

## Desenvolvimento do algoritmo de classificação hierárquica (*Winclus*) e aplicação à série de derivados indol-2-carboxilato

Irwin R.A. Menezes<sup>1</sup>, Adriano A.S. Araújo<sup>2</sup>, Sócrates C. H. Cavalcanti<sup>3</sup>,  
Nereide S. Santos-Magalhães<sup>4</sup>, Maria da Paz N. Moreno<sup>4</sup>, Hélio M. de Oliveira<sup>5</sup>,  
Antonio J. Alves<sup>4\*</sup>

<sup>1</sup> Departamento de Química, Universidade Federal de Minas Gerais, <sup>2</sup> Departamento de Farmácia, Universidade de São Paulo, <sup>3</sup> Departamento de Farmácia, Universidade Tiradentes, <sup>4</sup> Departamento de Farmácia, Universidade Federal de Pernambuco, <sup>5</sup> Departamento de Eletrônica de Sistemas, Universidade Federal de Pernambuco

*A Classificação Hierárquica ou Cluster Analysis é um tipo de reconhecimento de padrões aplicável ao planejamento de fármacos. Nesse trabalho, apresenta-se um modelo de classificação aglomerativa, de forma a agrupar substituintes com base nas similaridades dos seus parâmetros físico-químicos. Uma vez agrupados, estuda-se a relação entre o modelo de classificação hierárquica e as respostas biológicas dos compostos, a partir da Relação Quantitativa Estrutura-Atividade (QSAR). O método é representado por um novo algoritmo, denominado Winclus, para o qual foi desenvolvido um programa destinado à sua aplicação. Neste trabalho, é abordado o desenvolvimento do software e a aplicação do mesmo à série, descrita na literatura, de derivados indol-2-carboxilato, com atividade anticonvulsivante. O estudo foi realizado buscando-se estabelecer uma relação entre os agrupamentos dos substituintes, proposto pelo modelo, e as atividades biológicas dos compostos já sintetizados. Para tal, utilizou-se a forma gráfica de dendrograma e observou-se que o programa desenvolvido constitui uma ferramenta de apoio no planejamento de novos compostos biologicamente ativos.*

### Unitermos:

- Classificação hierárquica
- Planejamento de fármacos
- Derivados indol-2-carboxilato
- Winclus

### \*Correspondência:

A. J. Alves  
Universidade Federal de Pernambuco  
Departamento de Farmácia  
Av. Prof. Artur Sá, s/n  
52720-550 - Cidade Universitária,  
Recife - PE  
E-mail: leac@nlink.com.br

## INTRODUÇÃO

Em 1972, Topliss *et al.* descreveram uma *árvore de decisão ou árvore de agrupamentos* como auxílio na escolha racional de grupos substituintes. O objetivo era a seleção de substituintes químicos a partir de uma série inicial, em que eram evitados substituintes bioisósteros, visto que estes seriam providos de informações semelhantes. Um outro método de análise por agrupamento foi proposto por Hansch *et al.* (1973) na tentativa de estabelecer uma relação entre substituintes com base nas suas caracte-

terísticas físico-químicas. Observa-se, assim, que a busca de métodos racionais com o objetivo de selecionar substituintes químicos e estabelecer uma correlação entre eles vem sendo fonte de estudo há algumas décadas.

Com o avanço dos estudos de Relação Quantitativa Estrutura-Atividade (QSAR), tornou-se possível avaliar a influência dos parâmetros físico-químicos de cada grupo substituinte frente à atividade biológica e, assim, racionalizar a síntese de fármacos. Nesse sentido, foram realizados os primeiros trabalhos aplicando cálculo numérico com a finalidade de selecionar substituintes químicos e

planejar novos compostos. Um desses métodos matemáticos é conhecido como Classificação Hierárquica ou *Cluster Analysis*, um tipo de reconhecimento de padrões aplicado à química (Kowalski *et al.*, 1972; Basak *et al.*, 1995). Para este fim, o termo *clustering* é dado como sinônimo de “taxonomia numérica” e “classificação” (Fukunaga, 1972; Mardia *et al.*, 1979).

Uma das maiores dificuldades no desenvolvimento de métodos aplicáveis ao planejamento de fármacos é a análise de banco de dados. Nos últimos anos, diversos modelos de classificação foram propostos com o intuito de reduzir erros de extrapolação, aproximação e aqueles relacionados à dimensão dos bancos em estudo, podendo estes modelos ser classificados em dois grupos: a classificação supervisionada e a não-supervisionada.

A classificação por métodos supervisionados utiliza como principais ferramentas a Regressão Linear Múltipla (MRL), a Regressão por Mínimos Quadrados Parciais (PLS) e a Análise Estatística Isolinear de Multicategorias (SIMCA). Estes aplicativos apresentam a vantagem de produzir modelos capazes de fazer previsões para os dados em análise, uma vez que utiliza variáveis dependentes, como por exemplo a atividade biológica. Porém, apresentam a desvantagem das análises serem afetadas pela correlação entre os dados ou o chamado “efeito por chance” (Levingstone, 1995). Por sua vez, a classificação por métodos não-supervisionados é muito utilizada no estágio preliminar à análise dos dados, principalmente quando o número de variáveis excederem o número de amostras do banco em análise. Entre os diversos aplicativos existentes destacam-se: *k*-NN (*Nearest-Neighbour*), o Fator de Análise e a Análise de *Cluster* ou Classificação Hierárquica (Tomimaga, 1999).

Em 1987, Giacca *et al.* utilizaram um algoritmo de Classificação Hierárquica conhecido como *Average Linkage Clustering*, escrito em linguagem *Basic*, no qual os substituintes testados eram agrupados à medida em que suas similaridades aumentavam. Nesse modelo, matriz era construída e permitia encontrar a mais alta semelhança entre os grupos substituintes. O procedimento era realizado de modo que à cada etapa uma nova matriz de semelhanças era reduzida e computada e, assim, colunas e linhas eram formadas pela média das anteriores. Os resultados dessa seleção de grupos eram apresentados na forma de dendrogramas, ou seja, um diagrama em forma de árvore, onde as ramificações terminais eram dispostas progressivamente de acordo com a entrada dos dados. Cada ramificação que sustentava um grupo de dados isolados era denominada *cluster*.

A partir desses estudos, a Classificação Hierárquica passou a ser aplicada em vários campos do conhecimento científico, sendo muito útil em Planejamento e Controle de Infecção Hospitalar (Culasso *et al.*, 1991), Medicina (Craig

*et al.*, 1971), Planejamento de Fármacos (Topliss, 1993; Hansch *et al.*, 1995; Allen *et al.*, 1992) e Reconhecimento de Padrões (Kowalski *et al.*, 1972, 1973; Diday *et al.*, 1976).

No planejamento de fármacos, os modelos de Classificação Hierárquica vêm sendo aplicados para a seleção racional de grupos substituintes e em estudos de QSAR (Hansch *et al.*, 1979; Kubinyi, 1993). Uma das estratégias para tal fim consiste em se encontrar uma relação entre substituintes ou grupos de substituintes buscando-se obter compostos de alta atividade biológica. Um novo modelo para seleção destes grupos foi proposto por Santos-Magalhães *et al.* (1999), consistindo em um algoritmo que aponta o substituinte ou grupo de substituintes de maior atividade para uma determinada molécula protótipo.

Inicialmente descrito por Cavalcanti (1994), o Algoritmo de Classificação Hierárquica teve como princípio a aplicação da estatística, da matemática e da computação voltada ao planejamento de fármacos. Denominado, na sua primeira versão, por *Newclus* e posteriormente por *Winclus* (versão em Windows®), o programa executa os cálculos estatísticos e matemáticos necessários para agrupar substituintes de acordo com as similaridades dos seus parâmetros físico-químicos. Além disso, o *software* realiza as etapas necessárias para obtenção do dendrograma e seleciona o substituinte que promoverá teoricamente maior atividade à molécula protótipo (Moreno, 1996; Menezes, 2000).

O desenvolvimento de métodos racionais voltados à seleção de grupos substituintes e à síntese de novos fármacos mostra-se como um grande desafio na área de química fina e no planejamento de compostos biologicamente ativos. Nessa perspectiva, esse trabalho tem por objetivo introduzir um modelo modificado de classificação hierárquica visando orientar o pesquisador no estágio precedente à síntese de novas moléculas. Deve-se ressaltar que por meio da metodologia proposta é possível agrupar dois ou mais substituintes em uma só etapa, o que simplifica o procedimento e a análise dos dados.

## MÉTODO

### Modelo matemático e desenvolvimento do Algoritmo de Classificação Hierárquica

No método de classificação hierárquica, considera-se um número *K* de parâmetros e um número *N* de possíveis substituintes por sítio de substituição da molécula protótipo. Os dados dos parâmetros físico-químicos de todos os substituintes são, assim, descritos por matriz  $X(i, j)$ , onde *i* varia de 1 a *N* e *j* de 1 a *K*. Para avaliação das distância entre dois substituintes adotou-se a noção de distância Euclidiana, pelo fato dessa ser descrita na literatura com maior frequência (Hansch *et al.*, 1979; Levingstone, 1995).

No entanto, outros métodos também poderiam ser utilizados em procedimentos hierárquicos, como por exemplo: Karl Pearson, Mahalanobis e Battacharya (Moreno, 1995).

Sabendo-se que os parâmetros físico-químicos de um determinado grupo de substituintes não necessariamente estão numa mesma escala original e que cada conjunto de parâmetros define um ponto no espaço bidimensional ou multidimensional, torna-se necessário a normalização dos seus valores. Com a normalização, os parâmetros passam a estar numa mesma escala e cada um desses pontos pode ser comparado em termos de distância.

### Normalização

A normalização é realizada a cada etapa de agrupamento, ou seja, durante todo o processo de construção do dendrograma. Os cálculos para normalização são efetuados a partir das médias e desvios-padrão dos parâmetros físicos-químicos.

- Cálculo das médias e desvios-padrão dos valores de X, ou seja, para cada parâmetro determinado por  $j$  (Equações I e II):

$$\bar{X}_j = \frac{\sum_{i=1}^N X_{(i,j)}}{N} \quad \text{e} \quad d.p.j = \sqrt{\frac{\sum_{i=1}^N (X_{(i,j)} - \bar{X}_j)^2}{N}} \quad (\text{Equações I e II})$$

- Cálculo dos valores normalizados para cada X ( $i, j$ ) (Equação III):

$$N_{(i,j)} = \frac{X_{(i,j)} - \bar{X}_j}{d.p.j} \quad (\text{Equação III})$$

onde N é o número de substituintes,  $X_{(i,j)}$  é o valor a ser normalizado,  $\bar{X}_j$  é o valor da média para cada  $j$ ,  $N_{(i,j)}$  é o valor normalizado e  $d.p.j$  é o desvio-padrão para os valores, sendo  $j$  fixo.

Uma vez realizada a normalização, obtém-se uma nova matriz (N x K) com N objetos como substituintes e K variáveis como parâmetros. A matriz representativa é

mostrada a seguir (Figura 1).

$$N_{(n,m)} = \begin{bmatrix} N_{(1,1)} & N_{(1,2)} & N_{(1,3)} & \dots & N_{(1,m)} \\ N_{(2,1)} & N_{(2,2)} & N_{(2,3)} & \dots & N_{(2,m)} \\ N_{(3,1)} & N_{(3,2)} & N_{(3,3)} & \dots & N_{(3,m)} \\ \dots & \dots & \dots & \dots & \dots \\ N_{(n,1)} & N_{(n,2)} & N_{(n,3)} & \dots & N_{(n,m)} \end{bmatrix}$$

FIGURA 1 - Matriz de dados normalizados.

### Matriz de distâncias

A matriz formada pelo programa representa todas as distâncias entre os substituintes químicos, em relação às suas similaridades físico-químicas. Para obtê-la, é calculada a distância Euclidiana entre todos os substituintes e, por consequência, tem-se matriz de diagonal nula devido à distância entre um mesmo indivíduo ser igual a zero. Por outro lado, a mesma distância relativa entre dois substituintes ( $X_{n,m}$  e  $X_{m,n}$ ) conduz a matriz simétrica. Assim, a matriz de distâncias possui três características fundamentais: diagonal principal nula, simétrica e quadrática de N x N, em que N é o número de substituintes.

As distâncias são calculadas com base na definição de distância Euclidiana, que é dada pela Equação IV.

$$d_{(i,j)} = \sqrt{\sum_{k=1}^N (X_{(i,k)} - X_{(j,k)})^2} \quad (\text{Equação IV})$$

onde  $X_{(i,k)}$  e  $X_{(j,k)}$  são os pontos para os quais são calculadas as distâncias, sendo  $d_{(i,j)}$  a distância entre dois substituintes.

Sabendo-se que  $d_{(i,j)}$  são as distâncias, tem-se: para  $i = j$ ,  $d_{(i,j)} = 0$  e para  $d_{(k,w)} = d_{(w,k)}$ , como mostra a Figura 2.

$$d(K, W) = \begin{bmatrix} d_{(1,1)} & d_{(1,2)} & d_{(1,3)} & \dots & d_{(1,N)} \\ d_{(2,1)} & d_{(2,2)} & d_{(2,3)} & \dots & d_{(2,N)} \\ d_{(3,1)} & d_{(3,2)} & d_{(3,3)} & \dots & d_{(3,N)} \\ \dots & \dots & \dots & \dots & \dots \\ d_{(N,1)} & d_{(N,2)} & d_{(N,3)} & \dots & d_{(N,N)} \end{bmatrix} \quad \text{ou} \quad \begin{bmatrix} 0 & d_{(1,2)} & d_{(1,3)} & \dots & d_{(1,N)} \\ d_{(2,1)} & 0 & d_{(2,3)} & \dots & d_{(2,N)} \\ d_{(3,1)} & d_{(3,2)} & 0 & \dots & d_{(3,N)} \\ \dots & \dots & \dots & \dots & \dots \\ d_{(N,1)} & d_{(N,2)} & d_{(N,3)} & \dots & 0 \end{bmatrix}$$

FIGURA 2 - Matriz de distâncias d(K,W).

### Distância admissível de similaridades

Esta etapa define o critério de agrupamento dos substituintes, em que a matriz de distâncias é analisada em busca de um limiar denominado de *distância admissível*. Esta análise é feita com a finalidade de se obter a distância “limite” para que os substituintes sejam considerados absolutamente próximos e tidos como *vizinhos*.

Um substituinte  $Si_1$  é dito *vizinho* de um outro substituinte  $Si_2$  se e somente se a distância entre eles for menor ou igual a *distância admissível*. Esta distância é definida como uma quota limite abaixo da qual todos os substituintes podem ser agrupados. Maneira de calcular a *distância admissível* consiste em considerar os menores valores das distâncias de  $S_i$  para cada linha da matriz e, dentre esses, o maior valor como sendo a distância mínima admissível para que um substituinte possa ser considerado *vizinho* e, assim, agrupado.

### Vizinhanças

São considerados *vizinhos* os objetos que possuem distância menor ou igual à distância mínima admissível. Assim, são tidos como *vizinhos* aqueles substituintes cujo  $d \leq d_{\min}$ , sendo considerado o mais próximo de determinado objeto X o próprio X. As vizinhanças são representadas da seguinte maneira (Figura 3):

$$\begin{aligned} \Gamma 1 &= \{1, v_i = 2, \dots, n\} \\ \Gamma 2 &= \{2, v_i = 2, \dots, n\} \\ &\cdot \\ &\cdot \\ &\cdot \\ \Gamma n &= \{n, v_i = 2, \dots, n\} \end{aligned}$$

**FIGURA 3** - Representação da matriz de vizinhança entre os substituintes.

em que  $n$  representa o número de vizinhanças, dispostas em ordem crescente de proximidade. Observa-se, ainda, que  $v_i$  (primeiro *vizinho*) é o próprio substituinte e  $Gn$  representa a vizinhança do substituinte químico de número  $n$ .

### Obtenção do dendrograma e seleção do(s) composto(s) potencialmente ativo(s)

A árvore de agrupamentos possui forma gráfica, cujo eixo horizontal representa os substituintes e o eixo

vertical os níveis de redução da árvore. Os agrupamentos são realizados a cada etapa, tendo como consequência a redução proporcional do número de linhas e colunas da matriz. É importante destacar que para cada sítio de substituição da molécula protótipo é construído um dendrograma.

Uma vez obtido o dendrograma e de posse da equação de QSAR, é possível selecionar o caminho a ser percorrido na árvore para atingir o substituinte ou grupo de substituintes de maior atividade. Para tal, deve-se substituir os valores dos parâmetros dos *pseudo-substituintes* na equação de QSAR e dirigir o caminho da árvore sempre para o *pseudo-substituinte* de maior atividade. O programa *Winclus* conta com um algoritmo em linguagem *Basic*, denominado *Newbasic*, que calcula as *pseudo-atividades* de todos os nós da árvore direcionando o caminho de cada ramificação em direção àquele que apresenta maior atividade.

### Apresentação do software (Winclus)

O *Winclus* foi desenvolvido utilizando-se como ferramenta de programação a linguagem Delphi® 1.0, visto que, dessa forma, estaria apto a ser executado tanto nas versões antigas do Windows® (Windows 3.X) quanto nas versões mais atualizadas (Windows 95 e Windows 98). Além de fornecer os resultados necessários à construção da árvore, o *Winclus* descreve o programa de forma interativa e didática.

Numa fase inicial aplicou-se o programa a pouco mais de dez séries químicas, descritas na literatura, com a finalidade de comparar os resultados obtidos e comprovar a eficácia do modelo. Entre as séries testadas pode-se destacar: 2-aril-1,3-indanodionas (Berg *et al.*, 1975), derivados do ácido propiônico 2,4-(tiazol-2-il)fenil (Naito *et al.*, 1992), fenetilaminas (Unger *et al.*, 1973), ácidos *trans*-3-benzoilacrilícos (Bowden *et al.*, 1972), trifluorometanosulfonânílicas (Yapel, 1972), *N*-2-fenilguanidinas (Gambino *et al.*, 1992) e guanidiotiazolcarboxamidas (Schnur *et al.*, 1991).

### Etapas do Algoritmo de Classificação Hierárquica (Winclus)

- 1 - Dimensionar a entrada de dados e exibir o número limite de  $i$  e  $j$ , dada a matriz de dados  $X(i, j)$ ;
- 2 - Entrada de dados digitados em formato de matriz, exposto na tela ou por meio de arquivo externo;
- 3 - Armazenar em arquivo externo a matriz de dados utilizando formato ASCII;
- 4 - Obtenção dos valores normalizados da matriz de da-

dos original que contém os valores dos parâmetros físico-químicos para cada substituinte, sendo esses parâmetros os mesmos da equação de QSAR (devem ser calculadas médias e desvios);

- 5 - Obtenção da matriz de distâncias, calculando as distâncias Euclidianas entre todos os pares de substituintes;
- 6 - Obtenção de *vizinhos*, classificando em ordem decrescente de aproximação absoluta todos os substituintes;
- 7 - Armazenar em arquivo externo o conjunto de *vizinhos* (formato ASCII);
- 8 - Realizar agrupamentos ordenados, calculando as médias entre os *vizinhos* absolutamente próximos;
- 9 - Continuar agrupando os substituintes próximos até que se encerre o processo de redução (conjunto de

agrupamentos de substituintes ou *pseudo-substituintes*);

- 10 - Finalizados os agrupamentos de uma redução, armazenar em arquivo externo a matriz de dados resultante;
- 11 - Iniciar nova redução executando os passos de 4 a 8;
- 12 - Finalizar ao ser obtida a matriz de dados de dimensão  $X(2,2)$ ;
- 13 - Construir e exibir na tela o dendrograma de todo o processo de classificação hierárquica dos substituintes.

O fluxograma do *Winclus* está dividido em três etapas, que são representadas pelas Figuras 4, 5 e 6. Essas

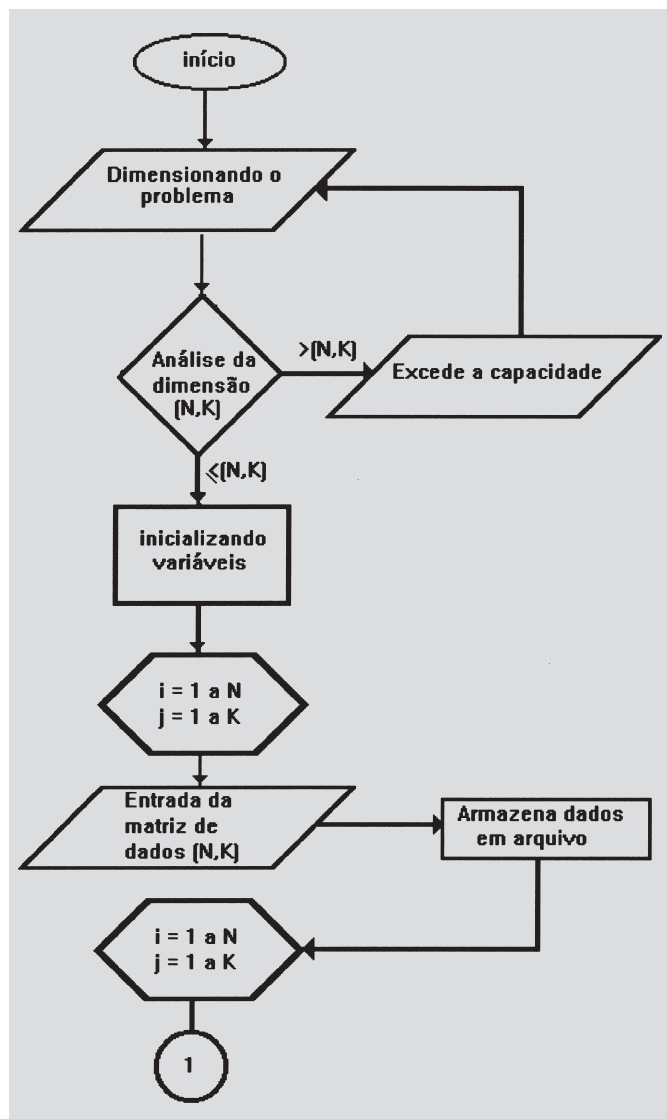


FIGURA 4 - Primeira parte do fluxograma do programa *Winclus*.

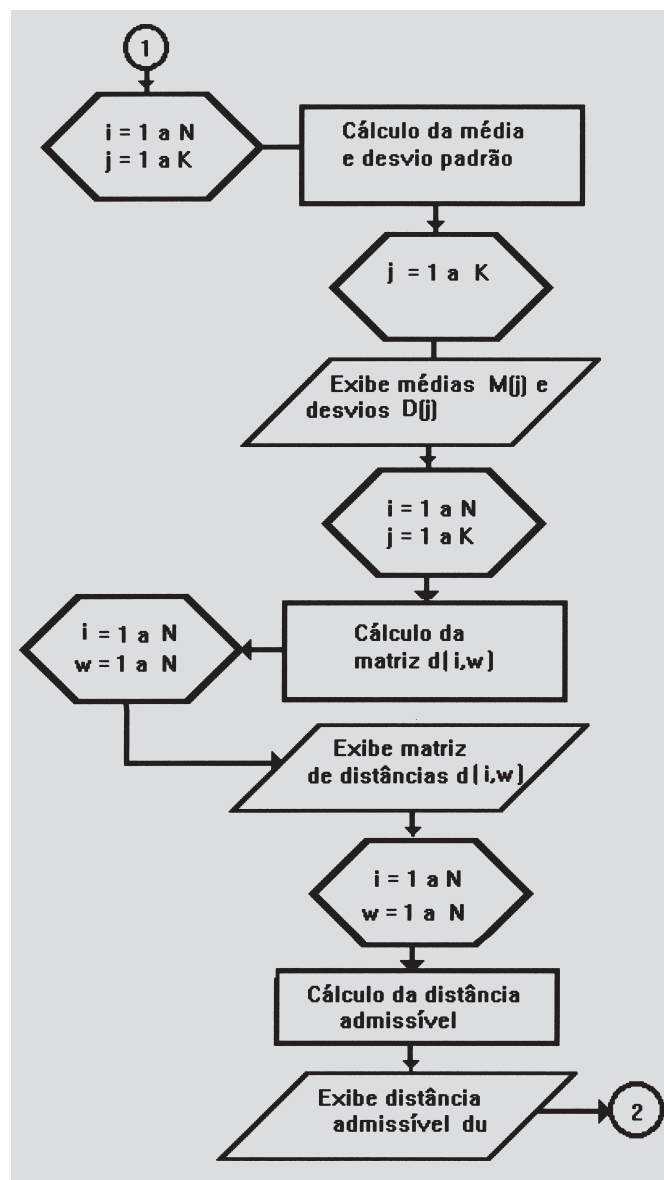


FIGURA 5 - Segunda parte do fluxograma do programa *Winclus*.

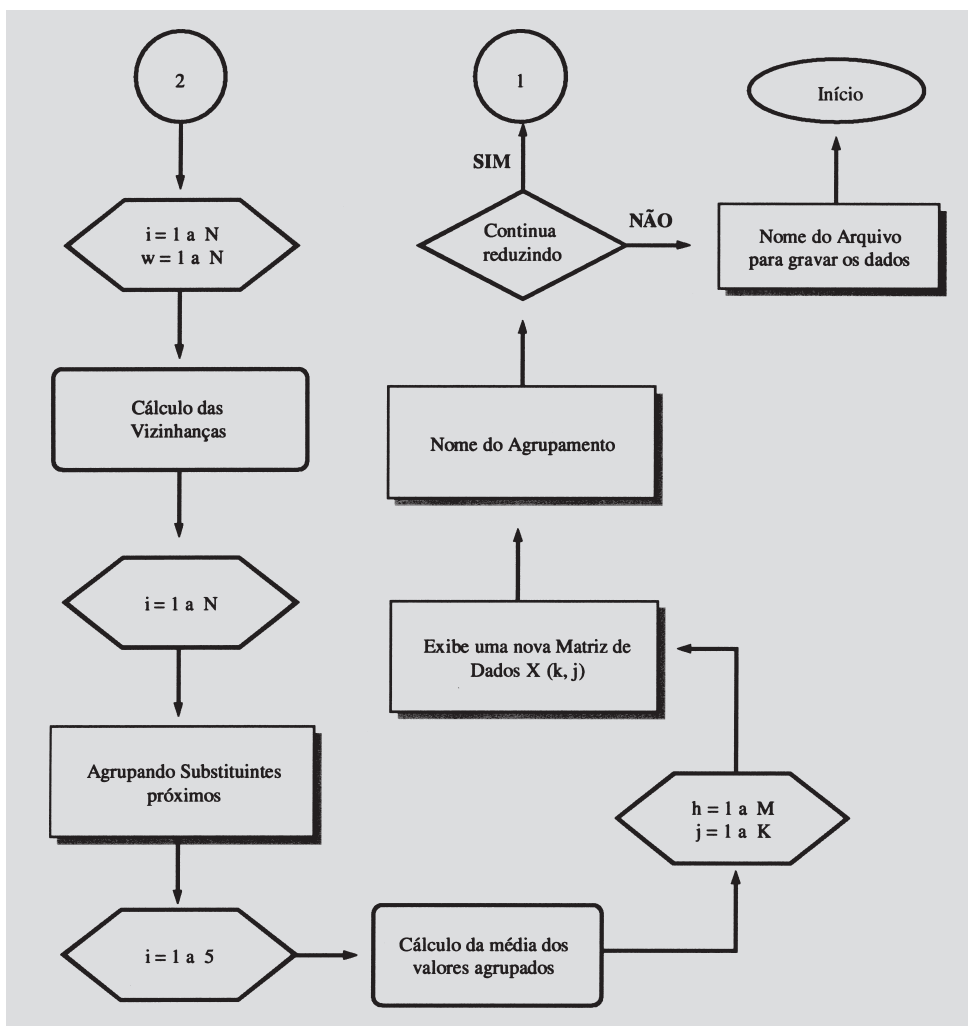


FIGURA 6 -Terceira parte do fluxograma do programa *Winclus*.

figuras estão ligadas entre si pelos números indicados no final de cada fluxograma.

A tela de execução inicia com o *Menu*, onde o usuário escolhe as opções desejadas para a introdução dos dados e finaliza com o último agrupamento, retornando à tela de opções. A Figura 7 ilustra a apresentação e execução do *Winclus* em suas diversas etapas.

## RESULTADOS E DISCUSSÃO

Com a finalidade de demonstrar e discutir os procedimentos efetuados pelo programa, propôs-se a aplicação à série de derivados indol-2-carboxilato (Figura 8), com potencial atividade anticonvulsivante. Essa série foi selecionada da literatura pelo fato de já ter sido sintetizada e ter suas atividades conhecidas, facilitando, assim, a comparação entre os dados experimentais e teóricos. O estudo foi realizado inicialmente por Romano *et al.* (1997),

baseando-se no modelo de Hansch para analisar a contribuição dos parâmetros lipofílicos, eletrônicos e estéreos, sendo considerados, respectivamente,  $p$ ,  $s$  e  $MR$  e adotado o sítio de substituição na posição *para*.

A Equação V foi descrita por Romano *et al.* (1977), sendo estabelecida por meio da Análise de Regressão Múltipla (MRA), de modo a representar os efeitos dos derivados indólicos em associação com o complexo receptor ionotrópico *N*-metil-D-aspartato (NMDA), responsável pelo efeito farmacológico. Os valores de  $pK_i$  representam a afinidade observada experimentalmente dos compostos frente ao sítio de ligação com a glicina. Essa equação relaciona-se, quantitativamente, à influência dos parâmetros físico-químicos frente à atividade biológica:

$$pK_i = -0.53 MR_{omp} - 0.39 \pi_{omp} - 0.82 \sigma_{para} + 8.23 \text{ (Eq. V)}$$

$$n = 25 \quad R^2 = 0.84 \quad s = 0.28 \quad F = 37$$

$$p < 0.0001 \quad R^2_{cv} = 0.76$$

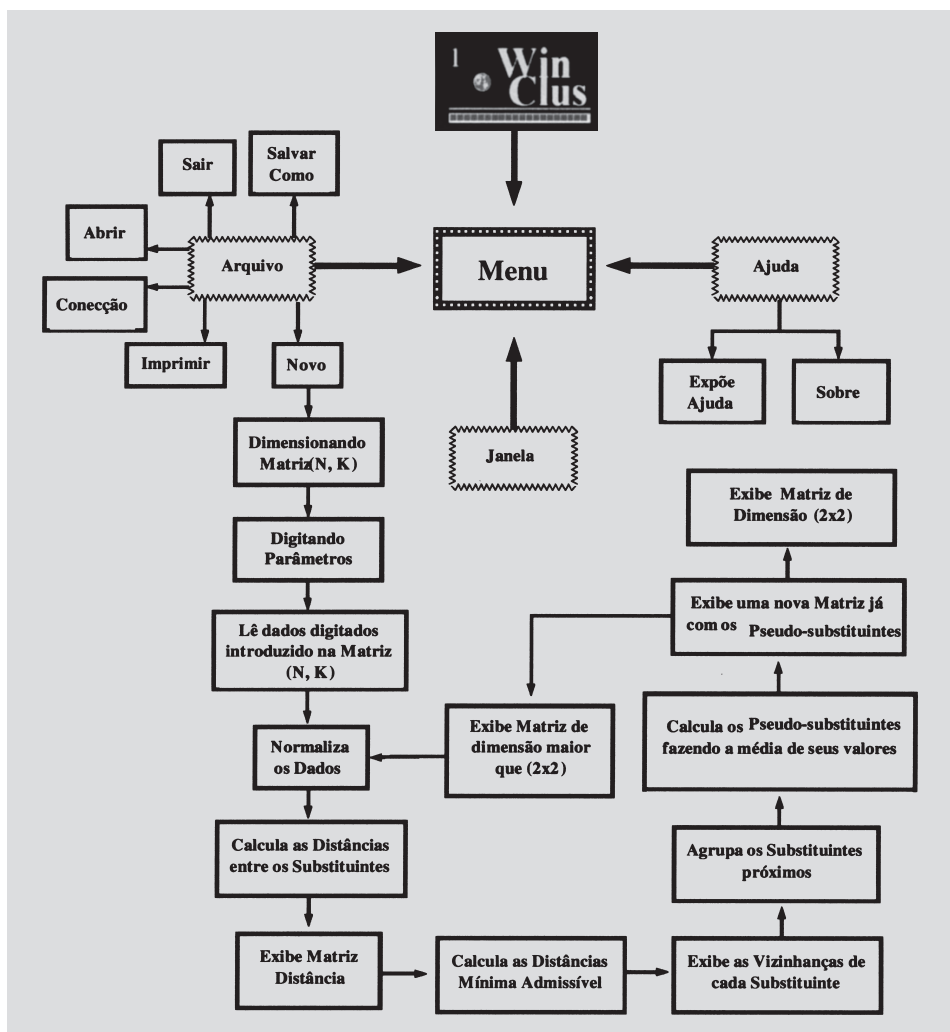


FIGURA 7 - Esquema de apresentação do programa *Winclus*.

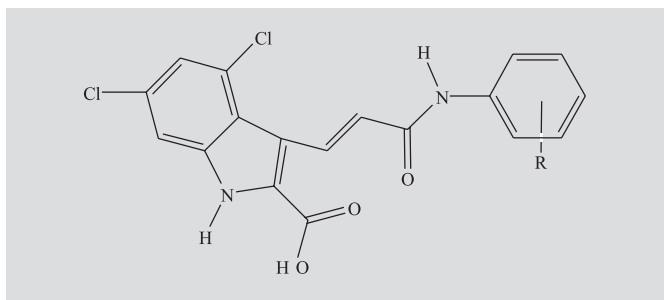


FIGURA 8 - Estrutura química da série de derivados indol-2-carboxilato.

em que,  $MR_{omp}$  representa o efeito estéreo dos substituintes para as posições *orto*, *meta* e *para*,  $\pi_{omp}$  é a contribuição do caráter lipofílico nas posições substituição *orto*, *meta* e *para*, e, finalmente,  $\sigma_{para}$  corresponde ao efeito eletrônico do grupo substituinte no sítio de substituição *para* (Hansch *et al*, 1979, 1975; Hammett, 1970). Os da-

dos estatísticos da equação apresentam o número de compostos ( $n$ ), o coeficiente de correlação ( $R^2$ ) para um intervalo de confiança de 90%, a probabilidade de  $R$  ser zero ( $p$ ), o coeficiente de validação cruzada ( $R^2_{cv}$ ), o desvio padrão da estimativa ( $s$ ) e a significância estatística global da equação ( $F$ ) (Wold, 1991).

Para aplicação ao programa *Winclus* foram realizados os seguintes procedimentos: 1) entrada de dados com um número  $N=12$  de substituintes e  $K=3$  de parâmetros (Tabela I); 2) os dados dos parâmetros físico-químicos de todos os substituintes foram descritos como matriz  $X(i, j)$ , onde  $i$  varia de 1 a 12 e  $j$  varia de 1 a 3 (Tabela II); 3) a distância entre os pontos foi dada pela distância Euclidiana e, assim, calculada a *distância admissível* para que os grupos fossem considerados *vizinhos* e conseqüentemente agrupados (Tabela III).

A Tabela I mostra a entrada dos dados e os resultados dos agrupamentos de substituintes na posição *para*,

em que os 12 substituintes são enumerados segundo sua seqüência original. Com a formação dos *pseudo-substituintes*, a tabela de dados é modificada (reduzida) e utilizada para calcular uma nova matriz de distâncias, ou

seja, à medida que os substituintes são agrupados, uma nova tabela de dados passa a ser fonte para calcular a matriz de distâncias seguinte e, assim, o número de grupos considerado (N) diminui até um único grupo.

**TABELA I** - Tabela de parâmetros físico-químicos dos substituintes e as correspondentes reduções a *pseudo-substituintes* da aplicação do *Winclus* à série de indol-2-carboxilato na posição *para*

| Nº  | Substituintes   | MR <sub>omp</sub> | $\pi_{omp}$ | $\sigma_{para}$ |
|---|---|-------------------|-------------|-----------------|
| 1   | H   | 0,0000            | 0,0000      | 0,0000          |
| 2   | NHCONH <sub>2</sub>   | 1,2400            | -1,3000     | -0,2400         |
| 3   | NH <sub>2</sub>   | 0,3700            | -1,2300     | -0,6600         |
| 4   | OH  | 0,1500            | -0,6700     | -0,3700         |
| 5   | COOH  | 0,6500            | 0,0400      | 0,4500          |
| 6   | N(CH <sub>3</sub> ) <sub>2</sub>  | 1,2900            | 0,2000      | -0,8300         |
| 7   | NO <sub>2</sub>   | 0,7200            | 0,3000      | 0,7800          |
| 8   | F   | 0,0100            | 0,4000      | 0,0600          |
| 9   | O(C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub>  | 1,0800            | 0,6100      | -0,2400         |
| 10  | N(C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub>  | 2,2200            | 1,0900      | -0,9000         |
| 11  | CF <sub>3</sub>   | 0,5100            | 1,3300      | 0,5400          |
| 12  | NHPh  | 2,8800            | 1,4800      | -0,4000         |
| Primeiro Agrupamento com os pseudo-substituintes formados |   |                   |             |                 |
| 1   | H / F   | 0,0050            | 0,2000      | 0,0300          |
| 2   | NHCONH <sub>2</sub>   | 1,2400            | -1,3000     | -0,2400         |
| 3   | NH <sub>2</sub> / OH  | 0,2600            | -0,9500     | -0,5150         |
| 4   | COOH / NO <sub>2</sub>  | 0,6850            | 0,1700      | 0,6150          |
| 5   | N(CH <sub>3</sub> ) <sub>2</sub> / O(C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub>   | 1,1850            | 0,4050      | -0,5350         |
| 6   | N(C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub> / NHPh   | 2,5500            | 1,2850      | -0,6500         |
| 7   | CF <sub>3</sub>   | 0,5100            | 1,3300      | 0,5400          |
| Segundo Agrupamento com os pseudo-substituintes formados  |   |                   |             |                 |
| 1   | H / F   | 0,0050            | 0,2000      | 0,0300          |
| 2   | NHCONH <sub>2</sub> / NH <sub>2</sub> / OH  | 0,7500            | -1,1250     | -0,3775         |
| 3   | COOH / NO <sub>2</sub> / CF <sub>3</sub>  | 0,5975            | 0,7500      | 0,5775          |
| 4   | N(CH <sub>3</sub> ) <sub>2</sub> / O(C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub>   | 1,1850            | 0,4050      | -0,5350         |
| 5   | N(C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub> / NPh  | 2,5500            | 1,2850      | -0,6500         |
| Terceiro Agrupamento com os pseudo-substituintes formados |   |                   |             |                 |
| 1   | H / F / COOH / NO <sub>2</sub> / CF <sub>3</sub>  | 0,3013            | 0,4750      | 0,3038          |
| 2   | NHCONH <sub>2</sub> / NH <sub>2</sub> / OH  | 0,7500            | -1,1250     | -0,3775         |
| 3   | N(CH <sub>3</sub> ) <sub>2</sub> / O(C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub>   | 1,1850            | 0,4050      | -0,5350         |
| 4   | N(C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub> / NPh  | 2,5500            | 1,2850      | -0,6500         |
| Quarto Agrupamento com os pseudo-substituintes formados   |   |                   |             |                 |
| 1   | H / F / COOH / NO <sub>2</sub> / CF <sub>3</sub>  | 0,3013            | 0,4750      | 0,3038          |
| 2   | NHCONH <sub>2</sub> / NH <sub>2</sub> / OH / N(CH <sub>3</sub> ) <sub>2</sub> / O(C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub>  | 0,9675            | -0,3600     | -0,4563         |
| 3   | N(C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub> / NPh  | 2,5500            | 1,2850      | -0,6500         |
| Quinto Agrupamento com os pseudo-substituintes formados   |   |                   |             |                 |
| 1   | H / F / COOH / NO <sub>2</sub> / CF <sub>3</sub> / NHCONH <sub>2</sub> / NH <sub>2</sub> / OH / N(CH <sub>3</sub> ) <sub>2</sub> / O(C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub> | 0,6344            | 0,0575      | -0,0763         |
| 2   | N(C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub> / NPh  | 2,5500            | 1,2850      | -0,6500         |



As matrizes de distâncias dos substituintes em *para* são mostradas na Tabela II. O cálculo da distância admissível é ilustrado considerando a matriz inicial, que corresponde ao primeiro grupo de vizinhanças (Tabela III). Para exemplificar, é descrito o cálculo da *distância admissível* para a matriz inicial, em que os menores valores não nulos de distâncias de cada uma das linhas da primei-

ra matriz são: 0,4552; 1,2566; 0,8568; 0,8568; 0,6821; 1,2103; 0,6821; 0,4552; 1,2103; 1,2665; 1,2434; 1,2665. Dentre estes, a maior distância é considerada como sendo a admissível deste nível de agrupamento. Sendo assim, os substituintes que possuem uma distância inferior a “1,2665” são agrupados. Os menores valores de cada linha estão em negrito e o maior dentre eles em negrito sublinhado.

**TABELA II** - Matrizes de distância obtidas pela aplicação do *Winclus* à série de indol-2-carboxilato na posição *para*

|           |               |        |               |               |               |               |               |               |               |               |        |               |
|-----------|---------------|--------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|--------|---------------|
|           | 1             | 2      | 3             | 4             | 5             | 6             | 7             | 8             | 9             | 10            | 11     | 12            |
| <b>1</b>  | 0,0000        | 2,0522 | 1,8766        | 1,0239        | 1,1135        | 2,1328        | 1,6952        | <b>0,4552</b> | 1,4614        | 3,2430        | 1,8686 | 3,7108        |
| <b>2</b>  | 2,0522        | 0,0000 | <b>1,2566</b> | 1,4328        | 2,0675        | 1,9862        | 2,6567        | 2,3984        | 2,1145        | 3,1110        | 3,3464 | 3,5935        |
| <b>3</b>  | 1,8766        | 1,2566 | 0,0000        | <b>0,8568</b> | 2,5149        | 1,9143        | 3,1902        | 2,2783        | 2,3175        | 3,3319        | 3,6028 | 4,1453        |
| <b>4</b>  | 1,0239        | 1,4328 | <b>0,8568</b> | 0,0000        | 1,8049        | 1,8183        | 2,4765        | 1,4344        | 1,7754        | 3,1924        | 2,8100 | 3,8869        |
| <b>5</b>  | 1,1135        | 2,0675 | 2,5149        | 1,8049        | 0,0000        | 2,4940        | <b>0,6821</b> | 1,0977        | 1,5092        | 3,2836        | 1,4414 | 3,3688        |
| <b>6</b>  | 2,1328        | 1,9862 | 1,9143        | 1,8183        | 2,4940        | 0,0000        | 3,0648        | 2,2075        | <b>1,2103</b> | 1,4426        | 2,9699 | 2,4185        |
| <b>7</b>  | 1,6952        | 2,6567 | 3,1902        | 2,4765        | <b>0,6821</b> | 3,0648        | 0,0000        | 1,5642        | 1,9700        | 3,6586        | 1,2434 | 3,5279        |
| <b>8</b>  | <b>0,4552</b> | 2,3984 | 2,2783        | 1,4344        | 1,0977        | 2,2075        | 1,5642        | 0,0000        | 1,3496        | 3,1592        | 1,4722 | 3,5538        |
| <b>9</b>  | 1,4614        | 2,1145 | 2,3175        | 1,7754        | 1,5092        | <b>1,2103</b> | 1,9700        | 1,3496        | 0,0000        | 1,8549        | 1,7745 | 2,2652        |
| <b>10</b> | 3,2430        | 3,1110 | 3,3319        | 3,1924        | 3,2836        | 1,4426        | 3,6586        | 3,1592        | 1,8549        | 0,0000        | 3,3111 | <b>1,2665</b> |
| <b>11</b> | 1,8686        | 3,3464 | 3,6028        | 2,8100        | 1,4414        | 2,9699        | <b>1,2434</b> | 1,4722        | 1,7745        | 3,3111        | 0,0000 | 3,1984        |
| <b>12</b> | 3,7108        | 3,5935 | 4,1453        | 3,8869        | 3,3688        | 2,4185        | 3,5279        | 3,5538        | 2,2652        | <b>1,2665</b> | 3,1984 | 0,0000        |

|          |        |               |               |               |               |        |               |
|----------|--------|---------------|---------------|---------------|---------------|--------|---------------|
|          | 1      | 2             | 3             | 4             | 5             | 6      | 7             |
| <b>1</b> | 0,0000 | 2,1478        | 1,5809        | <b>1,3817</b> | 1,7761        | 3,4440 | 1,6065        |
| <b>2</b> | 2,1478 | 0,0000        | <b>1,3166</b> | 2,2963        | 1,7901        | 3,1014 | 3,1369        |
| <b>3</b> | 1,5809 | <b>1,3166</b> | 0,0000        | 2,4939        | 1,7338        | 3,5053 | 3,0584        |
| <b>4</b> | 1,3817 | 2,2963        | 2,4939        | 0,0000        | 2,3015        | 3,4611 | <b>1,1814</b> |
| <b>5</b> | 1,7761 | 1,7901        | <b>1,7338</b> | 2,3015        | 0,0000        | 1,8439 | 2,3994        |
| <b>6</b> | 3,4440 | 3,1014        | 3,5053        | 3,4611        | <b>1,8439</b> | 0,0000 | 3,3191        |
| <b>7</b> | 1,6065 | 3,1369        | 3,0584        | <b>1,1814</b> | 2,3994        | 3,3191 | 0,0000        |

|          |               |        |        |               |        |
|----------|---------------|--------|--------|---------------|--------|
|          | 1             | 2      | 3      | 4             | 5      |
| <b>1</b> | 0,0000        | 1,8570 | 1,3983 | <b>1,6885</b> | 3,2253 |
| <b>2</b> | 1,8570        | 0,0000 | 2,8325 | <b>1,7917</b> | 3,3245 |
| <b>3</b> | <b>1,3983</b> | 2,8325 | 0,0000 | 2,3368        | 3,2467 |
| <b>4</b> | <b>1,6885</b> | 1,7917 | 2,3368 | 0,0000        | 1,7480 |
| <b>5</b> | 3,2253        | 3,3245 | 3,2467 | <b>1,7480</b> | 0,0000 |

|          |        |               |               |        |
|----------|--------|---------------|---------------|--------|
|          | 1      | 2             | 3             | 4      |
| <b>1</b> | 0,0000 | 2,2992        | <b>2,1651</b> | 3,3149 |
| <b>2</b> | 2,2992 | 0,0000        | <b>1,6277</b> | 3,0947 |
| <b>3</b> | 2,1651 | <b>1,6277</b> | 0,0000        | 1,6765 |
| <b>4</b> | 3,3149 | 3,0947        | <b>1,6765</b> | 0,0000 |

|          |               |               |        |
|----------|---------------|---------------|--------|
|          | 1             | 2             | 3      |
| <b>1</b> | 0,0000        | <b>1,9069</b> | 2,8878 |
| <b>2</b> | <b>1,9069</b> | 0,0000        | 2,4545 |
| <b>3</b> | 2,8878        | <b>2,4545</b> | 0,0000 |

|          |               |               |
|----------|---------------|---------------|
|          | 1             | 2             |
| <b>1</b> | 0,0000        | <b>2,4495</b> |
| <b>2</b> | <b>2,4495</b> | 0,0000        |

**TABELA III** - Distâncias admissíveis obtidas pela aplicação do *Winclus* à série de indol-2-carboxilato na posição *para*

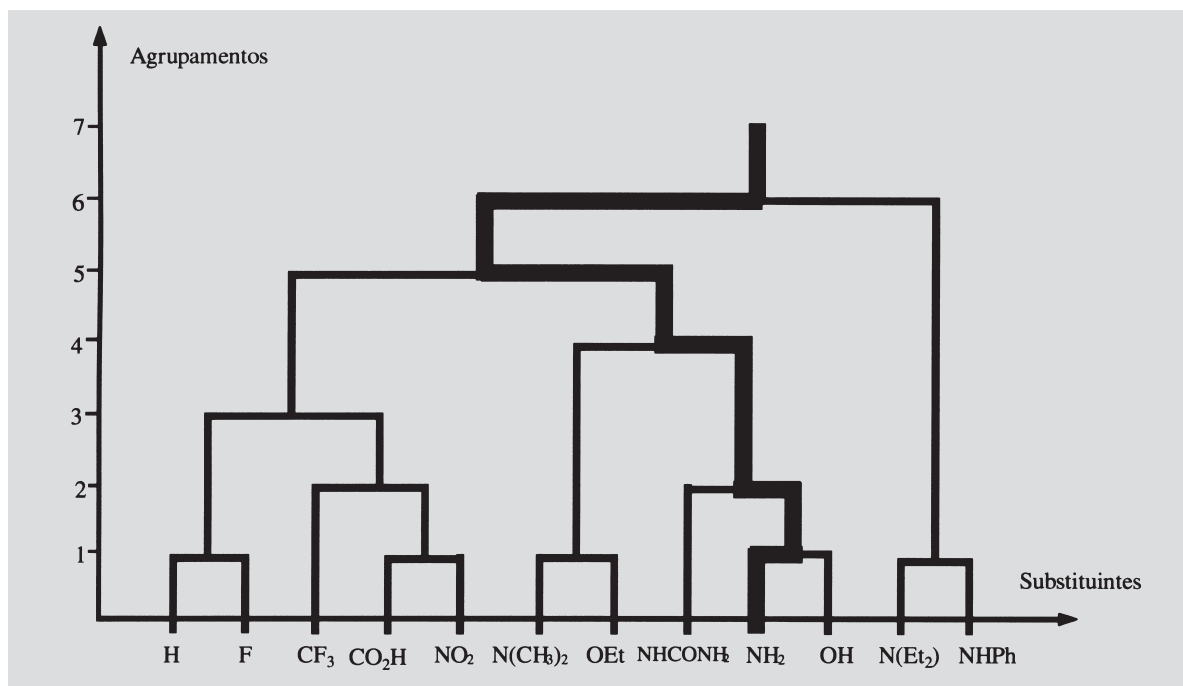
| Agrupamentos | Substituintes agrupados                       | Nº de grupos | Distância Admissível ( <i>du</i> ) |
|--------------|---|--------------|------------------------------------|
| 1º           | (1, 8) (2) (3, 4) (5, 7) (6, 9) (10, 12) (11) | 7            | 1.2665                             |
| 2º           | (1, 8) (2, 3, 4) (5, 7, 11) (6, 9) (10, 12)   | 5            | 1.8439                             |
| 3º           | (1, 8, 5, 7, 11) (2, 3, 4) (6, 9) (10, 12)    | 4            | 1.7917                             |
| 4º           | (1, 8, 5, 7, 11) (2, 3, 4, 6, 9) (10, 12)     | 3            | 2.1651                             |
| 5º           | (1, 8, 5, 7, 11, 2, 3, 4, 6, 9) (10, 12)      | 2            | 2.4545                             |

A Tabela III ilustra os resultados dos agrupamentos dos substituintes na posição *para*, obedecendo a ordem da Tabela I quanto à numeração. Pode ser observado que para cada redução (conjunto de agrupamentos), tem-se uma *distância admissível* correspondente. Nesta tabela os substituintes agrupados encontram-se entre parênteses.

A Figura 9 apresenta o dendrograma obtido com a simulação computacional à série de indol-2-carboxilato. Nesse, pode ser traçado o caminho que representa a seleção do substituinte ou grupo de substituintes, que promoverão teoricamente maior atividade biológica quando introduzidos na molécula protótipo. Essa seleção foi realizada substituindo-se os valores dos parâmetros dos *pseudo-substituintes* de cada ramificação na equação de QSAR (Equação V) e decidindo-se a partir do maior valor de *pseudo-atividade* que caminho percorrer até a base da árvore.

Na aplicação à série de indol-2-carboxilato, com sítio de substituição em *para*, observa-se que o substituinte selecionado ( $\text{NH}_2$ ), a partir do cálculo de *pseudo-atividade*, está entre aqueles de maior atividade observada experimentalmente (Tabela IV). Assim, acredita-se ser válida a representação dos grupos de similaridades nos dendrogramas calculados pelo programa.

A Tabela IV mostra os valores das atividades biológicas calculadas e obtidas experimentalmente para cada composto. Como pode ser visto, a partir dos resultados de  $\Delta p_i$  não se observam diferenças significativas na variação entre os cálculos das atividades. Esses dados podem ser comparados, também, de acordo com o Gráfico I, em que o maior erro encontrado entre as atividades não foi superior a 7,0%, fortalecendo a confiabilidade do método no planejamento de novas moléculas.

**FIGURA 9** - Dendrograma e seleção do substituinte mais ativo obtido pelo programa *Winclus* para a série de derivados indol-2-carboxilato, com substituição em *para*.

De acordo com a Tabela IV e o Gráfico I, pode-se observar a existência de boa correlação entre resultados das atividades teóricas e observadas experimentalmente para os compostos. Cerca de 81% dos compostos apresentaram resultados teóricos semelhantes àqueles observados experimentalmente.

**TABELA IV** - Valores experimentais e calculados para o  $pK_i$  na série de derivados indol-2-carboxilatos, com ação anticonvulsivante

| Substituintes                                  | Atividade Biológica |                    | $\Delta p_i$ |
|--|---------------------|--------------------|--------------|
|  | Experimental        | Calculada          |              |
| H  | 8,5 ( $\pm 0,19$ )  | 8,2 ( $\pm 0,19$ ) | 0,27         |
| F  | 8,2 ( $\pm 0,12$ )  | 8,0 ( $\pm 0,19$ ) | 0,18         |
| CF <sub>3</sub>                                | 6,8 ( $\pm 0,13$ )  | 6,9 ( $\pm 0,19$ ) | -0,19        |
| CO <sub>2</sub> H                              | 7,2 ( $\pm 0,21$ )  | 7,5 ( $\pm 0,19$ ) | -0,30        |
| NO <sub>2</sub>                                | 7,0 ( $\pm 0,06$ )  | 7,1 ( $\pm 0,19$ ) | -0,09        |
| N(CH <sub>3</sub> ) <sub>2</sub>               | 7,9 ( $\pm 0,18$ )  | 8,2 ( $\pm 0,19$ ) | -0,25        |
| O(C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub> | 8,3 ( $\pm 0,48$ )  | 7,6 ( $\pm 0,19$ ) | 0,68         |
| NHCONH <sub>2</sub>                            | 8,8 ( $\pm 0,37$ )  | 8,3 ( $\pm 0,19$ ) | 0,52         |
| NH <sub>2</sub>                                | 8,9 ( $\pm 0,11$ )  | 9,1 ( $\pm 0,19$ ) | -0,16        |
| OH   | 8,7 ( $\pm 0,04$ )  | 8,8 ( $\pm 0,19$ ) | -0,05        |
| N(C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub> | 7,0 ( $\pm 0,26$ )  | 7,4 ( $\pm 0,19$ ) | -0,37        |
| NHPh   | 6,7 ( $\pm 0,17$ )  | 6,5 ( $\pm 0,19$ ) | 0,24         |

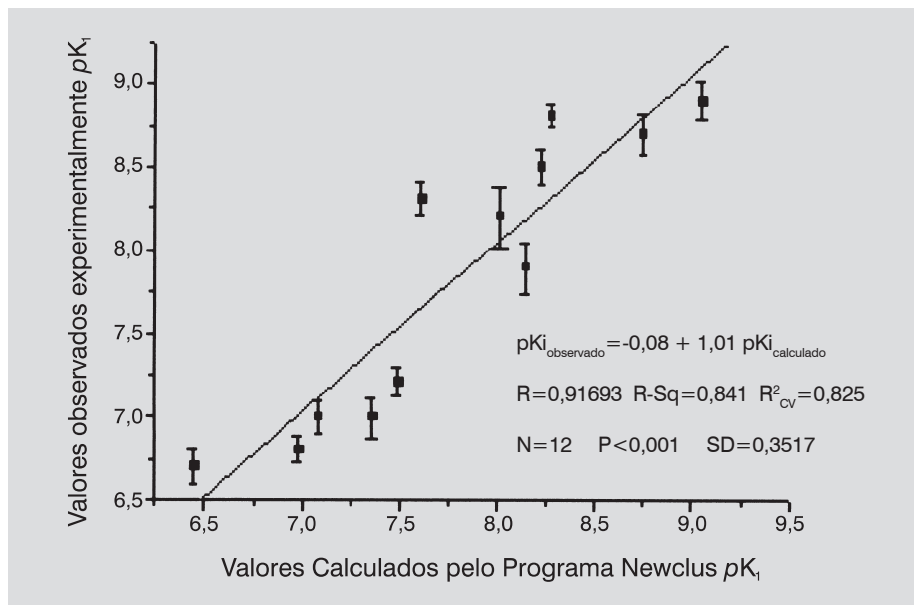
$\Delta p_i$  representa a diferença entre a atividade experimental e a calculada pelo programa.

## CONCLUSÃO

O Algoritmo de Classificação Hierárquica apresentado, aliado aos estudos de QSAR, constitui ferramenta importante na seleção de uma série de substituintes e, consequentemente, na procura de fármacos potencialmente mais ativos. A partir de diversas simulações ao programa pode-se observar que as similaridades entre os substituintes crescem com a proximidade da base do dendrograma, visto que os agrupamentos tendem a um único grupo cada vez mais heterogêneo. Além disso, a aplicação do *Winclus* à série de derivados indol-2-carboxilato, descrita na literatura por Romano *et al.* (1997), mostrou direcionar as ramificações do dendrograma ao substituinte ou grupo de substituintes que promoverão maior atividade à série.

A partir da implementação do algoritmo, abrem-se novas perspectivas voltadas ao planejamento de moléculas potencialmente mais ativas. O desenvolvimento do *software* e os resultados obtidos até o momento estimulam a criação de modelos cada vez mais sofisticados e que possam elucidar a influência de parâmetros físico-químicos sobre atividade, toxicidade e outros parâmetros de importância para a relação estrutura-atividade.

Além disso, o *Winclus* possibilita a introdução da interdisciplinaridade entre as áreas de Química Medicinal, Matemática, Estatística e Informática. Pode, ainda, ser utilizado no ensino de graduação e pós-graduação para ajudar o estudante de Farmácia e Química a entender e utilizar métodos computacionais como auxílio à pesquisa de síntese e planejamento de fármacos.



**GRÁFICO I** - Relação entre as atividades calculadas e experimentais para a série de derivados indol-2-carboxilato substituídos na posição *para*.

**ABSTRACT****Development of a Hierarchical Classification Algorithm (Winclus) and its application to an indole-2-carboxylate series**

*In this work we wish to present the application of a hierarchical clustering approach to drug design. Computer software first described by Cavalcanti is used as a tool to cluster substituents based on their physicochemical parameters. Once clustered, these substituents are used to study the relationships between this classification method and the biological activities of the studied compounds. With this objective, a chemical series, to which a Quantitative Structure-Activity Relationship (QSAR) had been previously performed, was used. A new software, named Winclus, was developed to facilitate our work. With this method it is also possible to analyze the relationship between an agglomerative clustering and the biological response of compounds. The procedure must be applied to series having known QSAR based on hydrophobic, electronic and/or stereochemical parameters. The model was applied to a series of substituted indole-2-carboxylates obtained from the literature. Our intent was to mainly test the model, making it executable and verifying at the same time its efficiency when clustering data sets is furnished.*

**UNITERMS:** Cluster Analysis. Drug design. Substituted indole-2-carboxylates. Winclus.

**AGRADECIMENTOS**

Os autores agradecem a Nelson Moreno pela colaboração na transcrição do programa, resultando no software Winclus e aos órgãos financiadores do projeto: FAPESP, CNPq e CAPES.

**REFERÊNCIAS BIBLIOGRÁFICAS**

- ALLEN, M. S., LALOGGIA, A. J., DORN, L. J., MARTIN, M. J., COSTANTINO, G., HAGEN, T. J., KOEHLER, K. F., SKOLNICK, P., COOK, J. M. Predictive binding of beta-carboline inverse agonists and antagonists via the CoMFA/GOLPE approach. *J. Med. Chem.*, v. 35, n. 22, p. 4001-4010, 1992.
- BASAK, S. C., GRUNWALD, G. D. Molecular Similarity and Estimation of Molecular Properties. *J. Chem. Inf. Comput. Sci.*, v. 35, n.3, p. 366-372, 1995.
- BERG, G. V., BULTSMA, T., REKKER, R. F., TOMAS, W. Structure-Activity-Relationships of 2-aryl-1,3-indandiones. *Eur. J. Med. Chem.-Chimica Ther.*, v. 10, n.3, p. 242-246, 1975.
- BOWDEN, K., HENRY, M. P. Structure activity relations. II. Antibacterial activity of trans-3-benzoylacrylic acids and esters. In: VAN VALKENBURG, W., ed. *Biological correlation's-the Hansch approach*. Washington: American Chemical Society, 1972. p. 130-142.
- CAVALCANTI, S. C. H. *Pesquisa de compostos ativos pela determinação de caminhos em árvore de agrupamento*. Recife, 1994. 116p. [Dissertação de Mestrado – Departamento de Ciências Farmacêuticas. Centro de Ciências da Saúde. Universidade Federal de Pernambuco].
- CRAIG, P. N., HANSCH, C. H., MCFARLAND, J. W., MARTIN, Y. C., PURCELL, W. P., ZAHRADNIK, R. Minimal statistical data for structure-function correlation's. *J. Med. Chem.*, v. 14, n. 5, p. 447, 1971.
- CULASSO, F., LENZI, A., FAVILIL, S., DONDERO, F. Statistical analysis in andrology. *Arch. Androl.*, v. 26, n. 3, p. 163-172, 1991.
- DIDAY, E., SIMON, J. C. Clustering analysis. In: *Communication and Cybernetics 10-digital pattern recognition*. Heidelberg: Springer-Verlag, 1976. p. 47-56.
- FUKUNAGA, K. *Introduction to Statistical Pattern Recognition*. London: Academic Press, 1972. p. 369-381.
- GAMBINO, J., FOCHER, F., HILDEBRAND, C., MAGA, G., NOONAN, T., SPADARI, S., WRIGHT, G. Quantitative Structure-Activity-Relationships of *N*-2-phenylguanines as inhibitors of herpes simplex virus thymine kinases. *J. Med. Chem.*, v. 35, p. 2979-2983, 1992.
- GIACCA, M., MENZO, S., TROJAN, S., MONTI-BRAGADIN, C. Cluster Analysis of Antibiotics Susceptibility patterns of clinical isolates as a tool in nosocomial infection surveillance. *Eur. J. Epidemiol.*, v. 2, n. 3, p. 155-163, 1987.
- HAMMETT, L. P. *Physical Organic Chemistry*. New York: McGraw-Hill, 1970. 404 p.
- HANSCH, C., UNGER, S. H., FORSYTHE, A. B. Strategy in drug design. Cluster analysis as an aid in the selection of substituents. *J. Med. Chem.*, v. 11, n.16, p. 1217-1222, 1973.

- HANSCH, C., LEO, A. *Substituent constants for correlation Analysis in Chemistry and Biology*. New York: John Wiley & Sons, 1979. 327 p.
- HANSCH, C., LEO, A. *Substituent constants for correlation analysis in chemistry and biology*. Washington: A.C.S. Professional Ref. Book, D.C., 1995. 557 p.
- HANSCH, C., LEO, A., HOEKMAN, D. *Exploring Q.S.A.R.: Hydrophobic, electronic and steric constants*. Washington: A.C.S. Professional Ref. Book, 1995. 348 p.
- KOWALSKI, B. R., BENDER, C. F. - Pattern Recognition. A powerful approach to interpreting chemical data. *J. Amer. Chem. Soc.*, v. 94, p. 5632-5639, 1972.
- KOWALSKI, B. R., BENDER, C. F. - Patter Recognition II – Linear and Nonlinear methods for displaying chemical data. *J. Amer. Chem. Soc.*, v. 95, p. 686-693, 1973.
- KUBINYI, H. QSAR: *Hansch analysis and related approaches*. Weinheim: VCH, 1993. 240 p.
- LEVINGSTONE, D. *Data Analysis for Chemistry*. New York: Oxford University Press, 1995. 96 p.
- MARDIA, K., KENT, J. T., BIBBY, J. M. *Multivariate Analysis*. London: Academic Press, 1979. 518 p.
- MENEZES, I. R. A. *WINCLUS: aplicação de reconhecimento de padrões na procura de compostos potencialmente bioativos*. Recife, 2000. 106 p. [Dissertação de Mestrado. Departamento de Ciências Farmacêuticas. Centro de Ciências da Saúde. Universidade Federal de Pernambuco].
- MORENO, M. N. P. – *Planejamento de fármacos utilizando método de classificação hierárquica*. Recife, 1995. 109p. [Dissertação de Mestrado. Departamento de Ciências Farmacêuticas. Centro de Ciências da Saúde. Universidade Federal de Pernambuco].
- MORENO, M. P. N., MAGALHÃES, N. S. S., CAVALCANTI, S. C. H., ALVES, A. J. Algoritmo de Classificação Hierárquica aplicado ao planejamento de fármacos. *Quím. Nova*, v. 19, n.6, p. 594-599, 1996.
- NAITO, Y., YAMAURA, Y., INOUE, Y., FUKAYA, C., YOKOYAMA, K., NAKAGAWA, Y., FUJITA, T. Quantitative Structure-Activity Relationships of 2-[4-(thiazol-2-yl)phenyl] propionic acid-derivatives inhibiting cyclooxygenase. *Eur. J. Med. Chem.*, v. 27, p. 645-654, 1992.
- ROMANO, F., CAPELLI, A. M., CONTI, N.A., CUGOLA, A., DONATI, D., FERIANI, A., GASTALDI, P., GAVIRAGHI, G., HEWKIN, C.T., MICHELI, F., MISSIO, A., MUGNAINI, M., PECUNIOSO, A., QUAGLIA, A.M., RATTI, E., ROSSI, L., TEDESCO, G., TRIST, D.G., REGGIANI, A. Substituted Indol-2-carboxylates as in vivo potent antagonist acting as the Strychnine-insensitive glycine binding site. *J. Med. Chem.*, v. 40, p. 841-850, 1997.
- SANTOS-MAGALHAES, N. S., HOLANDA-CAVALCANTI, S. C., MENEZES, I. R. A., SOUZA-ARAÚJO, A. A., OLIVEIRA, H. M., ALVES, A. J. Automated Search for Potentially Active Compounds by Using Cluster Trees. *Eur. Jour. Med. Chem.*, v. 34, n.2, p. 83-92, 1999.
- SCHNUR, R. C., GALLASCHUN, R. J., SINGLETON, D. H., GRISSOM, M., SLOAN, D. E., GOODWIN, P., McNIFF, P. A., FLIRI, A. F. J., MANGANO, F. M., OLSON, T. H., POLLACK, V. A. Quantitative-Structure-Activity-Relationships of antitumor guanodinothiazolecarboxamides with survival enhancement for therapy in the 3LL Lewis Lung-carcinoma model. *J. Med. Chem.*, v. 34, n.7, p. 1975-1982, 1991.
- TOMINAGA, Y. Comparative study of class data analysis with PCA-LDA, SIMCA, PLS, ANNs and *k*-NN. *Chemimet. Intell. Lab. Syst.* v. 49, p. 105-115, 1999.
- TOPLISS, J. G., COSTELLO, R. J. Change Correlation's Structure-Activity Studies using multiple regression analysis. *J. Med. Chem.*, v. 15, n.10, p. 1066-1068, 1972.
- TOPLISS, J. G. *Some Observations on Classical Q.S.A.R. – Perspective in Drug Discovery and Design*. v. 1, p. 253-257, 1993.
- UNGER, S. H., HANSCH, C. Model-Building in Structure-activity Relationships re-examination of adrenergic blocking activity of beta-halo-beta-arylalkylamines. *J. Med. Chem.* v. 16, p. 745-749, 1973.
- YAPEL JUNIOR, A. F. Structure-activity correlation's for *meta* and *para* substituted trifluoromethane sulfonamide pre-emergence herbicide. In: *Biological correlation's-The Hansch approach*. Washington: American Chemical Society, 1972. p. 183-194.
- WOLD, S. Validation of Q.S.A.R. *Quantitative Structure and activity Relationships*, v. 10, p. 191-193, 1991.