*, ou seja, um bom ajuste dos dados devido à presença de diversas entradas que combinadas de uma determinada maneira explica a variância dos dados variável dependente como ocorre na aprendizagem supervisionada. Pode-se comparar SOM com análise de componentes principais (PCA). O uso de SOM (*"Self Organization Maps")* tem

sido aplicado aos dados de propriedades de diversos compostos. (Manallack & Livingstone 1999).

Os neurônios da camada de saída estão interconectados por uma relação de vizinhança que descreve a estrutura do mapa. Por exemplo, na Figura 1.7.3.1. tem-se um mapa com a camada de saída, bidimensional, retangular. Nesta figura somente estão representados os vetores de código w, conectados ao neurônio j.

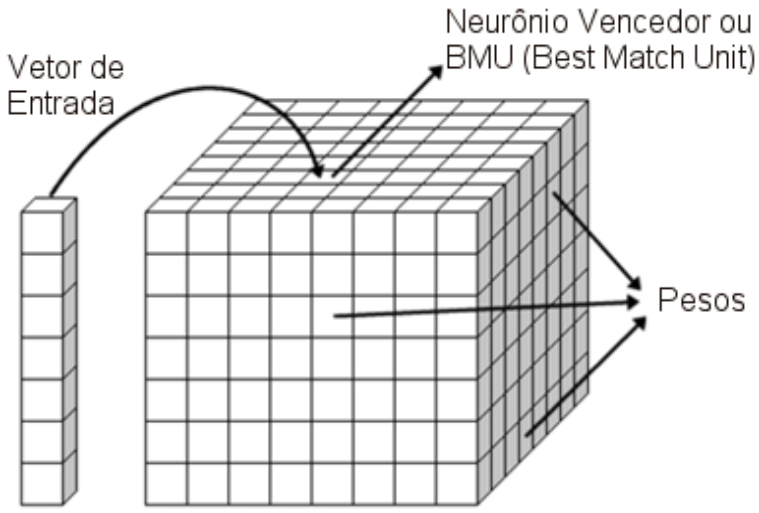


Figura 1.7.3.1. Representação de uma rede neural Kohonen. O vetor de entrada (amostra) é comparado com todos os vetores de pesos. O vetor peso mais semelhante com o vetor de entrada, elege o neurônio vencedor.

O mapa auto-organizável foi idealizado a partir da analogia com a região do córtex cerebral humano. Descobriu-se que esta parte do cérebro aloca regiões específicas para atividades específicas e que, para uma determinada ativação cerebral, o grau de ativação dos neurônios diminuía à medida que se aumentava a distância da região de ativação inicial (Kohonen, 2001).

Existem diferentes topologias para estruturação de um Mapa Auto-Organizável, sendo que a estrutura mais comum é a de duas dimensões. A organização dos neurônios pode ser hexagonal (6 vizinhos), ou retangular (4 ou 8 vizinhos) (figura 1.7.3.2).

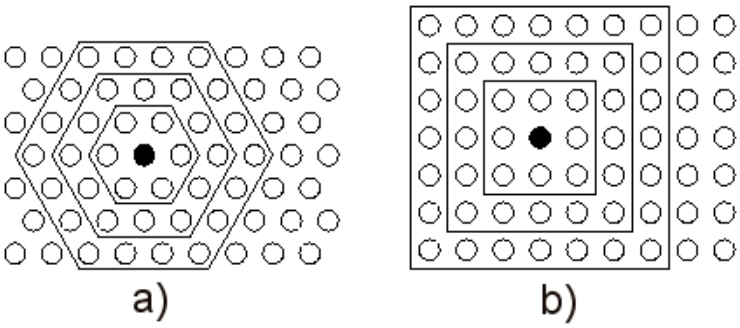


Figura 1.7.3.2. Topologias dos mapas auto-organizáveis com relação à vizinhança.

O SOM vem sendo aplicado numa ampla variedade de problemas em diversas áreas. Destacam-se as potencialidades de visualização de dados multivariados, análise de agrupamentos, mineração de dados, descoberta de conhecimento e compressão de dados (Kohonen, 2001).

1.7.3.1. Treinamento Padrão ou Seqüencial

Para o algoritmo de aprendizagem seqüencial as apresentações dos padrões x têm de ser de forma aleatória. O algoritmo básico de treinamento do SOM consiste de três fases. Na primeira fase, competitiva, os neurônios da camada de saída competem entre si, segundo algum critério, geralmente a distância Euclideana, para encontrar um único vencedor, também chamado de BMU (“Best Match Unit”). Este neurônio contém valores dos pesos, que foram inicialmente determinados de forma randômica, mais próximos do valor do vetor de entrada (Kohonen, 2001).

Portanto o neurônio cujo vetor de pesos *m* é mais próximo do vetor de entrada *x* (equação 1.7.3.1.1), ou seja, é o neurônio o qual os valores dos pesos são mais próximos dos valores dos dados de entrada (variáveis) para uma determinada amostra.

$$\|x - m_c\| = min\{\|x - m_i\|\} \qquad \textbf{Equação 1.7.3.1.1}$$

A distância pode ser a Euclidiana como citada anteriormente ou como mostrado na equação 1.7.3.1.2:

$$\|x - m_c\| = \sum_{k \in K} w_k (x_k - m_k)^2 \qquad \textbf{Equação 1.7.3.1.2}$$

onde: *K* é a série de variáveis do vetor da amostra *x*, x_k e m_k, são o k^{th} componente do vetor amostra e peso respectivamente e wk é uma forma preeliminar para excluir w_k=0 ou incluir w_k=1 a variável no processo de se achar o neurônio vencedor.

Na segunda fase, cooperativa, é definida a vizinhança deste neurônio. Esta vizinhança pode ser determinada pela distância topológica mostrada na equação 1.7.3.1.3, relacionando-se a alguma função. Uma das funções mais utilizadas é a gaussiana de vizinhança $\mathrm{h_{ci}}$:

$$h_{ci}(t) = exp\left(-\frac{\|r_c - r_i\|^2}{2\sigma^2(t)}\right) \qquad \textbf{Equação 1.7.3.1.3}$$

Onde : $\sigma(t)$é uma função monotonicamente decrescente (equação 1.7.3.1.4);

$$\sigma(t) = \sigma(0).\exp\left(\frac{t}{\tau_i}\right) \qquad \textbf{Equação 1.7.3.1.4}$$

Onde: τ_i é uma constante.

A função de vizinhança tem como objetivo controlar o nível de atuação dos neurônios em torno do neurônio vencedor do processo competitivo. Seguindo o modelo neurobiológico tem-se que o nível de atuação dos neurônios vizinhos decai à medida que o mesmo se distancia do BMU (Kohonen, 2001).

Na última fase, adaptativa, os vetores de peso do neurônio vencedor e de sua vizinhança são ajustados (equação 1.7.3.1.5).

$$m_i(t+1) = m_i(t) + \alpha(t)h_{ci}(t)[x(t) - m_i(t)] \quad \textbf{Equação 1.7.3.1.5}$$

Onde: $\alpha(t)$ é a função de taxa de aprendizagem, t é a coordenada discreta de tempo.

A função da taxa de aprendizagem pode ser linear exponencial ou inversamente proporcional ao fator tempo (t) (equação 1.7.3.1.6) (Kohonen 2001).

$$\alpha(t) = \frac{\alpha_0}{\left(1+100\frac{t}{T}\right)} \quad \textbf{Equação 1.7.3.1.6}$$

Onde: T é uma constante.

1.7.3.2. Treinamento em Lote ("batch training")

O método de treinamento em lote também é um método iterativo como o seqüencial. Porém em vez de utilizar uma amostra por vez, toda a série de dados é apresentada ao mapa antes de serem feito os devidos ajustes (Kohonen, 2001) (Vensanto *et al.* 1999).

Em cada treinamento, a série de dados é dividida de acordo com as regiões de Voronoi dos vetores de peso do mapa, ou seja, cada vetor de dados pertence a uma série de dados do mapa da unidade ao qual está mais próximo. Os pesos são calculados como mostrado na equação 1.7.3.2.1 (Vensanto *et al.* 1999)..

$$m(t+1) = \frac{\sum_{j=1}^{n} h_{ic(j)}(t)x_j}{\sum_{j=1}^{n} h_{ic(j)}(t)} \quad \textbf{Equação 1.7.3.2.1}$$

Onde: *c(j)* é o BMU do vetor da amostra x_j, $h_{i,c(j)}$ é a função de vizinhança aqui utilizado como um fator de ponderação, e *n* é o número de vetores das amostras;

Portanto os valores do vetor peso é atualizado simplesmente sendo substituídos pelos valores médio dos valores de todas as amostras, cada amostra é ponderada pelos valores de função de vizinhança.

Outra forma de calcular (Equação 1.7.3.2.2) os pesos atualizados dos neurônios é primeiramente calcular primeiramente a soma dos valores dos vetores das amostras de cada série de Voronoi (Vensanto *et al.* 1999).

$$s_i(t) = \sum_{j=1}^{n_{Vi}} x_j$$ **Equação 1.7.3.2.2**

Onde: n_{Vi} é o número de amostras na série de Voronoi da unidade i.

Portanto os novos valores dos vetores de ponderação dos neurônios podem ser calculados pela equação 1.7.3.2.3 (Vensanto *et al.* 1999)..

.

$$m(t+1) = \frac{\sum_{j=1}^{m} h_{ij}(t) s_j(t)}{\sum_{j=1}^{m} n_{Vi} h_{ij}(t) s_j(t)}$$ **Equação 1.7.3.2.3**

Onde: m é o número de unidades do mapa (neurônios);

Com relação ao desempenho, o método de treinamento em lote é muito mais rápido fornecendo resultados tão significativos como o método de treinamento seqüencial (Kohonen 2001).

A determinação dos parâmetros de aprendizagem em geral é empírica, baseada na experiência do usuário e em métodos de tentativa e erro. A dimensionalidade do mapa auto-Organizável e seu tamanho (m) dependerá do tipo de problema e propósito. A literatura mostra que a determinação do tamanho do SOM é um processo empírico (Kohonen, 2001). Em geral, o SOM bidimensional

NxM é usado devido sua capacidade de projeção dos dados de dimensão p num Mapa bidimensional. Para grandes volumes de dados, Mapas razoavelmente grandes são mais adequados. Todavia, grandes Mapas comprometem o desempenho do algoritmo e Mapas muito pequenos comprometem a integridade da formação topológica do SOM (Kohonen 2001).

O mapa de características calculado pelo algoritmo SOM é ordenado topologicamente, no sentido de que a localização espacial de um neurônio na grade corresponde a um domínio particular ou características dos padrões de entrada. O inverso nem sempre é verdadeiro (Kohonen, 2001).

O SOM características reflete variações na estatística da distribuição da entrada, embora a distribuição das unidades do SOM não seja exatamente a mesma da distribuição dos dados amostrais.

Pode-se afirmar que os Mapas Auto-Organizáveis fornecem uma aproximação discreta das assim chamadas curvas principais, e podem, portanto, ser vistos como uma generalização não-linear da análise de componentes principais (Silva, 2004).

Os mapas auto-organizáveis foram utilizados com sucesso em diversas aplicações em análise de banco de dados químicos, como na classificação de reações fotoquímicas (Zhang *et al.* 2005), quimiotaxonomia da família Asteraceae (Costa *et al.* 2005; Hristov *et al.* 2007), em relações entre estrutura química e atividade biológica (Gasteiger *et al.* 2003; Wagner *et al.* 2006; Fernandes *et al.* 2008), classificação de metabólitos (Gupta & Aires-de-Souza 2007), e predição de esqueletos diterpênicos (Emerenciano *et al.* 2006).

1.8. "Data-Mining" (Gasteiger *et al.* 2003)

O avanço na aquisição de dados para os sistemas tanto químicos como biológicos gerou um grande número de informações. Como conseqüência, nos últimos anos, procuram-se ferramentas, fundamentalmente matemáticas (Todeschini *et al.* 2004; Baroni *et al.* 1993; Kubinyi 1994), que permitam decodificar este volume imenso de informações, em termos estruturais e biológicos, ou seja, necessitou-se de criar um processo para analisar os dados e identificar/diferenciar as características e relações contidas neste. Estas abordagens, que se propõem extrair conhecimento de uma grande série de dados com o objetivo de fazer predições de novos eventos é denominado na língua inglesa como *data mining* (Gasteiger *et al.* 2003).

Considerando-se a seleção de variáveis e de compostos disponíveis em extensos bancos de dados, diversos algoritmos foram também utilizados e/ou desenvolvidos como, por exemplo, o algoritmo genético para a primeira (Leardi *et al.* 1992; Leardi 1994). Estes procedimentos devem, em princípio, gerar modelos que se apliquem não somente à série de treinamento, ou seja, devem gerar modelos robustos. Pode-se citar como exemplo de sucesso, a regra de seleção de compostos proposta por Lipinski (Lipinski *et al.* 1997). Nesta regra, os compostos são selecionados considerando-se as faixas de variação das propriedades que são importantes para a farmacocinética do composto.

1.8.1. Pré-tratamento dos Dados

O pré-tratamento de dados é recomendado ao se gerar um grande número de variáveis (Livingstone 1995), excluindo-se aquelas que não fornecem informações relevantes sobre o sistema, no entanto, contribuindo apenas para aumentar a quantidade de dados e de ruídos a serem tratados.

Na literatura sugere-se (Livingstone 1995) que uma maneira de se reduzir os dados é excluir as variáveis com valores constantes e aquelas com apenas um valor diferente na série. Tal situação ocorre quando há alguma propriedade mal escolhida para a série de compostos, ou seja, a variável é pouco representativa para aquela série. Atualmente, existem alguns pacotes de softwares que facilmente (com baixo custo computacional) identificam e/ou removem estas variáveis. Após a remoção destas, o escalonamento das variáveis e a matriz de correlação (análise da intercorrelação das variáveis restantes) podem ser então feitos (Livingstone 1995).

Deste modo, numa determinada série de dados, uma matriz de correlação pode ser construída entre cada par de variáveis. Em seguida, através da inspeção da matriz de correlação pode-se verificar e avaliar as características altamente correlacionadas, na série. A escolha do valor do nível máximo de corte entre as variáveis correlacionadas depende do método de análise aplicado a estas.

Alguns métodos, como por exemplo, a regressão linear múltipla – MLR, são sensíveis à presença de colinearidade na série de dados, podendo-se observar “overfit”. (ajuste em excesso). Considera-se que uma equação de regressão linear múltipla pode ser entendida como sendo uma série de variáveis, que explicam alguma ou toda variação da variável dependente (y). Assim sendo, se as variáveis independentes são correlacionadas em pares (apresentam colinearidade) ou em

forma de combinações lineares (multicolinearidade), então diferentes combinações podem explicar a mesma variação (grandeza e natureza) na variável dependente. A presença de duas variáveis colineares em uma equação pode gerar dados estatísticos de ajuste aparentemente válidos. O modelo gerado, porém, apresenta valores de coeficientes de regressão instáveis e, conseqüentemente acompanhados dos respectivos erros padrões altos (Livingstone 1995).

Efeito análogo de “overfit” pode ser observado ao se incluir muitas variáveis em uma equação de regressão. Desta forma, adiciona-se ruído ao modelo e, a equação resultante apresenta um bom ajuste apenas para as amostras aplicadas ao treinamento, apresentando um baixo poder de predição e de ajuste para outras amostras.

Outro aspecto a ser considerado na manipulação e tratamento de grande número de dados se refere à homogeneidade na distribuição dos dados na população estudada e a presença de “outliers”. Estes afetam as análises de regressão e, sua presença pode ocasionar erros na análise de regressão.

1.9. Índices Estatísticos Recentes

Apesar do conhecimento de diversos coeficientes estatísticos e de diversos métodos de seleção de modelos e, conseqüentemente das variáveis, ainda são encontrados estudos na literatura recente (Todeschini *et al.* 2004; Mattioni & Jurs 2002; Golbraikh & Tropsha 2002; Gasteiger *et al.* 2003), propondo-se novas ferramentas para avaliar e garantir a qualidade de predição do modelo bem como a elucidação de determinado mecanismo a partir do modelo gerado (Golbraikh & Tropsha 2002; Gasteiger *et al.* 2003). Esta necessidade aparece, pois encontram-se

com freqüência modelos que apresentam bom ajuste mas baixo poder de predição. Estes são algumas vezes resultados de uma correlação ao acaso e geralmente apresentam características indesejáveis como multicolinearidade, *overfitting* e, inclusão de variáveis que são apenas *"ruídos"* (Todeschini *et al.* 2004).

1.9.1. Regra QUIK

A regra QUIK (Q^2 *Under Influence of K*) proposta em 1998 (Todeschini *et al.* 1999; Todeschini *et al.* 2004) é um simples critério que permite a rejeição de modelos com alta colineariedade, o que pode ocasionar uma correlação ao acaso (Topliss & Costello 1972)(Topliss & Edwards 1979). A regra QUIK é ba seada no índice de correlação K (Todeschini *et al.* 1999; Todeschini 1997) que mede a correlação total de uma série de variáveis expresso na equação 1.9.1.1.

$$K = \frac{\sum_j \left| \left(\lambda_j / \sum_j \lambda_j \right) - (1/p) \right|}{2(p-1)/p} \qquad \textbf{Equação 1.9.1.1}$$

Onde: j = 1, ...,p e
$0 \leq K \leq 1$
λ_j são os auto-valores obtidos da matriz de correlação da série de dados de **X**(n,p);
n o número de objetos;
p o número de variáveis;

Essa regra é derivada da suposição evidente que a correlação total em uma série é dada pelas variáveis **X** independentes mais a variável dependente **Y** (K_{XY}), e esta deve ser sempre maior que a correlação medida apenas entre as variáveis independentes (K_X).

Desta forma, a regra QUIK determina que apenas modelos com correlação entre as variáveis independentes mais a variável dependente K_{XY} maior que a

correlação entre as variáveis independentes K_X podem ser aceitos (equação 1.9.1.2).

$$K_{XY} - K_X < \delta K \rightarrow \text{rejeite o modelo} \qquad \textbf{Equação 1.9.1.2}$$

Onde: δK é um limite definido (entre 0,01 a 0,05);

O δK pode ser zero se deseja um limite menos rigoroso. De qualquer forma limites menores que zero não são permitidos, ou seja, a diferença entre $K_{XY} - K_X$ não deve ser negativa.

A regra QUIK demonstrou-se eficiente em evitar modelos com multicolinearidade sem poder de predição. De outro lado essa regra não é eficiente para evitar variáveis que são apenas ruídos, desde que estas variáveis não são correlacionadas, portanto apresentando um valor de KX baixo. Nesse caso, mesmo uma baixa correlação entre a variável dependente com as variáveis independentes pode ser considerada significante através desta regra (Todeschini *et al.* 1999; Todeschini *et al.* 2004).

Adicionalmente a regra QUIK, propôs-se calcular o índice de degeneração multivariada D (equação 1.9.1.3). Nesta equação SR, S, e S+ correspondem ao índice de entropia relativa multivariada, ao índice de entropia multivariada, e ao índice de entropia total multivariada (equações 1.9.1.4 a 1.9.1.6). Estes índices medem a variabilidade contida numa série de dados. Nestas equações, n é o número de amostras, p é o número de variáveis independentes, nx é o número de valores iguais presentes na mesma variável, e K é o índice de correlação multivariada definida na equação Equação 1.9.1.1. (Todeschini *et al.* 1999).

$$D = \frac{S^+ - S}{S^+} = 1 - S_R \qquad \textbf{Equação 1.9.1.3}$$

$$S_R = \frac{S}{S^+}$$ **Equação 1.9.1.4**

$$S = \left[1 + (p-1)(1-K)\log_2 n\right] \times \frac{\sum_{j=1}^{p}\left(-\sum_x \frac{n_x}{n}\log_2 \frac{n_x}{n}\right)}{p}$$ **Equação 1.9.1.5**

$$S^+ = p\log_2 n$$ **Equação 1.9.1.6**

1.9.2. Regra do Q^2 Assintótico

Um modelo significantemente estatístico deve ter uma pequena diferença entre o valor do coeficiente de correlação (r^2) e a habilidade preditiva (Q_{cv}^2). De fato diferenças marcantes entre os valores r^2 e Q_{cv}^2 (Todeschini *et al.* 2004) podem ser devidos ao "*overfitting*" (fornecendo altos valores de r^2) ou por algum caso não predito (fornecendo baixos valores de Q_{cv}^2).

Mattioni e Jurs (Mattioni & Jurs 2002) propuseram uma função a qual contabiliza o custo na seleção do modelo, expresso na equação 1.9.2.1:

$$\cos t = rms_T + 0{,}4\left|rms_T - rms_{CV}\right|$$ **Equação 1.9.2.1**

Onde: rms_T é a raiz quadrada da média dos valores dos erros da série de treinamento;
rms_{CV} é a raiz quadrada da média dos valores dos erros da série de teste;
o valor 0.4 é um parâmetro empírico de ponderação da diferença entre a habilidade de ajuste e de predição;

Com o objetivo de se evitar este parâmetro empírico de ajuste, foi proposto um critério como uma regra de exclusão baseado no critério no comportamento

assintótico do Q_{cv}^2. Foi demonstrado que o Q^2 (Miller 1990) é relacionado assintóticamente ao coeficiente de correlação (r^2), desta forma um valor assintótico de Q^2 pode ser calculado pela equação 1.9.2.2. expressa:

$$Q^2{}_{ASYM} = 1 - \left(1 - r^2\right) \times \left(\frac{n}{n - p'}\right)^2 \qquad \textbf{Equação 1.9.2.2}$$

Onde: n é o número de objetos;
p' é o número de parâmetros do modelo;

A regra do Q assintótico é baseada na diferença entre o valor do coeficiente de predição Q^2_{cv} e o valor do Q^2_{ASYM} expresso na equação 1.9.2.3:

$$\text{se } Q^2{}_{cv} - Q^2{}_{ASYM} < \delta Q \longrightarrow \text{rejeite o modelo} \qquad \textbf{Equação 1.9.2.3}$$

Onde: δQ é o valor limite determinado;

Os autores desta regra assumiram que um modelo com um valor de coeficiente de predição Q^2_{cv} menor que uma quantidade δQ do valor do coeficiente de predição assintótico Q^2_{ASYM} deve ser rejeitado. Um limite simples δQ pode ser zero, um limite menos rigoroso pode ser –0,005, um limite mais rigoroso poderia ser 0.005 (Todeschini *et al.* 2004).

1.9.3. Regras Baseadas nas Funções R^P e R^N.

Os objetivos das duas regras apresentadas a seguir, são os de detectar *"overfitting"* devido presença de variáveis no modelo que estão explicando a mesma parte da variação da variável dependente e/ou devido a presença de variáveis no modelo que são apenas "ruídos" (Todeschini *et al.* 2004). Ambas as regras estão baseadas no parâmetro M_j o qual é obtido através da equação 1.9.3.1.

$$M_j = \frac{R_{jy}}{R} - \frac{1}{p}$$ **Equação 1.9.3.1**

Onde: $-\frac{1}{p} \leq M_j \leq \frac{p-1}{p}$;

p é o número de variáveis independentes presentes no modelo;

R_{jy} é o valor do coeficiente de correlação absoluta entre a variável independente j e a variável dependente y;

R é o valor do coeficiente de ajuste do model;

Nesta equação está implícito que se todas as variáveis independentes contribuírem na mesma proporção para explicar a variação contida na variável dependente, esta porção será de 1/p para a correlação múltipla R.

Cada contribuição R_{jy}/R do modelo é comparada com o valor 1/p e tem o objetivo de avaliar a contribuição de uma única variável no modelo. Os valores positivos de M_j são utilizados para o cálculo de R^P expresso na equação 1.9.3.2, e os valores negativos são utilizados para o cálculo de R^N expresso na equação 1.9.3.3.

$$R^P = \prod_{j=1}^{p^+} \left(1 - M_j \times \left(\frac{p}{p-1}\right)\right)$$ **Equação 1.9.3.2**

Onde: $M_j > 0$;

$0 \leq R^P \leq 1$;

R^P é calculado através das variáveis p^+, responsáveis pelas diferenças positivas M_j;

p é o número de variáveis independentes presentes no modelo;

$$R^N = \sum_{j=1}^{p^-} M_j$$ **Equação 1.9.3.3**

Onde: $M_j<0$;

$-1 < R^N \leq 0$;

R^N é calculado através das variáveis p^-, responsável pelas diferenças negativas M_j;

p é o número de variáveis independentes presentes no modelo.

Cada termo do produto de R^P representa o complemento de 1 de cada diferença positiva (M_j>0) escalonada para um valor máximo (p-1)/p. Dessa maneira é obtida uma espécie de penalidade para as variáveis presentes no modelo. O valor é baixo se a variável apresenta uma alta correlação absoluta com a resposta, caindo a zero quando o valor da correlação absoluta (R_{jy}) entre a variável independente e a dependente se iguala ao valor do coeficiente de ajuste do modelo (R) e com um número de parâmetros maior que 1. A função R^P é o produto destas penalidades. Um baixo valor de R^P, ocorre quando uma variável do modelo apresentar um valor de correlação absoluta muito próxima do valor do coeficiente de ajuste do modelo, portanto as outras não são significativas, desde que não contribuem para o aumento da correlação múltipla linear. Neste caso o modelo é demasiado complexo em relação a sua qualidade. Ao contrário se cada variável independente explicar uma fração 1/p do total do coeficiente de ajuste do modelo, o valor de R^P é igual a 1(Todeschini *et al.* 2004).

A regra validação de modelos através da função R^P é definida pela equação 1.9.3.4.

$$R^P < t^p \rightarrow \text{rejeite o modelo} \qquad \textbf{Equação 1.9.3.4}$$

Onde: t^p é um limite pré-definido de 0.01 a 0.1 dependendo dos dados. Um valor sugerido para t^p é 0.05. (Todeschini *et al.* 2004)

Suponha que um modelo apresente um valor de coeficiente de ajuste de R = 0,9, e que haja três variáveis independentes deste modelo, as quais apresentem coeficientes de correlação absoluta com a variável dependente respectivamente de R_{1y} = 0,9, R_{2y} = 0,1 R_{3y} = 0,1, valor da função R^P será zero. Portanto o modelo seria rejeitado.

A função R^N expressa na equação 1.9.3.3 é a soma das diferenças negativas M_j, obtida através das variáveis independentes as quais o valor da razão entre o valor de coeficiente de correlação absoluta e o valor do coeficiente de ajuste do modelo é igual a 1/p. A função R^N considera que um valor baixo de coeficiente de correlação absoluta da variável independente com a variável dependente pode ser um indício de uma variável não significante. A função RN indica o excesso de variáveis não significantes, e pode ser considerado como uma medida de *"overfitting"* devido a presença de variáveis que agregam apenas ruídos ao modelo (Todeschini *et al.* 2004).

Assumindo que em um modelo todas as variáveis apresentem um baixo valor de correlação absoluta com a variável dependente de ε, então o valor Mj de cada uma destas variáveis é expresso na equação 1.9.3.5.

$$\frac{\varepsilon}{R} - \frac{1}{p} = \frac{p\varepsilon - R}{pR} \qquad \textbf{Equação 1.9.3.5}$$

Onde: $\varepsilon << R$
ε é o valor de correlação absoluta entre a variável independente e a variável dependente;
R é o valor do coeficiente de ajuste do modelo;

O valor de ε pode ser alterado pelo usuário dependendo do conhecimento do ruído contido na variável dependente. Além disso, presume-se que não é permitida mais de uma variável que agregue somente ruído no modelo. Portanto o limite tn para a função RN pode ser estimado pela equação 1.9.3.6:

$$t^N(\varepsilon) = \frac{p\varepsilon - R}{pR} \qquad \textbf{Equação 1.9.3.6}$$

Onde: p é o número de variáveis no modelo;
ε é o valor determinado pelo usuário;
R é o valor do coeficiente de ajuste do modelo;

A escolha de aceitar eventualmente uma variável com baixo valor de correlação absoluta com a variável independente se deve a impossibilidade de saber se uma variável é apenas "ruído" ou se "explica" os resíduos do modelo. Por fim, regra validação de modelos através da função RN é definida pela equação 1.9.3.7.

se $R^N < t^N(\varepsilon) \rightarrow$ rejeite o modelo **Equação 1.9.3.7**

Onde: $t^N(\varepsilon)$ é um limite pré-definido.

Ao contrário de R^N que só pode ser negativo, o valor do limite t^N pode ser positivo. Neste caso, qualquer valor diferente de zero no R^N será rejeitado pela regra que independe das correlações entre as variáveis independentes e a variável dependente, e o modelo deve ser rejeitado devido ao baixo valor do coeficiente de ajuste R com relação ao nível de ruído ε escolhido. Isto indica que a correlação entre as variáveis independentes e a variável dependente ocorreu ao acaso (Todeschini *et al.* 2004).

Aumentando os valores de ε, aumentam-se também os valores do limite para a função R^N. Para um modelo com coeficiente de ajuste de 0,6 e os valores de ε iguais a 0,01, a 0,05 e a 0,1, resultam respectivamente em valores de limite para a função de RN de –0,317, de –0,250 e de –0,167 respectivamente. Com o valor de ε igual a 0, o valor limite para a função RN fica limitado a 1/p. Por exemplo: para um modelo com três variáveis, o valor de limite para a função de RN é de –0,333 (Todeschini *et al.* 2004). Alguns exemplos do comportamento das funções R^P e R^N são mostrados na tabela 1.9.3.1.

Tabela 1.9.3.1. Valores das funções R^P e R^N para alguns modelos teóricos com três variáveis independentes.

ID[a]	R_{1Y}[b]	R_{2Y}[c]	R_{3Y}[d]	R[e]	R^P[f]	R^N[g]	Modelo Aceito
1	0,90	0,90	0,90	0,90	**0**[h]	0	**Não**
2	0,90	0,10	0,10	0,90	**0**[h]	**-0,444**[i]	**Não**
3	0,89	0,50	0,10	0,90	**0,011**[h]	-0,222	**Não**
4	0,80	0,80	0,10	0,90	**0,028**[h]	-0,222	**Não**
5	0,80	0,70	0,10	0,90	0,056	-0,222	Sim
6	0,80	0,20	0,10	0,90	0,167	**-0,333**[i]	**Não**
7	0,60	0,40	0,10	0,90	0,417	-0,222	Sim
8	0,60	0,40	0	0,90	0,417	**-0,333**[i]	**Não**
9	0,50	0,30	0,10	0,90	0,667	-0,222	Sim
10	0,40	0,40	0,10	0,90	0,694	-0,222	Sim
11	0,30	0,30	0,30	0,90	1	0	Sim

[a] Número de identificação do modelo;
[b] Valor da correlação absoluta entre a variável independente 1 e a variável dependente;
[c] Valor da correlação absoluta entre a variável independente 2 e a variável dependente;
[d] Valor da correlação absoluta entre a variável independente 3 e a variável dependente;
[e] Valor do coeficiente de ajuste do modelo;
[f] Valor da função RP obtido;
[g] Valor da função RN obtido;
[h] Valores em negrito por serem menores do valor do limite de 0,05 estabelecido para a função R^P;
[i] Valores em negrito serem menores do valor do limite de -0,261 (ε = 0,01) estabelecido para a função R^N;

2. Objetivos

- Verificar e quantificar a relação existente do número de oxidação dos compostos pertencentes aos SLs nas tribos da família Asteraceae com os descritores obtidos das estruturas em 3D.
- Verificar se estes descritores contribuem para a diferenciação das tribos segundo a classificação de Bremer (figura 1.2.4), mais especificamente no caso da tribo Heliantheae; a fim de separar os principais ramos de subtribos segundo Stuessy, baseando-se no número de cromossomos e na morfologia das plantas (figura 1.2.7).
- Estudar as relações entre estrutura química e citotoxicidade, a fim de se delinear os requerimentos estruturais para esta atividade biológica e prever o potencial citotóxico de SLs.

3. METODOLOGIAS

3.1. Obtenção e Cadastro das Estruturas dos Sesquiterpenos Lactonizados e Respectivas Ocorrências Botânicas

Foram adicionados ao SISTEMATX os dados da estrutura da molécula investigada e respectiva ocorrência botânica, a partir da revisão bibliográfica (Seaman 1982). No SISTEMATX, deve-se associar a molécula a sua classe (sesquiterpeno lactonizado) e também ao respectivo esqueleto. Para indicar em quais espécies foi isolado o composto, primeiramente deve-se cadastrar a família Asteraceae (como neste trabalho nos interessa um nível hierárquico mais baixo que família), as tribos associados a esta, os respectivos gêneros e por fim as espécies associadas. O cadastro botânico é feito no módulo "**Dados Botânicos"** do SISTEMATX (figura 3.1.1). Neste estudo, seguiu-se a classificação de Bremer (Bremer 1996).

Foram adicionados ao SISTEMATX os dados da estrutura da molécula investigada respectiva ocorrência botânica, a partir da revisão bibliográfica (Seaman, 1982). No SISTEMATX, deve-se associar a molécula a sua classe (sesquiterpeno lactonizado) e também ao respectivo esqueleto.

O programa então automaticamente calcula, para cada composto, os valores do número de oxidação (NOX) de acordo com as regras de (Hendrickson *et al.* 1970), como foi descrito no **item 1.2.1**.

Para incluir, alterar ou excluir famílias, tribos e subtribos foram utilizados os seguintes módulos, respectivamente: "**Dados Botânicos Família", "Dados Botânicos Tribo"** e **"Dados Botânicos Subribo" ,** que formam a figura 3.1.1. Os botões presentes são padrões para todas as telas e são detalhados na tabela 3.1.1.

Para incluir, alterar ou excluir famílias foi utilizado o seguinte o módulo: “**Dados Botânicos Família”** figura 3.1.1. Os botões presentes são padrões para todas as telas e são detalhados na tabela 3.1.1.

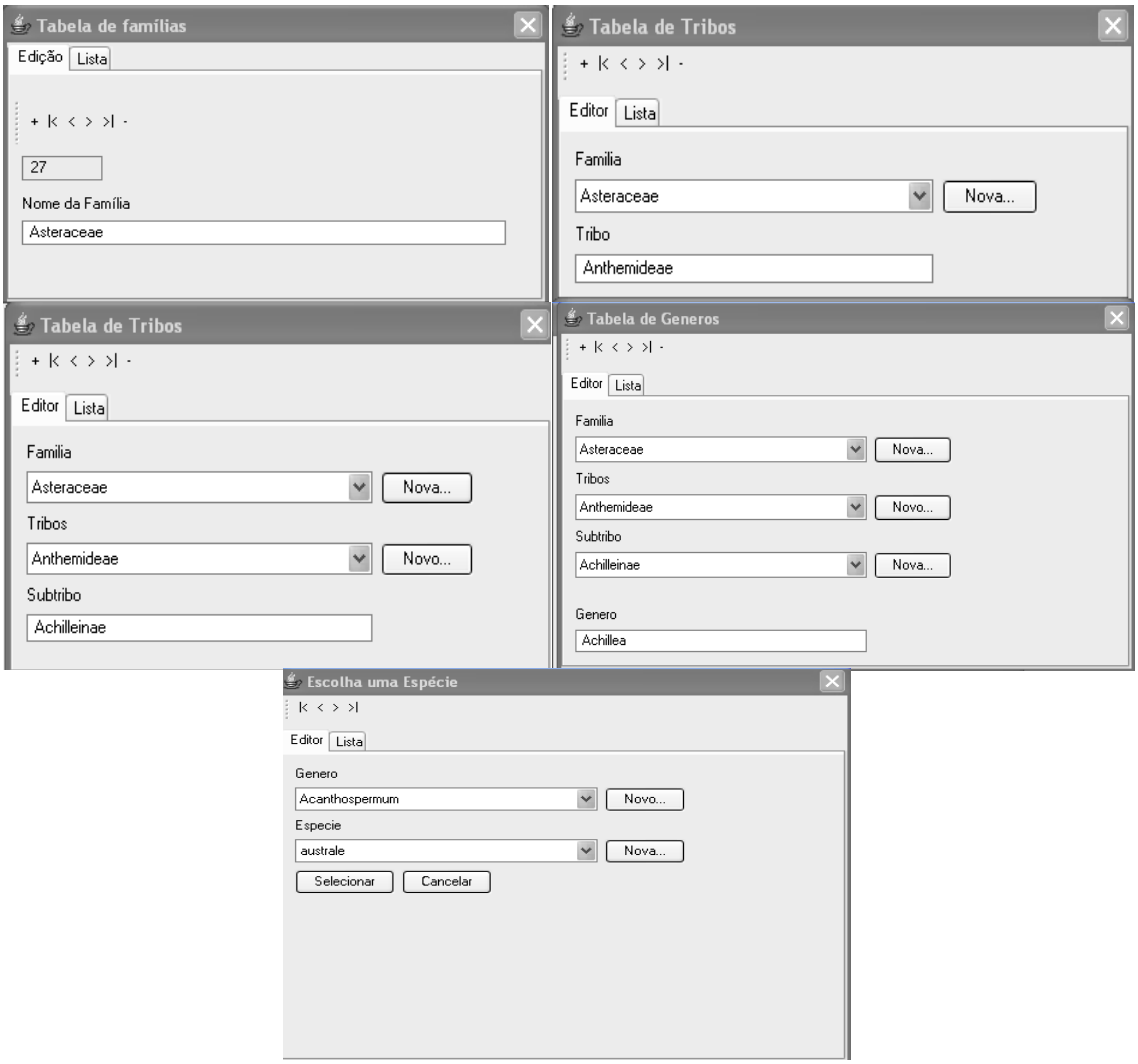


Figura 3.1.1. Telas dos módulos de cadastro botânico do SISTEMATX. A ordem de escolha deve ser Família, Tribo, Subtribo, Gênero, Espécie.

Tabela 3.1.1. Os botões e suas funções nos módulos de inserção de dados botânicos.

Botão	Função
+	Inserir ou salvar um registro
|<	Vai para o primeiro registro
<	Vai para o registro anterior
>	Vai para o próximo registro
>|	Vai para o último registro
-	Exclui um registro

Além desses botões, duas abas independentes podem ser observadas na tela de famílias, uma para “Edição” e outra para “Lista”. Na aba de “Edição” é possível inserir, alterar ou excluir uma família. Na aba de “Lista”, por sua vez, é possível ver as famílias já cadastradas e clicando-se em cima de uma delas, ela automaticamente vai para a aba de edição. A aba de lista é útil para encontrar determinada família.

Para incluir, alterar ou excluir Tribos foi utilizado o seguinte módulo: “Dados Botânicos Tribo”. Como pode ser visto na figura 3.1.1, a tela é muito semelhante à do módulo: “**Dados Botânicos Família”**.

Caso esteja cadastrando uma tribo cuja família ainda não tenha sido cadastrada, pode-se cadastrar diretamente a partir desta tela, basta clicar no botão “Nova” e o módulo de cadastramento de família será aberta. Após o cadastramento da nova família, ela aparecerá na lista. Os outros botões possuem funções idênticas aos botões da tela de família e “Dados Botânicos Família”. A exclusão, alteração e inserção são feitas da mesma forma que para famílias.

Para incluir, alterar ou excluir Subtribos foi utilizado o seguinte módulo: **“Dados Botânicos Subtribo”** (figura 3.1.1). As operações neste módulo são idênticas ao **“Dados Botânicos Tribo”** não se pode mudar a tribo ou família a qual uma subtribo pertence, mas pode-se trocar o nome da subtribo. A razão pela qual não é permitido trocar a Tribo a qual uma subtribo pertence ou então a sua família é manter o banco íntegro.

Da mesma forma para cadastrar um gênero, foi utilizado o módulo **“Dados Botânicos Gênero”** e para cadastrar espécies selecione **“Dados Botânicos Espécie”**. Ambos os módulos apresentam os mesmos recursos dos módulos descritos anteriormente

Para cadastrar classes e esqueletos de substâncias, acesse respectivamente, os módulos **“Dados Substâncias Classe”** e **“Dados Substâncias Esqueleto”**. A operação nestes módulos segue o mesmo padrão dos módulos **“Dados Botânicos”**. Na figura 3.1.2 nota-se que as telas destes módulos são semelhantes.

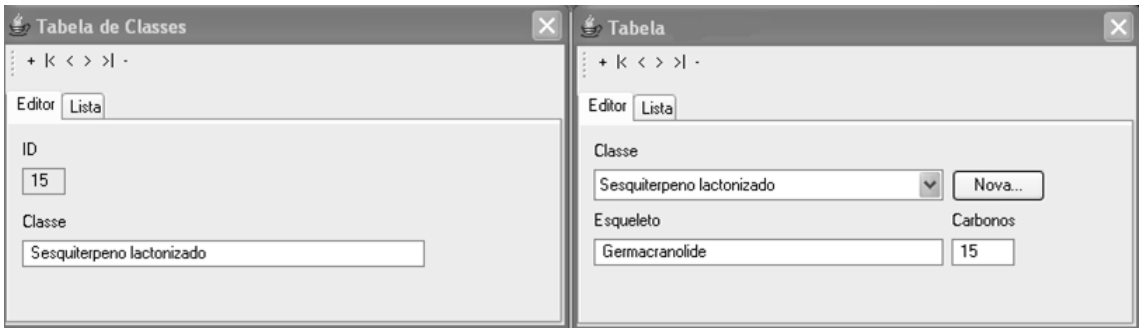


Figura 3.1.2. Telas dos módulos de cadastro de classes e esqueletos no SISTEMATX. A ordem de escolha deve ser Classe, Esqueleto.

Para se desenhar uma molécula, utilizamos o módulo “Dados Substâncias” (figura 3.1.3), o qual associamos a uma classe, um esqueleto, ambos previamente cadastrados, e por fim um nome. Este módulo carrega e exibe diversos dados. O cadastramento e gerenciamento de todos os dados do SISTEMATX podem ser feitos a partir deste módulo.

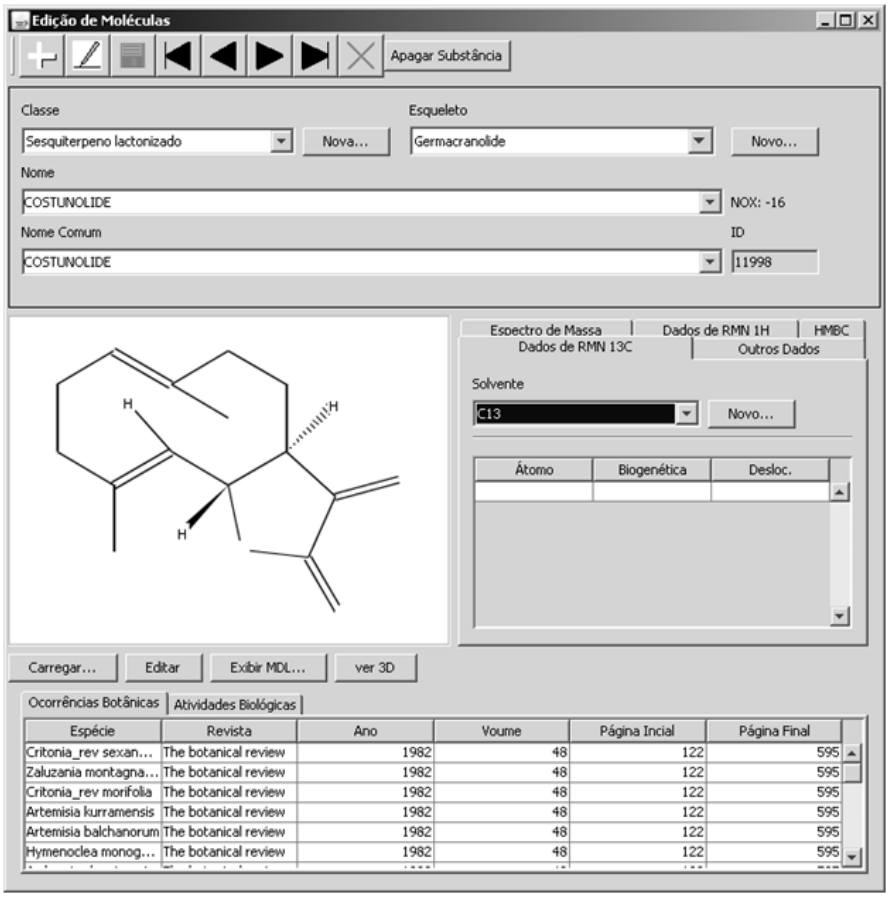


Figura 3.1.3. Tela do módulo de cadastro de substâncias no SISTEMATX. Neste módulo podemos associar diversas propriedades.

No "**Dados Substâncias**" (figura 3.1.3) pode-se visualizar classe, esqueleto, nomes das substâncias, o conjunto dos dados físicos químicos e a de ocorrências botânicas, diversas atividades e o número de oxidação que é calculado instantaneamente após desenhar a molécula. As caixas de listas são utilizadas para selecionar a classe, o esqueleto e até mesmo a substância a ser pesquisada/editada.

Para buscar uma substância basta clicar sobre o nome de uma substância ou sobre o seu nome trivial. O computador irá selecionar a substância correspondente. É possível fazer uma busca por semelhança estrutural, ou seja, após desenhar uma molécula, o sistema informa se a mesma já foi cadastrada, informando o nome, a classe e o esqueleto da mesma (figura 3.1.4).

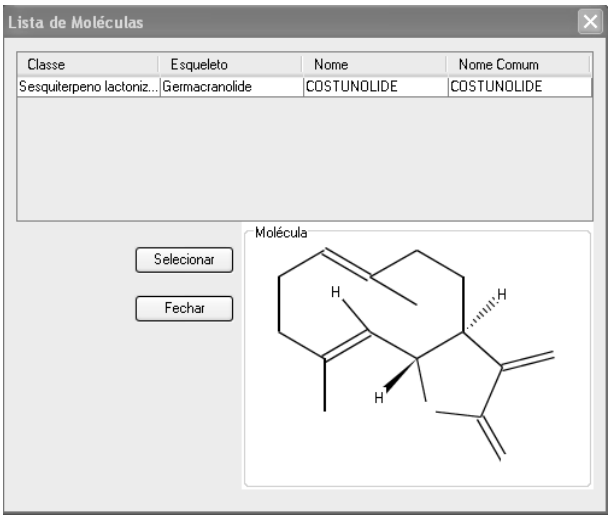


Figura 3.1.4. Tela que informa se uma estrutura já foi cadastrada no SISTEMATX, informando a classe, o esqueleto e o seu respectivo nome.

Os botões do navegador (quatro botões), que estão detalhados na tabela 3.1.2, são utilizados para visualizar as diversas estruturas.

Tabela 3.1.2. Os botões e suas funções nos módulos de inserção de substâncias.

Botão	Função
⏮	Primeira substância
◀	Substância Anterior
▶	Próxima substância
⏭	Última Substância

3.2. Obtenção de Estruturas em Três Dimensões dos Sesquiterpenos Lactonizados e Exportação dos Dados Botânicos

As coordenadas 3D dos SLs foram geradas através do programa SISTEMATX, a partir de dados de constituição 2D das moléculas desenhadas diretamente no sistema, com o módulo **"Exportar Dados Botânicos"** (figura 3.2.1), utilizando o "software" CORINA 3.2 (Sadowski & Gasteiger 1993; Schonberger *et al.*

2000). Foram selecionados todos os sequirterpenos lactonizados da família Asteraceae cadastrados. As moléculas são salvas em arquivos formato MDL (.mol).

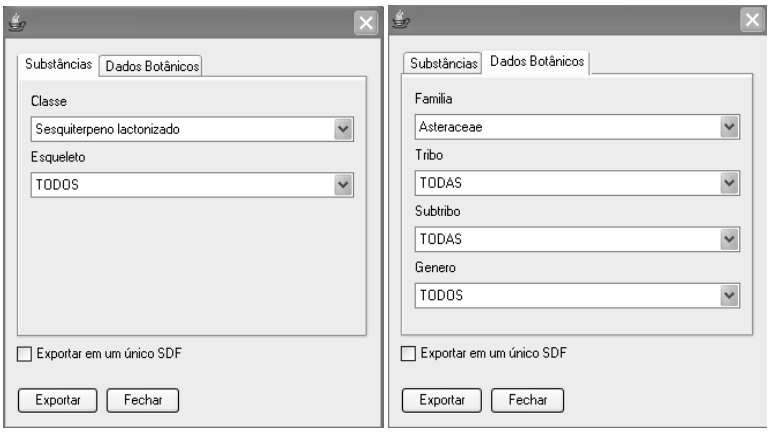


Figura 3.2.1. Módulo de exportação das estruturas das moléculas em 3D (em três dimensões). Podem-se selecionar as estruturas exportadas por classe e/ou esqueleto e as ocorrências por família, tribo, subtribo, gênero.

O programa CORINA (COoRdINAtes) (Schonberger *et al.* 2000), é uma ferramenta que utiliza linhas de comando, que automaticamente gera coordenadas no espaço em três dimensões a partir da molécula expressa em tabela de conectividade, como por exemplo arquivos no formato MDL (com extensão .mol) ou por uma representação linear como o código SMILES (Livingston 1995). Este programa combina fragmentos com comprimentos de ligações ângulos padrões e utilizando apropriados ângulos diedros. Em sistemas de anéis são considerados apenas os ângulos de torções que permitem o fechamento apropriado do anel. No CORINA, também as interações entre os átomos não ligados são minimizadas.

No momento que as estruturas das moléculas são exportadas, também é exportado um arquivo com os dados botânico. Este arquivo é gerado no formato ASCII que pode ser aberto no Excel (tabela 3.2.1).

Tabela 3.2.1. Dados extraído do SISTEMATX a partir do módulo “Exportar Dados Botânicos”. São gerados para cada molécula: o número identificador, sua respectiva classe, esqueleto, número de oxidação, a(s) espécie(s) a(s) qual(is) a molécula foi isolada, e os respectivos gênero , subtribo, tribo e família.

ID	CLASSE	ESQUELETO	NOX	FAMILIA	TRIBO	SUBTRIBO	GENERO	ESPECIE
12004	Sesquiterpeno lactonizado	Germacranolide	-14	Asteraceae	Eupatorieae	Eupatoriinae	Eupatorium	cannabinum
12004	Sesquiterpeno lactonizado	Germacranolide	-14	Asteraceae	Eupatorieae	Eupatoriinae	Eupatorium	formosanum
12003	Sesquiterpeno lactonizado	Germacranolide	-14	Asteraceae	Anthemideae	Artemisiinae	Artemisia	balchanorum
12002	Sesquiterpeno lactonizado	Germacranolide	-16	Asteraceae	Heliantheae	Helianthinae	Tithonia	rotundifolia
12002	Sesquiterpeno lactonizado	Germacranolide	-16	Asteraceae	Heliantheae	Helianthinae	Helianthus	pumilus
12001	Sesquiterpeno lactonizado	Germacranolide	-14	Asteraceae	Eupatorieae	Eupatoriinae	Eupatorium	cuneifolium
12001	Sesquiterpeno lactonizado	Germacranolide	-14	Asteraceae	Eupatorieae	Eupatoriinae	Eupatorium	semiserratum
12000	Sesquiterpeno lactonizado	Germacranolide	-14	Asteraceae	Eupatorieae	Eupatoriinae	Eupatorium	mikanioides
12000	Sesquiterpeno lactonizado	Germacranolide	-14	Asteraceae	Eupatorieae	Eupatoriinae	Eupatorium	semiserratum
12000	Sesquiterpeno lactonizado	Germacranolide	-14	Asteraceae	Heliantheae	Helianthinae	Helianthus	decapetalus
12000	Sesquiterpeno lactonizado	Germacranolide	-14	Asteraceae	Heliantheae	Helianthinae	Helianthus	mollis
11998	Sesquiterpeno lactonizado	Germacranolide	-16	Asteraceae	Eupatorieae	Critoniinae	Critonia	sexangularis

As estruturas das moléculas em três dimensões geradas pelo programa CORINA tiveram suas geometrias otimizadas com o emprego do programa Hyperchem Professional 6.03 (Hyperchem 2001). Inicialmente foi empregado o método de mecânica molecular MM+ (Hocquet & Langgård 1998; Leach 2001).

A mecânica molecular descreve as moléculas como um conjunto de “átomos ligados”, e é baseada num modelo de interações de processos como: estiramento de ligações, abertura e fechamento de ângulos e rotações sobre ligações. Também são consideradas as interações eletrostáticas e de volume de Van der Walls entre átomos não ligados. Diversos campos de força de mecânica molecular foram desenvolvidos tais MM2, MM3, Amber, entre outros (Leach 2001). O campo de força MM+ foi desenvolvido principalmente para ser utilizado em moléculas orgânicas e é uma extensão do campo de força MM2 desenvolvido por Allinger e colaboradores (Allinger *et al.* 1977). Os parâmetros implementados no campo de força MM+

superam o problema de falta de alguns parâmetros no MM2 (Hocquet & Langgård, 1998).

A seguir, as moléculas foram submetidas à otimização de geometria usando o método quântico semi-empírico AM1 (Austin Model 1) (Dewar *et al*. 1985). Os métodos semi-empíricos utilizam como base os modelos Hartree-Fock e diferem, principalmente, dos métodos “ab initio” por considerarem apenas os elétrons de valência do sistema, exigindo assim menores recursos computacionais. O método semi-empírico AM1 foi desenvolvido para eliminar os problemas do método semi-empírico MNDO, o qual superestimava as repulsões entre átomos separados por uma distância aproximadamente igual a soma dos respectivos raios de van de Walls (Leach 2001).

Tanto para mecânica molecular, como para semi-empírico, foi utilizado o método de minimização de energia conhecido como “gradiente conjugado Polak-Ribiere”. Os métodos de gradiente conjugados não apresentam os comportamentos oscilatórios dos gradientes de primeira ordem “steepest descents” (Leach 2001). Como condição de finalização do processo de otimização foi utilizado o valor de raiz média quadrática do gradiente de 0,1 kcal/mol.

3.3. Obtenção dos Descritores Moleculares

Para a obtenção dos descritores moleculares, foi utilizado o programa DRAGON 5.4. Os arquivos de entrada, que são as coordenadas dos átomos de cada molécula, foram selecionados na opção *“Calculate Descriptors”.* Na opção *“Descriptor Selection”* selecionou-se os seguintes blocos de descritores mencionados detalhadamente no **item 1.6**:

1. Constitucionais (gerados a partir de estruturas das moléculas em uma dimensão)
2. Grupos funcionais (gerados a partir de estruturas das moléculas em uma dimensão)
3. Átomo Centrado (gerados a partir de estruturas das moléculas em uma dimensão)
4. BCUT (gerados a partir de estruturas das moléculas em duas dimensões)
5. Auto-correlação 2D (gerados a partir de estruturas das moléculas em duas dimensões)
6. Topológicos (gerados a partir de estruturas das moléculas em duas dimensões)
7. Geométricos (gerados a partir de estruturas das moléculas em três dimensões)
8. RDF (gerados a partir de estruturas das moléculas em três dimensões)
9. 3D-MoRSE (gerados a partir de estruturas das moléculas em três dimensões)
10. GETAWAY (gerados a partir de estruturas das moléculas em três dimensões)
11. WHIM (gerados a partir de estruturas das moléculas em três dimensões)

A obtenção dos parâmetros é relativamente rápida, ou seja, em torno de 10 minutos obtêm-se os descritores para 1111 moléculas (sesquiterpenos lactonizados), utilizando-se um PC com processador Pentium IV (3.0 GHz) com 1 Gb de memória RAM com sistema operacional Windows XP.

3.4. Pré-tratamento de Dados

Considerando o texto apresentado e discutido anteriormente (**item 1.8**), os critérios de pré-tratamento de dados utilizado para cada bloco de descritores foram:

1. Retirada dos descritores que apresentavam valores iguais na série;

2. Retirada dos descritores que apresentavam apenas um valor diferente na série;

3. Retirada dos descritores que apresentavam correlação maior que 0,99 com outras variáveis, sendo que é retirado o maior número de variáveis intercorrelacionadas permanecendo a variável independente que apresenta maior correlação com a variável dependente;

4. A seguir, os cada bloco de descritores das respectivas moléculas é salvo em um arquivo formato ASCII (texto).

Os valores dos descritores são adicionados ao arquivo de ocorrências botânicas utilizando o programa Excel 2007 e salvo em um arquivo com formato texto. Portanto cada sesquiterpeno lactonizado está associado a uma série de descritores, ao seu número de oxidação e suas ocorrências botânicas em um arquivo texto (tabela 3.4.1). Foram gerados 11 arquivos, cada um correspondendo a um bloco de descritor descrito na seção 3.3. Com os descritores constitucionais, obtemos o número de carbonos presente em cada sesquiterpeno lactonizado, possibilitando calcular o grau de oxidação dividindo o número de oxidação (NOX) pelo número de carbonos como descrito na equação 1.2.1.2 (Gottlieb *et al*. 1996).

Tabela 3.4.1. Parte do arquivo gerado a partir da união dos arquivos de descritores GETAWAY, gerado pelo programa DRAGON 5.4, e de ocorrência botânica gerado pelo programa SISTEMATX. As variáveis ISH, HIC, HGM, H1u, e H2u são descritores gerados pelo programa DRAGON e NOX/nC é o grau de oxidação calculado a partir da divisão do número de oxidação (NOX) pelo número de carbonos (nC).

ID	FAMILIA	TRIBO	SUBTRIBO	GENERO	ESPECIE	ISH	HIC	HGM	H1u	H2u	NOX/nC
13211	Asteraceae	Anthemideae	Achilleinae	Achillea	asplenifolia	0.939	5.243	5.541	2.031	2.64	-0.706
12533	Asteraceae	Anthemideae	Achilleinae	Achillea	atrata	0.973	4.949	6.759	2.02	2.706	-0.800
12263	Asteraceae	Anthemideae	Achilleinae	Achillea	biebersteenii	0.95	5.091	6.32	1.996	2.552	-0.800
12962	Asteraceae	Anthemideae	Achilleinae	Achillea	biebersteenii	0.905	5.097	6.483	2.166	2.721	-0.400
12518	Asteraceae	Anthemideae	Achilleinae	Achillea	cartilaginea	0.893	4.96	6.856	1.949	2.575	-0.800
13211	Asteraceae	Anthemideae	Achilleinae	Achillea	collina	0.939	5.243	5.541	2.031	2.64	-0.706
12518	Asteraceae	Anthemideae	Achilleinae	Achillea	eriophora	0.893	4.96	6.856	1.949	2.575	-0.800
12520	Asteraceae	Anthemideae	Achilleinae	Achillea	lanulosa	0.95	5.006	6.7	1.953	2.523	-0.667
12533	Asteraceae	Anthemideae	Achilleinae	Achillea	lanulosa	0.973	4.949	6.759	2.02	2.706	-0.800
12534	Asteraceae	Anthemideae	Achilleinae	Achillea	lanulosa	0.95	4.992	6.607	2.034	2.663	-0.667
12535	Asteraceae	Anthemideae	Achilleinae	Achillea	lanulosa	0.842	5.156	5.732	2.075	2.828	-0.588
12131	Asteraceae	Anthemideae	Achilleinae	Achillea	millefolium	0.948	5.41	4.741	2.344	3.486	-0.737
12134	Asteraceae	Anthemideae	Achilleinae	Achillea	millefolium	0.905	5.291	5.573	2.225	3.208	-0.824

3.5. Correlação entre o Grau de Oxidação Médio dos Sesquiterpenos Presentes nas Tribos da Família Asteraceae e Descritores Moleculares

Para cada um dos 11 arquivos gerados, com os valores dos descritores do sesquiterpenos lactonizados, suas respectivas ocorrências botânicas (tribo, subtribo, gênero), e o seu grau de oxidação, foram calculadas as médias por tribo dos valores dos descritores e dos graus de oxidação utilizando-se o Excel 2007. Obteve-se 11 arquivos (um para cada bloco de descritores) como o mostrado na tabela 3.5.1.

Tabela 3.5.1. Representação parcial do arquivo gerado a partir da união dos descritores GETAWAY. Para cada tribo foi calculado a média dos valores dos descritores (ITH, ISH, HIC, HGM, H1u, H2u, NOX/nC)[a] e do grau de oxidação dos sesquiterpenos presentes em cada tribo.

TRIBO	ITH	ISH	HIC	HGM	H1u	H2u	NOX/nC
Anthemideae	76.144	0.906	5.106	6.347	2.108	2.806	-0.794
Arctoteae	76.092	0.918	5.087	6.469	2.149	2.807	-0.819
Astereae	66.242	0.925	5.003	6.718	2.075	2.756	-0.983
Cardueae	96.726	0.892	5.308	5.368	2.205	2.880	-0.655
Eupatorieae	109.362	0.861	5.454	4.932	2.382	3.398	-0.593
Gnaphalieae	74.041	0.897	5.100	6.283	2.127	2.839	-0.847
Helenieae	91.161	0.893	5.268	5.683	2.218	3.031	-0.698
Heliantheae	97.851	0.884	5.330	5.425	2.289	3.207	-0.692
Inuleae	72.635	0.896	5.125	6.358	2.129	2.860	-0.934
Lactuceae	93.396	0.899	5.206	5.852	2.181	2.924	-0.620
Liabeae	75.352	0.924	5.039	6.587	2.098	2.626	-0.807
Mutisieae	93.660	0.865	5.336	5.356	2.288	3.266	-0.753
Plucheeae	112.211	0.918	5.477	4.704	2.420	3.564	-0.700
Senecioneae	87.237	0.901	5.293	5.548	2.264	3.290	-0.936
Vernonieae	107.630	0.876	5.396	5.075	2.335	3.324	-0.560

[a] Média dos valores dos descritores GETAWAY e do grau de oxidação para cada tribo.

Para cada um dos 11 arquivos gerados como o demonstrado na tabela 3.5.1, foram geradas equações que correlacionam os valores da média dos descritores com os da média do grau de oxidação dos sesquiterpenos lactonizados por tribo. Neste processo utilizou-se o programa MOBYDIGS v. 1.0 (Talete 2004) que possibilita selecionar as variáveis por algoritmo genético, usando o comando "Variable Subset Selection - Genetic Algorithm (VSS-GA)".

O algoritmo genético baseia-se na evolução de uma população de modelos, ou seja, um conjunto de modelos classificados de acordo com alguma função objetivo, neste caso o Q_{cv}^2. No algoritmo genético do programa MobyDigs v 1.0, cada indivíduo é determinado por um cromossomo que é um vetor binário, onde cada posição (um gene) corresponde a uma variável, ou seja, descritor (1 se for incluído no modelo e 0 caso contrário). Por isso, cada cromossomo representa uma equação definida por um subconjunto de variáveis (descritores).

No programa MobyDigs v1.0, pode-se determinar o tamanho da população, ou seja o número de equações retidas e o número máximo permitido de variáveis (descritores) em um modelo. O número mínimo de variáveis permitido é um. Em uma população, os “crossovers” e mutações são repetidos até um número máximo de iterações, ou o processo é encerrado arbitrariamente.

A inicialização do algoritmo genético é formada pela população aleatória modelos com um número de variáveis (descritores) entre 1 e o número máximo determinado pelo usuário. O número de indivíduos (equações), também é definido pelo usuário e para este trabalho foi escolhido como 100. Neste estudo, o número máximo de variáveis (descritores) permitidas nas equações foi 5.

O valor da função selecionada (Q_{cv}^2), ou seja, utilizada para classificar os indivíduos (equações), é calculado em um processo denominado avaliação. Os modelos (equações) são, então, ordenados no que diz respeito ao valor do coeficiente de predição pelo método de *“full cross-validation”* (Q_{cv}^2) (Leardi *et al*. 1992). Também foi definido o número mínimo de equações, neste estudo 3, que deve ser retidas para cada número de variáveis (de 1 a 5).

Em uma população, os pares de modelos são selecionados com uma probabilidade que é calculada em função da sua qualidade, ou seja, pelo seu valor de Q_{cv}^2. No caso é utilizado o processo de seleção por roleta.

Na seleção por roleta, quanto maior o valor de Q_{cv}^2 da equação (indivíduo), maior a sua chance de ser selecionado. Em uma “**roleta”** são colocados todos os indivíduos da população. O lado de cada seção da roleta é proporcional ao valor de Q_{cv}^2 de cada indivíduo: maior for esse valor, mais larga a seção.

Em seguida, a partir de cada par de modelos selecionados (pais), é gerado um novo modelo, preservando as características comuns destes, misturando-os de

acordo com a probabilidade de “crossover”. Se o “filho” gerado coincide com um dos indivíduos já presentes na própria população, este simplesmente é rejeitado (Leardi *et al*. 1992), caso contrário, a equação é avaliada. Se o valor da função objetivo (Q_{cv}^2) da equação é melhor do que o pior valor presente na população, esta será incluída no lugar correspondente na população à sua classificação, caso contrário, não é considerada.

Cada indivíduo da população pode ter seu gene aleatoriamente mudado, alterando o valor de 0 para 1, de 1 para 0, ou mesmo deixado inalterado. Os indivíduos alterados por mutação são avaliados da mesma forma que os ”filhos” no processo *“crossover”*. Este processo é controlado por probabilidade que é normalmente fixado em valores baixos, permitindo assim poucas mutações.

Depois de certo número de iterações definida pelo usuário, uma nova geração da população pode ser criada aleatoriamente. No programa MobyDigs v 1.0, 50% dos indivíduos da população são recriados de forma aleatória, substituindo os 50% indivíduos da população com os piores valores de Q_{cv}^2. Foi determinado neste trabalho que este processo é realizado a cada 1000 iterações.

As equações geradas são modelos lineares (MLR – regressão linear múltipla) com até 5 variáveis, selecionando-se a equação com maior valor do coeficiente de predição gerado pelo método *full cross-validation (*Q_{cv}^2*)* (equação deste parâmetro estatístico encontra-se na equação 3.5.9 na tabela 3.5.2).

Tabela 3.5.2. Alguns parâmetros estatísticos selecionados para avaliar a validade estatística das correlações/modelos gerados.

Expressão Matemática	Explicação, Incógnitas, Onde:	Equação
$SS_T = \sum_{i=1}^{n} (y_i - \bar{y})^2$	*- SST é a soma total dos quadrados;* *- yi é o valor da variável dependente observado;* - $\bar{y}$ *é o valor médio da variável dependente observado na série;*	Equação. 3.5.1
$SS_R = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$	*- SSR é a soma dos quadrados dos resíduos;* *- yi é o valor da variável dependente observado;* - $\hat{y}_i$ *é o valor calculado da variável dependente através do modelo de regressão;*	Equação. 3.5.2
$PRESS = \sum_{i=1}^{n} (y_i - \hat{y}_i^{*})^2$	*- PRESS é a soma dos quadrados dos erros residuais de predição;* *- yi é o valor observado da variável dependente do composto da série de treinamento, o qual não participou da equação de regressão;* - $\hat{y}_i^{*}$ *é o valor calculado da variável dependente através do modelo de regressão do respectivo composto;*	Equação. 3.5.3
$RMSE = \sqrt{\frac{SS_R}{n-p-1}}$	*- RMSE é a raiz da média quadrática dos erros;* *- SSR é a soma dos quadrados dos resíduos;* *- n é o número de amostras;* *- p é o número de variáveis;*	Equação. 3.5.4
$SEP_{cv} = \sqrt{\frac{PRESS}{n}}$	*- SEPcv é a raiz da média quadrática dos erros de predição do cross-validation;* *- PRESS é a soma dos quadrados doserros residuais de predição;* *- n é o número de amostras;*	Equação. 3.5.5
$SEP = \sqrt{\frac{SS_R^{*}}{n'}}$	- SEP é a raiz da média quadrática dos erros de predição; - SSR* é a soma dos quadrados dos resíduos da série de teste ; - n' é o número de amostras da série de teste;	Equação. 3.5.6
$r^2 = 1 - \frac{\sum_{i=1}^{n} (y_{pi} - \hat{y}_{pi})^2}{\sum_{i=1}^{n} (y_{pi} - \bar{y}_p)^2}$	- r^2 é o coeficiente de correlação entre os y calculados e os y observados; - y_{pi} é o valor calculado da variável dependente do composto através do modelo; - $\hat{y}_{pi}$ é o valor calculado da variável dependente através da equação da reta de ajuste entre os valores observados e os valores calculados de pIC50; - $\bar{y}_p$ é a média dos valores calculados de y pelo modelo;	Equação 3.5.7
$Q^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$	- Q^2 é o coeficiente de predição; - y_i é o valor observado da variável dependente do composto da série de teste; - $\hat{y}_i$ é o valor calculado através do modelo de regressão do respectivo composto; - $\bar{y}$ é a média dos valores observados da variável dependente na série de teste;	Equação 3.5.8
$Q_{cv}^{\ 2} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i^{*})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$	- Q_{cv}^2 é o coeficiente de predição pelo método de *full cross-validation*; - y_i é o valor observado da variável dependente do composto da série de treinamento o qual não participou da equação de regressão; - $\hat{y}_i^{*}$ é o valor da calculado variável dependente através do modelo de regressão do respectivo composto; - $\bar{y}$ é a média dos valores observados da variável dependente ;	Equação 3.5.9

3.6. Uso de Mapas Auto-organizáveis (Kohonen) e Descritores Moleculares na Quimiotaxonomia das Tribos da Família Asteraceae

Cada um dos 11 arquivos (como o mostrado na tabela 3.4.1), correspondendo ao bloco de descritores (citados no **item 3.3**) com a ocorrência botânica, foi utilizado como dados de entrada na rede neural Kohonen. Nesta rede, os descritores moleculares são as variáveis de entrada, ou seja, cada amostra (sesquiterpeno lactonizado) corresponde a um vetor, o qual é constituído pelos valores dos descritores deste composto.

Como foi citado no **item 1.7.3** a rede neural Kohonen utiliza a aprendizagem não supervisionada. Os dados de ocorrência botânica, no caso as tribos, são utilizados no SOM apenas para "rotular" áreas do mapa, não participando do treinamento. Nesta fase apenas os descritores moleculares foram utilizados como dados de entrada.

Para gerar os SOMs foi utilizado o aplicativo SOM toolbox 2.0 (Vesanto *et al*. 2005) para Matlab 6.5. Todas as estruturas dos mapas foram geradas em 2 dimensões, e os neurônios forma organizados de forma retangular, no qual cada neurônio tem 4 vizinhos. A equação 1.7.3.1.2 foi utilizada para medir a semelhança entre o vetor de entrada (composto) e o vetor de ponderação do neurônio, possibilitando encontrar o BMU ("Best Match Unit"). Na determinação da vizinhança foi utilizada a função gaussiana (equação 1.7.3.1.3) e o treinamento foi realizado em lote, como descrito no **item 1.7.3.2**.

Após o treinamento da rede Kohonen, os neurônios do mapa são rotulados pelo maior número de ocorrências, ou seja, se a maioria das ocorrências for de uma tribo (por exemplo: Heliatheae), este neurônio será rotulado como uma "região" de Heliantheae. Todas as ocorrências neste determinado neurônio são considerados

acertos se forem de Heliantheae, caso contrário serão considerados erros. As dimensões do mapa foram determinadas empiricamente, minimizando o erro.

Foram utilizadas nesta análise as tribos com o maior número de ocorrências. Para a tribo Heliantheae as subtribos foram classificadas nos ramos A, B e C, como demonstrado na figura 1.2.7, e utilizados como dados de entrada no SOM.

Na figura 3.6.1, está esquematizado resumidamente todo o processo utilizado para a obtenção das equações de regressão linear por algoritmo genético, correlacionando o grau de oxidação com os descritores moleculares para as tribos da família Asteraceae (Bremer 1996), como também a obtenção dos mapas auto-organizáveis para as tribos e para os ramos da tribo Heliantheae (Stuessy 1977).

Figura 3.6.1. Esquema do procedimento de regressão linear múltipla utilizando algoritmo genético (MLR-GA) correlacionando os valores médios de grau de oxidação das tribos com os dos descritores, e de análise para a obtenção dos mapas auto-organizáveis (Kohonen NN) para as ocorrências das tribos da famíla Asteraceae (Bremer, 1996), e ramos da tribo Heliantheae (Stuessy, 1977).

3.7. Relação entre Estrutura Química e Atividade Biológica de Sesquiterpenos Lactonizados

O estudo da relação entre a estrutura química e a atividade citotóxica não era um dos objetivos iniciais deste trabalho, porém foi realizado este estudo com série de sesquiterpenos lactonizados selecionados da literatura.

3.7.1 Selecão dos dados da literatura.

A série de sesquiterpenos lactonizados investigada **1-37** (figura 3.7.1.1) foi selecionada da literatura (Kupchan *et al*. 1969a; Kupchan *et al*. 1969b; Kupchan *et al*. 1971; Kupchan *et al*. 1973), utilizando-se os seguintes critérios:

1. Os experimentos foram realizados pelo mesmo grupo de pesquisa;
2. Seguiram-se os mesmos protocolos experimentais para cada uma das medidas de atividade biológica

Os autores realizaram ensaios experimentais com a finalidade de comprovar atividade inibitória, destes compostos, *in vitro,* contra células derivadas de carcinoma humano da nasofaringe (células KB).

Os valores de atividade biológica reportados na literatura estão em ED_{50}, ou seja, 50% da dose efetiva para a atividade citotóxica obtida por meio de uma curva dose-resposta (tabela 3.7.1.1). Os valores de ED_{50} estão reportados em µg/mL. Os dados foram convertidos para concentração molar e finalmente calculados os valores de $pED_{50} = -\log ED_{50}$.

Tabela 3.7.1.1. Série de sesquiterpenos lactonizados selecionados da literatura com seu número de identificação, seu respectivo nome original da literatura, esqueleto e valores de atividade biológica. Entre parêntesis está a identificação do composto na literatura o qual foi extraído.

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

[1](Kupchan *et al*. 1969 A) [2](Kupchan *et al*. 1969 B) [3](Kupchan *et al*. 1971) [4](Kupchan *et al*. 1973)

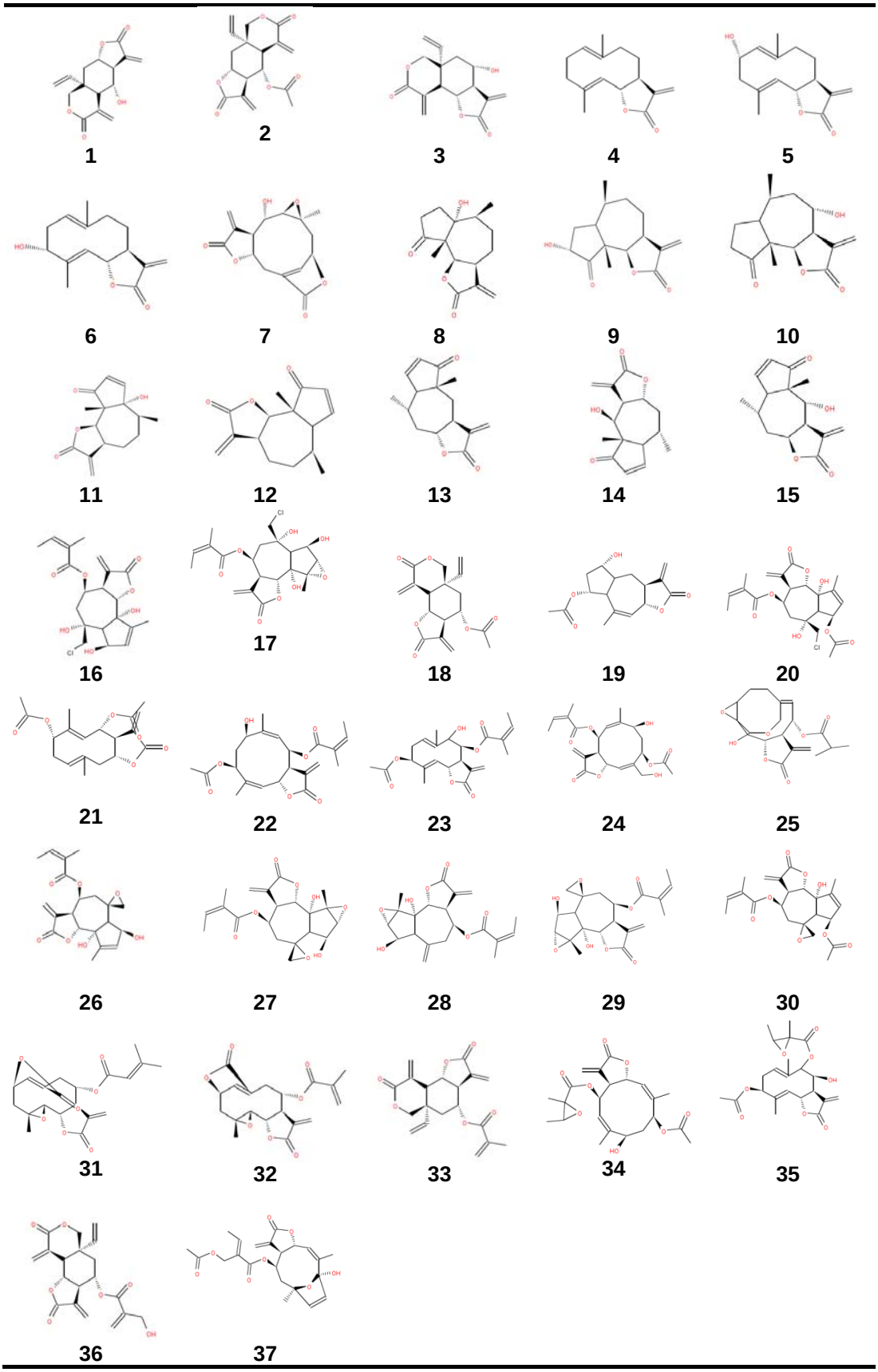


Figura 3.7.1.1. Estruturas dos sesquiterpenos lactonizados, com atividade citotóxica frente a células KB, e respectivos números de identificação.

Utilizando-se o SISTEMATX toda a informação estrutural foi extraída, como explicado nos itens 3.1 e 3.2. Após a geração das estruturas em 3 dimensões pelo programa CORINA, estas foram otimizadas no programa SPARTAN para Windows v. 4.0 (Wavefunction, Inc., Irvine, Calif.). As moléculas foram submetidas à mecânica molecular MMFF, a seguir, o método químico quântico semi-empírico AM1 (Austin Model 1) (Dewar 1981) foi empregado. O valor de gradiente da raiz quadrada média (RMS) de 0,001 kcal/mol foi estabelecido como condição de término. Moléculas com a energia minimizada foram salvas como MDL MolFiles para o cálculo de vários descritores moleculares usando o programa DRAGON Professional version 5.4.

3.7.2. Obtenção dos Descritores Moleculares para a Série de 37 Sesquiterpenos Lactonizados com Atividade Citotóxica

Os grupos de descritores gerados pelo programa DRAGON 5.4 foram (descritores 3D):

1. RDF (150 descritores);
2. 3D-MoRSE (160 descritores);
3. GETAWAY (197 descritores);
4. WHIM (99 descritores);
5. Descritores geométricos (74 descritores).

Estes totalizam 680 descritores calculados baseados na representação em 3 dimensões das moléculas. A seleção das variáveis foi avaliada para cada grupo e para o todo conjunto total. Para cada bloco de descritores, as variáveis constantes foram excluídas. Para os descritores remanescentes, uma análise de correlação em

pares foi feita para a exclusão daqueles altamente correlacionados (r < 0,99). Portanto, o número de descritores DRAGON usado em nossos cálculos foi reduzido a 396:

1. RDF (71 descritores);
2. 3D-MoRSE (104 descritores);
3. GETAWAY (128 descritores);
4. WHIM (66 descritores);
5. Descritores geométricos (22 descritores).

3.7.3. Cálculo dos Modelos de Regressão

O programa MobyDigs 1.0 foi usado para o cálculo dos modelos de regressão através de algoritmo genético. Os compostos foram inicialmente divididos em dois subconjuntos: um conjunto de treinamento composto por 28 substâncias e um grupo de teste externo, composto por 9 substâncias, selecionado de forma aleatória, porém abrangendo toda a faixa de valores de pED_{50} utilizado na série de treinamento. Os modelos para atividade citotóxica foram construídos baseados no grupo de treinamento e validados interna e externamente. O critério de seleção dos modelos foi o maior valor do coeficiente de predição gerado pelo método *full cross-validation (*Q_{cv}^2*)* (equação 3.5.9).

4. Resultados

4.1 Dados Gerados das Estruturas em Três Dimensões e as Respectivas Ocorrências Químicas, Utilizando o SISTEMATX.

Após a geração dos dados botânicos foram obtidos os dados mostrados na tabela 4.1.1. Têm-se 1111 sesquiterpenos lactonizados diferentes correspondendo a 1979 ocorrências químicas para 15 tribos, 63 subtribos, 161 gêneros e 658 espécies da família Asteraceae.

Tabela 4.1.1. Tribos, respectivos acrônimos e os dados botânicos adicionados e utilizados no SISTEMATX.

Tribo	Acrônimo	Subtribo	Gênero	Espécies	Ocorrências	Compostos	Ocorrências /compostos
Anthemideae	Ant	9	15	130	363	154	2,16
Arctoteae	Arc	3	4	7	14	12	1,16
Astereae	Ast	2	2	3	8	8	1,00
Cardueae	Car	3	15	55	118	63	1,87
Eupatorieae	Eup	12	19	62	201	165	1,22
Gnaphalieae	Gna	3	5	5	7	7	1,00
Helenieae	Hel	3	14	73	209	123	1,69
Heliantheae	Hln	9	39	163	612	385	1,59
Inuleae	Inu	1	5	21	104	69	1,51
Lactuceae	Lac	5	6	13	28	17	1,64
Liabeae	Lia	1	4	7	15	12	1,25
Mutisieae	Mut	2	7	11	27	22	1,23
Plucheeae	Plu	1	1	1	1	1	1,00
Senecioneae	Sen	2	9	22	57	37	1,54
Vernonieae	Ver	7	16	85	215	116	1,85
Total		63	161	658	1979	1111	1,78

4.2. Obtenção dos Descritores Moleculares

As 1111 moléculas obtidas foram utilizadas como dados de entrada no programa DRAGON 5.4 no cálculo dos seguintes descritores:

1. Constitucionais: 48 descritores;
2. Grupos funcionais: 154 descritores;

3. Átomo Centrado: 120 descritores;
4. Auto-correlação 2D: 96 descritores;
5. Autovalores Burden: 64 descritores;
6. Topológicos: 119 descritores;
7. Geométricos: 74 descritores;
8. RDF: 150 descritores;
9. 3D-MoRSE: 160 descritores;
10. GETAWAY: 197 descritores;
11. WHIM: 99 descritores.

Totalizando 1281 descritores. Durante o pré-tratamento de dados foi retirado, para cada bloco, os descritores que apresentavam valores iguais na série, os que apresentavam apenas um valor diferente na série e os que apresentavam correlação maior que 0,99 com outras variáveis, restando:

1. Constitucionais: 32 descritores;
2. Grupos funcionais: 35 descritores;
3. Átomo Centrado: 42 descritores;
4. Auto-correlação 2D: 38 descritores;
5. Autovalores Burden (BCUT): 40 descritores;
6. Topológicos: 59 descritores;
7. Geométricos: 42 descritores;
8. RDF: 150 descritores;
9. 3D-MoRSE: 160 descritores;
10. GETAWAY: 188 descritores;
11. WHIM: 99 descritores.

Totalizando 885 descritores restantes. Gerados em 11 arquivos, ou seja, um arquivo para cada bloco de descritores.

4.3. Correlação entre o Grau de Oxidação Médio dos Sesquiterpenos Presentes nas Tribos da Família Asteraceae e Descritores Moleculares

Com os valores dos descritores calculados, cada um dos 11 arquivos gerados foi anexado a sua respectiva ocorrência botânica (tribo, subtribo, gênero), e ao seu grau de oxidação (NOX/nC), ou seja, o valor do número de oxidação dividido pelo número de carbonos presentes em cada molécula. A seguir, para cada tribo foi calculada a média dos valores dos descritores e do grau de oxidação. O arquivo resultante foi utilizado como dado de entrada no MobyDigs 1.0 para selecionar os descritores e gerar as equações lineares múltiplas.

Obtiveram-se diversas equações estatisticamente significativas para cada bloco de descritores que explicasse a variância dos valores do grau de oxidação entre as tribos. Foram selecionadas as que apresentassem os valores mais altos de Q_{cv}^2 com apenas um descritor, pois ao obtermos equações com índices significativos semelhantes, escolhe-se aquela com o menor número de variáveis. As equações, os índices estatísticos, e os descritores são mostrados na tabela 4.3.1 e nas equações 4.3.1 a 4.3.11.

Tabela 4.3.1. Bloco de descritores utilizados, respectivos descritores selecionados nas regressões lineares múltiplas, e seus coeficientes de regressão (r^2) e de predição interna (Q_{cv}^2).

Bloco de Descritores	Descritores Selecionados	r^2	Q_{cv}^2
Constitucionais	AMW	0,981	0,975
Grupos Funcionais	nHAcc	0,856	0,820
Átomo Centrado	O-058	0,740	0,674
Auto Correlação 2D	ATS4m	0,792	0,726
Autovalores Burden	BELv4	0,725	0,647
Topológicos	DELS	0,861	0,822
Geométricos	G(O..O)	0,872	0,832
RDF	RDF045m	0,812	0,764
3D- MoRSE	Mor07u	0,840	0,803
GETAWAY	H5m	0,871	0,839
WHIM	L2v	0,883	0,857

Constitucionais:

Nox/nC = 0,501 (±0,042) AMW - 4,209 (0,290) **Equação 4.3.1**

(n=15; r^2=0,981; s=0,019; F=659,76; Q_{cv}^2=0,975; SPRESS=0,020)

Grupos Funcionais:

Nox/nC = 0,092 (±0,022) nHAcc - 1,189 (±0,109) **Equação 4.3.2**

(n=15; r^2=0,856; s=0,051; F=77,27; Q_{cv}^2=0,820; SPRESS=0,053)

Átomo Centrado:

Nox/nC = 0,221 (±0,078) O-058 - 1,178 (0,153) **Equação 4.3.3**

(n=15; r^2=0,740; s=0,069; F=37,05; Q_{cv}^2=0,674; SPRESS=0,072)

Auto Correlação 2D:

Nox/nC = 0,505 (±0,155) ATS4m - 2,668 (0,587) **Equação 4.3.4**

(n=15; r^2=0,792; s=0,061; F=49,57; Q_{cv}^2=0,726; SPRESS=0,066)

Autovalores Burden:

Nox/nC = 2,104 (±0,776) BELv4 - 4,020 (±1,203) **Equação 4.3.5**

(n=15; r^2=0,725; s=0,070; F=34,30; Q_{cv}^2=0,647; SPRESS=0,075)

Topológicos:

Nox/nC = 0,012 (±0,003) DELS – 1,211 (±0,112) **Equação 4.3.6**

(n=15; r^2=0,861; s=0,050; F=80,19; Q_{cv}^2=0,822; SPRESS=0,053)

Geométricos:

Nox/nC = 0,004 (±0,001) G(O..O) - 0,953 (±0,052) **Equação 4.3.7**

(n=15; r^2=0,872; s=0,048; F=88,73; Q_{cv}^2=0,832; SPRESS=0,051)

RDF:

Nox/nC = 0,034 (±0,010) RDF045m – 1,155 (±0,119) **Equação 4.3.8**

(n=15; r^2=0,812; s=0,058; F=46,26; Q_{cv}^2=0,764; SPRESS=0,060)

3D- MoRSE:

Nox/nC = 0,151 (±0,018) Mor07u - 1,356 (±0,159) **Equação 4.3.9**

(n=15; r^2=0,840; s=0,054; F=67,37; Q_{cv}^2=0,803; SPRESS=0,056)

GETAWAY:

Nox/nC = 1,789 (±0,413) H5m - 1,060 (±0,075) **Equação 4.3.10**

(n=15; r^2=0,871; s=0,049; F=87,50; Q_{cv}^2=0,839; SPRESS=0,050)

WHIM:

Nox/nC = 0,300 (±0,065) L2v - 1,659 (±0,198) **Equação 4.3.11**

(n=15; r^2=0,883; s=0,046; F=98,22; Q_{cv}^2=0,857; SPRESS=0,047)

Na tabela 4.3.2 estão: os valores da média do grau de oxidação de cada tribo; os valores obtidos pela equação 4.3.1, que apresenta o maior valor de coeficiente de predição interna por validação cruzada utilizando um descritor constitucional e os respectivos erros. Na figura 4.3.1 está o gráfico obtido da tabela 4.3.2.

Tabela 4.3.2. Média dos valores de grau de oxidação (NOX/nC) real para 15 tribos da família Asteraceae, os valores de grau de oxidação calculado a partir da equação 4.3.1 e os respectivos erros.

Tribo	NOX/nC Real	NOX/nC Calculado	Erro (Real - Calculado)
Anthemideae	-0,794	-0,798	0,004
Arctoteae	-0,819	-0,817	-0,002
Astereae	-0,983	-0,983	0,000
Cardueae	-0,655	-0,626	-0,029
Eupatorieae	-0,593	-0,603	0,009
Gnaphalieae	-0,847	-0,859	0,012
Helenieae	-0,698	-0,717	0,018
Heliantheae	-0,692	-0,706	0,014
Inuleae	-0,934	-0,942	0,008
Lactuceae	-0,620	-0,605	-0,015
Liabeae	-0,807	-0,772	-0,035
Mutisieae	-0,753	-0,774	0,021
Plucheeae	-0,700	-0,723	0,023
Senecioneae	-0,936	-0,929	-0,007
Vernonieae	-0,560	-0,582	0,022

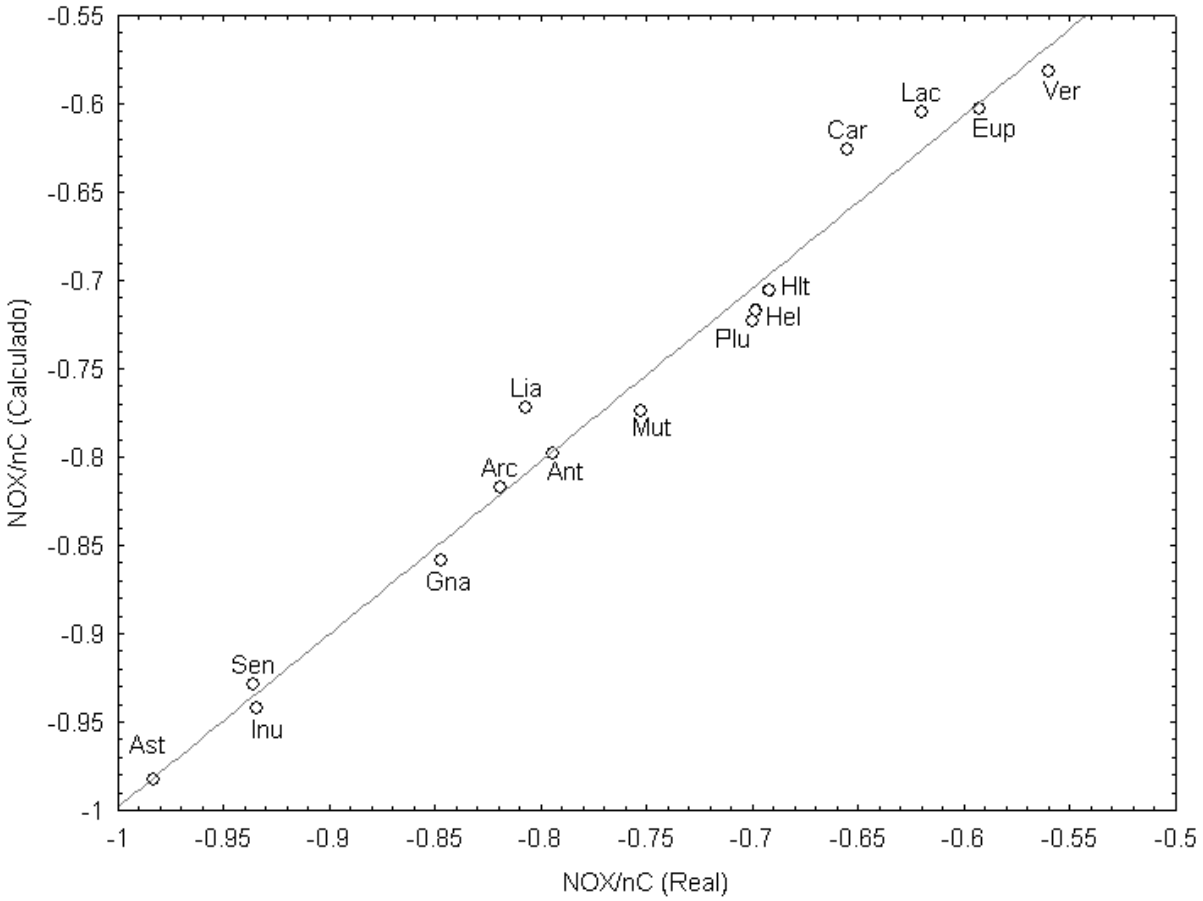


Figura 4.3.1. Gráfico do número do grau de oxidação (NOX/nC) real da média das tribos *versus* o calculado pela equação 4.3.1.

4.4 Mapas Auto-organizáveis (Kohonen) e Descritores Moleculares na Quimiotaxonomia das Tribos da Família Asteraceae

Utilizamos os 11 arquivos de ocorrências botânicas (1 para cada bloco de descritores) para 9 tribos juntamente com os valores dos descritores, como dados de entrada no software SOM toolbox 2.0. Os mapas auto-organizáveis gerados estão apresentados na figura 4.4.1. OS valores de acertos para cada tribo e as dimensões dos mapas são mostrados na tabela 4.4.1. As 9 tribos selecionadas para análise, foram aquelas com os maiores valores de ocorrências botânicas (tabela 4.4.1). Nos mapas representados as ocorrências químicas de determinadas tribos ocupam regiões que são rotuladas pelas seguintes cores:

1. Anthemideae: azul,
2. Cardueae: marrom,
3. Eupatorieae: amarelo,
4. Helenieae: laranja,
5. Heliantheae: vermelho,
6. Inuleae: rosa,
7. Lactuceae: cinza,
8. Senecioneae: azul claro,
9. Vernonieae: verde.

Tabela 4.4.1. Resultados dos Mapas Auto-Organizáveis, e suas respectivas dimensões, com os valores das ocorrências, os números de acertos absolutos e relativos para 9 tribos da família Asteraceae utilizando os blocos de descritores gerados pelo programa DRAGON 5.4.

		Constitucionais - 40x30[a]		Grupos funcionais - 35x35[a]		Átomo Centrado - 40x30[a]		Auto-correlação 2D - 40x30[a]	
Tribo	**Ocorrências**	**Nº de acertos**	**% de acerto**	**Nº de acertos**	**% de acerto**	**Nº de acertos**	**% de acerto**	**Nº de acertos**	**% de acerto**
Anthemideae	363	268	73,8	334	92,0	338	93,1	336	92,6
Cardueae	118	81	68,6	103	87,3	101	85,6	95	80,5
Eupatorieae	201	130	64,7	148	73,6	157	78,1	151	75,1
Helenieae	209	110	52,6	159	76,1	171	81,8	176	84,2
Heliantheae	612	451	73,7	517	84,5	511	83,5	522	85,3
Inuleae	104	29	27,9	32	30,8	46	44,2	61	58,7
Lactuceae	28	18	64,3	20	71,4	21	75,0	14	50,0
Senecioeae	57	38	66,7	50	87,7	53	93,0	47	82,5
Vernonieae	215	147	68,4	174	80,9	175	81,4	178	82,8
Total	1907	1272	66,7	1537	80,6	1573	82,5	1580	82,9

		BCUT - 40x30[a]		Topológicos - 40X30[a]		Geométricos - 40x30[a]		RDF - 40X30[a]	
Tribo	**Ocorrências**	**Nº de acertos**	**% de acerto**	**Nº de acertos**	**% de acerto**	**Nº de acertos**	**% de acerto**	**Nº de acertos**	**% de acerto**
Anthemideae	363	342	94,2	324	89,3	339	93,4	341	93,9
Cardueae	118	105	89,0	99	83,9	96	81,4	97	82,2
Eupatorieae	201	154	76,6	160	79,6	156	77,6	161	80,1
Helenieae	209	183	87,6	175	83,7	180	86,1	180	86,1
Heliantheae	612	513	83,8	497	81,2	487	79,6	529	86,4
Inuleae	104	59	56,7	57	54,8	49	47,1	54	51,9
Lactuceae	28	19	67,9	22	78,6	20	71,4	17	60,7
Senecioeae	57	52	91,2	46	80,7	44	77,2	48	84,2
Vernonieae	215	165	76,7	168	78,1	164	76,3	167	77,7
Total	1907	1592	83,5	1548	81,2	1535	80,5	1594	83,6

		3D-MoRSE-40x30[a]		GETAWAY- 40x35[a]		WHIM – 40x30[a]	
Tribo	**Ocorrências**	**Nº de acertos**	**% de acerto**	**Nº de acertos**	**% de acerto**	**Nº de acertos**	**% de acerto**
Anthemideae	363	290	79,9	341	93,9	339	93,4
Cardueae	118	83	70,3	104	88,1	106	89,8
Eupatorieae	201	128	63,7	168	83,6	149	74,1
Helenieae	209	121	57,9	173	82,8	158	75,6
Heliantheae	612	430	70,3	498	81,4	517	84,5
Inuleae	104	43	41,3	50	48,1	57	54,8
Lactuceae	28	17	60,7	17	60,7	19	67,9
Senecioeae	57	33	57,9	46	80,7	44	77,2
Vernonieae	215	143	66,5	164	76,3	166	77,2
Total	1907	1288	67,5	1561	81,9	1555	81,5

[a] blocos de descritores utilizados e dimensões dos mapas auto-organizáveis.

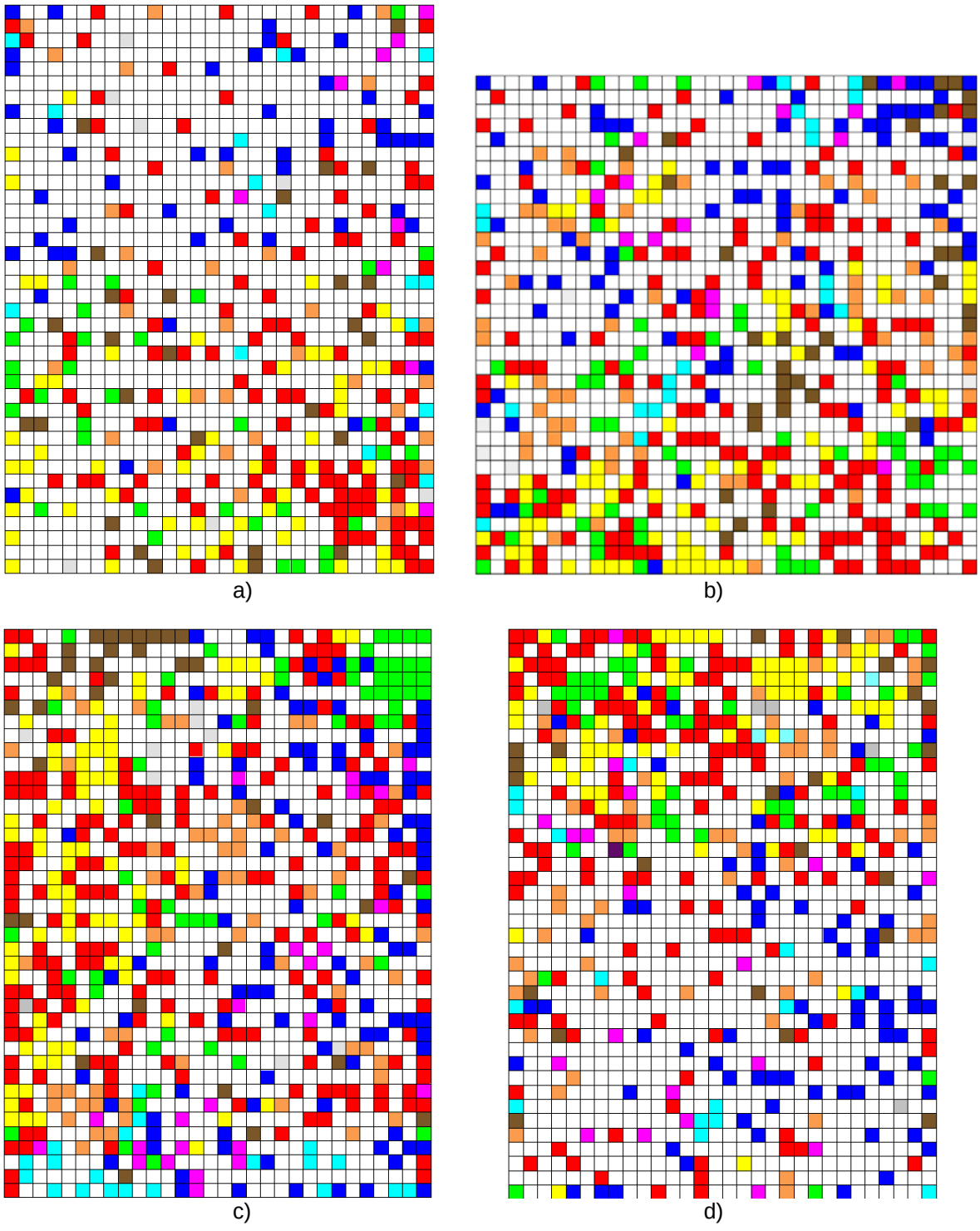


Figura 4.4.1. Mapas Auto-Organizáveis obtidos classificando 9 tribos da família Asteraceae (tabela 4.4.1). Mapas: a) Utilizando o bloco de descritores constitucionais, dimensão de 40 por 30 neurônios; b) Utilizando o bloco de descritores de grupos funcionais, dimensão de 35 por 35 neurônios; c) Utilizando o bloco de descritores de átomo centrando, dimensão de 40 por 30 neurônios; d) Utilizando o bloco de descritores auto-correlação 2D, dimensão de 40 por 30 neurônios. Onde: vermelho: Heliantheae; azul: Anthemideae; amarelo: Eupatorieae; verde: Vernonieae; Rosa Inuleae; Cinza: Lactuceae; marrom: Cardueae; laranja: Heliantheae; Azul claro: Senecioneae.

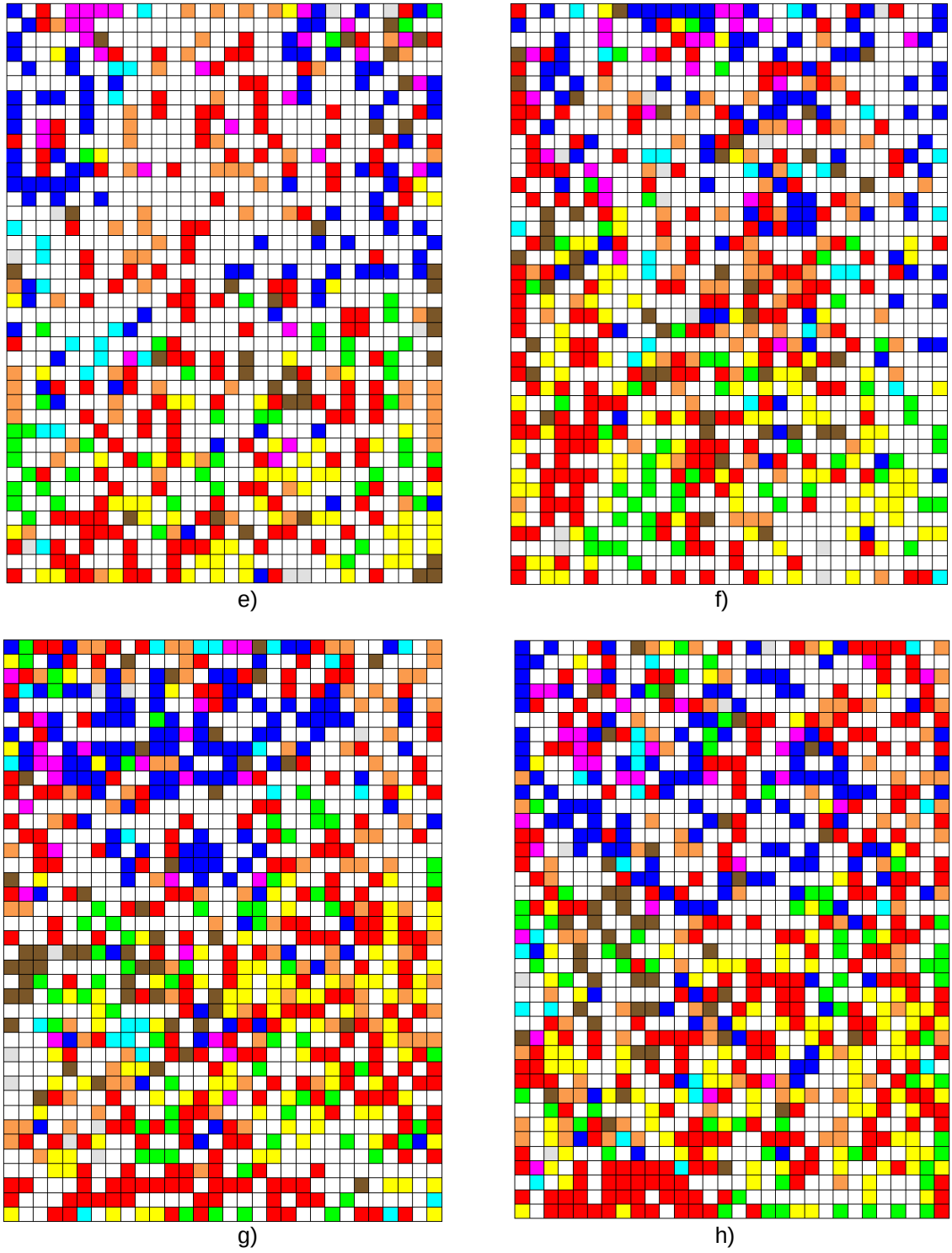
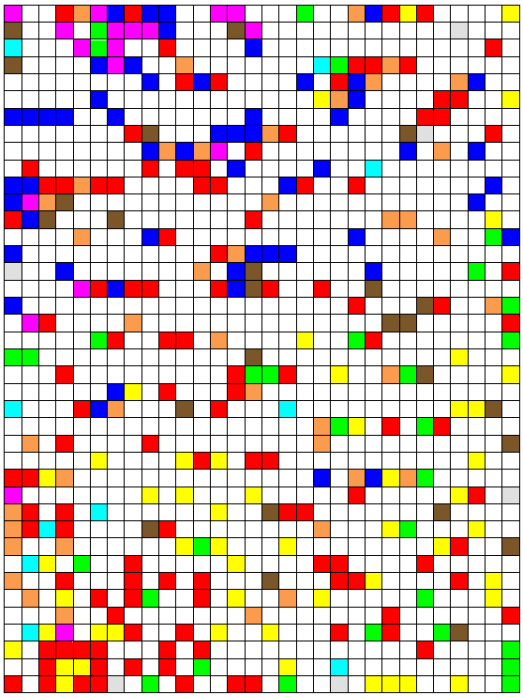
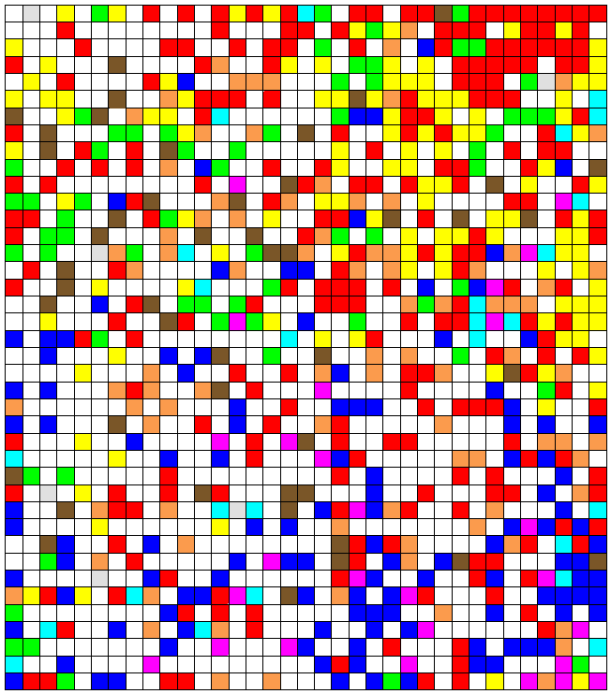


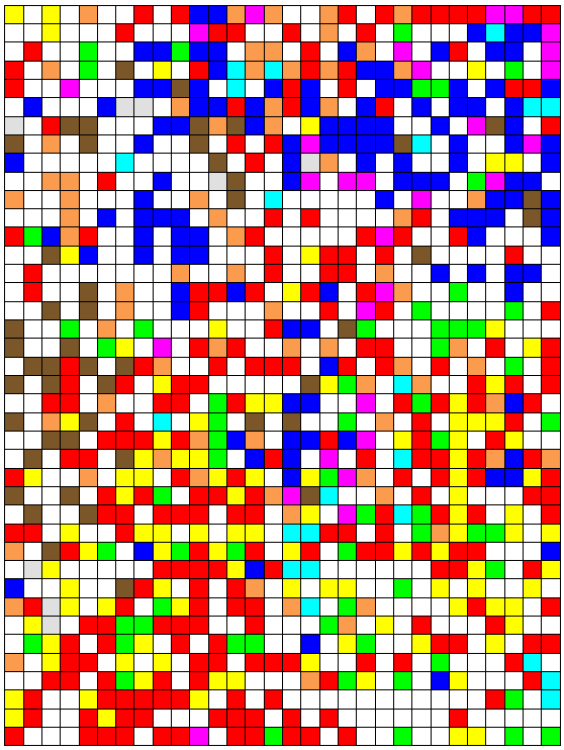
Figura 4.4.1. Continuação Mapas: e) Utilizando o bloco de descritores BCUT, dimensão de 40 por 30 neurônios; f) Utilizando o bloco de descritores topológicos, dimensão de 40 por 30 neurônios; g) Utilizando o bloco de descritores geométricos, dimensão de 40 por 30 neurônios; h) Utilizando o bloco de descritores RDF, dimensão de 40 por 30 neurônios. Onde: vermelho: Heliantheae; azul: Anthemideae; amarelo: Eupatorieae; verde: Vernonieae; Rosa Inuleae; Cinza: Lactuceae; marrom: Cardueae; laranja: Heliantheae; Azul claro: Senecioneae.



i)



j)



k)

Figura 4.4.1. Continuação Mapas: i) Utilizando o bloco de descritores 3D MoRSE, dimensão de 40 por 30 neurônios; j) Utilizando o bloco de descritores GETAWAY, dimensão de 40 por 35 neurônios; K) Utilizando o bloco de descritores WHIM, dimensão de 40 por 30 neurônios. Onde: vermelho: Heliantheae; azul: Anthemideae; amarelo: Eupatorieae; verde: Vernonieae; Rosa Inuleae; Cinza: Lactuceae; marrom: Cardueae; laranja: Heliantheae; Azul claro: Senecioneae.

4.5 Mapas Auto-organizáveis (Kohonen) e Descritores Moleculares na Quimiotaxonomia dos Ramos da Tribo Heliantheae

Utilizamos os 11 arquivos de ocorrências botânicas para os ramos A e C, juntamente com os valores dos descritores, como dados de entrada no software SOM toolbox 2.0. Os mapas auto-organizáveis gerados estão apresentados na figura 4.5.1. Os valores de acertos para cada ramo e as dimensões dos mapas são mostrados na tabela 4.5.1. O ramo B foi excluído da análise por apresentar apenas 4 ocorrências. Nos mapas representados as ocorrências químicas dos ramos ocupam regiões que são rotuladas pelas cores:

- Ramo A: azul
- Ramo C: vermelho

Tabela 4.5.1. Resultados dos Mapas Auto-Organizáveis, suas respectivas dimensões, valores das ocorrências e números de acertos absolutos e relativos para os ramos A e C da tribo Heliantheae (Stuessy, 1977), utilizando os blocos de descritores gerados pelo programa DRAGON 5.4.

		Constitucionais - 13x11[a]		Grupos funcionais - 14x10[a]		Átomo Centrado - 14x10[a]		Auto-correlação 2D - 21x7[a]	
Ramos	**Ocorrências**	**Nº de acertos**	**% de acerto**	**Nº de acertos**	**% de acerto**	**Nº de acertos**	**% de acerto**	**Nº de acertos**	**% de acerto**
A	505	439	86,93	446	88,32	476	94,26	486	96,24
C	253	188	74,31	222	87,75	209	82,61	215	84,98
Total	758	627	82,72	668	88,13	685	90,37	701	92,48
		BCUT - 29x5[a]		**Topológicos - 24X6[a]**		**Geométricos - 13x11[a]**		**RDF - 24X6[a]**	
Tribo	**Ocorrências**	**Nº de acertos**	**% de acerto**	**Nº de acertos**	**% de acerto**	**Nº de acertos**	**% de acerto**	**Nº de acertos**	**% de acerto**
A	505	480	95,05	462	91,49	463	91,68	463	91,68
C	253	221	87,35	211	83,40	180	71,15	205	81,03
Total	758	701	92,48	673	88,79	643	84,83	668	88,13
		3D-MoRSE- 13x11[a]		**GETAWAY- 36x4[a]**		**WHIM – 13x11[a]**			
Tribo	**Ocorrências**	**Nº de acertos**	**% de acerto**	**Nº de acertos**	**% de acerto**	**Nº de acertos**	**% de acerto**		
A	505	395	78,22	449	88,91	443	87,72		
C	253	232	91,70	225	88,93	177	69,96		
Total	758	627	82,72	674	88,92	620	81,79		

[a] blocos de descritores utilizados e dimensões dos mapas auto-organizáveis.

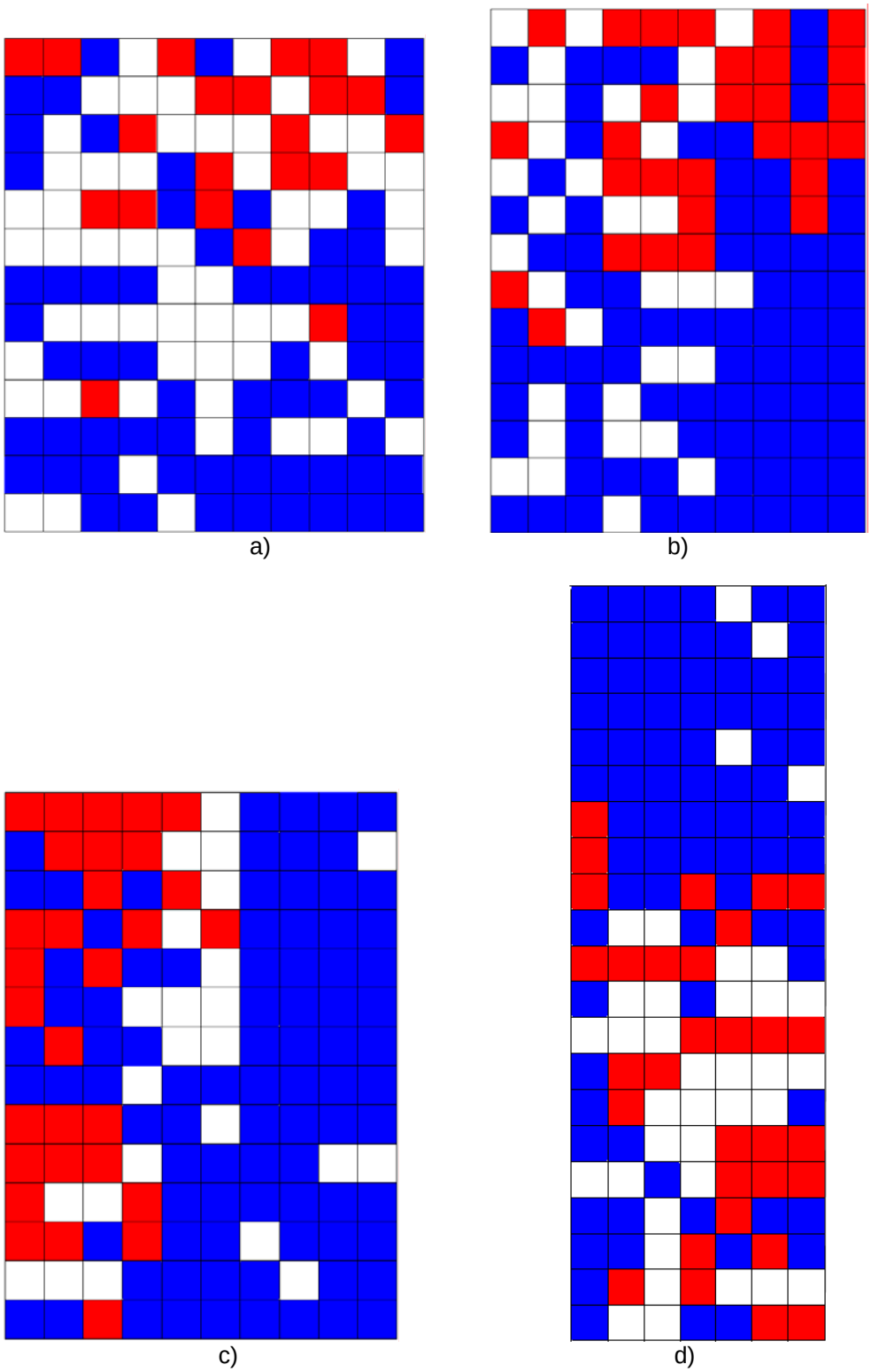


Figura 4.5.1. Mapas Auto-Organizáveis obtidos classificando os ramos A e C da tribo Heliantheae (tabela 4.5.1) segundo Stuessy. Mapas: a) Utilizando o bloco de descritores constitucionais, dimensão de 13 por 11 neurônios; b) Utilizando o bloco de descritores de grupos funcionais, dimensão de 14 por 10 neurônios; c) Utilizando o bloco de descritores de átomo centrando, dimensão de 14 por 10 neurônios; d) Utilizando o bloco de descritores auto-correlação 2D, dimensão de 21 por 7 neurônios. Onde: azul- ramo A; vermelho- ramo C.

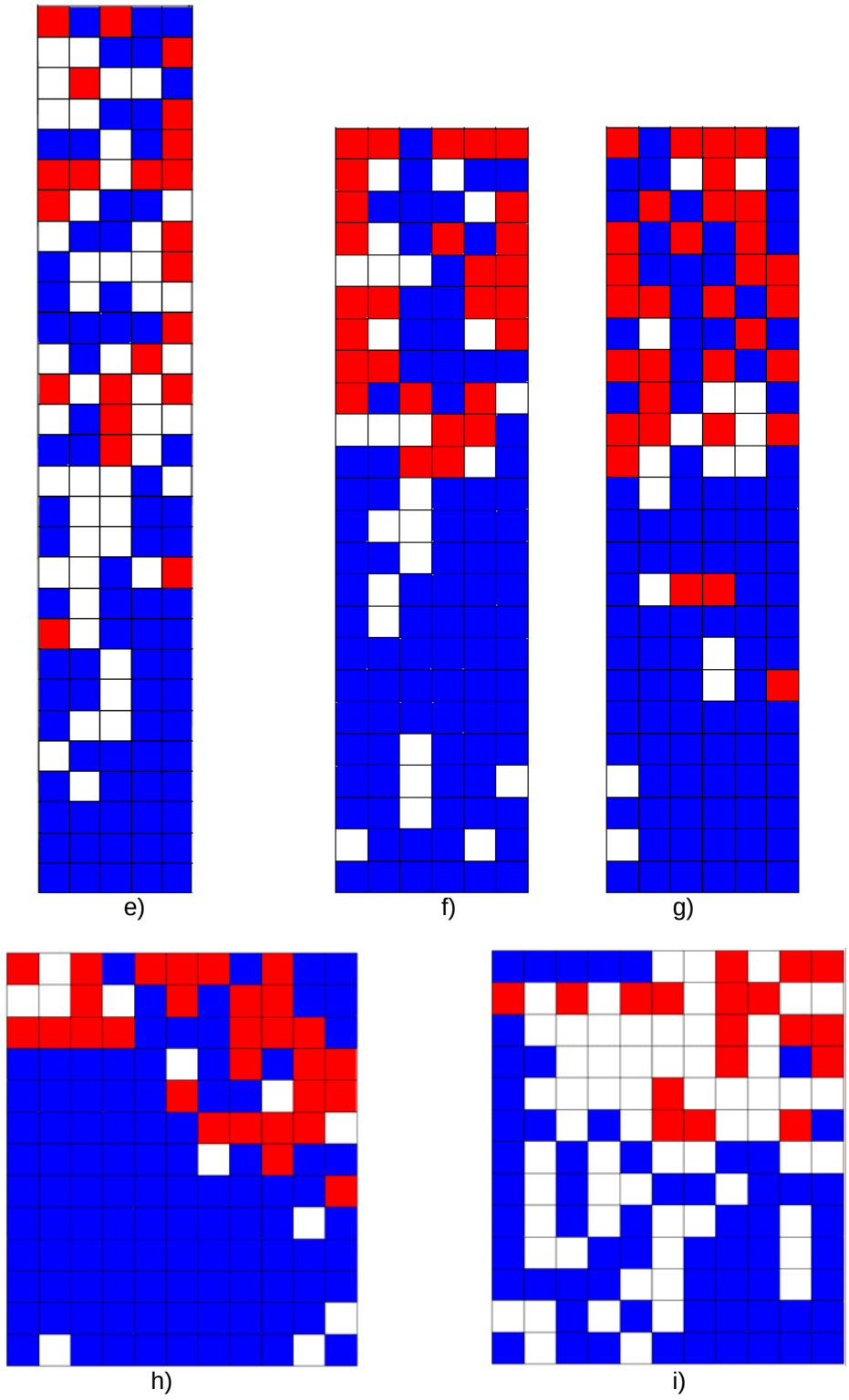
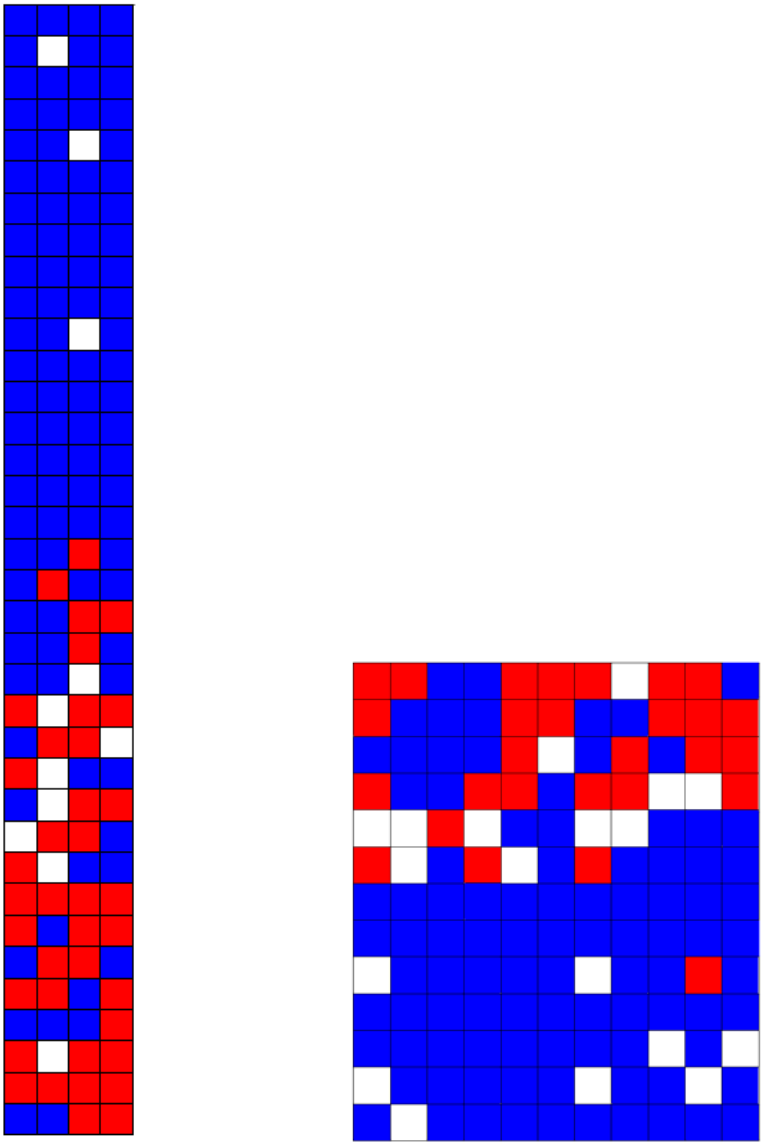


Figura 4.5.1. Continuação. Mapas: e) Utilizando o bloco de descritores BCUT, dimensão de 29 por 5 neurônios; f) Utilizando o bloco de descritores topológicos, dimensão de 24 por 6 neurônios; g) Utilizando o bloco de descritores RDF, dimensão de 24 por 6 neurônios; h) Utilizando o bloco de descritores geométricos, dimensão de 13 por 11 neurônios; i) Utilizando o bloco de descritores 3D-MoRSE, dimensão de 13 por 11 neurônios. Onde: azul- ramo A; vermelho-ramo C.



j) k)

Figura 4.5.1. Continuação. Mapas: j) Utilizando o bloco de descritores GETAWAY, dimensão de 36 por 4 neurônios; k) Utilizando o bloco de descritores WHIM, dimensão de 13 por 11 neurônios. Onde: azul- ramo A; vermelho-ramo C.

4.6. Série de 37 Sesquiterpenos Lactonizados com Atividade Citotóxica

Após a análise de regressão da série de treinamento, foi selecionada a equação 4.6.1, a qual contém os descritores SPAN, G(O…O), Mor15u, Mor13m e $R8e^{+}$. Estes são capazes de explicar 82,6% da variância na atividade citotóxica.

pED_{50} = + 0,484 (±0,232) SPAN - 0,011(±0.005) G(O..O) + 0,791(±0,459) Mor13m + 0,297 (±0,260) Mor15u - 84,459 (±27.104) $R8e^{+}$ + 6.250 (±1,456) **Equação 4.6.1**

(n=28; r^2=0,826; s=0,258; F=21,04; Q^2_{cv}=0,743; $S\text{-}_{PRESS}$=0,314; n_{ext}=9; r^2_{ext}= 0,800; Q^2_{ext} = 0,704)

Na tabela 4.6.1, estão os valores calculados de pED_{50} a partir da equação 4.6.1, e os valores de pED_{50} experimental e os respectivos erros para a série de treinamento.

Tabela 4.6.1. Valores experimentais de pED_{50}, valores calculados através da equação 4.6.1 e seus respectivos erros para as substâncias pertencentes ao grupo de treinamento.

Composto	pED_{50} Experimental	pED_{50} Calculado	Erro (Calculado – 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

A partir dos valores da tabela 4.6.1 foi feito o gráfico dos valores de pED_{50} experimental versus o valor de pED_{50} calculado (figura 4.6.1) e o gráfico dos valores dos erros (pED_{50} calculado – experimental) versus os valores de pED_{50} experimental (figura 4.6.2).

A figura 4.6.1 mostra o ajuste de uma linha reta conforme a distribuição dos pontos usados para a calibração do modelo, esta aproximação linear é validada pela observação da figura 4.6.2, o qual apresenta a distribuição randômica dos pontos.

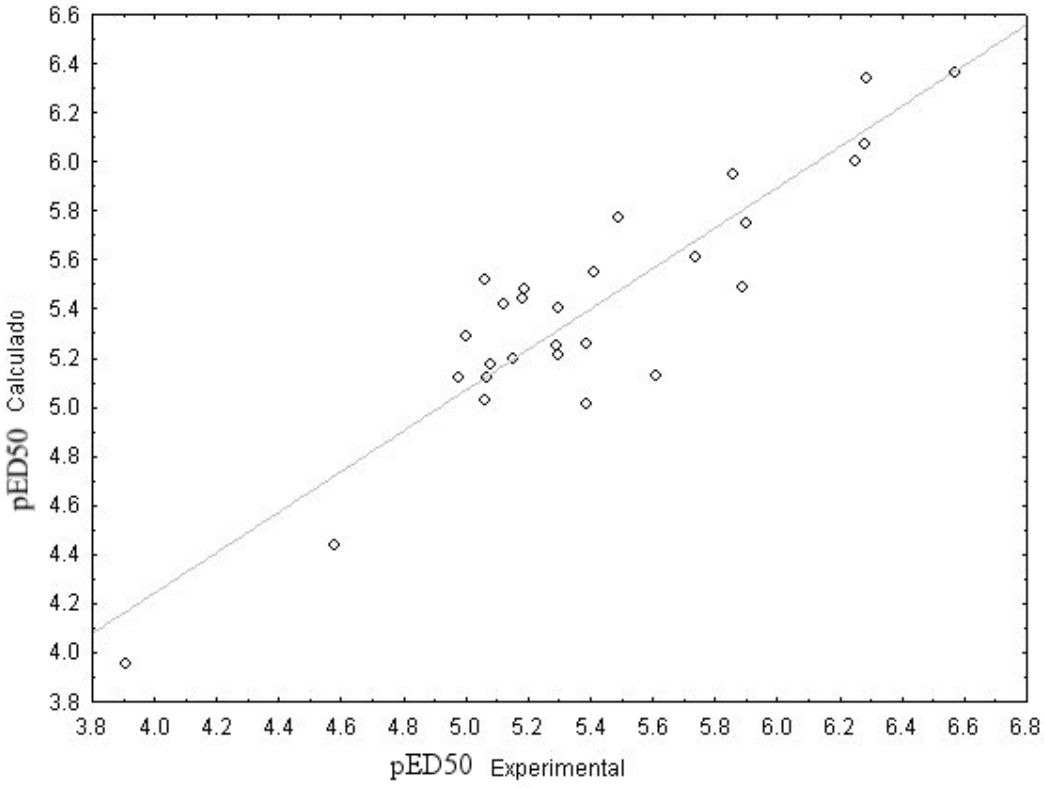


Figura 4.6.1. Gráfico dos valores de atividade experimental (pED_{50}) versus os valores de atividade calculada para a série de treinamento.

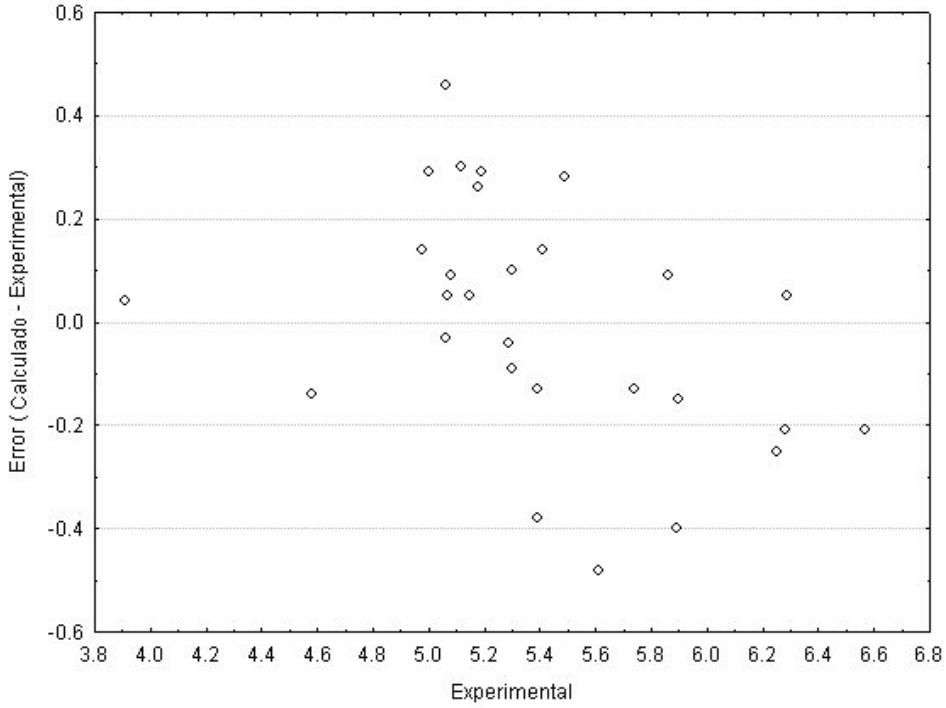


Figura 4.6.2. Gráfico dos valores de atividade experimental (pED_{50}) versus seus respectivos erros (valor calculado – valor experimental) para a série de treinamento.

A tabela 4.6.2 e o respectivo gráfico feito a partir destes valores (figura 4.6.3), mostram os resultados obtidos com a série de validação externa composta de 9 moléculas. Há um ajuste linear considerável e a equação 4.6.1 mostrou-se capaz de diferenciar os compostos mais ativos dos menos ativos.

Tabela 4.6.2. Valores experimentais de pED_{50}, valores previstos pela equação 4.6.1 e seus respectivos erros para as substâncias pertencentes a série de teste.

Composto	pED_{50} Experimental	pED_{50} Previsto	Erros (Previsto – 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

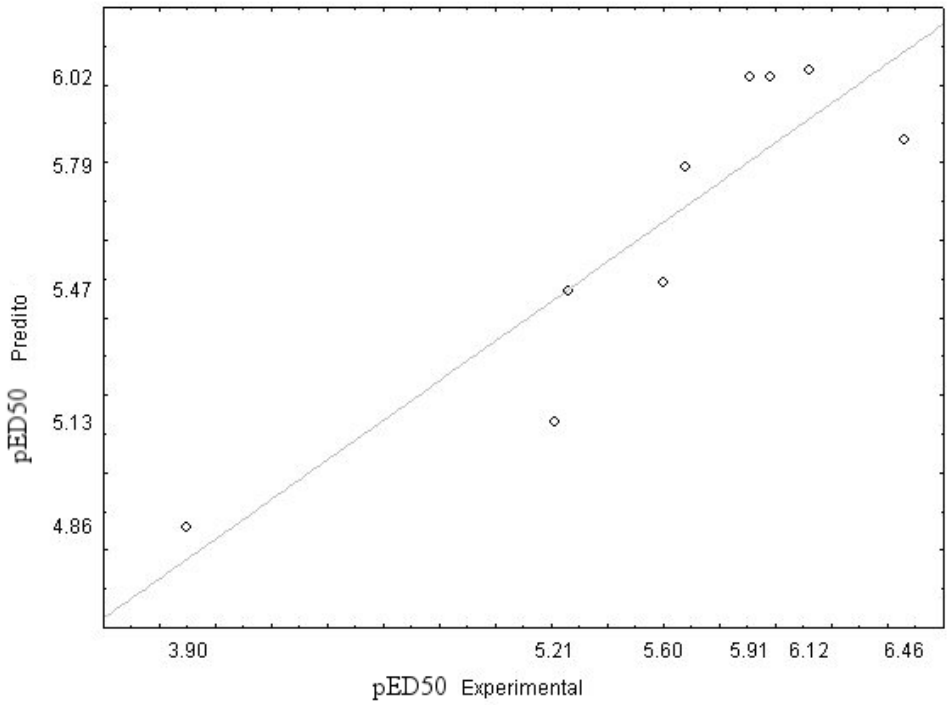


Figura 4.6.3. Gráfico dos valores de atividade experimental (pED_{50}) versus os valores de atividades preditas para a série de teste.

5. DISCUSSÂO

5.1. Dados Gerados das Estruturas em Três Dimensões e as Respectivas Ocorrências Químicas, Utilizando o SISTEMATX

A análise da tabela 4.1.1 verificou que Heliantheae é a tribo em que foi cadastrado o maior número de gêneros, espécies, ocorrência e compostos. Outras tribos com números significativos foram Anthemidae, Eupatorieae, Vernonieae e Helenieae.

5.2. Obtenção dos Descritores Moleculares

Os 11 blocos de descritores obtidos que foram utilizados nas análises de regressão linear múltipla e nos mapas auto-organizáveis, usam a informação das moléculas através de dados constitucionais de massa atômica (descritores constitucionais), fragmentos (grupos funcionais, átomo centrado), topologia (auto-correlação 2D, BCUT, topológicos) e conformacionais (Geométricos, RDF, 3D-MoRSE, GETAWAY, WHIM).

5.3. Correlação entre o Grau de Oxidação Médio dos Sesquiterpenos Presentes nas Tribos da Família Asteraceae e Descritores Moleculares

A análise das equações 4.3.1 a 4.3.11 e da tabela 4.3.1 verificou a presença de índices estatísticos significativos, em todos os blocos de descritores, utilizando-se apenas uma variável. Em adição, a contribuição de todos os descritores selecionados (equações 4.3.1 a 4.3.11) com relação ao

grau de oxidação foi positiva, ou seja, quanto maior o valor do descritor, maior é o valor do grau de oxidação observado.

A equação 4.3.1, que contém um descritor constitucional (AMW), tem os valores de coeficientes estatísticos mais elevados (r^2, Q_{cv}^2 e F). O descritor AMW representa a média do peso molecular, envolvendo os pesos atômicos da molécula. Os sesquiterpenos lactonizados, majoritariamente apresentam apenas átomos de carbono, hidrogênio e oxigênio, portanto SLs com poucos átomos de hidrogênio, mais ligações duplas e com mais átomos de oxigênio, têm valores maiores de AMW e, conseqüentemente, maior valores de grau de oxidação (NOX/nC).

A equação 4.3.2 contém um descritor de grupo funcional (nHAcc). Este indica o número de átomos aceptores de ligações de hidrogênio (flúor, oxigênio, nitrogênio). Para os SLs, os aceptores de ligação de hidrogênio são principalmente os átomos de oxigênio. Quanto maior o número destes, maior o valor do grau de oxidação. Os coeficientes estatísticos desta equação apesar de extremamente altos, foram inferiores aos observados na equação 4.3.1.

A equação 4.3.3, que contém um descritor de átomo centrado (O-058) (Viswanadhan *et al.* 1989), apresenta valores de coeficientes estatísticos baixos (r^2, Q_{cv}^2 e F) em relação às outras equações obtidas, com exceção da equação 4.3.5. O descritor O-058 representa o número de grupos carbonilas na molécula. Por esta equação (equação 4.3.3), quanto maior for o número destes grupos nos SLs, maior é o valor do grau de oxidação observado.

A equação 4.3.4 contém um descritor de auto-correlação 2D (ATS4m). Estes descritores são obtidos por meio da representação de uma molécula em 2 dimensões (Broto *et al.* 1984), como detalhado no **item 1.6.5**. O valor do

descritor ATS4m é obtido utilizando a equação 1.6.5.3. Pode-se notar que é gerado pelo resultado da soma dos produtos das massas dos átomos que estão a uma distância topológica quatro. A presença de carbonila α,β-insaturada e epóxidos vizinhos as lactonas apresentam maiores valores de ATS4m.

A equação 4.3.5 contém um descritor BCUT (BELv4) e apresenta os valores de coeficientes estatísticos mais baixos (r^2, Q_{cv}^2 e F). O descritor BELv4 é o quarto autovalor mais baixo da matriz de conectividade (Todeschini & Consonni 2000), como descrito no **item 1.6.8,** na qual os valores dos elementos diagonais são constituídos pelos valores de volume de van der Waals. Como foi descrito anteriormente sobre o descritor ATS4m, a presença de carbonila α,β-insaturada e epóxidos vizinhos as lactonas contribuem no aumento dos valores de BELv4.

A equação 4.3.6 contém um descritor topológico (DELS). Este descritor é obtido pela soma dos valores absolutos das diferenças dos estados intrínsecos dos átomos de uma molécula. O estado intrínseco é diretamente proporcional aos elétrons de valência e inversamente proporcional aos elétrons de ligação sigma (Todeschini & Consonni 2000). Os sesquiterpenos lactonizados, que apresentam mais átomos de oxigênio e principalmente grupos carbonila, tem valores maiores deste descritor.

A equação 4.3.7 apresenta um descritor geométrico (GO..O), que representa a soma das distâncias geométricas entre todos os pares de átomos de oxigênio. Os SLs que tiverem o maior número de oxigênios na molécula e/ou maiores as distâncias entre os mesmos, tem maiores valores de G(O…O) (Randic *et al.* 1994; Todeschini & Consonni 2000).

A equação 4.3.8 contém um descritor RDF (RDF045m), obtido utilizando-se a equação 1.6.3.1. Quanto maior a massa de átomos que estejam a uma distância aproximadamente de 4,5 angstroms, maior o valor deste descritor. SLs que apresentam ramificações ricas em ligações duplas com oxigênios possuem valores maiores de RDF045m, pois os átomos de oxigênio destas estão a uma distância aproximada faixa entre 4 e 5 angstroms.

A equação 4.3.9 apresenta um descritor 3D-MoRSE (Mor07u), gerado pela equação 1.6.4.1. Neste caso o valor de s é 6 $Å^{-1}$. É um descritor estritamente de caráter estérico, no qual não é utilizada nenhuma propriedade atômica como peso. Foi verificado que SLs com maior número de ramificações, porém curtas e/ou com mais de 2 anéis, exibem valores maiores de Mor07u.

A equação 4.3.10 contém um descritor GETAWAY (H5m), calculado pela equação 1.6.1.2. Quanto maior o grau de acessibilidade entre os átomos separados por uma distância topológica 5 e suas massas, maior será o valor de H5m. SLs que apresentam ramificações ricas de ésteres tem valores de H5m altos.

Um descritor WHIM (L2v) está presente na equação 4.3.11, que contém coeficientes estatísticos mais elevados (r^2, Q_{cv}^2 e F), com exceção da equação 4.3.1. O descritor L2v é segundo autovalor da matriz de covariância obtida através da equação 1.6.2.1 ponderada pelo volume de van der Waals. SLs que apresentam configuração das duplas ligações do anel invertidas apresentam ramificações com ângulos mais próximos de 90º com relação ao plano do anel. Foi verificado que estas ramificações são ricas de ésteres e hidroxilas. Também foi verificado SLs epoxidados apresentam valores altos de L2v.

Analisando a tabela 4.3.2 e a figura 4.3.1 obtidas através da equação 4.3.1, que apresenta os melhores índices estatísticos, verifica-se o ajuste dos pontos à reta de regressão. Para todas as tribos o valor calculado do grau de oxidação, utilizando a equação 4.3.1, é muito próximo ao real.

Na figura 4.3.1 verifica-se que as tribos das subfamílias Cichorioideae e Asteroideae não estão agrupadas. Comparando as classificações de Bremer (Bremer 1996), Jansen (Kim & Jansen 1996) e Funk (Funk *et al.* 2005) (figuras 1.2.4 a 1.2.6) com o a distribuição das tribos com relação ao grau de oxidação não é verificado nenhuma corroboração entre as árvores e o gráfico (figura 4.3.1). As tribos como Heliantheae e Helenieae estão próximas e também Liabeae e Arctoteae, como esperado.

As tribos Vernonieae, Cardueae e Lactuceae apresentam valores mais altos de grau de oxidação médios que tribos consideradas mais evoluídas como Heliantheae, Helenieae, Eupatorieae e Anthemideae. A tribo Arctoteae (subfamília Cichorioideae) tem valor de grau de oxidação próximo ao da Anthemidae, (subfamília Asteroideae). As tribos Astereae, Inuleae e Senecioneae apresentam baixos valores de grau de oxidação. Aparentemente para os sesquiterpenos lactonizados não há nenhuma relação entre seu o grau de oxidação e evolução das tribos.

5.4. Mapas Auto-organizáveis (Kohonen) e Descritores Moleculares na Quimiotaxonomia das Tribos da Família Asteraceae

A análise da tabela 4.4.1 verificou que todos os blocos de descritores, com exceção dos constitucionais (66,7%) e dos 3D-MoRSe (67,5%),

apresentam índices de acerto global acima de 80%. Os mapas têm dimensões de 40 por 30, com 1200 neurônios, exceto os mapas obtidos utilizando descritores funcionais (35 por 35, com 1225 neurônios) e descritores GETAWAY (40 por 35, com 1400 neurônios).

No mapa auto-organizável utilizando os descritores RDF observou-se o maior índice de acerto global (83,6%) e o SOM obtido com os descritores constitucionais apresenta o pior índice (tabela 4.4.1). Não melhora nos valores dos índices de acerto ao utilizar descritores que são obtidos através das estruturas das moléculas representadas em três dimensões (WHIM, GETAWAY, RDF, 3D-MoRSE, Geométricos) com relação aos obtidos em duas dimensões (Auto-correlação 2D, Topológicos, BCUT) e aos que envolvem no cálculo fragmentos moleculares (Grupos Funcionais, Átomo Centrado).

A tribo Inuleae tem o pior índice de acerto, ou seja, os mapas não diferenciaram esta tribo das demais. Os índices de acerto variaram de 27,6% com o mapa obtido por meio dos descritores constitucionais e 56,7% com o obtido utilizando os descritores BCUT. Em todos os mapas, as áreas desta tribo (neurônios em rosa) estão próximas da tribo Anthemideae (neurônios em azul) (figura 4.4.1). Em todas as três classificações Bremer (Bremer 1996), Jansen (Kim & Jansen 1996) e Funk (Funk *et al.* 2005), estas 2 tribos estão próximas (figuras 1.2.4 a 1.2.6). O baixo índice de acerto da tribo Inuleae deve-se ao fato de que alguns dos neurônios ocupados por esta tribo estão misturados aos neurônios das Anthemideae (regiões em azul) e Heliantheae (regiões em vermelho) (figura 4.4.1).

A tribo Anthemideae apresentou os maiores índices de acerto, o menor valor de 73,8% no mapa utilizando descritores constitucionais e o maior de

94,2% obtido com os descritores BCUT. Em todos os mapas obtidos, as regiões ocupadas por esta tribo (em azul) foram distintas das Eupatorieae (amarelo), Vernonieae (verde), e em menor escala de Heliantheae (vermelho) e Helenieae (laranja) (figura 4.4.1).

A tribo Senecioneae apresenta altos índices de acerto, acima de 75%, exceto para os mapas obtidos com os descritores constitucionais e com os descritores 3D-MoRSE (tabela 4.4.1). Na maioria dos mapas obtidos (figura 4.4.1) as regiões ocupadas por esta tribo (azul claro) estão nas proximidades da Anthemideae (azul), o que corrobora com as classificações propostas por Bremer (Bremer 1996), Jansen (Kim & Jansen 1996) e Funk (Funk *et al.* 2005) (figuras 1.2.4 a 1.2.6).

A tribo Eupatorieae tem valores de índice de acerto acima de 73% para todos os mapas, exceto para os obtidos com os descritores constitucionais e 3D-MoRSE (tabela 4.4.1). As regiões ocupadas por esta tribo nos mapas auto-organizáveis são próximas das Heliantheae (vermelho) e Helenieae (laranja) (figura 4.4.1). A proximidade destas três tribos é verificada por Bremer (Bremer 1996) e Funk (Funk *et al.* 2005) (figuras 1.2.4 e 1.2.6).

A tribo Heliantheae tem altos valores de índice de acerto para todos os mapas obtidos (tabela 4.4.1). Neste estudo é a tribo com o maior número de compostos e ocorrências, sua diversidade estrutural com relação aos SLs, é visualizado nos mapas auto-organizáveis (figura 4.4.1). Apesar de estar concentrada principalmente em uma determinada região (vermelho), esta tribo ocupa extensas áreas nos SOMs. Como mencionado anteriormente as regiões da tribo Heliantheae estão próximas das Helenieae (laranja) e Eupatorieae (amarelo).

A tribo Helenieae apresenta altos índices de acerto em todos os mapas obtidos, exceto para aqueles utilizando descritores constitucionais (52,6%) e 3D-MoRSE (57,9%) (tabela 4.4.1). Na figura 4.4.1 verifica-se que a região ocupada por esta tribo (laranja) nos mapas Kohonen é próxima das Heliantheae (vermelho) e Eupatorieae (amarelo) como observado anteriormente.

A tribo Vernonieae gerou altos valores de índice de acerto, exceto para os mapas que utilizam descritores constitucionais (68,4) e 3D-MoRSE (66,5). Com exceção do mapa que utiliza o bloco de descritores átomo centrado (figura 4.4.1. c), as regiões ocupadas por esta tribo (verde) estão mais próximas das Eupatorieae (amarelo) e Heliantheae (vermelho). Estas duas tribos pertencem à subfamília Asteroideae e não estão próximas à Vernonieae em nenhuma das classificações (Bremer 1996; Kim & Jansen 1996; Funk *et al.* 2005) (figuras 1.2.4 a 1.2.6).

Não foi verificada proximidade da tribo Vernonieae com a Lactuceae (cinza), que também pertence à mesma subfamília (Cichorioideae). Este fato pode ser explicado pelo baixo número de ocorrências (28) e compostos (17) da tribo Lactuceae utilizados neste estudo. Isto ajuda elucidar os baixos valores de índice de acerto desta tribo em todos os mapas obtidos, exceto naqueles que utilizaram os descritores de grupos funcionais (71,4%), topológicos (78,6%) e geométricos (71,4%). Esta tribo (cinza) ocupou poucos neurônios que estão distribuídos por todos os mapas gerados, não possibilitando determinar uma região predominante (figura 4.4.1).

A tribo Cardueae (subfamília Carduoideae) apresenta, em todas as redes geradas, altos índices de acerto, exceto para o SOM obtido com os

descritores constitucionais (68,6%). As regiões ocupadas por esta tribo (marrom) estão distribuídas entre as Heliantheae (vermelho), Eupatorieae (amarelo), Vernonieae (verde), e Anthemideae (azul).

5.7. Mapas Auto-organizáveis (Kohonen) e Descritores Moleculares na Quimiotaxonomia das Tribos da Família Asteraceae

Na tabela 4.5.1, podemos observar altos valores de índice de acerto total em todos os mapas obtidos. Índices com valores inferiores foram obtidos com os descritores WHIM (81,79%), constitucionais e 3D-MoRSE (ambos com 82,72%). Não foi verificado aumento nos valores de índices de acerto total nos mapas gerados com os descritores obtidos com as representações das estruturas das moléculas em três dimensões (Geométricos, RDF, 3D-MoRSE, GETAWAY e WHIM). Apesar dos valores extremamente significativos obtidos com estes descritores, todos acima de 80% de índice de acerto total, nenhum destes cinco blocos de descritores obteve valores de acerto total acima de 90%.

Os descritores que apresentam os maiores valores de índice de acerto são os que utilizam as estruturas representadas em duas dimensões: descritores BCUT e Auto-correlação 2D, ambos com 92,48%. Os mapas Kohonen que foram gerados com os descritores obtidos dividindo as moléculas em fragmentos apresentam valores significativamente altos de índice de acerto, sendo 88,13% para os descritores de grupos funcionais e 90,37% para os de átomo centrado.

Os mapas de menores dimensões foram obtidos a partir dos descritores de grupos funcionais e os descritores de átomo centrado, ambos de 14 por 10 neurônios, totalizando 140 unidades. O mapa de maior dimensão foi originado com os descritores de auto-correlação 2D, de 21 por 7, totalizando 147 neurônios.

Na figura 4.5.1, nota-se em todos os mapas que o ramo A das subtribos de Heliantheae (Stuessy 1977), detalhada no **item 1.2** e na figura 1.2.7, ocupa maior área (regiões em azul). Tal fato deve-se, obviamente, por conter maior número de ocorrências e compostos. O ramo C (regiões em vermelho) apresenta metade do numero de ocorrências do ramo A, ocupando assim um menor de neurônios em todos os mapas. Nestes, os ramos A e C ocupam regiões distintas, confirmando os valores obtidos para estes mapas na tabela 4.5.1. O ramo B não foi utilizado neste estudo por ter apenas 4 ocorrências.

Em ambos os ramos foram obtidos altos valores de índices de acerto. O ramo A obteve o maior valor (96,24%) no mapa gerado a partir de descritores de auto-correlação 2D e o menor foi observado com os descritores 3D-MoRSE (78,22%). O ramo B apresenta o maior valor com o SOM gerado com os descritores 3D-MoRSE (91,70%) e o menor com os descritores WHIM (69,96%).

5.6. Relações entre Estrutura Química e Atividade Biológica de Sesquiterpenos Lactonizados

Os tipos de descritores presentes na equação (equação 4.6.1) são os que apresentam maior significância estatística e que, conseqüentemente,

melhor relacionam as estruturas dos sesquiterpenos lactonizados com sua atividade citotóxica são: 3D MoRSE (Mor15u e Mor13m), Geométricos (SPAN e G(O...O)) e GETAWAY ($R8e^{+}$).

Analisando-se a equação 4.6.1 pode-se verificar que o valor do coeficiente de predição interno Q_{cv}^{2} é significativo (0,743), indicando um modelo robusto. O valor de F é altamente significante, pois para 95% de confiança com 5 e 22 graus de liberdade, o valor mínimo necessário é 2,66.

A atividade biológica é intimamente ligada à estrutura tridimensional e às propriedades eletrônicas dos sítios específicos da molécula. O potencial do descritor 3D-MoRSE de considerar simultaneamente a estrutura 3D e as propriedades atômicas, como as cargas parciais, faz dele um descritor particularmente apropriado para o estudo de informações biológicas (Schuur *et al.* 1996; Gasteiger *et al.* 1996).

Os descritores Mor15u e Mor13m são estritamente relacionados com a estereoquímica dos compostos. Entretanto, o último também considera também o peso dos átomos nos cálculos. Ambos são calculados a partir da equação 1.6.4.1.

O descritor SPAN é um descritor geométrico (**item 1.6.6**) e tem seu cálculo baseado na escolha do raio da menor esfera, centrada no centro de massa, englobando completamente todos os átomos da molécula (equação 5.6.1). Dessa forma, compostos os quais possuem um maior número de ramificações e grupos capazes de deslocar o centro de massa, têm seus raios aumentados (Todeschini *&* Consonni 2000).

$$SPAN = max_i(r_i) \qquad \textbf{Equação 5.6.1}$$

Onde: ri é a distância entre o átomo i e o centro de gravidade da molécula.

O descritor G(O…O), também é um descritor geométrico e representa a soma das distâncias entre todos os pares de átomos de oxigênio. Quanto maior o número destes átomos na molécula e/ou maiores as distâncias entre os mesmos, maior será esta soma (Randic *et al.* 1994; Todeschini & Consonni 2000).

Os descritores GETAWAY (*Geometric Topology and Atom Weights Assembly)* estão relacionados com a influência dos átomos na determinação da forma molecular (*leverages*) e com a distância entre eles, como explicado no **item 1.6.1**. O descritor $R8e^+$ é o valor máximo do cálculo (equação 1.6.1.5), que multiplica os "leverages*"* entre dois átomos e o valor das respectivas eletronegatividades de Sanderson, com distância topológica igual a 8, divididos pela distância geométrica entre os mesmos. Quanto maior a influência dos átomos na forma molecular, maior a eletronegatividade, e mais próxima a distância entre eles, maior será o valor de $R8e^+$ (Consonni *et al.* 2002a; Consonni *et al.* 2002b).

A equação 4.6.1 utiliza os descritores previamente citados, todos calculados por representação 3D das moléculas, e revela que os parâmetros relacionados à conformação e estereoquímica são os mais importantes no que diz respeito à atividade citotóxica destes sesquiterpenos lactonizados.

Analisando-se as estruturas destes compostos, separando-os em grupos que obtiveram os maiores e menores valores de atividade biológica e comparando-os pelos elementos que possuem, os quais podem ser responsáveis por uma alta ou baixa atividade, algumas considerações podem ser feitas, como relatadas a seguir:

Primeiramente, é importante destacar que todos os compostos, apresentando alta atividade ou não, possuíam a estrutura α-metileno-γ-lactona.

Os compostos com os maiores valores atividades biológicas (pED_{50} > 6,0) foram **15, 16, 20, 23, 26, 28** e **30** (figura 3.7.1.1). As substâncias **15, 16, 20, 26** e **30** (figura 3.7.1.1) continham uma dupla ligação no anel de cinco membros na posição 3 (figura 5.x.1). Os compostos **16, 20, 26, 28** e **30** (figura 3.7.1.1) têm um grupo hidroxila no carbono que liga o ciclo-heptano ao ciclopentano (posição 5).

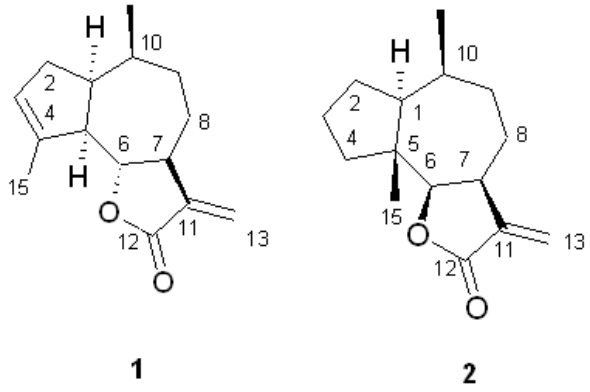


Figura 5.6.1. Esqueletos Guaianolídeo (1) e Pseudoguaianolídeo (2)

Os SLs que apresentam ambas as estruturas (**16, 20, 26** e **30**) não possuem grande variação dos valores de atividade biológica, sendo que o composto de número **20** apresenta o maior valor (figura 3.7.1.1).

As substâncias mais ativas pertencem ao esqueleto dos tipos guaianolídeo (**16, 20, 26, 28** e **30**) e pseudoguaianolídeo (**11, 12, 13, 14** e **15**) (figura 5.4.1), sendo que valores superiores de atividades foram observadas no primeiro. Este fato mostra que o grupo metil na posição 4 pode ser mais importante para a atividade do que na posição 5. A exceção é o composto **28** (pED_{50} = 5,98), o qual possui atividade menor que o **15** (pED_{50} = 6,12), provavelmente porque é o único entre os guaianolídeos a não apresentar dupla

ligação na posição 3. Pode-se notar que os outros compostos do tipo guaianolídeo (**17, 19, 27** e **29**) possuem um grupo hidrofílico nesta mesma posição.

Os compostos **17** e **28** apresentam estruturas similares,uma estrutura muito similar; entretanto, a presença de um grupo metilênico ao invés de uma hidroxila e um cloreto de metila na posição 10, contribui para um considerável aumento na atividade em **28**. Provavelmente a dupla ligação nesta posição é importante para a atividade.

Esta importância pode ser igualmente visualizada comparando-se os compostos **26, 27, 28, 29** e **30.** O primeiro (**26**), o qual possui um valor de atividade (pED_{50}) considerável, todavia menor que a do composto **30**, que tem uma dupla ligação no anel de 5 membros e um grupo hidroxila na posição 5. A substância **28**, que apresenta atividade menor que ambos os compostos acima, possui a ligação dupla na forma de um grupo metileno na posição 10 e um grupo hidroxila na posição 5. Os compostos **27** e **29**, os quais têm as menores atividades entre este grupo, exibem um grupo epóxi na posição 10, ao invés de um grupo metileno; um grupo hidroxila na posição 5 e não têm uma ligação dupla no anel de 5 membros (posição 3), como o composto **28**. Por fim, a substância (**30**) mostra a maior atividade, tendo ligação dupla na posição 3 e o grupo epóxi na posição 10. Verifica-se que a dupla ligação (assim como o grupo metileno na posição 10 ou na posição 3) é importante para a atividade. Provavelmente, isto se deve ao mecanismo de alquilação através do qual os sesquiterpenos lactonizados exercem sua atividade biológica.

Outra importante característica nos compostos que possuem o esqueleto do tipo guaianolídeo é a presença de um 8β-angelato e, conseqüentemente, a

6,12-lactonização em suas estruturas. Estes compostos (**16, 17, 20, 26-30**) exibiram as maiores atividades citotóxicas. A partir desta observação pode-se supor que estas características estruturais são relevantes para a atividade biológica.

Como já previamente citado, pode-se notar que entre os compostos **9-15,** os quais se apresentam como pseudoguaianolídeos do tipo ambrolídeo, a presença de uma dupla ligação na ciclopentanona (carbonila α,β-insaturada) aumenta a atividade, principalmente em **11-15**. Comprovando a importância destes grupos na atividade biológica dos sesquiterpenos lactonizados que possuem o esqueleto pseudoguaianolídeo.

Sendo a estereoquímica um importante fator para a atividade biológica e sobre SLs pudemos observar que compostos com maiores atividades foram os guaianolídeos e pseudoguaianolídeos, podendo-se supor que estes tipos de esqueletos têm um conjunto de fatores mais adequado para a atividade citotóxica do que os outros tipos.

As características eletrônicas, as quais também estão implícitas no descritor $R8e^{+}$ (equação 4.6.1), podem ser associadas à presença de duplas ligações. Estas estruturas aumentam a nuvem eletrônica, assim como podem influenciar a forma molecular, originando uma nova conformação. A presença dos átomos de oxigênio também aumenta a nuvem eletrônica, porém estas estruturas parecem ter importância secundária na atividade, uma vez que há compostos com um número elevado destes átomos que não apresenta uma considerável atividade como, por exemplo, os compostos **7**, **36** e **37**. Portanto, o tipo dos descritores envolvidos na equação estatisticamente mostra uma concordância com as características acima analisadas.

6. CONCLUSÕES

Não foi estabelecida nenhuma relação entre o grau de oxidação dos sesquiterpenos lactonizados e a evolução das tribos da família Asteraceae. Diversas equações apresentaram coeficientes altamente significativos com apenas um descritor de diversos blocos Puderam-se identificar algumas características estruturais relacionadas ao grau de oxidação interpretando os descritores moleculares.

Os mapas auto-organizáveis obtidos para as 9 tribos tiveram altos índices de acerto, separando as tribos ao utilizar os descritores moleculares de acordo com as classificações já propostas. Conclui-se que os SOMs (“Self-Organizing Maps”) combinados com os descritores moleculares podem ser utilizados como uma ferramenta para classificação em baixos níveis hierárquicos como tribos.

Os mapas auto-organizáveis obtidos com os descritores moleculares dividiram os ramos da tribo Heliantheae com altos índices de acerto e nitidamente. Em um nível hierárquico mais baixo que o estudo anterior, novamente a combinação de redes Kohonen com descritores moleculares obtiveram resultados que corroboram a classificação proposta da literatura.

As diferentes rotas metabólicas são caracterizadas por uma série de compostos e suas estruturas 3D devem ser utilizadas no processo de diferenciação, pois os metabólitos são formados nas cavidades de enzimas, inerentemente 3D na natureza. Pequenas alterações nas cavidades destas podem produzir metabólitos com pequenas diferenças em suas estruturas em 3D. Partindo das considerações citadas anteriormente, um descritor molecular

3D pode representar o avanço evolutivo ocorrido durante as transformações metabólicas e as médias destes valores para grupos de táxons (ex. tribos), separando-os em um espaço bidimensional e servindo como base para um novo tipo de quimiotaxonomia. Porém para ambos os estudos, tribos e ramos da tribo Heliantheae, com mapas auto-organizáveis, os descritores obtidos por fragmentos ou pela representação da estrutura dos sesquiterpenos lactonizados em duas dimensões foram suficientes para obtermos resultados satisfatórios. Não houve melhora nos resultados com os descritores que utilizam a representação das estruturas em três dimensões.

Com relação ao estudo da estrutura dos sesquiterpenos lactonizados e sua atividade citotóxica, verificou-se que os descritores selecionados na equação estatisticamente mais significativa gerada representam uma descrição global de propriedades estéricas e características eletrônicas de cada molécula. As características estruturais, presentes nos sesquiterpenos lactonizados deste estudo, são muito importantes para a atividade biológica, como a dupla ligação no ciclopentano, bem como na posição 10, assim como o grupo hidroxila na 5 e o grupo angelato na 8. Através deste fato pôde-se constatar que os compostos mais ativos são aqueles os quais apresentam os tipos guaianolídeo e pseudoguaianolídeo como esqueleto.

Um estudo mais extensivo é necessário para se comparar um número maior de compostos, incluindo sesquiterpenos lactonizados, os quais não possuam a estrutura α–metileno–γ–lactona. Assim maiores informações poderão ser obtidas a fim de se elucidar se a mesma possui relevância para a atividade citotóxica ou não, bem como sesquiterpenos lactonizados apresentando uma maior variedade de tipos de esqueletos, ou seja, estrutural,

para confirmar se os tipos guaianolídeo e pseudoguaianolídeo são mais ativos que os outros.

7. Referências

Allinger N Conformational Analysis. 130. MM2. A hydrocarbon force field utilizing vi and v2 torsional terms. Journal of The American Chemical Society. 1977; 99(25):8127-34

Balaban AT, Devillers J. Topological Indices and Related Descriptors in QSAR and QSPR. Amsterdam: Gordon and Breach Science Publishers; 1999.

Baroni M, Constantino G, Cruciani G, Riganelli DLV, Clementi S. Generating optimal linear pls estimations (golpe): an advanced chemometric tool for handling 3d-qsar problem. Quantitative Structure-Activity Relationships. 1993;12:9-20.

Belvisi L, Bravi G, Scolastico C, Vulpetti A, Salimbeni A, Todeschini R. A 3d qsar approach to the search for geometrical similarity in a series of nonpeptide angiotensin-ii receptor antagonists. Journal of Computer-Aided Molecular Design. 1994;8(2):211-20.

Bentham G. Notes on the classification, history, and geographical distribution of the Compositae. Journal of the Linnean Society, Botany. 1873;13:335-557.

Bremer K. Asteraceae - Cladistic & classification. Portland (OR): Timber Press; 1994.

Bremer K. Tribal interrelationships of the Asteraceae. Cladistics. 1987;3:210-53.

Bremer K, Jansen RK, Karis PO, Kallersjo M, Keeley SC, Kim KJ, Michaels HJ, Palmer JD, Wallace RS. A review of the phylogeny and classification of Asteraceae. Nordical Journal of Botany. 1992;12:141-8.

Bremer K Major clades and grades of the Asteraceae. In Compositae: Systematics In Hind DJN, Beentje H, editors. Compositae: Systematics. Proceedings of the International Compositae Conference. Vol. 1. Kew: Royal Botanic Garden; 1996 p. 1-7.

Jansen RK, Kim K. Implications of chloroplast DNA data for the classification and phylogeny of the Asteraceae. In Hind DJN, Beentje H, editors. Compositae Systematics. Proceedings of the International Compositae Conference. Vol. 1. Kew: Royal Botanic Garden; 1996 p.317-39.

Bohm BA, Stuessy TF. Flavonoids of the sunflower family. New York: Springer-Wien; 2001.

Broto P, Moreau G, Vandycke C. Molecular-structures - perception, auto-correlation descriptor and sar studies - perception of molecules - topological-structure and 3-dimensional structure. European Journal of Medicinal Chemistry. 1984;19(1):61-5.

Burden FR. A chemically intuitive molecular index based on the eigenvalues of a modified adjacency matrix. Quantitative Structure-Activity Relationships. 1997;16(4):309-14.

Burden FR. Molecular-Identification Number for Substructure Searches. Journal of Chemical Information and Computer Sciences. 1989;29(3):225-7.

Calabria, ML. Emerenciano, VP. Ferreira, MJP. Scotti, MT. Mabry, TJ. Phylogenetic analysis of tribes of the Asteraceae based on phytochemical data. Natural Products Communications, Estados Unidos. 2007; 2(3): 277-85.

Carbo R, Leyda L, Arnau M. How similar is a molecule to another - an electron-density measure of similarity between 2 molecular-structures. International Journal Of Quantum Chemistry .1980;17(6):1185-9.

Cassini H. Tableau exprimant les affinities des tribus naturelles de famile des Synanthérées. In Cloquet MH, editor. Dictionnaire des Sciences Naturelles, vol. 3. 2nd. Le Normant (Paris): Ed. G. Cuvier; 1816.

Cleva C, Cachet C, Cabrol-bass D. Clustering of infrared spectra with Kohonen networks. Analusis. 1999;27: 81-90.

Consonni V, Todeschini R, Pavan M. Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 1. Theory of the novel 3D molecular descriptors. Journal of Chemical Information and Computer Sciences. 2002a;42(3):682-92.

Consonni V, Todeschini R, Pavan M, Gramatica P. Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 2. Application of the novel 3D molecular descriptors to QSAR/QSPR studies. Journal of Chemical Information and Computer Sciences. 2002b;42(3):693-705.

Cordell GA. Biosynthesis of sesquiterpenes. Chemical Reviews. 1976;76(4):425-60.

Cronquist A. The evolution and classification of flowering plants. 2º ed. Portland (Oregon): New York Botanical Garden Press; 1988. p. 1-555.

Da Costa FB, Terfloth L, Gasteiger J. Sesquiterpene lactone-based classification of three Asteraceae tribes: a study based on self-organizing neural networks applied to chemosystematics. Phytochemistry. 2005;66;345-53.

Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP. The development and use of quantum-mechanical molecular-models .76. AM1 - A new general-purpose quantum-mechanical molecular-model. Journal of the American Chemical Society. 1985;107:3902-9.

Dirsch VM, Stuppner H, Vollmar AM. Helenalin triggers a CD95 death receptor-independent apoptosis that is not affected by overexpression of Bcl-xL or Bcl-2. Cancer Research. 2001;61(15):5817-23.

Diudea MV, Horvath D, Graovac A. Molecular topology .15. 3d distance matrices and related topological indexes. Journal of Chemical Information and Computer Sciences. 1995;35(1):129-35.

Doucet J, Panaye A, Feuilleaubois E, Ladd P. Neural networks and ^{13}C NMR chemical shift prediction. Journal of Chemical Information and Modeling. 1993;33: 320-4.

Emerenciano VP, Ferreira MJP, Branco MD, Dubois JE. The applications of Bayes theorem in natural products as a guidefor skeletons identifications. Chemometrics and Intelligent Laboratory Systems. 1998a; 4: 83-92..

Emerenciano VP, Rodrigues GV, S. Alvarenga SAV, Macari PAT, Kaplan MAC. Um método para união de vários marcadores quimiotaxonômicos. Química Nova. 1998b;21:125-9.

Emerenciano VP, Alvarenga SAV, Scotti MT, Ferreira MJP, Stefani R , Nuzillard J-M. Automatic identification of terpenoid skeletons by feed-forward neural networks. Analytica Chimica Acta. 2006;579:217–26.

Emerenciano VP, Scotti MT, Stefani R, Alvarenga SAV, Nuzillard JM, Rodrigues GV. Diterpene Skeletal Type Classification and Recognition using Self-Organizing Maps. Internet Electronic Journal of Molecular Design. 2006;5(4):213-23.

Fernandes MB, Scotti MT, Ferreira MJP, Emerenciano VP. Use of self-organizing maps and molecular descriptors to predict the cytotoxic activity of sesquiterpene lactones. European Journal of Medicinal Chemistry. 2008 Jan:1-9.

Fraser LA, Mulholland A, Fraser DD. Classification of limonoids and protolimonoids using neural networks. Phytochemical analysis. 1997;8:301-11.

Funk V, Bayer RJ, Keeley S, Chan R, Watson L, Gemeinholzer B, Schilling E,Panero JL, Baldwin BG, Garcia-Jacas N, Susanna A, Jansen RK.. Everywhere but Antarctica: using a supertree to understand the diversity and distribution of the Compositae. Biologiske Skrifter. 2005;55: 343–74.

Gasteiger J, Sadowski J, Schuur J, Selzer P, Steinhauer L, Steinhauer V. Chemical information in 3D space. Journal of Chemical Information and Computer Sciences. 1996;36(5):1030-7.

Gasteiger J, Teckentrup A, Terfloth L, Spycher S. Neural networks as data mining tools in drug design. Journal of Physical Organic Chemistry. 2003;16:232-245.

Gastmans JP, Furlan M, Lopes MN, Borges JHG, Emerenciano VP. A inteligência artificial aplicada à química de produtos naturais. O programa Sistemat. Parte I – Bases Teóricas. Química Nova. 1990;13:10-15. **a**

Gastmans JP, Furlan M, Lopes MN, Borges JHG, Emerenciano VP. A inteligência artificial aplicada à química de produtos naturais. O Programa Sistemat. Parte II – Organização do Programa e Aplicativos. Química Nova. 1990;13:75-80. **b**

Geary RC. The Contiguity Ratio and Statistical Mapping. The Incorpored Statistician. 1954;5(3):115-45.

Geissman TA, Crout D.H.G. Organic Chemistry of Secondary Plant Metabolism. San Francisco: Freeman Cooper & Company; 1969

Ghose AK, Pritchett A, Crippen M. Atomic Physicochemical Parameters for Three Dimensional Structure Directed Quantitative Structure-Activity relatioships III: Modeling Hydrophobic Interactions. Journal Of Computational Chemistry. 1988;9(1):80-90.

Golbraikh A, Tropsha A. Beware of q2!. Journal of Molecular Graphics and Modelling. 2002;20(4):269-76.

Goldstein JL, Brown MS. Regulation of the mevalonate pathway. Nature. 1990;343(6257):425-30.

Good AC. The calculation of molecular similarity - alternative formulas, data manipulation and graphical display. Journal of Molecular Graphics. 1992;10(3):144-151.

Good AC, So SS, Richards WG. Structure-activity-relationships from molecular similarity- matrices. Journal of Medicinal Chemistry. 1993;36(4):433-8.

Gottlieb OR. The role of oxygen in phytochemical evolution towards diversity. Phytochemistry. 1989;28:2545-2558.

Gottlieb OR, Kaplan MAC. Micromolecular evolution: The redox theory. Natural Products Letters. 1993;2:171-177.

Gottlieb OR, Kaplan MAC, Borin, MRMB. Biodiversidade: um enfoque químico-biológico. Rio de Janeiro: Editora da UFRJ; 1996.

Guha R, Serra JR, Jurs PC. Generation of QSAR sets with a self-organizing map. Journal Of Molecular Graphics & Modelling. 2004;23(1):1-14.

Gupta S, Aires-de-Sousa J. Comparing the chemical spaces of metabolites and available chemicals: models of metabolite-likeness. Molecular Diversity. 2007;11(1):23-36.

Hansch C. Comprehensive medicinal chemistry: the rational design, mechanistic study and theropeutic application of chemical compounds. Oxford: Pergamon; 1990.

Harborne JB, Mabry TJ, Mabry H. The flavonoids. London: Champman & Hall; 1975.

Harborne JB. Ecological Biochemistry. 3rd. London: Academic Press; 1988. Emerenciano VP, Ferreira MJP, Branco MD, Dubois JE. The applications of Bayes theorem in natural products as a guidefor skeletons identifications. Chemometrics and Intelligent Laboratory Systems. 1998;4:83-92.

Heilmann J, Wasescha MR, Schmidt TJ. The influence of glutathione and the cysteine levels on the cytotoxicity of helenanolide type sesquiterpene lactones against KB cells. Bioorganic & Medicinal Chemistry. 2001, 9(8), 2189-94.

Hemmer MC, Steinhauer V, Gasteiger J. Deriving the 3D structure of organic molecules from their infrared spectra. Vibrational Spectroscopy. 1999;19(1):151-64.

Hendrickson JB, Cram DJ, Hammond GS. Organic Chemistry. 3º ed. New York: Ed. McGraw-Hill; 1970. P. 1-1280.

Heywood VH, Harborne JB, Turner BL. The Biology and Chemistry of the Compositae, Vols I and II New York: Academic Press; 1977.

Hind DJN, Beentje HJ. Compositae: Systematics. Proceedings of the International Compositae Conference. Vol.1, 2, Kew: Royal Botanic Gardens; 1994.

Hocquet A, Langgård M. An Evaluation of the MM+ Force Field Journal of Molecular Modeling [Electronic Publication]. 1998; 4(3):94 – 112.

Hodgkin EE, Richards WG. Molecular Similarity Based on Electrostatic Potential and Electric Field. International Journal of Quantum Chemistry: Quantum Biology Symposium. 1987;14:105-10.

Hoffman O. Compositae. In : Engler A, Prantl K, editors. Die Natiirlichen Pflanzenfamilien . Vol . 4 Leipzig: Engelmann; 1890 . P. 87–381.

Hristozov D, Da Costa FB, Gasteiger J. Sesquiterpene Lactones-Based Classification of the Family Asteraceae Using Neural Networks and k-Nearest Neighbors. Journal of Chemical Information and Modeling. 2007;47(1):9-19.

Hyperchem available from Hypercube Inc., Gainesville, Florida, USA, 2001.

Jansen RK, Kim K. Implications of chloroplast DNA data for the classification and phylogeny of the Asteraceae. In Hind DJN, Beentje H, editors. Compositae Systematics. Proceedings of the International Compositae Conference. Vol. 1. Kew: Royal Botanic Garden; 1996 p.317-339.

Kelsey RG, Shafizadeh F. Sesquiterpene lactones and systematics of the genus Artemisia. Phytochemistry. 1979;18(10):1591-611.

Klebe G, Abraham U, Mietzner T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. Journal of Medicinal Chemistry. 1994;37(24):4130-46.

Kier LB, Hall LH, Frazer JW. An index of electrotopological state for atoms in molecules. Journal of Mathematical Chemistry. 1991;7(1-4):229-41.

Kim K, Jansen RK. ndhF sequence evolution and the major clades of the sunflower family. Proceedings of the National Academy of Science USA. 1995;92:10379-10383.

Kohonen T. Self-Organizing Maps, volume 30 of Springer Series in Information Sciences. 3rd. Heidelberg (Berlin): Springer; 2001.

Kubinyi, H. QSAR: Hansch Analysis and Related Approaches. Weinheim: VCH; 1993a.

Kubinyi, H. 3D QSAR in Drug Design. Theory, Methods and Application. Leiden: ESCOM; 1993b.

Kubinyi H. Variable selection in qsar studies .1. An evolutionary algorithm. Quantitative Structure-Activity Relationships. 1994;13(3):285-94.

Kubinyi H, Hamprecht FA, Mietzner T. Three-dimensional quantitative similarity-activity relationships (3D QSiAR) from SEAL similarity matrices. Journal of Medicinal Chemistry. 1998;41(14):2553-64.

Kupchan, SM, Kelsey, JE, Maruyama M, Cassady JM, Hemingway JC, Knox J R, Tumor Inhibitors. XLI. Structural elucidation of tumor-inhibitory sesquiterpene lactones from Eupatorium rotundifolium. Journal of Organic Chemistry 1969; 12:3876-83a.

Kupchan SM, Hemingway RJ, Karim A, Werner D. Tumor Inhibitors. XLVII. Vernodalin and Vernomygdin, to new cytotoxic sesquiterpene lactones from Vernonia amygdalina Del. Journal of Organic Chemistry. 1969; 12, 3908-3911b.

Kupchan SM, Fessler DC, Eakin MA, Giacobbe TJ. Reactions of alpha methylene lactone tumor inhibitors with model biological nucleophiles. Science. 1970;168(3929): 376-8.

Kupchan SM. Recent advances in the chemistry of tumor inhibitors of plant origin. Transactions of the New York Academy of Sciences. 1970;32(1):85-106.

Kupchan SM, Eakin MA. Thomas AM. Tumor inhibitors. 69. Structure-Cytotoxicity Relationships among the Sesquiterpene Lactones. Journal of Medicinal Chemistry. 1971, 14(12), 1147-52.

Kupchan SM, Maruyama M, Hemingway RJ, Hemingway JC, Shibuya S, Fujita T. Structural elucidation of novel tumor-inhibitory sesquiterpene lactones from Eupatorium cuneifolium. Journal of Organic Chemistry.1973; 12: 2189-96.

Lange BM, Croteau R. Isopentenyl diphosphate biosynthesis via a mevalonate-independent pathway: Isopentenyl monophosphate kinase catalyzes the terminal enzymatic step. Proceedings of the National Academy of Sciences of the United States of America. 1999;96(24):13714-9.

Leardi R, Boggia R, Terrile M. Genetic algorithms as a strategy for feature-selection. Journal Of Chemometrics. 1992;6(5):267-81.

Leardi R. Application of a genetic algorithm to feature-selection under full validation conditions and to outlier detection. Journal of Chemometrics. 1994;8(1):65-79.

Lichtenthaler HK. The 1-deoxy-D-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. Annual Review of Plant Physiology and Plant Molecular Biology. 1999;50:47-65.

Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews. 1997;23(1-3):3-25.

Livingstone D. Data Analysis for Chemists. New York: Oxford Science Publications; 1995.

Lohninger H, Stanci F. Comparing the performance of neural networks to well-established methods of multivariate data analysis: the classification of mass spectral data. Fresenius' Journal of Analytical Chemistry. 1992;344:188-89.

Macari PAT, Gastmans JP, Rodriguez GV, Emerenciano VP. An expert system for structure elucidation of triterpenes. Spectroscopy-An International Journal. 1994;12:139-66.

Manallack DT, Livingstone DJ. Neural networks in drug discovery: have they lived up to their promise?. European Journal of Medicinal Chemistry. 1999;34:195-208.

Mattioni BE, Jurs PC. Development of quantitative structure-activity relationship and classification models for a set of carbonic anhydrase inhibitors. Journal of Chemical Information and Computer Sciences. 2002;42(1):94-102.

Mcculloch WS, Pitts W. A Logical Calculus of the Ideas Immanent in Nervous Activity," Bulletin of Mathematical Biophysics. 1943;5:115-37.

Miller AJ. Subset Selection in Regression. London: Chapman and Hall; 1990.

Minsky M, Papert S. Perceptrons. Cambridge: MIT Press; 1969.

Moran PAP. Notes on continuous stochastic phenomena. Biometrika. 1950;37(1-2):17-23.

Moreau G, Broto P. The auto-correlation of a topological-structure - a new molecular descriptor. Nouveau Journal de Chimie-New Journal of Chemistry. 1980;4(6):359-60.

Murrray RDH. The natural coumarins. Occurrence, chemistry and biochemistry. New York: John Wiley & Sons, 1982.

Pearlman RS, Smith KM. Metric validation and the receptor-relevant subspace concept. Journal of Chemical Information and Computer Sciences. 1999;39(1):28-35.

Petitjean M. Applications of the radius diameter diagram to the classification of topological and geometrical shapes of chemical-compounds. Journal of Chemical Information and Computer Sciences. 1992;32(4):331-7.

Picman AK. Biological activities of sesquiterpene lactones. Biochemical Systematics and Ecology. 1986;14(3): 255-81.

Proksch P, Rodriguez E. Chromenes and benzofuranes of the Asteraceae, their chemistry and biological significance. Phytochemistry. 1983;22:2335-48.

Randic M, Kleiner AF, Dealba L M. Distance matrices. Journal of Chemical Information and Computer Sciences. 1994; 34: 277-86.

Reynolds CA, Burt C, Richards WG. A linear molecular similarity index. Quantitative Structure-Activity Relationships. 1992;11(1):34-5.

Rohmer M, Knani M, Simonin P, Sutter B, Sahm H. Isoprenoid biosynthesis in bacteria - a novel pathway for the early steps leading to isopentenyl diphosphate. Biochemical journal. 1993;295:517-24.

Sadowski J, Gasteiger J. From atoms and bonds to 3-dimensional atomic coordinates - automatic model builders. Chemical Reviews. 1993;93(7):2567-81.

Sadowski J, Gasteiger J, Klebe G. Comparison of automatic 3-dimensional model builders using 639 x-ray structures. Journal of Chemical Information and Computer Sciences. 1994;34(4):1000-8.

Schmidt TJ, Helenanolide-type sesquiterpene lactones – III. Rates and stereochemistry in the reaction of helenalin and related helenanolides with sulfhydryl containing biomolecules. Bioorganic & Medicinal Chemistry. 1997;5(4):645-53.

Schmidt TJ. Toxic activities of sesquiterpenes lactones – Structural and biochemical aspects. Current Organic Chemistry. 1999;3(6):577-608.

Schuur JH, Selzer P, Gasteiger J. The coding of the three-dimensional structure of molecules by molecular transforms and its application to structure-spectra correlations and studies of biological activity. Journal of Chemical Information and Computer Sciences. 1996;36(2):334-44.

Seaman FC. Sesquiterpene lactones as taxonomic characters in the Asteraceae. Botanical Review.1982;48:123–551.

Seaman F, Bohlmann F, Zdero C, Mabry TJ. Diterpenes of flowering plants – Compositae (Asteraceae). New York: Springer-Verlag; 1990.

Serilevy A, Salter R, West S, Richards WG. Shape similarity as a single independent variable in qsar. European Journal of Medicinal Chemistry. 1994;29(9):687-94.

Silva MA. Mapas Auto-Organizáveis na Análise Exploratória de Dados Geoespaciais Multivariados [Dissertação de Mestrado]. São José dos Campos: Instituto Nacional de Pesquisas Espaciais; 2004.

Smith JRM. Exploring the possibilities of applying artificial networks on problems in analytical chemistry [Tese de Doutorado]. Katholieke Unviveristeit, Nijmegen; 1993.

Stuessy TF. Heliantheae – systematic review. In: Harborne JB, Turner BL, editors. The biology and chemistry of the Compositae. vol. 2 London: Academic Press; 1977.

Talete, s. R. L. Dragon for windows (software for molecular descriptor calculations). Version 5.4 – 2006 – http://www.talete.mi.it.

Talete, S. R. L. Mobydigs Academic version - Version 1.0 – 2004 – http://www.talete.mi.it.

Todeschini R. Gramatica P. 3D-modelling and prediction by WHIM descriptors .5. Theory development and chemical meaning of WHIM descriptors. Quantitative Structure-Activity Relationships. 1997a;16(2):113-9.

Todeschini R, Gramatica P. 3D-modelling and prediction by WHIM descriptors .6. Application of WHIM descriptors in QSAR studies. Quantitative Structure-Activity Relationships. 1997b;16(2):120-5.

Todeschini R. Data correlation, number of significant principal components and shape of molecules. The K correlation index. Analytica Chimica Acta. 1997;348(1-3):419-30.

Todeschini R, Consonni V, Maiocchi A. The K correlation index: theory development and its application in chemometrics. Chemometrics and Intelligent Laboratory Systems. 1999;46(1):13-29.

Todeschini R, Consonni V. Handbook of Molecular Descriptors. Weinheim, (Germany): WILEY - VCH; 2000.

Todeschini R, Consonni V, Mauri A, Pavan M. Detecting "bad" regression models: multicriteria fitness functions in regression analysis. Analytica Chimica Acta. 2004;515(1):199-208.

Topliss JG, Costello RJ. Chance correlations in structure-activity studeis using multiple regression analysis. Journal of Medicinal Chemistry. 1972;15(10):1066-8.

Topliss JG, Edwards RP. chance factors in studies of quantitative structure-activity-relationships. Journal of Medicinal Chemistry. 1979;22(10):1238-44.

Vesanto J, Himberg J, Alhoniemi E, Parhankangas J. Self-Organizing Map in Matlab: the SOM Toolbox. Proceeding of the Matlab DSP Conference. 1999; 35–40.

Vesanto, J. Himberg, E. Alhoniemi and J. Parhankangas, SOM Toolbox 2.0 for Matlab 5, 2005, http://www.cis.hut.fi/projects/somtoolbox.

Viswanadhan VN, Ghose AK, Revankar GR, Robins RK. Atomic physicochemical parameters for 3 dimensional structure directed quantitative structure - activity relationships .4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally-occurring nucleoside antibiotics. Journal of Chemical Information and Computer Sciences. 1989;29(3):163-72.

Wagenitz G. Systematics and phylogeny of the Compositae (Asteraceae). Plant Systematics and Evolution. 1976;125:29-46.

Wagner S, Hofmann A, Siedle B, Terfloth L, Merfort I, Gasteiger J. Development of a Structural Model for NF- κB Inhibition of Sesquiterpene Lactones Using Self-Organizing Neural Networks. Journal of Medicinal Chemistry. 2006;49(7):2241-52.

Wold S, Esbensen K, Geladi P. Principal component analysis. Chemometrics and Intelligent Laboratory Systems. 1987;2(1-3):37-52.

Yoshioka H, Mabry TJ, Timmermann BN. Sesquiterpene lactones: Chemistry, NMR, and Plant Distribuition. Japan: University of Tokyo Press; 1973.

Zdero C, Bohlmann F. Systematics and evolution within the Compositae, seen with the eyes of a chemist. Plant Systematics Evolution. 1990;171:1-14.

Zhang Q-Y, Aires-de-Souza J. Structure-Based Classification of Chemical Reactions without Assignment of Reaction Centers. Journal of Chemical Information and Modeling. 2005;45:1775-83.

Zupan E, Gasteiger J. Neural Networks for Chemists – An Introduction. Weinhein: VCH; 1993.

SÚMULA CURRICULAR

DADOS PESSOAIS

Nome: Marcus Tullius Scotti
Local e data de nascimento: Alagoa Grande - PB, 12/09/1975.

EDUCAÇÃO

Colégio Etapa, São Paulo, 1992.

Universidade de São Paulo, São Paulo, 1999.
Engenheiro Químico

Universidade de São Paulo, São Paulo, 2005.
Mestrado em Química Orgânica

FORMAÇÃO COMPLEMENTAR

Especialização em Administração Industrial, Fundação Carlso Alberto Vanzolini, São Paulo, 2002.

OCUPAÇÃO

Bolsista de Doutorado, CAPES, 2005

PUBLICAÇÕES

Artigos Completos

1. CALABRIA, Maria Lalita ; EMERENCIANO, Vicente de Paulo ; FERREIRA, Marcelo J P ; SCOTTI, Marcus Tullius ; MABRY, Tom J . Phylogenetic analysis of tribes of the Asteraceae based on phytochemical data.. Natural Products Communications, Estados Unidos, v. 2, n. 3, p. 277-285, 2007.
2. SCOTTI, Marcus Tullius ; FERNANDES, Mariane B ; FERREIRA, Marcelo J P ; EMERENCIANO, Vicente de Paulo . Quantitative structure activity relationship of sesquiterpene lactones with cytotoxic activity. Bioorganic & Medicinal Chemistry, Estados Unidos, v. 15, n. 8, p. 2927-2934, 2007.
3. SCOTTI, Luciana ; SCOTTI, Marcus Tullius ; ISHIKI, Hamilton Mitsugo ; FERREIRA, Marcelo J P ; EMERENCIANO, Vicente de Paulo ; MENEZES, Carla Maria de ; FERREIRA, Elizabeth Igne . Quantitative elucidation of the structure-bitterness relationship. Food Chemistry, v. 115, p. 77-83, 2007.
4. SCOTTI, Luciana ; SCOTTI, Marcus Tullius ; CARDOSO, Carmen Lucia ; PAULETTI, Patrícia Mendonça ; GAMBOA, Ian Castro ; BOLZANI, Vanderlan da Silva ; VELASCO, Maria Valéria Robles ; FERREIRA, Elizabeth Igne . Modelagem molecular aplicada ao desenvolvimento de moléculas com atividade antioxidante visando ao uso cosmético. RBCF. Revista Brasileira de Ciências Farmacêuticas, v. 43, p. 153-166, 2007.

5. EMERENCIANO, Vicente de Paulo ; BARBOSA, Karina ; SCOTTI, Marcus Tullius ; FERREIRA, Marcelo J P . Self-organizing Maps in Chemotaxonomic Studies of Asteraceae: a Classification of Tribes using Flavonoid Data. Journal of the Brazilian Chemical Society, v. 18, p. 891-899, 2007.
6. EMERENCIANO, Vicente de Paulo ; Diego D. G. ; FERREIRA, Marcelo J P ; SCOTTI, Marcus Tullius ; RODRIGUES, Gilberto V . Computer-Aided Prediction of 125Te and 13C NMR Chemical Shifts of Diorgano Tellurides. Journal of the Brazilian Chemical Society, v. 18, p. 1183-1188, 2007.
7. EMERENCIANO, Vicente de Paulo ; SCOTTI, Marcus Tullius ; STEFANI, Ricardo ; NUZILLARD, Jean Marc ; ALVARENGA, Sandra A V ; RODRIGUES, Gilberto V . Diterpene Skeletal Type Classification and Recognition using Self-Organizing Maps. Internet Eletronic Journal Of Molecular Design, v. 5, n. 4, p. 213-223, 2006.
8. FERREIRA, Marcelo J P ; BARBOSA, Karina ; SCOTTI, Marcus Tullius ; MAGENTA, Mara ; STEFANI, Ricardo ; EMERENCIANO, Vicente de Paulo . Principal Component Analysis Of Heliantheae (Asteraceae) Sensu Stuessy And Karis and Ryding Based On Chemical Data. Natural Products An Indian Journal, v. 2, n. 2, p. 35-44, 2006.
9. EMERENCIANO, Vicente de Paulo ; CABROLBASS, D ; FERREIRA, Marcelo J P ; ALVARENGA, Sandra A V ; BRANT, Antonio J C ; SCOTTI, Marcus Tullius ; BARBOSA, Karina . Chemical Evolution in the Asteraceae. The Oxidation-Reduction Mechanism and Production of Secondary Metabolites. Natural Product Communications, Westerville, OH - USA, v. 1, n. 6, p. 495-507, 2006.
10. EMERENCIANO, Vicente de Paulo ; ALVARENGA, Sandra A V ; SCOTTI, Marcus Tullius ; FERREIRA, Marcelo J P ; STEFANI, Ricardo ; NUZILLARD, Jean Marc . Automatic identification of terpenoid skeletons by feed-forward neural networks. Analytica Chimica Acta, v. 579, n. 2, p. 217-226, 2006.

Resumos em Congressos

1. CORREIA, Mauro Vicentine ; SCOTTI, Marcus Tullius ; EMERENCIANO, Vicente de Paulo . Redes Neurais não supervisionadas utilizadas no estudo Quimiotaxonômico da tribo Heliantheae (Asteraceae). . In: 30ª Reunião Anual da Sociedade Brasileira de Química, 2007, Águas de Lindóia, 2007.
2. FERNANDES, Mariane B ; SCOTTI, Marcus Tullius ; FERREIRA, Marcelo J P ; EMERENCIANO, Vicente de Paulo . Relação quantitativa estrutura - atividade de sesquiterpenos lactonizados com atividade citotóxica. In: 30ª Reunião Anual da Sociedade Brasileira de Química, 2007, Águas de Lindóia, 2007.
3. SCOTTI, Luciana ; SCOTTI, Marcus Tullius ; CARDOSO, Carmen Lucia ; PAULETTI, Patrícia Mendonça ; GAMBOA, Ian Castro ; BOLZANI, Vanderlan da Silva ; VELASCO, Maria Valéria Robles ; MENEZES, Carla Maria de ; FERREIRA, Elizabeth Igne . Análise das superfícies eletrônicas obtidas em compostos de atividade antioxidante extraídos da espécie nacional Arrabidaea samydoides. In: 30ª Reunião Anual da Sociedade Brasileira de Química, 2007, Águas de Lindóia, 2007.
4. SCOTTI, Luciana ; SCOTTI, Marcus Tullius ; CARDOSO, Carmen Lucia ; PAULETTI, Patrícia Mendonça ; GAMBOA, Ian Castro ; BOLZANI, Vanderlan da Silva ; VELASCO, Maria Valéria Robles ; MENEZES, Carla Maria de ; FERREIRA, Elizabeth Igne . Estudo quimiométrico de compostos extraídos de plantas nacionais com atividade antioxidante utilizando-se o programa VolSurf. In: 30ª Reunião Anual da Sociedade Brasileira de Química, 2007, Águas de Lindóia, 2007.
5. ISHIKI, Hamilton Mitsugo ; SCOTTI, Marcus Tullius ; ISHIKI, Renata R. ; SCOTTI, Luciana ; EMERENCIANO, Vicente de Paulo . Estudo de Relação-Quantitativa Estrutura Química-Atividade Biológica de Flavonóides com Atividade Anti-tripanossoma.. In: 30ª Reunião Anual da Sociedade Brasileira de Química, 2007, Águas de Lindóia, 2007.
6. SCOTTI, Marcus Tullius ; EMERENCIANO, Vicente de Paulo ; SCOTTI, Luciana ; ISHIKI, Renata R. ; ISHIKI, Hamilton Mitsugo . Emprego de algoritmo genético em estudos de QSAR de O-(2-fenóxi)etil-N-aralquilcarbamatos com atividade herbicida. In: 30ª Reunião Anual da Sociedade Brasileira de Química, 2007, Águas de Lindóia, 2007.
7. SCOTTI, Luciana ; SCOTTI, Marcus Tullius ; ISHIKI, Hamilton Mitsugo ; FERREIRA, Marcelo J P ; FERREIRA, Elizabeth Igne ; EMERENCIANO, Vicente de Paulo . Quantitative elucidation of the structure-bitterness relationship in sesquiterpene lactone. In: 6th International Congress of Pharmaceutical Sciences, 2007, Ribeirão Preto. 6th International Congress of Pharmaceutical Sciences, 2007.
8. SCOTTI, Marcus Tullius ; FERNANDES, Mariane B ; FERREIRA, Marcelo J P ; EMERENCIANO, Vicente de Paulo . Use of self-organizing maps and molecular descriptors to predict the cytotoxic activity

of sesquiterpene lactones. In: 6th International Congress of Pharmaceutical Sciences, 2007, Ribeirão Preto. 6th International Congress of Pharmaceutical Sciences, 2007.
9. SCOTTI, Marcus Tullius ; FERREIRA, Marcelo J P ; STEFANI, Ricardo ; EMERENCIANO, Vicente de Paulo . Quantitative Relationship Between Oxidation of Diterpenes and 3D Molecular Descriptors in Asteraceae Family. In: 1st BRAZILIAN CONFERENCE ON NATURAL PRODUCTS (1st BCNP), 2007, São Pedro. 1st BRAZILIAN CONFERENCE ON NATURAL PRODUCTS (1st BCNP), 2007. v. 1.
10. EMERENCIANO, Vicente de Paulo ; SCOTTI, Marcus Tullius ; FERREIRA, Marcelo J P ; CORREIA, Mauro Vicentine ; ALVARENGA, Sandra A V ; RODRIGUES, Gilberto V . Self-Organizing Maps as Tool For Taxonomic Classifications at Lower Hierarchical Level. In: 1st BRAZILIAN CONFERENCE ON NATURAL PRODUCTS (1st BCNP), 2007, São Pedro. 1st BRAZILIAN CONFERENCE ON NATURAL PRODUCTS, 2007.
11. Souza, Amanda ; SCOTTI, Marcus Tullius ; Young, Maria Cláudia Marx ; EMERENCIANO, Vicente de Paulo ; Moreno, Paulo Roberto H. . Principal Components Analysis for Determination of the Seasonal Variation in the Volatile Oil Composition from Myrcia macropoda DC (Myrtaceae). In: 1st BRAZILIAN CONFERENCE ON NATURAL PRODUCTS (1st BCNP), 2007, São Pedro. 1st BRAZILIAN CONFERENCE ON NATURAL PRODUCTS (1st BCNP), 2007.
12. FERREIRA, Marcelo J P ; SCOTTI, Marcus Tullius ; EMERENCIANO, Vicente de Paulo . Prediction of Anti-Inflammatory Activity of Sesquiterpene Lactones Using Self-Organizing Maps and 13C NMR Data. In: 1st BRAZILIAN CONFERENCE ON NATURAL PRODUCTS (1st BCNP), 2007, São Pedro. 1st BRAZILIAN CONFERENCE ON NATURAL PRODUCTS (1st BCNP), 2007.
13. SCOTTI, Luciana ; SCOTTI, Marcus Tullius ; Pasqualoto, Kerly Fernanda ; FERREIRA, Elizabeth Igne ; EMERENCIANO, Vicente de Paulo . Use of Self-Organizing Maps of Flavonoids and Analogues with Antiprotozoal Activities. In: 1st BRAZILIAN CONFERENCE ON NATURAL PRODUCTS (1st BCNP), 2007, São Pedro. 1st BRAZILIAN CONFERENCE ON NATURAL PRODUCTS (1st BCNP), 2007.
14. Rossini, Michelle ; SCOTTI, Marcus Tullius ; FERREIRA, Marcelo J P ; STEFANI, Ricardo ; EMERENCIANO, Vicente de Paulo . Self-Organizing Maps to Predict the Anti-Viral Activity of Sesquiterpene Lactones in the Subgenomic HCV Replicons System. In: 1st BRAZILIAN CONFERENCE ON NATURAL PRODUCTS (1st BCNP), 2007, São Pedro. 1st BRAZILIAN CONFERENCE ON NATURAL PRODUCTS (1st BCNP), 2007.
15. SCOTTI, Marcus Tullius ; Muramatsu, Eric ; Rossini, Michelle ; FERREIRA, Marcelo J P ; EMERENCIANO, Vicente de Paulo . 13C NMR Spectral Data and Molecular Descriptors to Predict the Antioxidant Activity of Flavonoids. In: 1st BRAZILIAN CONFERENCE ON NATURAL PRODUCTS (1st BCNP), 2007, São Pedro. 1st BRAZILIAN CONFERENCE ON NATURAL PRODUCTS (1st BCNP), 2007.
16. SCOTTI, Marcus Tullius ; ISHIKI, Hamilton Mitsugo ; Muramatsu, Eric ; CORREIA, Mauro Vicentine ; FERREIRA, Marcelo J P ; EMERENCIANO, Vicente de Paulo . Use of Self-Organizing Maps and 13C NMR Spectral Data to Predict Aldose Reductase Activity of Flavonoids. In: 1st BRAZILIAN CONFERENCE ON NATURAL PRODUCTS (1st BCNP), 2007, São Pedro. 1st BRAZILIAN CONFERENCE ON NATURAL PRODUCTS (1st BCNP), 2007.
17. Cruz, Ana Valéria de Mello ; FERREIRA, Marcelo J P ; SCOTTI, Marcus Tullius ; Kaplan, Maria Auxiliadora C. ; EMERENCIANO, Vicente de Paulo . Chemotaxonomic Relationship Relationships in Celastraceae. In: 1st BRAZILIAN CONFERENCE ON NATURAL PRODUCTS (1st BCNP), 2007, São Pedro. 1st BRAZILIAN CONFERENCE ON NATURAL PRODUCTS (1st BCNP), 2007.
18. EMERENCIANO, Vicente de Paulo ; SCOTTI, Marcus Tullius ; FERREIRA, Marcelo J P ; CORREIA, Mauro Vicentine ; ALVARENGA, Sandra A V ; RODRIGUES, Gilberto V . Chemosystematics of Asteraceae Tribes Using Principal Component Analysis. In: 1st BRAZILIAN CONFERENCE ON NATURAL PRODUCTS (1st BCNP), 2007, São Pedro. 1st BRAZILIAN CONFERENCE ON NATURAL PRODUCTS (1st BCNP), 2007.
19. EMERENCIANO, Vicente de Paulo ; FERREIRA, Marcelo J P ; SCOTTI, Marcus Tullius ; CORREIA, Mauro Vicentine ; ALVARENGA, Sandra A V ; RODRIGUES, Gilberto V . Use of Backpropagation Artificial Neural Networks to Predict the Ocurrences of Chemical Classes in Asteraceae. In: 1st BRAZILIAN CONFERENCE ON NATURAL PRODUCTS (1st BCNP), 2007, São Pedro. 1st BRAZILIAN CONFERENCE ON NATURAL PRODUCTS (1st BCNP), 2007.
20. SCOTTI, Luciana ; SCOTTI, Marcus Tullius ; CARDOSO, Carmen Lucia ; PAULETTI, Patrícia Mendonça ; GAMBOA, Ian Castro ; BOLZANI, Vanderlan da Silva ; VELASCO, Maria Valéria Robles ; MENEZES, Carla Maria de ; FERREIRA, Elizabeth Igne . Estudo Quimiométrico de Compostos Extraídos de Plantas Nacionais com atividade Antioxidante Utilizando-se o Programa Volsurf. In: XII Semana Farmacêutica de Ciência e tecnologia, 2007, São Paulo. XII Semana Farmacêutica de Ciência e tecnologia, 2007.

21. ISHIKI, Hamilton Mitsugo ; SCOTTI, Luciana ; SCOTTI, Marcus Tullius ; ISHIKI, Renata R. ; EMERENCIANO, Vicente de Paulo . Estudo de Relação Quantitativa Estrutura Química-Atividade Biológica de Compostos Naturais Anti-Tripanossoma. In: Encontro de Ensino, Pesquisa e Extensão da Unoeste, 2007, Presidente Prudente. Encontro de Ensino, Pesquisa e Extensão da Unoeste, 2007.
22. CALABRIA, Lalita Maria ; EMERENCIANO, Vicente de Paulo ; SCOTTI, Marcus Tullius ; MABRY, Tom J . Secondary Chemistry of the Compositae Family. In: The International Compositae Alliance, 2006, Barcelona. Secondary Chemistry of the Compositae Family, 2006.
23. SCOTTI, Marcus Tullius ; ISHIKI, Hamilton Mitsugo ; EMERENCIANO, Vicente de Paulo . Estudo da Relação Quantitativa Entre Estrutura Química e Atividade Biológica. In: XI Encontro Anual e de Pesquisa Institucional e de Iniciação Científica, 2006, Presidente Prudente, 2006.

24. SCOTTI, Marcus Tullius ; SCOTTI, Luciana ; EMERENCIANO, Vicente de Paulo ; FERREIRA, Elizabeth Igne ; MENEZES, Carla . Estudos de modelagem molecular e QSAR de structure-bitterness relationship em sesquiterpenos lactonizados. In: 29ª Reunião Anual da Sociedade Brasileira de Química, 2006, Águas de Lindóia, 2006.
25. SCOTTI, Marcus Tullius ; SCOTTI, Luciana ; FERREIRA, Elizabeth Igne ; MENEZES, Carla ; VELASCO, Maria Valéria ; BOLZANI, Vanderlan da Silva . Estudos de QSAR de compostos com atividade antioxidante extraídos de plantas pertencentes às famílias Chimarrhis turbinata e Arrabidea semydoides. In: 29a Reunião Anual da Sociedade Brasileira de Química, 2006, Águas de Lindóia, 2006.
26. SCOTTI, Marcus Tullius ; ISHIKI, Hamilton Mitsugo ; EMERENCIANO, Vicente de Paulo . Redes neurais não supervisionadas para classificação de séries de compostos extraídos de plantas com atividade anti-câncer. In: 29ª Reunião Anual da Sociedade Brasileira de Química, 2006, Águas de Lindóia, 2006.
27. SCOTTI, Marcus Tullius ; ISHIKI, Hamilton Mitsugo ; EMERENCIANO, Vicente de Paulo . Aplicação de descritores moleculares em estudos de QSAR de flavonóides com atividade anticâncer. In: 29ª Reunião Anual da Sociedade Brasileira de Química, 2006, Águas de Lindóia, 2006.
28. SCOTTI, Marcus Tullius ; SCOTTI, Luciana ; CARDOSO, Carmen Lucia ; PAULETTI, Patrícia Mendonça ; GAMBOA, Ian Castro ; BOLZANI, Vanderlan da Silva ; VELASCO, Maria Valéria Robles ; MENEZES, Carla Maria de ; FERREIRA, Elizabeth Igne . Estudo de QSAR de Compostos com Atividade Antioxidante Extraídos de Plantas Pertencentes às Espécies Chimarris Turbinata e Arrabidaea Samydoides. In: XI Semana Farmacêutica de Ciência e Tecnologia, 2006, São Paulo, 2006.
29. SCOTTI, Marcus Tullius ; SCOTTI, Luciana ; CARDOSO, Carmen Lucia ; PAULETTI, Patrícia Mendonça ; GAMBOA, Ian Castro ; BOLZANI, Vanderlan da Silva ; MENEZES, Carla Maria de ; VELASCO, Maria Valéria Robles ; FERREIRA, Elizabeth Igne . Chemometrics Studies of Brazilian Natural Products via Volsurf Approach. In: The 3th Brazilian Symposium on Medicinal Chemistry, 2006, São Pedro, 2006.
30. SCOTTI, Marcus Tullius ; ISHIKI, Hamilton Mitsugo ; EMERENCIANO, Vicente de Paulo . Application of Molecular Descriptors to Predict Aldose Reductase Activity by Flavonoids Compounds. In: The 3th Brazilian Symposium on Medicinal Chemistry, 2006, São Pedro, 2006.
31. SCOTTI, Marcus Tullius ; SCOTTI, Luciana ; FERREIRA, Elizabeth Igne ; MENEZES, Carla ; VELASCO, Maria Valéria ; BOLZANI, Vanderlan da Silva . Avaliação quali- e qunatitativa de relação entre estrutura química e antioxidante de compostos da flora brasileira. In: 28ª Reunião Anual da Sociedade Brasileira de Química, 2005, Poços de Caldas, 2005.
32. SCOTTI, Marcus Tullius ; ISHIKI, Hmilton Mitsugo . Estudo de Relações Quantitativas Estrutura-Propriedade de Compostos Orgânicos de Baixo Peso Molecular. In: X Encontro Anual e de Pesquisa Institucional e de Iniciação Científica, 2005, Presidente Rpudente. ENAPI 2005, 2005.
33. SCOTTI, Marcus Tullius ; ISHIKI, Hamilton ; AMARAL, Antonia Tavares Do ; REZENDE, Leandro de . Critérios de seleção de parâmetros estruturais de inibidores da ribonucleotídeo redutase para estudos de QSAR através de análise PLS. In: 27ª Reunião Anual da Sociedade Brasileira de Química, 2004, Salvador - Ba. 27ª Reunião Anual da Sociedade Brasileira de Química, 2004.
34. SCOTTI, Marcus Tullius ; AMARAL, Antonia Tavares Do ; ISHIKI, Hamilton ; REZENDE, Leandro de . Selection Criteria of DRAGON Descriptors for QSAR PLS Models. In: The 2nd Brazilian Symposium on Medicinal Chemistry, 2004, Rio de Janeiro. The 2nd Brazilian Symposium on Medicinal Chemistry, 2004.
35. SCOTTI, Marcus Tullius ; ISHIKI, Hamilton Mitsugo ; REZENDE, Leandro de ; AMARAL, Antônia Tavares Do . Estudos de Relações Quantitativas Estrutura-Atividade de Inibidores da Ribonucleotídeo Redutase de Células Tumorais . In: IX Encontro Anual de Pesquisa Institucional da UNOESTE, 2004, Presidente Prudente. ENAPI 2004, 2004. v. 1. p. 230.

36. XAVIER ,Célio ; SCOTTI, Marcus Tullius . Efeito da adicao de cromia na resistencia mecanica de uma alumina-alfa. In: Congresso Brasileiro de Ceramica, 38, 1994, Blumenau. Congresso Brasileiro de Ceramica, 38. Anais. Sao Paulo : Associacao Brasileira de Ceramica, 1994., 1994. v. 40. p. 25.

Livros Grátis

(http://www.livrosgratis.com.br)

Milhares de Livros para Download:

Baixar livros de Literatura
Baixar livros de Literatura de Cordel
Baixar livros de Literatura Infantil
Baixar livros de Matemática
Baixar livros de Medicina
Baixar livros de Medicina Veterinária
Baixar livros de Meio Ambiente
Baixar livros de Meteorologia
Baixar Monografias e TCC
Baixar livros Multidisciplinar
Baixar livros de Música
Baixar livros de Psicologia
Baixar livros de Química
Baixar livros de Saúde Coletiva
Baixar livros de Serviço Social
Baixar livros de Sociologia
Baixar livros de Teologia
Baixar livros de Trabalho
Baixar livros de Turismo