

**U. PORTO**



**FACULDADE DE FARMÁCIA  
UNIVERSIDADE DO PORTO**

**Serviço de Bioquímica**

**Estudo *in silico* da relação estrutura-actividade de  
piperazinas nos receptores 5-HT<sub>1A</sub> e 5-HT<sub>2A</sub>**

**Curso de Mestrado em Toxicologia Analítica Clínica e Forense**

**António Machado**

**FFUP, 2006/2008**

**Este trabalho não inclui as críticas nem as sugestões feitas pelo júri**

**Orientadores:**

**Professora Doutora Irene Jesus Rebelo**

**Dr. Eduardo Tejera**

**Directora do Curso:**

**Professora Doutora Maria de Lourdes Bastos**

# *Agradecimentos*

Agradeço à Professora Doutora Maria Irene Jesus Rebelo e ao Dr. Eduardo Tejera, os meus orientadores neste projecto computacional, por todo o conhecimento e auxílio concedido ao longo deste ano, bem como pela simpatia e compreensão demonstrada, nos momentos mais árduos, tornando-os mais agradáveis e fáceis de ultrapassar.

Agradeço à Universidade do Porto, em particular à Faculdade de Farmácia, por me ter facultado as condições necessárias para a concretização deste trabalho no tempo previsto.

Aos meus colegas e grupo de trabalho do Serviço de Bioquímica da Faculdade de Farmácia da Universidade do Porto, um obrigado especial por todo o companheirismo demonstrado.

É com enorme prazer que agradeço ao meu grupo de amigos em Espinho, pela sua amizade única e momentos inesquecíveis que juntos compartilhamos e ao meu grupo de amigos de Cuba, em especial o Eduardo e a Valentina, por todo o apoio e companheirismo ao longo deste percurso académico.

Aos meus melhores amigos, aos meus pais, um muito obrigado por todo o apoio, carinho, compreensão e, acima de tudo, por esta oportunidade que nunca lhes foi concedida – a formação académica. Agradeço à minha mãe por toda a esperança depositada em mim e ao meu pai por todas as palavras sábias e ensinamentos que me transmitiu.

À minha doce irmã, que representa a minha esperança de um futuro melhor, um beijo especial.

**A todos a minha sincera gratidão,**

***António Machado***



## Índice

|  |            |
|--|------------|
| Índice de figuras .....  | ii         |
| Índice de tabelas .....  | iv         |
| Abreviaturas e acrónimos .....   | v          |
| <b>Resumo .....</b>  | <b>vii</b> |
| <br>   |            |
| <b>Parte I – Introdução</b> .....  | <b>1</b>   |
| <b>1 – Relação estrutura-actividade</b> .....  | <b>2</b>   |
| 1.1 – Abordagem do estudo de interacções ligando-receptor.....   | 3          |
| 1.2 – Metodologia geral para o estudo QSAR.....  | 5          |
| 1.2.1 – Selecção de ligandos e optimização estrutural .....  | 5          |
| 1.2.2 – Métodos gerais para a obtenção de parâmetros moleculares.....  | 8          |
| <i>Descritores da estrutura molecular</i> .....  | 8          |
| 1.2.3 – Campos de interacção molecular (MIF) .....   | 12         |
| 1.2.3.1 – Potenciais de interacção molecular (MIP).....  | 12         |
| 1.2.4 – Tratamento estatístico e métodos matemáticos aplicados em QSAR.....  | 18         |
| 1.2.4.1 – Simplificação e selecção de variáveis .....  | 19         |
| 1.2.4.2 – Modelos de regressão utilizados em QSAR .....  | 22         |
| <i>Regressão dos componentes principais e mínimos quadrados parciais</i> .....   | 23         |
| <i>Redes neuronais artificiais</i> .....   | 24         |
| 1.2.4.2 – Validação cruzada (Cross-validation) .....   | 26         |
| <b>2 – As piperazinas e os farmacóforos serotoninérgicos 1A e 2A</b> .....   | <b>28</b>  |
| 2.1 – Estudo de farmacóforos para o receptor 5-HT <sub>1A</sub> .....  | 29         |
| 2.2 – Estudo de farmacóforos para o receptor 5-HT <sub>2A</sub> .....  | 31         |
| <br>   |            |
| <b>Parte II – Estudo in silico da relação estrutura-actividade de piperazinas nos receptores 5-HT<sub>1A</sub> e 5-HT<sub>2A</sub></b> ..... | <b>33</b>  |
| <b>3 – Objectivos</b> .....  | <b>34</b>  |
| <b>4 – Materiais e métodos</b> .....   | <b>35</b>  |
| 4.1 – Materiais e suporte informático .....  | 35         |
| 4.2 – Selecção e optimização geométrica da família de piperazinas.....   | 35         |
| 4.3 – Obtenção de descritores estruturais .....  | 39         |
| 4.4 – Métodos estatísticos aplicados .....   | 41         |
| 4.4.1 – Análise de conjuntos (Clusters Analysis).....  | 41         |
| 4.4.2 – Selecção das variáveis para cada receptor .....  | 43         |
| 4.4.3 – Métodos de regressão e de validação cruzada .....  | 44         |
| <b>5 – Resultados e discussão</b> .....  | <b>45</b>  |
| 5.1 – Análise dos modelos preditivos para os receptores 5-HT .....   | 47         |
| 5.2 – Análise dos clusters para os receptores 5-HT <sub>1A</sub> .....   | 53         |
| 5.3 – Análise dos clusters para os receptores 5-HT <sub>2A</sub> .....   | 57         |
| 5.4 – Afinidade das piperazinas para os receptores 5-HT <sub>1A</sub> e 5-HT <sub>2A</sub> .....   | 60         |
| <b>6 – Conclusões</b> .....  | <b>64</b>  |
| <b>7 – Perspectivas futuras</b> .....  | <b>65</b>  |
| <br>   |            |
| <b>Bibliografia</b> .....  | <b>66</b>  |
| <b>Anexos</b> .....  | <b>77</b>  |

## Índice de figuras

|   |    |
|---|----|
| Figura 1 - Esquema ilustrativo dos passos essenciais na elaboração de um estudo QSAR e construção do modelo preditivo.....  | 5  |
| Figura 2 - Exemplo ilustrativo: A) dos ângulos $\Phi$ e $\Psi$ do pentano; e B) da paisagem conformacional ( <i>landscape</i> ) correspondente dos respectivos ângulos (Bajorath, 2004).....  | 7  |
| Figura 3 - Representação de uma molécula na forma de grafos e de duas das suas matrizes, sem considerar os átomos de hidrogénio.....  | 10 |
| Figura 4 - O cálculo do MIP através de uma grelha de estudo ( <i>grid box</i> ).....  | 13 |
| Figura 5 - Exemplo ilustrativo do cálculo do MIF com três moléculas de prova diferentes: A) ambiente hidrofóbico; B) grupo amida; C) grupo com carga.....   | 14 |
| Figura 6 - Obtenção de descritores gerados por <i>VolSurf</i> (Cruciani <i>et al.</i> , 2000a).....   | 15 |
| Figura 7 - Esquema simplificado de uma rede neuronal.....   | 25 |
| Figura 8 - Imagem ilustrativa do mecanismo de acção de drogas de abuso do tipo ecstasy (Adaptado de <a href="http://www.drugabuse.gov">www.drugabuse.gov</a> ).....   | 28 |
| Figura 9 - Esquema ilustrativo dos farmacóforos propostos para os receptores 5-HT <sub>1A</sub> (Mellin <i>et al.</i> , 1993; López-Rodríguez <i>et al.</i> , 2002).....  | 29 |
| Figura 10 - Ilustração do <i>software ChemOffice</i> utilizado para o desenho e para a optimização geométrica das moléculas da série em estudo.....   | 37 |
| Figura 11 - Ilustração do <i>software MOPAC</i> utilizado para a re-optimização geométrica das moléculas da série em estudo.....  | 37 |
| Figura 12 - Exemplificação dos resultados obtidos pela re-optimização das moléculas pertencentes à série em estudo pelo <i>software MOPAC</i> .....   | 38 |
| Figura 13 - Ilustração do <i>software e-Dragon</i> utilizado para a obtenção de descritores topológicos e geométricos das moléculas da série em estudo.....   | 39 |
| Figura 14 - Ilustração do <i>software Volsurf</i> utilizado para a obtenção de descritores de propriedades físico-químicas das moléculas da série em estudo.....  | 40 |
| Figura 15 - Representação gráfica da análise de componentes principais (PCA) na série em estudo para os receptores 5-HT <sub>1A</sub> (à esquerda) e 5-HT <sub>2A</sub> (à direita).....  | 45 |
| Figura 16 - Representação gráfica por MLR (à esquerda) dos valores de pK <sub>i</sub> teóricos <i>versus</i> experimentais e tabela (à direita) dos coeficientes do modelo preditivo geral para os receptores 5-HT <sub>1A</sub> .....  | 47 |
| Figura 17 - Representação gráfica (à esquerda) dos valores de leverage ( $h^* = 0.29$ ) <i>versus</i> desvio padrão das moléculas da série em estudo e representação gráfica (à direita) da validação cruzada para o modelo preditivo geral dos receptores 5-HT <sub>1A</sub> . Os compostos com círculos azuis apresentam um desvio padrão fora do intervalo normalizado ( $> \pm 2S$ )..... | 48 |
| Figura 18 - Representação gráfica por MLR (à esquerda) dos valores de pK <sub>i</sub> teóricos <i>versus</i> experimentais e tabela (à direita) dos coeficientes do modelo preditivo geral para os receptores 5-HT <sub>2A</sub> .....  | 49 |

- Figura 19 - Representação gráfica (à esquerda) dos valores de leverage ( $h^* = 0.36$ ) versus desvio padrão das moléculas da série em estudo e representação gráfica (à direita) da validação cruzada para o modelo preditivo geral dos receptores 5-HT<sub>2A</sub>. Os compostos com círculos vermelhos apresentam um valor de  $h > h^*$ , enquanto que os compostos com círculos azuis apresentam um desvio padrão fora do intervalo normalizado ( $> \pm 2S$ ).....50
- Figura 20 - Representação gráfica do número de variáveis latentes versus o coeficiente de preditividade ( $Q^2$ ) obtido no modelo preditivo construído para os receptores 5-HT<sub>1A</sub>.....51
- Figura 21 - Ilustração do conjunto de *clusters* obtidos pelo programa *Moloc* por similaridade estrutural a partir das moléculas pertencentes à série em estudo. Os conjuntos seleccionados (1 a 5) para este estudo estão assinalados a vermelho.....52
- Figura 22 - Ilustração do conjunto de gráficos leverage versus desvio padrão para cada um dos *clusters* dos modelos preditivos relativamente ao receptor 5-HT<sub>1A</sub>. Os leverages críticos ( $h^*$ ) por *clusters* (1 a 5) são, respectivamente, os seguintes: 0.38, 0.54, 0.38, 0.71 e 1.04.....54
- Figura 23 - Ilustração do conjunto de gráficos leverage versus desvio padrão para cada um dos *clusters* dos modelos preditivos para o receptor 5-HT<sub>2A</sub>. Os leverages críticos ( $h^*$ ) por *clusters* (1 a 5) são, respectivamente, os seguintes: 0.75, 0.52, 0.50, 0.75 e 1.09.....58
- Figura 24 - Representação dos farmacóforos para os receptores 5-HT<sub>1A</sub> e 5-HT<sub>2A</sub> encontram-se nas imagens à esquerda e à direita, respectivamente, obtidas através do *software Moloc*.....62

## Índice de tabelas

|  |    |
|--|----|
| Tabela 1 - Estratégias utilizadas no estudo das interações ligando-receptor.....   | 4  |
| Tabela 2 - Resumo dos descritores utilizados no estudo e na construção do modelo QSAR ( <i>Dragon</i> , <a href="http://www.disat.unimib.it/chm/Dragon">www.disat.unimib.it/chm/Dragon</a> ; Stewart, 1993).....                       | 9  |
| Tabela 3 - Resumo dos vários tipos de descritores que podem ser obtidos por <i>VolSurf</i> ( <i>Cruciani et al.</i> , 2000a).....  | 16 |
| Tabela 4 - O número de piperazinas que constituem cada um dos <i>clusters</i> seleccionados para a construção de modelos preditivos mais específicos para os receptores 5-HT <sub>1A</sub> e 5-HT <sub>2A</sub> .....                  | 52 |
| Tabela 5 - Os modelos preditivos obtidos para cada um dos cinco <i>clusters</i> por regressão múltipla a partir dos descritores seleccionados e os parâmetros estatísticos mais relevantes para os receptores 5-HT <sub>1A</sub> ..... | 53 |
| Tabela 6 - Os modelos preditivos obtidos para cada um dos cinco <i>clusters</i> por regressão múltipla a partir dos descritores seleccionados e os parâmetros estatísticos mais relevantes para os receptores 5-HT <sub>2A</sub> ..... | 57 |

## Abreviaturas e acrónimos

- 1D – Unidimensional
- 2D – Bidimensional
- 3D – Tridimensional
- 3D-MoRSE – *Molecule Representation of Structures based on Electron diffraction*
- 5-HT – Serotonina ou 5-hidroxitriptamina
- AM1 – Modelo de Austin 1 (*Austin Model 1*)
- ANN – Rede(s) Neuronal(is) Artificial(is) (*Artificial Neural Network*)
- CoMFA – Análise comparativa do campo molecular (*Comparative Molecular Field Analysis*)
- CoMMA – Análise comparativa do momento molecular (*Comparative Molecular Moment Analysis*)
- CORINA – Programa informático utilizado nos estudos de química teórica
- CV – Validação cruzada (*Cross-validation*)
- DiP – Método do perfil de distância (*Distance Profile*)
- DRY – Ensaio hidrofóbicos
- FFD – Desenho Factorial Fraccionado
- GA – Algoritmo(s) genético(s) (*Genetic Algorithm*)
- GRID – Descritores dependentes de potenciais de interacção
- IC<sub>50</sub> – Índice de ionização a 50 % (*Ionization Concentration at 50%*)
- K<sub>i</sub> – Constante de ionização a 50%
- LC<sub>50</sub> – Concentração letal a 50% (*Lethal Concentration at 50%*)
- LOO – *Leave-One-Out*
- LMO – *Leave-Multiple-out*
- BOO – *Bootstrap*
- LSD – Dietilamida do ácido lisérgico
- LV – Variáveis latentes
- MAO – Monoaminoxidase
- MDMA – 3,4-metilenodioximetanfetamina
- MEP – Potencial electrostático molecular (*Molecular Electrostatic Potencial*)
- MIF – Campo de interacção molecular (*Molecular Interaction Field*)
- MIP – Potenciais de interacção molecular (*Molecular Interaction Potentials*)
- MLR – regressão linear múltipla (*Multiple Linear Regression*)
- MM – Mecânica molecular
- MTD – Diferença mínima topológica (*Minimum Topological Difference*)
- PC – Componentes principais (*Principal Components*)
- PCA – Programa de análise de componente principal (*Principal Component Analysis*)
- PCR – Regressão de componente principal (*Principal Component Regression*)
- PLS – Quadrado mínimo parcial (*Partial Least Square*)
- PM3 – *Parameterization Method 3*
- Q<sup>2</sup> – Coeficiente de correlação preditiva
- QSAR – Relação de actividade - estrutura quantitativa (*Quantitative Structure-Activity Relationship*)

- QSPR – Relação de actividade - propriedade quantitativa (*Quantitative Structure-Property Relationship*)
- SCR – Soma quadrada dos resíduos
- SERT – Transportador pré-sináptico da serotonina
- SESP – Método patamar do mínimo Início-Fim geométrico (*Start-End Shortest Path*)
- VI – Variáveis independentes
- WHIM – *Weighted Holistic Invariant Molecular*
- Y – Variáveis dependentes

## Resumo

As piperazinas constituem uma família de compostos bastante ampla, com um potencial de interacção em diversos receptores serotoninérgicos, convertendo estes compostos em uma fonte de agentes altamente selectivos e com uma variedade de propriedades terapêuticas. Apesar disso, esta família de compostos também manifesta inúmeros efeitos toxicológicos, devido às suas interacções com diversos receptores, particularmente com diferentes subtipos do receptor serotoninérgico (5-HT).

O subtipo 1A dos receptores 5-HT (5-HT<sub>1A</sub>) é frequentemente um alvo terapêutico para o tratamento da depressão e da ansiedade. Por outro lado, o subtipo 2A dos receptores 5-HT (5-HT<sub>2A</sub>) é responsável por vários efeitos adversos de inúmeras drogas de abuso com propriedades psicoactivas e alucinogéneas, como o *ecstasy* e o *LSD*, devido ao estímulo exacerbado da secreção de serotonina e à inibição da degradação enzimática da serotonina no centro de recompensa (*reward system*).

Nas últimas duas décadas, muitos investigadores estudaram a estrutura química das arilpiperazinas, como a benzilpiperazina e das fenilpiperazinas, bem como as modificações estruturais capazes de alterar a afinidade das piperazinas para os diferentes subtipos dos receptores 5-HT. A possibilidade de identificar as propriedades químicas ou estruturais das piperazinas responsáveis pela resposta biológica específica nos receptores serotoninérgicos pode conduzir à elucidação da afinidade das piperazinas para os receptores 5-HT<sub>1A</sub> e 5-HT<sub>2A</sub> e à predição do pK<sub>i</sub> de novas piperazinas para a obtenção de uma máxima eficácia terapêutica e de uma mínima toxicidade.

Este trabalho consistiu no estudo *in silico* da afinidade para os receptores 5-HT<sub>1A</sub> e 5-HT<sub>2A</sub> em uma série de 129 piperazinas, através de técnicas Relação de Actividade-Estrutura Quantitativa (*Quantitative Structure-Activity Relationship*, QSAR). Com este objectivo foi construída uma base de dados com a conformação tridimensional das 129 piperazinas e dos seus valores de pK<sub>i</sub> para os dois receptores 5-HT. Em seguida, foi elaborada a optimização das estruturas tridimensionais da série em estudo e a criação de descritores moleculares para as mesmas. Por último, com base nos dados obtidos anteriormente, foram construídos os modelos preditivos e executado o tratamento estatístico para o ajuste e a validação correcta dos modelos alcançados.

## *Parte I – Introdução*



## ***1 – Relação estrutura-actividade***

Em 1893, um estudo demonstrou que a toxicidade de uma série de moléculas orgânicas estava inversamente relacionada com as respectivas solubilidades na água. Mais tarde, foi sugerido que a acção fisiológica de um grupo de compostos orgânicos, denominados por narcóticos depressivos, estava interrelacionada com o coeficiente de partilha óleo/água. Em 1939, introduziu-se o conceito termodinâmico que correlaciona a acção bioquímica com a saturação relativa dos compostos voláteis na via de administração escolhida (Barnett *et al.*, 1978; Selassie, 2003). Para além disto, uma investigação extensiva comprovou a importância da ionização de bases e ácidos fracos na actividade bacteriológica. Entretanto, Hammett (1970) delineou a importância dos efeitos induzidos pelos substituintes nas reacções bioquímicas dos compostos. Da mesma forma se constatou a necessidade de diferenciar os efeitos advindos da polaridade, ressonância e estereoquímica, introduzindo-se o conceito de parâmetros ou descritores na caracterização de compostos estudados (Kubinyi *et al.*, 2002; Selassie, 2003).

Hansch e Muir publicaram, em 1962, um estudo elaborado sobre a relação entre estrutura – actividade (*Structure-Activity Relationships, SAR*) de reguladores de crescimento em plantas, descrevendo a dependência da actividade biológica com a hidrofobicidade e as constantes de Hammett. Através do sistema octanol/água, foi definida uma série de coeficientes de partilha e, assim, foi introduzida uma nova escala hidrofóbica (Jürgen, 2000; Selassie, 2003).

Assim, as contribuições feitas por Hammett e outros investigadores possibilitaram o desenvolvimento do paradigma da relação de actividade – estrutura quantitativa (*QSAR*) por Hansch e Fujita em 1962.

A relação estrutura-propriedade quantitativa (*Quantitative Structure-Property Relationship, QSPR*) ou relação estrutura-actividade quantitativa (*Quantitative Structure-Activity Relationship, QSAR*) é o processo pelo qual a estrutura química de um composto é quantitativamente correlacionada com as suas propriedades intrínsecas ou actividade biológica, respectivamente. Por exemplo, a actividade biológica pode expressar-se quantitativamente como a quantidade de substância necessária para se obter determinada resposta biológica, do mesmo modo que as propriedades físico-

químicas ou estruturais podem ser convertidas em números, possibilitando a formulação de uma relação matemática entre estas duas variáveis. Esta expressão matemática pode ser utilizada para prever a resposta de outras estruturas químicas. Este processo, QSAR, é muitas vezes utilizado para criar modelos preditivos e o desenho de fármacos (Jürgen, 2000; Selassie *et al.*, 2003).

Na química clínica, a maioria das propriedades farmacológicas depende da interacção entre o fármaco com determinado receptor. A força das interacções está correlacionada com a afinidade para o receptor e, conseqüentemente, os descritores moleculares possibilitam, de certa forma, a quantificação dessa ligação (Livingstone, 2000; Selassie, 2003). Segundo a equação 1, a actividade biológica, usualmente em forma logarítmica, é descrita mediante um conjunto de parâmetros próprios ou descritores das moléculas em estudo.

$$\text{Actividade} = f(\text{propriedades físico-químicas e /ou estruturais}) \quad \text{eq. 1}$$

Como se pode observar, a equação anterior, na sua forma mais geral, define uma correlação através de uma determinada função matemática, não constituindo, no entanto, uma teoria. Em QSAR não existe uma teoria que caracterize a natureza intrínseca da relação entre a estrutura e a actividade. O fenómeno em estudo não é completamente conhecido, não sendo, por isso, possível descrever a ligação numa equação baseada numa teoria linear. Por esta razão, os modelos QSAR pertencem à família dos modelos empíricos, os quais proporcionam somente uma solução aproximada. Algumas suposições podem ser extrapoladas por observação, como a forma, a continuidade e o alcance da aplicação desta relação obtida pelo modelo empírico (Barnett *et al.*, 1978; Selassie *et al.*, 2002; Philip *et al.*, 2003).

### 1.1 – Abordagem do estudo de interacções ligando-receptor

O estudo por QSAR pode ser realizado basicamente por dois tipos de abordagens: o estudo pelo ligando, a partir do qual se analisa o conjunto ou família de compostos envolvidos na actividade desejada; ou, então, pelo estudo do receptor, onde se investiga as suas características intrínsecas e que permitem a sua activação ou inibição, possibilitando uma caracterização estrutural dos ligandos pretendidos (Selassie *et al.*, 2003; Philip *et al.*, 2003; Bajorath, 2004).

Na análise pelo ligando, assume-se que moléculas semelhantes possuem actividades semelhantes, hipótese esta que não é absoluta. Deste modo, o principal objectivo deste enfoque é a caracterização de um conjunto de compostos similares com o intuito de extrair os padrões que definem a actividade biológica. Assim, é imperativo que se estudem as mais pequenas diferenças a nível molecular (Selassie *et al.*, 2003; Thilagavathi & Chakraborti, 2005).

Por outro lado, a abordagem pelo receptor tem como propósito a avaliação das propriedades do ligando, a partir do espaço electro-topológico do centro activo do receptor. Esta avaliação não é realizada por uma só via ou de forma directa; pelo contrário, extrai-se uma série de propriedades que, posteriormente, são combinadas adequadamente para a construção ou o desenho de ligandos mais eficazes.

Utilizando as técnicas QSAR, por qualquer das duas abordagens, é possível identificar as estruturas químicas que podem possuir bons efeitos inibitórios num receptor específico, bem como uma baixa toxicidade (Selassie, 2003; Selassie *et al.*, 2003).

Na tabela 1, estão resumidas as estratégias empregues no estudo da interacção ligando-receptor, observando-se que, em qualquer das abordagens mais utilizadas, é necessário a utilização de métodos teóricos (*docking*, MD, MM, etc) para a análise das forças intermoleculares, em particular as ligações não covalentes.

**Tabela 1 - Estratégias utilizadas no estudo das interacções ligando-receptor.**

| <i>Estrutura do Ligando Conhecida</i> |                     | <i>Estrutura do Ligando Desconhecida</i> |  |
|---------------------------------------|---------------------|--|--|
| <i>Receptor conhecido</i>             |                     | <i>Receptor desconhecido</i>             | <i>Receptor conhecido</i>                            |
| Estrutura                             | Sequência           | QSAR, Previsão do Receptor               | QSAR, Previsão do Ligando, Química Combinatória, etc |
|                                       | Previsão estrutural |  |  |
| Simulações moleculares                |                     |  |  |
| (MD, MM, etc)                         |                     |  |  |
| <i>Docking</i> , MD, MM, etc          |                     |  |  |

Existem outras formas de criar modelos QSAR com a informação do receptor, como por exemplo, os modelos 5D-QSAR e FEFF-3D QSAR. Em ambos os métodos analisa-se a energia de interacção, quer pela simulação do local activo do receptor

mediante a criação de um espaço conformacional (5D-QSAR), quer pela conversão dos componentes de energia de interacção do campo de força em descritores específicos para um sistema ligando-receptor (FEFF-3D QSAR) (Ekins *et al.*, 2000; Ekins *et al.*, 2001; Moon *et al.*, 2000; Santos-Filho *et al.*, 2001; Selassie *et al.*, 2002; Ducki *et al.*, 2005).

## 1.2 – Metodologia geral para o estudo QSAR

A metodologia QSAR adopta um conjunto de procedimentos que podem seguir vias de análise computacional distintas, consoante a informação inicial disponível, apresentando, contudo, passos obrigatoriamente comuns. A figura 1 apresenta o esquema dos passos essenciais para a elaboração do estudo QSAR (Barnett *et al.*, 1978; Bajorath, 2004).

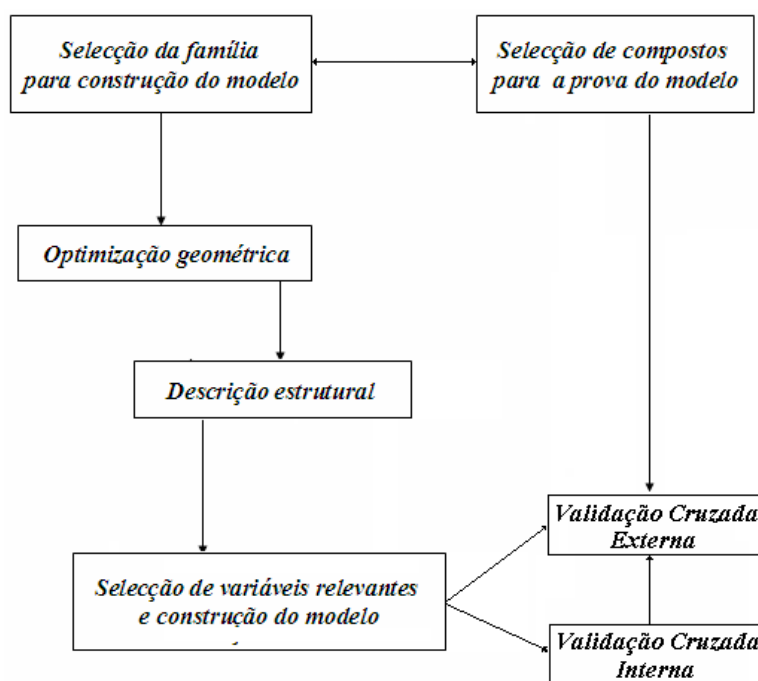


Figura 1 - Esquema ilustrativo dos passos essenciais na elaboração de um estudo QSAR e construção do modelo preditivo.

### 1.2.1 – Seleção de ligandos e otimização estrutural

A seleção de ligandos é o primeiro passo a realizar em estudos QSAR, efectuando-se uma base de dados de determinada família ou conjunto de compostos. É

obrigatório que esta recolha de dados seja criteriosa e elaborada, apresentando uma série de moléculas quimicamente similares ou com características semelhantes, como a actividade biológica, a constante de ionização ( $K_i$ ) e as cargas parciais (Hagadone, 1992; Bajorath, 2004).

As moléculas devem então apresentar similaridade para que o estudo seja válido e eficiente, pois uma série de moléculas pouco similares origina modelos não produtivos (Nikolova & Jaworska, 2003; Bajorath, 2004). O problema do cálculo da similaridade molecular será discutido com maior detalhe adiante.

Dado que os modelos QSAR das moléculas da série em estudo são criados através dos descritores extraídos da estrutura tridimensional dos compostos seleccionados, a obtenção de uma geometria correcta é de grande importância (Hagadone, 1992; Thorner *et al.*, 1997).

O cálculo computacional da geometria molecular é uma das funções mais essenciais dos programas de química computacional, embora não seja um processo trivial. O uso destes programas permite descrever a geometria da molécula em estudo, analisando as energias de interacção e os seus respectivos gradientes para determinar a geometria molecular correspondente ao estado de mínima energia (Hurst & Heritage, 1997; Jensen, 2001; Young, 2001).

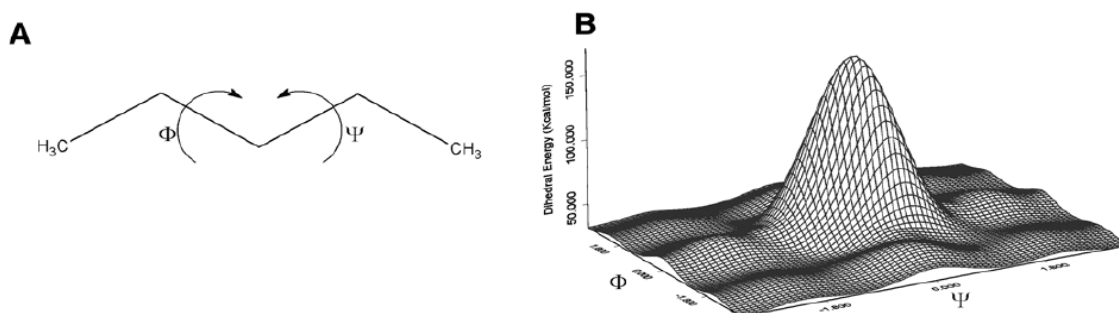
A utilização de programas com representação e construção gráfica de compostos é bastante frequente, possibilitando assim o desenho gráfico das moléculas pretendidas. A existência de diversos programas permite este tipo de operação, já que alguns deles facultam o desenho da estrutura bidimensional e transmutam-na para uma estrutura tridimensional, enquanto que outros programas, por seu lado, já permitem um desenho tridimensional adicionando, posteriormente, os hidrogénios necessários à estrutura final. Existem ainda outros programas que constroem a molécula tridimensional a partir de uma lista de elementos e de hibridações (Young, 2001; Popp *et al.*, 2002).

Muitos destes *softwares* possuem uma base de dados de grupos funcionais, os quais podem ser bastante convenientes para determinados projectos de investigação, ou empregam ainda campos de forças específicos, como por exemplo o *CORINA*, que possibilitam um cálculo mais rápido de uma geometria de baixa energia sem recorrer a métodos de mecânica quântica como os que utilizam o *software MOPAC*. Estes programas demonstram uma excelente *performance* com moléculas orgânicas (Stewart, 1993; Jensen, 2001; Young, 2001; Popp *et al.*, 2002).

Independentemente da eficiência do método utilizado, a geometria final obtida corresponde à energia do nível teórico usado pelo computador. De uma forma geral, há uma tendência para se utilizar métodos rápidos e estimativos ou métodos precisos mas computacionalmente intensivos e morosos. No entanto, existem métodos rápidos e precisos, todavia apenas aplicáveis a uma classe limitada de moléculas (Jensen, 2001; Young, 2001; Popp *et al.*, 2002).

Com o intuito de se obter resultados precisos o mais rápido possível, são frequentemente utilizadas com bastante êxito duas otimizações geométricas. Uma vez finalizada uma geometria correcta com um reduzido nível teórico, esta é novamente utilizada para uma segunda optimização a um nível teórico muito mais rigoroso (Young, 2001; Popp *et al.*, 2002).

A escolha do algoritmo de optimização geométrica é amplamente influenciada pelo tempo necessário que o computador leva a optimizar a geometria molecular. Métodos como o *AMI* (Modelo de Austin 1, *Austin Model 1*) e o *PM3* (*Parameterization Method 3*) utilizam meios semi-empíricos para obter resultados em moléculas orgânicas. Embora o *PM3* apresente algumas vantagens em relação ao *AMI*, este não é eficaz para compostos com alguns tipos de ligações C-N, embora estes juntamente com o método de *CORINA* sejam usados nos estudos do QSAR (Jensen, 2001; Young, 2001; Popp *et al.*, 2002).



**Figura 2 - Exemplo ilustrativo: A) dos ângulos  $\Phi$  e  $\Psi$  do pentano; e B) da paisagem conformacional (*landscape*) correspondente aos respectivos ângulos (Bajorath, 2004).**

É importante referir que permanece desconhecido se a geometria molecular calculada na optimização corresponde a um mínimo local ou absoluto (figura 2), embora se realizem certos procedimentos estatísticos para determinar um grau de confiança. Além disso, a optimização estrutural é realizada frequentemente em QSAR sem considerar qualquer tipo de interacção entre a molécula e algum solvente, o que é incorrecto pois haverá certamente alguma mudança na conformação final dos ligandos

quando se encontram em solução. No entanto, usualmente têm-se em conta os grupos carregados, protonados e os pares iónicos no processo de optimização. Na análise comparativa do campo molecular (*Comparative Molecular Field Analysis, CoMFA*) procede-se à tentativa de minimização deste tipo de erro ao considerar-se as energias de interacções dos ligandos com uma molécula de prova, como por exemplo, a água (Jensen, 2001; Kubinyi *et al.*, 2002; Bajorath, 2004; Kovalishyn *et al.*, 2007).

### 1.2.2 – Métodos gerais para a obtenção de parâmetros moleculares

Podem existir duas vias para a obtenção de parâmetros moleculares, as que somente consideram as propriedades inerentes à molécula (geometria, densidade de cargas, etc) e as que se baseiam no campo criado por esta, quer seja electrostático, como os potenciais electrostáticos moleculares (*Molecular Electrostatic Potential, MEP*), quer pelo produto da interacção com outras moléculas, como os potenciais de interacções moleculares (*Molecular Interaction Potentials, MIP*) (Kubinyi *et al.*, 2002; Bajorath, 2004). Estas duas formas de caracterização molecular serão elucidadas em seguida.

#### *Descritores da estrutura molecular*

Nos últimos anos tem-se proposto inúmeros descritores moleculares. Uma das classificações mais aceites para estes descritores faz referência ao tipo de informação molecular necessária à sua construção, isto é, dependendo da utilização das propriedades 1D, 2D, ou 3D da molécula (Tabela 2) (Bajorath, 2004; Tetko, 2005; Tetko *et al.*, 2005).

Os descritores 1D são obtidos a partir da informação molecular, não requerendo o conhecimento prévio da topologia ou estrutura tridimensional. Por isso, o peso molecular e o número de átomos de cada elemento são, provavelmente, os descritores mais utilizados nesta categoria, embora existam ainda muitos outros (Bonchev, 1980; Bonchev, 2000; Bajorath, 2004; Tetko *et al.*, 2005).

Por sua vez, os descritores 2D são construídos com base na conectividade molecular dos compostos. Os primeiros índices de conectividade molecular foram descritos por Randic (Randic *et al.*, 2004) e intensamente investigados por Kier, Hall e

outros colaboradores (Kier *et al.*, 1975; Kier *et al.*, 1975a; Bonchev, 2000; Vukičević & Trinajstić, 2003; Hall & Henry, 2006).

Existem, ainda, os descritores 3D, gerados a partir da estrutura tridimensional dos compostos cujas propriedades podem ser globais (Tetko *et al.*, 2001). Exemplos disso são o momento dipolar ou as energias HOMO e LUMO, as densidades de carga, volume e outros, na sua maioria obtida por *softwares* específicos, como por exemplo o *software Dragon* (Tetko *et al.*, 2005).

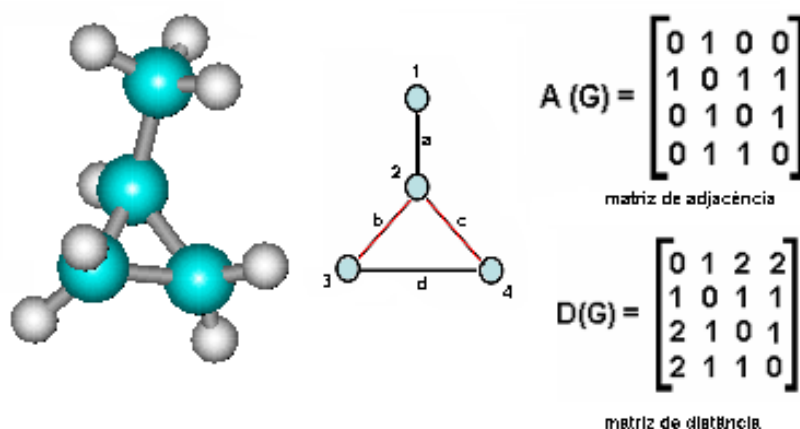
**Tabela 2 - Resumo dos descritores utilizados no estudo e na construção do modelo QSAR (*Dragon*, [www.disat.unimib.it/chm/Dragon](http://www.disat.unimib.it/chm/Dragon); Stewart, 1993).**

| <i>ID conjunto</i> | <i>Conjunto de descritores</i>       | <i>Dimensão dos Descritores</i> | <i>Nº de Descritores</i> |
|--------------------|--------------------------------------|---------------------------------|--------------------------|
| 1                  | Descritores constitucionais          | 1D                              | 48                       |
| 2                  | Descritores topológicos              | 2D                              | 119                      |
| 3                  | <i>Walk and path counts</i>          | 2D                              | 47                       |
| 4                  | Índices de conectividade             | 2D                              | 33                       |
| 5                  | Índices informativos                 | 2D                              | 47                       |
| 6                  | Autocorrelações 2D                   | 2D                              | 96                       |
| 7                  | Índices de pontes de adjacência      | 2D                              | 107                      |
| 8                  | Descritores <i>Burden eigenvalue</i> | 2D e 3D                         | 64                       |
| 9                  | Índices de carga topológicos         | 2D                              | 21                       |
| 10                 | Índices de <i>Eigenvalue-based</i>   | 2D e 3D                         | 44                       |
| 11                 | <i>Randic molecular profiles</i>     | 2D                              | 41                       |
| 12                 | Descritores geométricos              | 2D e 3D                         | 74                       |
| 13                 | Descritores RDF                      | 3D                              | 150                      |
| 14                 | Descritores 3D-MoRSE                 | 3D                              | 160                      |
| 15                 | Descritores WHIM                     | 3D                              | 99                       |
| 16                 | Descritores GETAWAY                  | 3D                              | 197                      |
| 17                 | Nº de grupos funcionais              | 1D e 2D                         | 154                      |
| 18                 | Fragmentos centrais de átomos        | 2D                              | 120                      |
| 19                 | Descritores de carga                 | 2D e 3D                         | 14                       |



|    |  |    |    |
|----|--|----|----|
| 20 | Propriedades moleculares                     | 3D | 29 |
| 21 | $\mu$ Momento bipolar<br>(Calculado por PM3) | 3D | 1  |
| 22 | AP ( <i>Average</i> de polaridade molecular) | 3D | 1  |
| 23 | NFL (Número de níveis orbitais ocupados)     | 3D | 1  |
| 24 | HOMO (Orbitais Homo moleculares)             | 3D | 1  |
| 25 | LUMO (Orbitais Lumo moleculares)             | 3D | 1  |

A maioria dos índices topológicos pode ser obtida a partir da aplicação da teoria de grafos nas estruturas 2D da molécula (Estrada, 1995; Gutman *et al.*, 1998). De modo que um gráfico  $G(V,E)$  é um conjunto de vértices  $V=(v_1, v_2, \dots, v_n)$  e pontes  $E=(e_1, e_2, \dots, e_m)$  representando respectivamente os átomos e as ligações (figura 3). Os grafos são expressos matematicamente por intermédio de diferentes matrizes como a de adjacência  $A(G)$  e a matriz de distâncias  $D(G)$  (figura 3) (Estrada, 1995; Gutman *et al.*, 1998; Trinajstić & Gutman, 2002).



**Figura 3 - Representação de uma molécula na forma de grafos e de duas das suas matrizes, sem considerar os átomos de hidrogénio.**

A diversidade dos gráficos moleculares é ilimitada, o que justifica a existência de um grande número de descritores moleculares. O uso da teoria de grafos na química computacional é de grande importância não só para a obtenção de descritores mas também para a resolução de muitos outros problemas, como a similaridade e identificação de substituintes comuns (Estrada, 1995; Trinajstić & Gutman, 2002).

Os descritores 3D moleculares apresentam uma invariante translacional e rotacional, ou seja, não necessita de uma superimposição estrutural dos compostos estudados, para poder logo compará-los entre as diferentes moléculas. Estes descritores podem ser obtidos de diversas formas e metodologias (Jensen, 1999; Tetko *et al.*, 2001; Young, 2001; Kubinyi *et al.*, 2002), tais como:

- a. *O método do perfil de distância (Distance Profile, DiP)*
- b. *O EVA e o EEVA*
- c. *O Start-End Shortest Path (SESP)*
- d. *A análise comparativa de espectro (Comparative Spectra Analysis, CoSA)*
- e. *Os descritores WHIM (Weighted Holistic Invariant Molecular)*
- f. *A análise comparativa do momento molecular (Comparative Molecular Moment Analysis, CoMMA)*
- g. *O 3D-MoRSE (Molecule Representation of Structures based on Electron diffraction)*
- h. *Análise da distância interna (Internal Distance Analysis, IDA)*
- i. *Métodos de autocorrelação (Autocorrelation methods)*

Os métodos como *SESP* (Baumann, 2002), *DiP* (Baumann, 2002a), *WHIM* (Todeschini & Lasagni, 1994), *IDA* (Baumann, 2002a) e *3D-MoRSE* (Schuur *et al.*, 1996) utilizam as distâncias inter-atômicas para a obtenção dos descritores 3D. Por outro lado, os métodos *EVA* (Ferguson *et al.*, 1997), *EEVA* (Tuppurainen *et al.*, 2002) e *CoSA* (Bursi *et al.*, 1999) estudam os espectros moleculares para originar famílias de descritores QSAR espectroscópicos. Nesta abordagem, as técnicas utilizam as coordenadas 3D da molécula para projectar uma série de matrizes e funções Gaussianas, proporcionando assim um conjunto de descritores 3D.

A análise comparativa do momento molecular (*Comparative Molecular Moment Analysis, CoMMA*) (Silverman, 1998) faz uso de momentos moleculares como os momentos principais de inércia e propriedades derivadas dos momentos dipolares e quádruplos para caracterizar os compostos em estudos de CoMFA.

Os métodos de autocorrelação (*Autocorrelation methods*) (Broto *et al.*, 1984; Wagener *et al.*, 1995; Fechner *et al.*, 2003; Kein *et al.*, 2004) transformam propriedades dependentes de alinhamento espacial em vectores independentes de orientação molecular, os quais podem ser correlacionados com a actividade biológica. Por exemplo, Wagener *et al.* (1995) utilizaram a autocorrelação de propriedades

moleculares superficiais para obter a informação contida no potencial hidrofóbico ou electrostático na superfície de Van der Waals.

A aplicação deste género de descritores pode ser observado em diversos trabalhos, como por exemplo Gute & Basak (1997), os quais criaram um modelo preditivo de toxicidade ( $LC_{50}$ ) a partir de um conjunto de 69 derivados do benzeno, segundo parâmetros estruturais, químicos, geométricos e quânticos. Também Randic & Basak (2001) publicaram um modelo para 21 éteres de alquilo, utilizando o número de átomos de hidrogénio como descritor molecular e, assim, lograram a obtenção de resultados mais apurados que Basak *et al.* (1999) que, por seu lado, utilizaram o índice de conectividade de 1ª ordem (tabela 2) para calcular a capacidade de inibição de 105 benzamidas.

### 1.2.3 – Campos de interacção molecular (MIF)

Nas técnicas QSAR parte-se do princípio que a actividade de um composto depende das suas interacções com o receptor específico, desenvolvendo-se métodos que permitem a análise computacional por comparação de interacções conhecidas de outros compostos com as da substância em estudo. Os potenciais de interacção molecular (*Molecular Interaction Potentials, MIP*) contêm as energias de interacção entre as moléculas de prova ou para testes, como por exemplo a água, e o composto examinado em todos os pontos espaciais. Em casos mais complexos, a prova molecular pode ser efectuada com uma molécula pequena ou um grupo químico, como o grupo amina de uma ligação peptídica (Kubinyi *et al.*, 2002; Bajorath, 2004).

#### 1.2.3.1 – Potenciais de interacção molecular (MIP)

O MIP é calculado utilizando uma grelha de estudo, conhecida por *grid box*, a qual possui, no mínimo, o tamanho do maior composto da série. A resolução da célula ou unidade da grelha possui uma dimensão entre 0.5 e 1 Å, sendo que em cada vértice da célula são colocados diferentes grupos ou moléculas de prova (grupo amida, grupo carbonilo, água, etc) (Bajorath, 2004), obtendo-se, assim, vários campos de interacção molecular (MIF) hidrofóbicos, hidrofílicos (Poda *et al.*, 2005), entre outros, consoante o tipo de molécula de prova (figura 4). É importante ainda referir que cada cálculo efectuado na grelha (*grid box*) corresponde a um descritor da molécula em estudo.

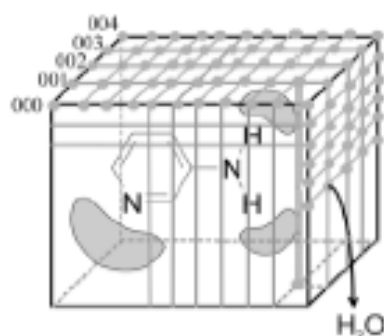


Figura 4 - O cálculo do MIP através de uma grelha de estudo (*grid box*).

A interacção entre a molécula de estudo e a molécula de prova, em cada vértice da célula, é calculada mediante um campo de forças definido, sendo constituído por uma série de equações que representam o potencial de energia superficial com as respectivas alterações na geometria da molécula (Learch, 2001). De uma forma simples, a energia potencial de interacção não covalente ( $E_{non-cov}$ ) pode ser calculada pela seguinte equação:

$$E_{non-cov} = \sum E_{vdW} + \sum E_{electrostática} + [H] \quad \text{eq. 2}$$

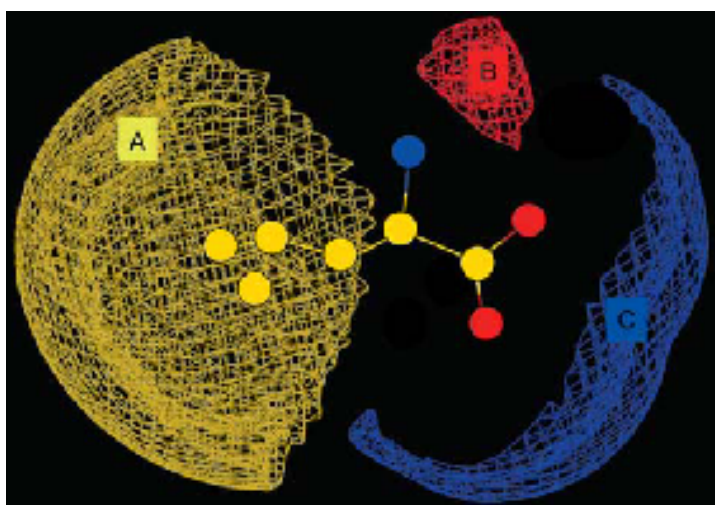
$$E_{non-cov} = \sum [C_{12}(i, j) / r_{ij}^{12} - C_6(i, j) / r_{ij}^6] + \sum q_i q_j / (4\pi\epsilon_0\epsilon_r r_{ij}) + [H] \quad \text{eq. 3}$$

onde, na equação 2:  $E_{vdw}$  e  $E_{electrostática}$  correspondem às energias de van der Waals e electrostáticas, respectivamente, para cada par de átomos que constituem a molécula em estudo e para a molécula de prova em questão; e, por fim,  $[H]$  simboliza a energia associada às pontes de hidrogénio determinada pela respectiva fórmula utilizada pelo investigador. Por sua vez, na equação 3:  $C_{12}$  e  $C_6$  correspondem a parâmetros de interacção de Van der Waals;  $q$  equivale a carga do átomo, sendo  $r$  a distância entre átomos  $i$  e  $j$ ; e, finalmente,  $\epsilon_0$  e  $\epsilon_r$  correspondem a constantes dieléctricas do meio em estudo. O conjunto formado pela equação 3 com os parâmetros constituintes denomina-se por campo de força no contexto da mecânica e dinâmica molecular (Jensen, 2001; Young, 2001; Rapaport, 2004).

Neste tipo de cálculo ( $E_{non-cov}$ ) somente se consideram as últimas três parcelas da equação 3, ou seja, as ligações não covalentes, em que o *software* responsável por este procedimento é denominado por *GRID* (Bitetti-Putzer *et al.*, 2001). Outra

particularidade deste campo é o facto do valor da constante dieléctrica se alterar com o meio ambiente do ensaio “experimental” ou teste (Learch, 2001).

O MIP contém uma série de informações relacionadas com a potencialidade de interacção de uma molécula (figura 5). Quando o MIP de duas moléculas difere numa região espacial da grelha, estas interagem de forma diferente nessa região específica. Esta diferença de interacção pode ser responsável pela actividade distinta de cada uma delas. Deste modo, o modelo obtido por CoMFA permite estudar uma correlação entre as diferenças de MIP e das actividades biológicas (Miranker & Karplus, 1991; Learch, 2001; Kubinyi *et al.*, 2002; Poda *et al.*, 2005).



**Figura 5 - Exemplo ilustrativo do cálculo do MIP com três moléculas de prova diferentes: A) ambiente hidrofóbico; B) grupo amida; C) grupo com carga.**

Pode-se afirmar que cada valor de MIP é sensível à orientação da estrutura utilizada, impossibilitando, assim, a comparação directa de uma dada série de ligandos com base nos valores de MIP. Deste modo, a comparação da série necessita de um passo prévio de alinhamento estrutural dos compostos para que a grelha de cálculo possa ser utilizada por todos (Kubinyi *et al.*, 2002). No entanto, o alinhamento estrutural é um passo complexo, especialmente se os ligandos são estruturalmente distintos uns dos outros. Assim, torna-se necessário possuir uma similaridade entre os ligandos da série de estudo seleccionada (Nikolova & Jaworska, 2003; Langer & Hoffmann, 2006).

Como o alinhamento ou a sobreposição representam um problema relevante, foram desenvolvidos índices ou descritores a partir de MIP (Kubinyi *et al.*, 2002) que não requerem um prévio alinhamento molecular (Miller, 1998).

O princípio básico deste processo consiste em extrair a informação presente no MIP e convertê-la num número reduzido de descritores quantitativos, os quais são de fácil compreensão e interpretação (figura 6). Alguns destes descritores são especialmente criados para caracterizar as propriedades globais da molécula (Tabela 3), como por exemplo descritores hidrofílicos, enquanto que outros são mais sujeitos às propriedades locais da molécula, como o momento *integy*, e mais sensíveis às alterações conformacionais (Cruciani *et al.*, 2000a).

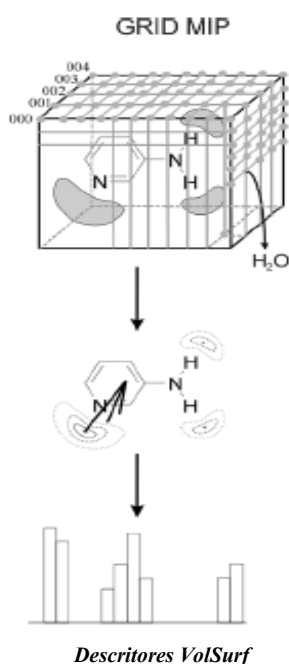


Figura 6 – Obtenção de descritores gerados por *VolSurf* (Cruciani *et al.*, 2000a).

Os descritores mais importantes, quando se pretende explicar se o composto possui a capacidade de atravessar a membrana celular, são as regiões hidrofílicas e os factores de capacidade. Estes correspondem à capacidade da molécula de interagir com as moléculas de água. As regiões hidrofílicas são definidas como invólucros moleculares acessíveis e atraídas pelas moléculas da água, enquanto que os factores de capacidade representam a proporção da superfície hidrofílica em relação à superfície total da molécula em estudo. Por outras palavras, as regiões hidrofílicas avaliam a hidrofilicidade total, enquanto que os factores de capacidade medem a hidrofilicidade relativa (Cruciani *et al.*, 2000a).

Tabela 3 - Resumo dos vários tipos de descritores que podem ser obtidos por *VolSurf* (Cruciani *et al.*, 2000a).

| Tipo de descritor | Nome dos descritores                     | Descrição e caracterização dos descritores   |
|-------------------|--|--|
| Tamanho           | MW                                       | Peso molecular   |
|                   | Volume molecular<br>Superfície molecular | Volume e superfície são analisados por computador a partir da componente positiva do MIP a uma energia <i>cutoff</i> de 0.2 kcal/mol |
| Forma             | Razão<br>Volume/Superfície               | A razão é uma medida de rugosidade ao nível molecular  |
|                   | Globularidade molecular                  | Definida como $S/S_{eq}$ , em que $S_{eq}$ representa a área superficial da esfera molecular (volume, V)                             |
| Hidrofílico       | Descritores hidrofílicos                 | Define o volume do revestimento hidrofílico a vários níveis de interação (desde -0.2 até -6 kcal/mol)                                |
|                   | Factores de capacidade                   | Razão da superfície hidrofílica emergente <i>versus</i> a superfície molecular total   |
|                   | Momento <i>integy</i>                    | Vector que advém do centro de massa estrutural para o centro das regiões hidrofílicas tendo em conta um dado nível energético        |
| Hidrofóbico       | Descritores hidrofóbicos                 | Define o volume de revestimento hidrofóbico em vários níveis de interação (desde 0.0 até -0.2 kcal/mol)                              |
|                   | Momento <i>integy</i>                    | Vector que resulta do centro de massa estrutural para o centro das regiões hidrofóbicas tendo em conta um dado nível energético      |
| Vários            | Interação local de energia mínima        | Energia dos melhores três locais de energia mínima   |

|        | Distâncias de energia mínima     | Distâncias entre as energias mínimas   |
|--------|----------------------------------|--|
| Vários | Balanço hidrofílico-lipofílico   | Descreve que tipo de efeito, hidrofílico ou hidrofóbico, domina na molécula                |
|        | Momento Anfílico                 | Vector que resulta do centro hidrofóbico dominante para o centro hidrofílico dominante     |
|        | Empacotamento crítico            | Parâmetro que prediz o empacotamento molecular, tal como a formação de micelas             |
|        | Ligação por pontes de hidrogénio | Capacidade de ligação entre uma molécula de prova polar com outra molécula, excepto a água |
|        | Polaridade                       | Estimativa da polaridade média molecular   |

A distribuição das interacções hidrofílicas e hidrofóbicas são codificadas nos denominados momentos *integy*. Em relação às regiões hidrofílicas, os momentos *integy* são vectores que resultam do centro de massa da molécula para o centro das suas regiões hidrofílicas num determinado nível de energia. A molécula possuidora de um grande momento *integy* tem uma concentração clara de regiões hidrofílicas em apenas uma das extremidades, enquanto que, pelo contrário, a molécula com um pequeno momento *integy* comporta duas regiões hidrofílicas perto do centro de massa ou em extremidades opostas. Assim, um elevado momento *integy* auxilia a molécula a atravessar as membranas biológicas (Cruciani *et al.*, 2000a).



#### **1.2.4 – Tratamento estatístico e métodos matemáticos aplicados em QSAR**

O processo de descrição da estrutura química, independentemente do método utilizado (MIP, descritores moleculares, etc), como foi anteriormente referido, origina um grande número de variáveis para cada uma das moléculas consideradas na construção do modelo. O CoMFA, por exemplo, utiliza os MIP para descrever o espaço de interações em cada um dos pontos que compõem a grelha, os quais representam os descritores de espaço, isto é, possíveis variáveis no modelo matemático em construção. De maneira semelhante, os índices gerados a partir da informação 1D, 2D e 3D das moléculas representam potenciais variáveis do modelo. Assim, toda esta informação transporta consigo um conjunto de situações que podem ser classificadas segundo os respectivos critérios:

- Variáveis com informação relevante
- Variáveis com informação irrelevante
- Variáveis que partilham o mesmo grau e tipo de informação

A presença de variáveis com informação irrelevante e, simultaneamente, a sobreposição de informação provocam um ruído de fundo no processamento dos dados, enquanto que a existência de um número muito maior de variáveis, em relação às moléculas em estudo, impede a aplicação directa dos métodos convencionais de regressão múltipla. Estas situações induzem a aplicação de métodos estatísticos e matemáticos que permitam a simplificação do número de variáveis e a busca de uma correlação mais adequada entre as novas variáveis e a propriedade biológica em estudo (Esbensen, 2002). Deste modo, o objectivo destes métodos estatístico-matemáticos é a resolução da equação 1, sendo importante destacar que a resolução desta equação passa por encontrar o menor número de variáveis “m” que garantam a maior preditividade possível. Por outro lado, as variáveis  $P_1, P_2, P_3, \dots, P_m$  não são obrigatoriamente os descritores anteriormente referidos; na maioria dos casos constituem transformações matemáticas.

A resolução da equação 1 consiste, então, em definir a quantidade e tipo de variáveis, assim como as suas contribuições ( $w_1, w_2, w_3, \dots, w_m$ ) para a actividade biológica.

### 1.2.4.1 – Simplificação e selecção de variáveis

A simplificação das variáveis é uma tarefa usual na exploração de dados, existindo muitos métodos delineados para este objectivo. Deste modo, na didáctica QSAR são utilizados (Tipping & Bishop, 1999; Cruciani *et al.*, 2000b; Fichera *et al.*, 2000; Gustafsson, 2001; Tetko *et al.*, 2001; Tetko *et al.*, 2005; Ragno *et al.*, 2006), principalmente, os seguintes métodos: a Análise de Componentes Principais (PCA); os Mínimos Quadrados Parciais (PLS); diferentes géneros de Redes Neurais Artificiais (ANN); o Desenho Factorial Fraccionado (FFD); e os Algoritmos genéticos (GA). Uma explicação detalhada de cada uma destas abordagens prolongaria demasiado o presente estudo e está além dos objectivos propostos para este trabalho de mestrado. No entanto, dada a sua importância e conceito explicaremos resumidamente alguns destes métodos.

#### *Análise de Componentes Principais*

Este método realiza uma simplificação da dimensão através da análise da covariância entre os diferentes factores ou variáveis, de maneira tal que as novas variáveis obtidas recolham a maior variabilidade possível dos dados iniciais (Tipping & Bishop, 1999; Cruciani *et al.*, 2000b; Esbensen, 2002; Vilca, 2004). Este processamento transforma o PCA numa ferramenta útil, não apenas na simplificação do tamanho das variáveis, como também possibilita a exploração dos dados para a identificação de padrões e subgrupos.

A matriz de componentes principais é obtida, então, através de:

$$T = XP \tag{eq. 4}$$

A matriz T (*scores*) (eq. 4) é composta pelas projecções das variáveis originais (matriz X) nas direcções principais, ou seja, são os componentes principais (PCs) obtidos por combinações lineares das variáveis originais e os coeficientes da matriz de carga (P) (*loading*), cujos elementos realmente representam o ângulo nas direcções que garantam o máximo ajuste, que correspondem aos autovectores obtidos pela matriz de correlações (Esbensen, 2002).

Podemos observar pela equação 4 que algumas das propriedades da matriz de PCs são as seguintes: cada PC é uma combinação linear das variáveis originais, os quais são ortogonais entre si, sendo que por isso não partilham nenhuma informação, podendo fazer referência a um sem considerar nenhum outro. Embora a obtenção de variáveis independentes e ortogonais seja de grande importância, este processo pode também tornar-se um obstáculo, devido à extrema dificuldade em entender o significado de cada PC obtido.

O método PCA é utilizado, em grande escala, na metodologia QSAR podendo ser encontradas algumas variantes deste que contribuem para ultrapassar as suas limitações (Tipping & Bishop, 1999; Lawrence, 2005).

### *Algoritmos genéticos*

O algoritmo genético (*Genetic Algorithm*, GA) concretiza o seu objectivo pela geração de populações de modelos preditivos, através de um processo reprodutivo, bem como através da respectiva optimização da função relacionada com a qualidade do modelo construído.

Na terminologia algorítmica genética, cada indivíduo da população inicial é denominado por cromossoma, consistindo num vector binário em que cada posição (gene) corresponde a uma variável. Por sua vez, cada gene apresenta o número 1 ou 0, caso esteja incluído ou ausente no modelo construído, respectivamente. Por fim, cada cromossoma representa um possível modelo com determinadas variáveis.

Uma vez estabelecidos os parâmetros anteriores, a evolução do algoritmo genético é constituído por quatro passos gerais:

- I. Iniciação da população;
- II. *Crossover*;
- III. Mutação;
- IV. Nova geração.

## I. Iniciação da população

A população inicial é constituída, em primeira instância, por modelos aleatórios com um número de variáveis entre 1 e  $L$ , sendo  $L$  o número máximo determinado pelo investigador. O valor da função seleccionada para cada modelo é calculado no processo denominado “avaliação”.

Os modelos calculados são ordenados por ordem decrescente e de acordo com a qualidade do respectivo modelo.

## II. *Crossover*

Os pares de modelos são seleccionados a partir do conjunto de cromossomas anteriormente calculado. Seguidamente, os respectivos pares permitem a geração de um novo modelo, o qual preserva as características comuns dos modelos ascendentes (os progenitores) e incorpora características diferentes de acordo com a probabilidade de cruzamento genético. Se o “modelo-filho” coincide com algum já existente na população actual ou se a sua qualidade é inferior a qualquer um deles, então será rejeitado. Caso contrário, será avaliado e ordenado na restante população, segundo o nível de qualidade.

Este conjunto de passos é repetido inúmeras vezes até à obtenção do máximo valor da função que descreve a qualidade do modelo construído; a função seleccionada para este estudo será explicada mais adiante.

## III. Mutação

Após o procedimento de *crossover*, a população encaminha-se para um processo de mutação em que cada um dos seus modelos sofre alterações aleatórias nos seus respectivos genes. Os modelos comutados são, então, avaliados e ordenados na população de forma semelhante ao *crossover*.

Este processo é controlado por probabilidade mutagénica, a qual permite um movimento dos mínimos anteriormente estabelecidos por *crossover*.

#### IV. Nova geração

A nova geração de uma população é obtida através da supressão no cromossoma de uma percentagem pré-estabelecida e pela sua recriação aleatória. Por norma, a supressão cromossómica emprega ronda aproximadamente os 50% dos modelos totais criados, sendo que os excluídos correspondem aos modelos com menor qualidade.

Esta metodologia permite uma melhor exploração das combinações de variáveis e, por conseguinte, um melhor estudo dos descritores estruturais.

É importante referir que o método GA proporciona um conjunto de modelos de regressão eficazes, bem como, uma caracterização pormenorizada das variáveis utilizadas em cada modelo construído e sua influência no mesmo. O carácter de avaliação ou de qualidade dos modelos criados é quantificado por um ou mais critérios estatísticos, tais como, o coeficiente de correlação ( $R$ ), o factor  $F$  e o nível de preditividade ( $Q^2$ ).

A etapa final consiste na aplicação de outros três métodos de regressão; mais especificamente, a regressão linear múltipla (MLR), os mínimos quadrados parciais (PLS) e as redes neuronais artificiais (ANN). A utilização destes três métodos deve-se às diferentes vantagens e desvantagens que cada um apresenta no processo preditivo.

##### 1.2.4.2 – Modelos de regressão utilizados em QSAR

Alguns modelos de regressão que iremos analisar, além de correlacionarem variáveis, reduzem simultaneamente o seu número, podendo ser encontrados modelos de regressão lineares e não lineares. Dentro dos processos lineares existem os métodos de regressão simples, múltiplos (*Multiple Linear Regression*, MLR) e os mínimos quadrados parciais (PLS), enquanto que, relativamente aos métodos não lineares predominam a aplicação de redes neuronais artificiais (ANN) e os variantes de PCA e de PLS.

O uso de métodos de regressão simples ou múltiplos, na maioria dos casos, é impraticável devido à quantidade de variáveis independentes ser muito superior ao número de moléculas utilizadas na construção do modelo; por esta razão, apenas pode ser utilizada depois de se realizar PCA ou AG. O excesso de variáveis provoca o sobreajuste (*overfitting*), originando assim um ajuste de grande qualidade, mas

comprometendo a preditividade desejada. Por esta razão é necessário recorrer ao PLS, PCR, GA ou ANN.

### *Regressão dos componentes principais e mínimos quadrados parciais*

O método de regressão dos componentes principais (PCR) consiste em encontrar os PCs da matriz de dados  $X$  e, posteriormente, fazer uma regressão entre os componentes e a variável dependente ( $Y$ ) que no QSAR seria a actividade biológica. Esta via elimina o problema de multi-colinearidade das variáveis independentes (VI), apesar de persistir a procura do conjunto óptimo de variáveis dentro dos componentes seleccionados, explicando bem a variabilidade de  $X$ , mas não das propriedades de  $Y$ . Deste modo, não existe nenhuma garantia que todos os componentes encontrados em  $X$  sejam responsáveis pelo comportamento da actividade biológica (Garthwaite, 1994; Cruciani *et al.*, 2000b; Cruciani *et al.*, 2002; Gustafsson, 2001; Esbensen, 2002; Vilca, 2004; Tetko *et al.*, 2005).

O método dos mínimos quadrados parciais (PLS) é um processo iterativo, onde existem várias variantes que, por sua vez, definem diferentes algoritmos, os quais serão discutidos de forma breve e sem grande pormenor matemático. Como foi já referido anteriormente, o PCR não extrai informação da variável dependente, corrigindo-se este problema, através do PLS, ao conformar um conjunto de variáveis latentes (LV) sobre a base da maximização da covariância entre  $X$  e  $Y$ .

O esquema simplificado do método PLS é o seguinte (Garthwaite, 1994; Esbensen, 2002; Vilca, 2004):

Para um conjunto de  $k$  factores, o processo seguinte é repetido  $k$  vezes.

1)  $w = \text{cov}(Y, X)$  seguido pela normalização de  $w$  tal que:  $\|w\| = 1$ .

2)  $T = X \cdot w$

3)  $v = \frac{T^T \cdot Y}{T^T \cdot T}$

4)  $b = \frac{T^T \cdot X}{T^T \cdot T}$

5)  $X = X - T \cdot b$

6)  $Y = Y - T \cdot v$

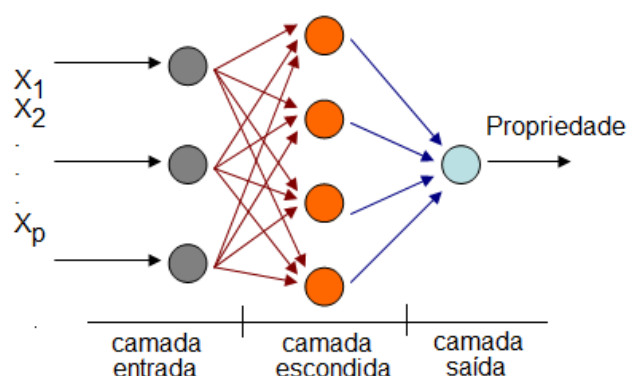
No passo 1, cada elemento da matriz  $w$  contém, basicamente, o coeficiente de regressão linear entre  $Y$  e cada variável  $X_i$ , sendo que, posteriormente esta matriz é normalizada na unidade. O passo 2 reporta-se à construção da matriz de variáveis latentes  $T$  ( $T_1 \dots T_k$ ), que contém a informação de  $X$ , como no PCR, e também de  $Y$ . Por sua vez, o passo 3 apresenta o cálculo do coeficiente de regressão simples entre  $T$  e  $Y$ , de maneira que  $Y_{pred} = v \cdot T$ . De forma semelhante ao passo 3, o passo 4 calcula o vector  $b$  cujos elementos correspondem ao coeficiente de regressão de cada variável  $X_i$  com respeito a  $T$ , pelo que se compreende que:  $X_{pred}^i = b_i \cdot T$ . Os dois últimos passos equivalem à actualização das matrizes originais para seguidamente repetir o mesmo procedimento  $k$  vezes. De referir que este é um processo iterativo onde, após a primeira iteração, as matrizes originais foram modificadas pela iteração anterior e em cada um destes ciclos é gerada uma nova variável latente  $T_k$ .

### *Redes neuronais artificiais*

O uso dos métodos anteriormente referidos, como o PCR, o PLS e o PCA, assumem *a priori* a forma de correlacionar os descritores moleculares e a actividade dos compostos, isto é, estes métodos baseiam-se nas correlações lineares, as quais podem não conduzir aos modelos ideais. De forma geral, a relação estrutura-actividade pode classificar-se como um problema de identificação de padrões e, portanto, qualquer ferramenta estatística utilizada nesta área matemática pode ser adaptada para a resolução deste problema.

Uma das ferramentas mais utilizada no reconhecimento de padrões e, mesmo, em QSAR são, sem sombra de dúvida, as redes neuronais, cujo fundamento matemático é extremamente complicado para realizar uma discussão detalhada. Além disso, as redes estruturais propostas para o desenho e previsão de fármacos são diferentes.

Uma rede neuronal artificial (*Artificial Neural Network*, ANN) é somente um modelo com o intuito de imitar as redes neuronais biológicas quanto à sua capacidade de analisar e adquirir conhecimento (Tetko, 2002), sendo esta aplicação muito versátil.



**Figura 7 - Esquema simplificado de uma rede neuronal.**

A estrutura geral de uma rede neuronal, como se pode observar na figura 7, consta normalmente de várias entradas neuronais interconectadas e uma saída; no entanto, pode também existir mais do que uma saída. É importante referir que cada neurónio da camada de entrada pode receber uma ou mais variáveis, ou seja,  $X_1 \dots X_p$  (Basak *et al.*, 1999a; Terfloth & Gasteiger, 2001; Kamphausen *et al.*, 2002; Tetko & Tanchuk, 2002; Novič & Roncaglioni, 2004).

A forma como cada neurónio processa o sinal de entrada é determinada por um conjunto de pesos iniciais e funcionais unidos à própria estrutura ou topologia da rede, permitindo obter vários níveis de processamento. No caso representado na figura 7, o conjunto de variáveis iniciais é processado para a obtenção de uma determinada propriedade, como por exemplo a actividade biológica ou a toxicológica (Novič & Roncaglioni, 2004).

Uma das características mais relevantes das ANN é a capacidade de “aprender”, ou seja, uma rede neuronal pode ser treinada, segundo determinado método, para a identificação de múltiplos padrões, sendo a técnica mais utilizada em QSAR para solucionar alguns problemas como:

- Análise de dados multidimensionais (como conjuntos de descritores moleculares ou mesmo os MIP).
- Classificação e previsão de propriedades físico-químicas ou biológicas.
- Comparação de compostos.



Deste modo, a aplicação de redes neuronais artificiais, assim como outros métodos não-lineares, tem vindo a desenvolver as capacidades computacionais e a melhorar os resultados na previsão e no desenho de fármacos (Basak *et al.*, 1999a; Terfloth & Gasteiger, 2001; Kamphausen *et al.*, 2002; Novič & Roncaglioni, 2004).

É importante referir que, apesar de este ser um método extremamente eficaz, as redes neuronais não apresentam uma função estrutura-actividade, mas sim um mecanismo matemático que não é exposto ao utilizador sendo que, conseqüentemente, a análise dos parâmetros estruturais se torna extremamente complexa.

#### 1.2.4.2 – Validação cruzada (*Cross-validation*)

A presença de um excesso de variáveis conduz, como já foi referido anteriormente, ao fenómeno de sobre-ajuste (*overfitting*) afectando assim as propriedades preditivas do modelo de regressão final. O método mais utilizado para a resolução deste problema é denominado por validação cruzada (*cross-validation*). Este procedimento consiste nos seguintes passos (Garthwaite, 1994; Esbensen, 2002; Vilca, 2004; Lawrence, 2005):

- 1) Os dados iniciais  $X$  ( $n \times p$ ) são permutados e divididos em  $k$  partes, de forma que cada parte  $V_j$ ,  $j = 1 \dots k$  tem aproximadamente  $n/k$  moléculas,  $k$  pode tomar vários valores incluindo  $n$ .
- 2) O modelo de regressão (PLS, ANN ou MLR) é calculado excluindo a  $j$ -ésima parte.
- 3) Com o modelo determinado, são calculados os valores da parte excluída:  $\hat{y}_i^j$ ,  $j=1 \dots k$ .
- 4) É calculada a soma de quadrados residuais (SCR) correspondentes:

$$SCR_j = \sum_i (y_i - \hat{y}_i^j)^2 \quad \text{ou coeficiente de preditividade } Q^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i^j)^2}{\sum_i (y_i - \bar{y})^2}$$

para  $j = 1 \dots k$ .

- 5) O SCR médio considerado é:  $\frac{1}{n} \sum_{j=1}^k SCR_j$

Como se verifica, o objectivo deste processo é comprovar a preditividade do modelo de regressão. Este procedimento permite determinar as condições óptimas para as capacidades preditivas, mas também possibilita a identificação de valores afastados

dentro da série de amostras, isto é, as moléculas com comportamentos muito diferentes (*outlier*).

É importante referir que esta validação do modelo construído pode ser executada de diversas maneiras, consoante o número de moléculas retiradas para o grupo de teste ( $V_j, j = 1 \dots k$ ). No passo 1 do procedimento pode-se retirar apenas uma molécula para a validação (validação *leave-one-out*) ou um grupo de moléculas  $k > 1$  (validação *leave-multiple-out*). O grupo  $k$  da validação pode conter compostos repetidos, devido a múltiplas escolhas de maneira aleatória, sendo que, neste caso o processo de validação chama-se *bootstrap*. Por último, caso se retire  $k$  moléculas (aproximadamente 30% da série em estudo) antes da construção do modelo preditivo, a validação denomina-se por externa.

## 2 – As piperazinas e os farmacóforos serotoninérgicos 1A e 2A

É largamente conhecido que as piperazinas possuem a capacidade de mimetizar os efeitos farmacológicos e tóxicos da MDMA, provocando assim os mesmos efeitos neurotóxicos (Michael *et al.*, 2005). Desta maneira, as piperazinas induzem lesões nos terminais nervosos serotoninérgicos, detectáveis através de biomarcadores bioquímicos e histológicos.

Os consumidores de drogas de abuso, como MDMA e piperazinas, apresentam uma redução significativa de receptores serotoninérgicos e do respectivo neurotransmissor, comprovando os efeitos neurodegenerativos em seres humanos (McCann *et al.*, 1998).



Figura 8 - Imagem ilustrativa do mecanismo de acção de drogas de abuso do tipo ecstasy (Adaptado de [www.drugabuse.gov](http://www.drugabuse.gov)).

De acordo com o mecanismo da acção biológica da MDMA, postula-se que as piperazinas actuam ao nível dos neurónios serotoninérgicos, promovendo a libertação da serotonina por permuta ao nível do transportador SERT (Transportador pré-sináptico de serotonina) e apresentando diversas fases excitatórias. Na primeira fase, as piperazinas penetram nas vesículas sinápticas e impulsionam o efluxo de serotonina (figura 8). Na segunda fase, o aumento de serotonina deve-se também à inibição da degradação metabólica do neurotransmissor pela monoaminoxidase (MAO) ao nível da mitocôndria. A terceira fase caracteriza-se por um conjunto de sinais celulares, tais

como: a inibição da biossíntese e armazenamento da serotonina nas vesículas, estimulação da libertação da serotonina na fenda sináptica, estimulação dos receptores pós-sinápticos e inibição do sistema de captação neuronal da serotonina. Por fim, a quarta e última fase representa um estado de taquifilaxia, ou seja, um esgotamento das reservas de serotonina, induzindo as alterações psicóticas referidas anteriormente (Carmo, 2007). Esta neurotoxicidade e possível neurodegeneração apresentam-se em diversas áreas como o córtex, o hipocampo, o tálamo, o gânglio basal e o centro de recompensa (Gee *et al.*, 2005; Capela *et al.*, 2006; Carmo, 2007).

O conhecimento de que o bloqueio de receptores 5-HT<sub>2</sub> e de que o papel modulador do receptor 5-HT<sub>1A</sub> contribuem para a terapêutica de alterações psicóticas, renova o interesse farmacológico e bioquímico no estudo da ligação destes receptores (Hoyer *et al.*, 2002). Deste modo, o estudo teórico torna-se uma ferramenta ideal e essencial na análise da afinidade entre os diferentes ligandos agonistas e antagonistas com os diferentes receptores, possibilitando um esclarecimento da especificidade de cada receptor e das suas funcionalidades em condições fisiológicas e fisiopatológicas.

## 2.1 – Estudo de farmacóforos para o receptor 5-HT<sub>1A</sub>

Nos receptores 5-HT<sub>1A</sub>, o modelo Mellin, como se pode observar na figura 9, possui um plano aromático e um átomo *dummy* ligado ao átomo N (do grupo amina), o qual está localizado a 2.1-2.6 Å abaixo do plano aromático a uma distância de 5.2-5.7 Å do centro aromático usual (Mellin *et al.*, 1991; Connie *et al.*, 1993). Esta foi uma das primeiras propostas de farmacóforo para o receptor 5-HT<sub>1A</sub>, tendo a construção deste modelo sido obtida a partir de uma família de ligandos lineares e anéis triciclos angulares como núcleo comum.

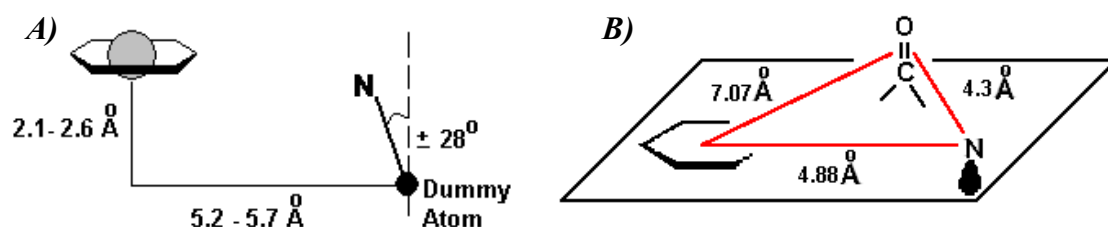


Figura 9 - Esquema ilustrativo dos farmacóforos propostos para os receptores 5-HT<sub>1A</sub> (Mellin *et al.*, 1993; López-Rodríguez *et al.*, 2002).

O denominado átomo *dummy* refere-se a um átomo não necessariamente ligado à molécula, como o átomo H externo advindo de uma ponte de hidrogénio que, no processo de alinhamento, funciona como ponto de referência para a pesquisa completa da conformação 3D. O alinhamento foi discutido anteriormente no capítulo 1, o qual é um passo necessário para a identificação do farmacóforo.

Este modelo farmacóforo foi otimizado, posteriormente, utilizando uma outra família de compostos (Mellin *et al.*, 1991; López-Rodríguez *et al.*, 2002) e propondo outro grupo aceitador de ligações de hidrogénio (figura 9B). Por outro lado, o átomo N correspondente à figura 9A foi também indicado como o aceitador central da ponte de hidrogénio. As distâncias entre o aceitador da ligação H e a estrutura linear ou cíclica, principalmente na vizinhança do átomo N, são um dos factores críticos para a afinidade deste receptor.

Por norma, aceita-se o modelo de 2 pontos (*2-point model*; figura 9A) como o ponto de partida para o reconhecimento molecular, tendo sido validado mesmo para a família de arilpiperazinas (Gaillard *et al.*, 1996; López-Rodríguez *et al.*, 2002); no entanto, o modelo de 3 pontos (*3-point model*; figura 9B) permanece válido, sendo aplicável a moléculas da família da buspirona (López-Rodríguez *et al.*, 2002).

Diversos estudos QSAR têm sido realizados, nos últimos dez anos, nas fenil- ou arilpiperazinas (Hibert *et al.*, 1989; Kuipers *et al.*, 1995; Gaillard *et al.*, 1996; López-Rodríguez *et al.*, 2002). Nas arilpiperazinas foi descoberta uma correlação entre a afinidade e a lipofilia, embora as propriedades estéricas e electrónicas permaneçam como os principais factores dominantes, representando, respectivamente, 47 e 53 % da correlação estrutura-actividade. Nas fenilpiperazinas, a contribuição das mesmas propriedades representam 98 e 2 %, respectivamente, em particular na vizinhança do átomo N4. O incremento do volume estérico pode ser quer positivo quer negativo para a actividade biológica, dependendo da acessibilidade relativamente ao átomo N4 (López-Rodríguez *et al.*, 2002).

Uma caracterização mais pormenorizada foi proposta para as N1-arilpiperazinas, em vez das N4-arilpiperazinas, onde as principais propriedades estruturais para o modelo farmacóforo são as seguintes:

1. Uma região estérica favorável e próxima do anel aromático;
2. Uma acessibilidade elevada para o grupo básico (grupo amina);
3. Uma carga negativa favorável e próxima da posição orto- do anel aromático;
4. Efeitos lipofílicos e electrostáticos opostos no substituinte de azoto.

Deste modo, conclui-se que as propriedades estéricas e electrónicas das moléculas são extremamente importantes, reflectindo-se, conseqüentemente, nos aspectos conformacionais, que devem ser considerados no desenho de novos fármacos, bem como na construção de um modelo farmacóforo. No caso dos derivados de piperazinas, um dos aspectos conformacionais mais importantes é a posição relativa do átomo N básico no respectivo anel aromático, dependendo do tipo de derivados, onde o átomo N pode estar ou não integrado na interacção  $\pi$ - $\pi$  com o receptor 5-HT<sub>1A</sub> (para uma informação mais detalhada ver as referências: Hibert *et al.*, 1989; Kuipers *et al.*, 1995; Gaillard *et al.*, 1996; López-Rodríguez *et al.*, 2002).

## 2.2 – Estudo de farmacóforos para o receptor 5 HT<sub>2A</sub>

A diversidade estrutural dos ligandos 5-HT<sub>2A</sub> representa um desafio em termos de definição de um modelo farmacóforo, embora os modelos propostos tenham concordado que a ligação com o respectivo receptor necessite de dois anéis aril e um grupo básico N. Segundo este modelo, Rowley *et al.*, em 2001, e Tammy *et al.*, em 2006, sugeriram que os antagonistas 5-HT<sub>2A</sub> podem ser divididos em duas classes: ligandos, nos quais os anéis aril e o grupo básico N possuem um arranjo triangular; e aqueles, cujos anéis aril e grupo básico N estão dispostos linearmente.

Os estudos de modificação molecular sobre centros de ligação demonstraram que o azoto (N) básico dos ligandos interage com o aminoácido carregado Asp155 localizado na hélice 3 do receptor 5-HT<sub>2A</sub>, sendo esta ligação importante, quer para agonistas quer para antagonistas. Investigações recentes sugerem que a interacção com o grupo N básico poderá ser mais complexo do que originalmente se supôs, devido a uma possível interacção adicional da amina protonada com a Ser159, localizada também na hélice 3 (Wang *et al.*, 1993; Rowley *et al.*, 2001). Apesar destas localizações no receptor 5-HT<sub>2A</sub>, são conhecidos outros fragmentos no mesmo receptor que podem estar envolvidos na interacção com o ligando (Michael *et al.*, 2006); como por exemplo, o Phe339 e o Phe340 (interacção  $\pi$ - $\pi$ ) presentes no domínio transmembranar directamente acoplados com o terceiro *loop* intracelular, o qual está relacionado com a capacidade do receptor activar a respectiva proteína G. Este facto confirmou a necessidade de uma análise mais pormenorizada das interacções ligando-proteína e do modelo farmacóforo.

O maior problema da estrutura da piperazina e, conseqüentemente, no correspondente modelo farmacóforo é o possível reconhecimento destas moléculas por outros receptores, tais como os receptores D<sub>2</sub> e os canais K<sup>+</sup> relacionados com o ritmo cardíaco. O grupo N básico e a interacção  $\pi$ - $\pi$  são substituintes estruturais comuns a inúmeras moléculas (incluindo os derivados de piperazinas), apresentando diferentes graus de afinidade com todos estes receptores, sendo que, em alguns casos, os efeitos colaterais destes fármacos podem justificar-se pelos diferentes tipos de interacções.

Num estudo recente (Tammy *et al.*, 2006), a presença do grupo N básico como ponto crítico para a interacção ligando-receptor foi colocado em questão. De facto, esta investigação demonstra que, pelo menos na família das piperidinas sulfona, a substituição do grupo N básico por outro grupo de estrutura não básica não afecta a afinidade e reduz mesmo a possibilidade de interagir com outros receptores. Esta experiência laboratorial cria uma nova perspectiva da interacção bioquímica que possibilita o desenho de novos fármacos e a simulação computacional.

*Parte II – Estudo in silico da relação estrutura-actividade  
de piperazinas nos receptores 5-HT<sub>1A</sub> e 5-HT<sub>2A</sub>*



### **3 – Objectivos**

O trabalho computacional desenvolvido teve como principais objectivos a construção de um modelo preditivo e a caracterização da afinidade das piperazinas para os receptores 5-HT<sub>1A</sub> e 5-HT<sub>2A</sub>. Deste modo, foram propostos os seguintes objectivos específicos:

- Obtenção de um conjunto de benzil- e de fenilpiperazinas com os respectivos valores de pK<sub>i</sub> para cada um dos receptores serotoninérgicos (subtipos 1A e 2A).
- Cálculo de descritores moleculares por aplicação de diversas metodologias.
- Aplicação de vários métodos matemático-estatísticos com o intuito de explorar diferentes modelos preditivos.
- Comparação e validação da preditividade dos modelos elaborados.
- Análise dos descritores e dos aspectos moleculares relacionados com a afinidade dos receptores em estudo.

## **4 – Materiais e métodos**

O trabalho computacional desenvolvido consiste essencialmente nas seguintes etapas: a elaboração de uma base de dados, a obtenção de descritores estruturais, a construção e a optimização dos modelos preditivos e análise das estruturas-actividade obtidas.

A metodologia QSAR adopta um conjunto de procedimentos que podem seguir vias de análise computacional distintas, consoante a informação inicial disponível, apresentando, contudo, passos obrigatoriamente comuns. Assim, a metodologia realizada em cada componente será explicada separadamente ao longo desta planificação experimental.

### **4.1 – Materiais e suporte informático**

Os *softwares* utilizados ao longo deste trabalho são expostos detalhadamente ao longo do procedimento experimental. Por sua vez, os cálculos computacionais necessários para a construção dos modelos empíricos estão indicados nas respectivas secções.

Todos os compostos utilizados correspondem à família das benzil- e das fenilpiperazinas, obtidos em artigos já publicados e em bases de dados públicas disponíveis até à data de conclusão desta dissertação. As diferentes piperazinas eleitas serão referidas em cada secção correspondente, apresentando-se uma listagem das mesmas no anexo I.

### **4.2 – Selecção e optimização geométrica da família de piperazinas**

A selecção de ligandos é o primeiro passo a realizar em estudos QSAR, procedendo-se à elaboração de uma base de dados de determinada família ou conjunto de compostos.

Esta consiste na selecção de piperazinas com valores experimentais de afinidade ( $IC_{50}$ ,  $K_i$  ou  $pK_i$ ) publicados em artigos científicos ou reunidos em outras bases de dados, como por exemplo a *Binding Database*<sup>2</sup> e *IUPHAR Database*<sup>3</sup>, de forma a obter

---

<sup>2</sup> Site: [www.bindingdb.org/](http://www.bindingdb.org/)

o maior número possível de piperazinas com afinidade para os receptores serotoninérgicos, mais propriamente os subtipos 5-HT<sub>1A</sub> e 5-HT<sub>2A</sub> (ver as tabelas I-VII em anexo). É importante referir que, embora as piperazinas seleccionadas estejam agrupadas nas tabelas como um só conjunto de moléculas, os modelos empíricos foram construídos para cada receptor específico, consoante os valores de pK<sub>i</sub> específicos disponíveis relativamente aos receptores 5-HT<sub>1A</sub> e 5-HT<sub>2A</sub>. Deste modo, as séries em estudo apresentam conjuntos de moléculas diferentes, apesar de existirem piperazinas em comum.

As moléculas devem apresentar similaridade para que o estudo possa ser considerado válido e eficiente, pois uma série de moléculas pouco similares origina modelos pouco produtivos; no entanto, a mesma série deve apresentar uma certa diversidade estrutural para que os modelos construídos tenham um amplo domínio de aplicação. O ideal seria partir de moléculas cujo valor de pK<sub>i</sub> fosse calculado pelo mesmo método e grupo de trabalho, de modo a reduzir os erros experimentais. No entanto, na prática, este procedimento é, muitas vezes, pouco valorizado quando se trabalha ou se constrói pequenas bases de dados, ao contrário do que se pode verificar neste trabalho experimental. Uma forma de reduzir estes erros parte da recolha de vários valores obtidos segundo diferentes métodos, bem como da realização de grupos de pesquisa e de cálculos das médias experimentais, estratégia essa que foi aplicada a muitas das moléculas estudadas, cujos valores de afinidades provieram de múltiplas fontes.

Os modelos empíricos em estudos QSAR são criados a partir dos descritores extraídos da estrutura tridimensional das moléculas seleccionadas, sendo que a obtenção de uma geometria é de grande importância.

Neste passo computacional, foram utilizados dois programas informáticos, mais propriamente, o *ChemOffice*<sup>4</sup> e o *MOPAC* (Stewart, 1993), embora existam ainda vários outros que permitem o estudo estrutural pretendido. O uso destes programas possibilita a descrição da geometria de cada molécula em estudo, analisando as energias de interacção e os seus respectivos gradientes para determinar a geometria molecular correspondente ao estado de mínima energia. Como foi já referido anteriormente, aplicam-se muitas vezes duas optimizações geométricas, de forma a obter resultados precisos o mais rapidamente possível.

---

<sup>3</sup> Site: [www.iuphar.org/](http://www.iuphar.org/)

<sup>4</sup> <http://www.cambridgesoft.com/software/ChemOffice/>

Após o desenho computacional de cada uma das piperazinas da série em estudo, obtidos graças ao *ChemOffice*, foi possível granjear uma geometria correcta com um reduzido nível teórico. No entanto, seguidamente e não obstante dos resultados antes alcançados, esse mesmo nível sofre uma optimização mais rigorosa através da utilização do *MOPAC*.

A mecânica molecular, mais especificamente o campo de força MM2, foi o método escolhido para a pré-optimização geométrica no *ChemOffice* (figura 10).

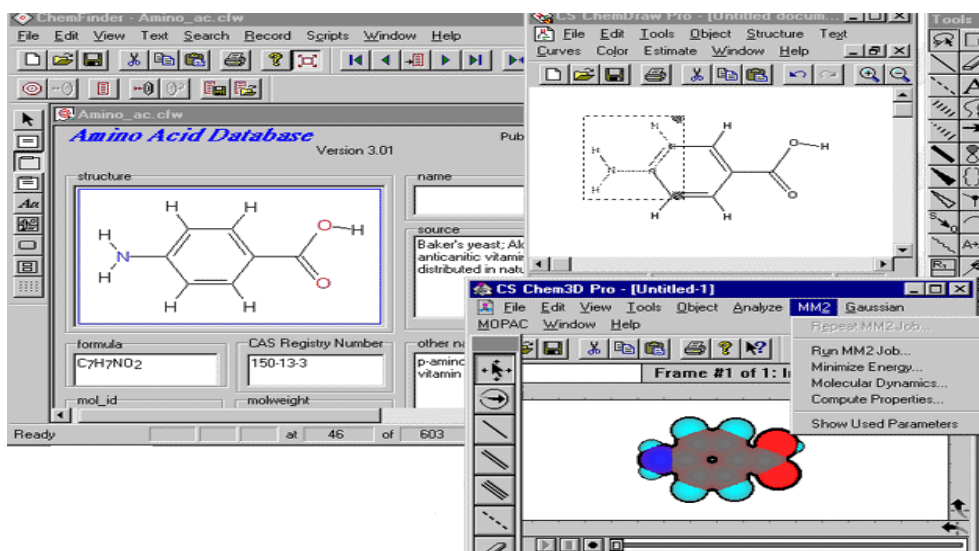


Figura 10 - Ilustração do software *ChemOffice* utilizado para o desenho e para a optimização geométrica das moléculas da série em estudo.

Em seguida, foi eleito o modelo Austin 1 (*AM1*) como o método a adoptar para o algoritmo de optimização geométrica no *MOPAC*, o qual utiliza meios semi-empíricos da Mecânica Quântica, de forma a obter resultados finais na optimização geométrica inicial das moléculas orgânicas (figura 11).

```

C:\Mopac6\MOPAC6.EXE
MOPAC v.6 for 3/486/Pentium PC's. Release 1.02, March 1997.
Windows 95 and NT environments
Universidad de La Habana, Cuba, 1994-1997

File management:
<fn>.MOP = input file name
<fn>.LOG = run data file name
<fn>.RSM = output resume file name
<fn>.OUT = main output file name
<fn>.RES = general data restart file name
<fn>.DEM = density matrix restart file name
<fn>.POP = Mulliken pop. file name
<fn>.CAR = Cartesian coordinate file name

Max. heavy atoms: 50; Max. hydrogen atoms: 50

Enter the input file name <fn> without .MOP extension => BZP
Do you want a trace on screen, although it takes a longer runtime? <Y/N> Y_
  
```

Figura 11 - Ilustração do software *MOPAC* utilizado para a re-optimização geométrica das moléculas da série em estudo.



Este procedimento é essencial para descrever correctamente a geometria molecular de cada uma das piperazinas; no entanto, permanece por esclarecer se a geometria calculada na optimização corresponde a um mínimo local ou absoluto.

### 4.3 – Obtenção de descritores estruturais

Nesta etapa computacional existem, como já foi referido na componente teórica, duas vias para a obtenção de parâmetros estruturais: a que somente considera as propriedades moleculares (geometria, densidade de cargas, etc); e aquela que se baseia no campo criado por esta, quer seja electrostático, quer seja o produto da interacção com outras moléculas (MIP). Para isso, foram utilizados, respectivamente, os *softwares e-Dragon* e *Volsurf* para cada uma das abordagens.

O *e-Dragon* corresponde a uma versão *website* do programa *Dragon*, o qual permite o cálculo de descritores moleculares desenvolvidos pelo grupo do Professor Todeschini (Tetko, 2005). Estes descritores permitem a avaliação das QSAR e QPSR de base de dados, como a série de piperazinas elaborada anteriormente.

O *e-Dragon* fornece 1666 descritores moleculares que se encontram divididos em 20 grupos lógicos (ver a secção 1.2.2), como se pode observar na figura 13. Assim, o investigador consegue calcular diversos tipos de descritores topológicos e geométricos (Tetko *et al.*, 2005).

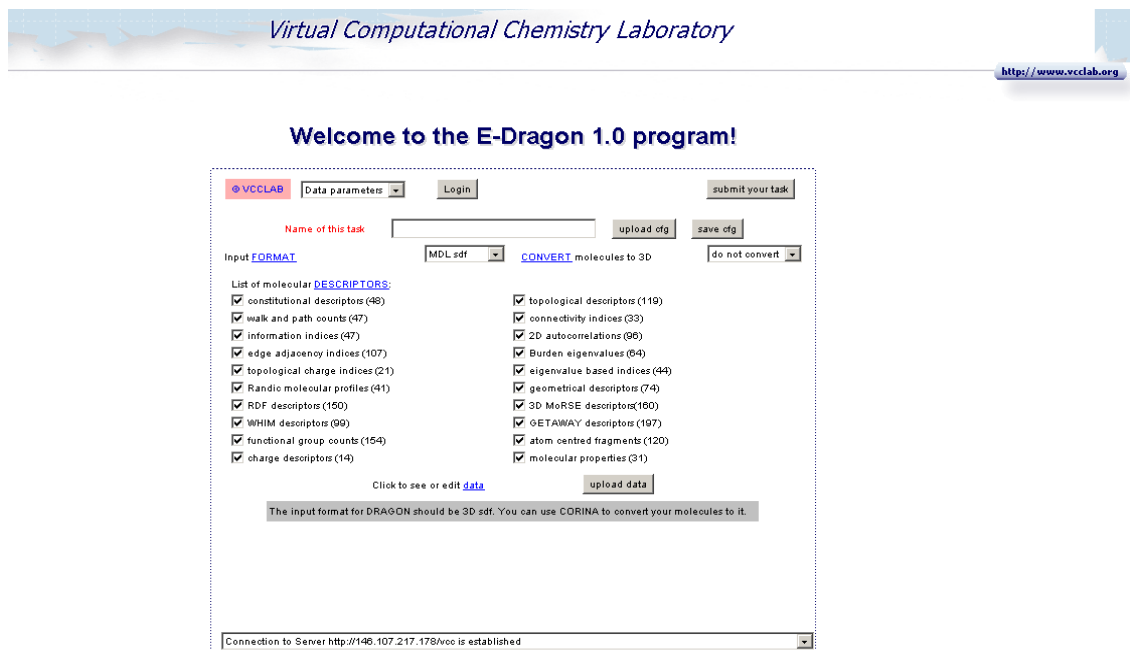


Figura 13 - Ilustração do *software e-Dragon* utilizado para a obtenção de descritores topológicos e geométricos das moléculas da série em estudo.

Para executar este programa é necessário que os ficheiros de dados estruturais<sup>5</sup> sejam obtidos por outros programas específicos de modelação molecular sendo que, neste caso específico, foram utilizados o *ChemOffice* e o *MOPAC* (Stewart, 1993), respectivamente, sendo que o *e-Dragon* aceita diversos formatos de ficheiros de dados estruturais, como por exemplo, SMILES, SDF (MDL) ou MOL2 (Sybyl).

Assim, os ficheiros estruturais das piperazinas foram convertidos em ficheiros com o formato SDF, através do *software Open Source Babel 2.1.1*<sup>6</sup>, antes de serem introduzidos no *e-Dragon* (figura 13).

Por sua vez, o *VolSurf*<sup>7</sup> é um programa que se baseia nos potenciais de interacção molecular (MIP) calculados a partir da série de moléculas em estudo, de forma a criar os seus descritores *VolSurf*, os quais perfazem um total de 94 descritores. Os *MIP* contêm as energias de interacção entre as moléculas de prova e os compostos examinados em todos os pontos espaciais da grelha de estudo (*grid box*). Neste trabalho aplicaram-se três moléculas de prova, mais concretamente, a H<sub>2</sub>O, a DRY e o grupo amida -NH, com uma resolução ou unidade de grelha correspondente a 0,5 Å (figura 14). A molécula de prova H<sub>2</sub>O fornece 48 descritores do número total obtido, enquanto que a DRY faculta 28 descritores, sendo os restantes pertencentes ao grupo amida -NH.

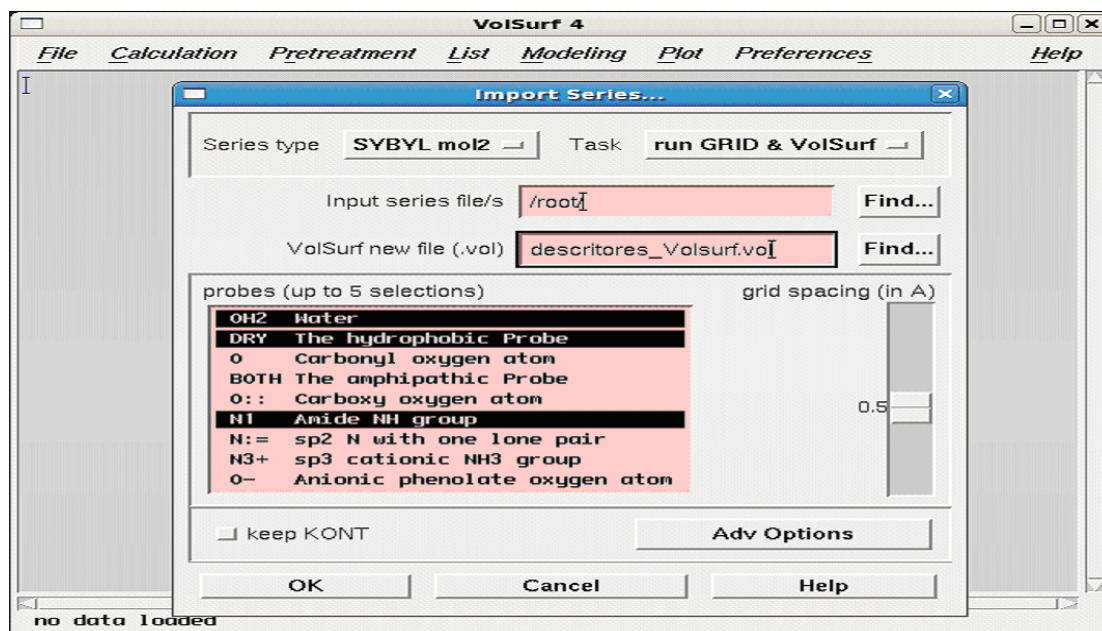


Figura 14 - Ilustração do *software Volsurf* utilizado para a obtenção de descritores de propriedades físico-químicas das moléculas da série em estudo.

<sup>5</sup> O *e-Dragon* requer que a optimização das moléculas seja realizada com estruturas contendo os respectivos átomos de hidrogénio.

<sup>6</sup> O *Open Source Babel 2.1.1* é o programa habitualmente utilizado para a conversão de formatos dos ficheiros de trabalho. (<http://openbabel.sourceforge.net/>)

<sup>7</sup> [http://moldiscovery.com/soft\\_volsurf.php/](http://moldiscovery.com/soft_volsurf.php/)

As propriedades farmacocinéticas das drogas dependem de uma variedade de parâmetros físico-químicos, necessitando, por isso, de uma descrição multivariada. De uma forma geral, o *VolSurf* possibilita a obtenção de descritores de propriedades físico-químicas significantes, a partir da estrutura 3D da molécula em estudo e sem a necessidade de um alinhamento inicial da mesma. Estes caracterizam quantitativamente o volume molecular, a polaridade, a hidrofobicidade e outras propriedades, bem como o balanço entre elas. A correlação entre os descritores resultantes é correctamente processada por intermédio de uma análise estatística multivariada e já integrada no *software* do *VolSurf*, embora também possa ser elaborado um tratamento independente deste programa através dos resultados obtidos.

#### **4.4 – Métodos estatísticos aplicados**

O processo de tratamento de dados é essencial para uma descrição correcta da estrutura química de cada uma das moléculas da série de estudo, independentemente do método computacional utilizado, dado que o número de descritores é extremamente elevado. Por conseguinte, torna-se necessário filtrar a informação obtida com o intuito de identificar e eliminar todos os dados irrelevantes ou repetidos e, assim, minimizar o ruído de fundo presente no processamento da predição. Para isso, torna-se necessário a aplicação de métodos estatísticos e matemáticos que permitam a simplificação do número de variáveis e a obtenção de uma correlação optimizada. Deste modo, foram aplicados diversos métodos estatísticos, entre os quais: a regressão linear múltipla (MLR), o método dos mínimos quadrados parciais (PLS), a análise de componentes principais (PCA) e os algoritmos genéticos (GA).

##### **4.4.1 – Análise de conjuntos (*Clusters Analysis*)**

O tratamento de dados é, como referido anteriormente, fundamental para a obtenção de um modelo preditivo adequado. No entanto, a descrição correcta da estrutura química de cada molécula e a filtração da informação obtida podem não ser suficientes para alcançar a predição pretendida através de métodos estatísticos. Este caso ocorre normalmente quando a série em estudo apresenta um elevado número de



compostos e uma diversidade estrutural relevante. Deste modo, torna-se necessária a formação de subconjuntos estruturais (*clusters*) da série em estudo, possibilitando assim uma análise mais profunda da diversidade estrutural presente e a construção de um modelo mais apropriado para cada subconjunto (*clusters*) elaborado.

A caracterização das estruturas químicas e seu potencial de interação tomam um papel central na formação de *clusters*. Quando as bases de dados ou séries de moléculas possuem um elevado número de informação, como acontece no nosso caso, torna-se complicado estudar a sua similaridade e, conseqüentemente, a elaboração de *clusters*. Este problema induz restrições significativas na complexidade dos algoritmos de pesquisa para a série em estudo; no entanto, existem vários modelos de pesquisa que combinam a velocidade e a minuciosidade do computador.

A caracterização das propriedades inerentes a cada estrutura química consiste em duas componentes: a descrição quantitativa dos termos relevantes de cada estrutura química em estudo e, em seguida, a determinação da forma como podemos comparar qualquer par de estrutura. Enquanto que, o primeiro componente atribui para cada estrutura uma série de valores, como o número de certo elemento presente ou determinado parâmetro físico-químico ou mesmo geométrico, o segundo componente determina a similaridade entre duas estruturas em estudo, através da comparação entre os dois conjuntos de parâmetros estabelecidos anteriormente. A similaridade entre as duas moléculas é representada por um número entre zero e um, onde o valor zero indica a inexistência de compatibilidade e o valor um aponta para uma homogeneidade completa.

A metodologia da similaridade implementada no *software Moloc*<sup>8</sup> baseia-se nas coordenadas atômicas e nas propriedades químicas das estruturas em estudo, correspondendo estas ao conjunto de descritores para a determinação da compatibilidade entre cada par de moléculas. Além disto, deve-se acrescentar que a função de similaridade é maximizada pela variação do par de estruturas em estudo até se encontrar o máximo de sobreposição entre as duas moléculas em questão. Deste modo, a análise de similaridade torna-se mais pormenorizada e rigorosa, em termos computacionais.

---

<sup>8</sup> <http://www.moloc.ch/>

#### 4.4.2 – Selecção das variáveis para cada receptor

O primeiro passo do procedimento estatístico consiste na simplificação do número de variáveis tendo, para isso, sido utilizada a análise dos componentes principais (PCA) e dos algoritmos genéticos (GA). O método PCA foi executado pelo programa *STATISTIC*<sup>9</sup>, enquanto que os GA foram efectuados no *software MatLab*<sup>10</sup> através de um programa desenvolvido com o auxílio do Dr. Eduardo Tejera. No entanto, foi previamente realizada a simplificação do número de variáveis através da exclusão de descritores que apresentavam uma gama de valores nulos superior a 80 % em toda a sua extensão, assim como uma intercorrelação igual ou superior a 90 %.

As características da aplicação computacional desenvolvida para a execução dos GA, bem como os parâmetros de cálculo, são as seguintes:

- Utilização do coeficiente de predictividade (*leave-one-out*),  $Q^2_{\text{LOO}}$ , como função de optimização.
- A população inicial consistiu em 600 cromosomas obtidos de forma probabilística. A probabilidade de cada variável é proporcional ao seu coeficiente de correlação parcial com o respectivo valor de afinidade.
- O cruzamento (*crossover*) é uniforme e com uma probabilidade de 0.5, ou seja, aproximadamente 50% das características de cada um dos pais são acopladas para a obtenção dos cromossomas-filhos, conservando-se os aspectos comuns dos progenitores.
- O número de ciclos nos cruzamentos efectuados, antes da mutação, é de 1000 iterações.
- A probabilidade de mutação é também de 50%, incrementando-se assim a possibilidade de explorar um espaço amplo de combinações entre variáveis seleccionadas.
- O critério de paragem baseia-se no limite de 100.000 iterações.
- Os métodos de regressão implementados para a construção dos modelos foram o MLR e o PLS, ambos com um número máximo de 5 variáveis.

---

<sup>9</sup> <http://awstats.sourceforge.net>

<sup>10</sup> <http://www.mathworks.com/>

Deste modo foram alcançados dez modelos finais possíveis para cada um dos receptores 5-HT (1A e 2A).

#### 4.4.3 – Métodos de regressão e de validação cruzada

Posteriormente ao PCA, foi utilizado o MLR como método de regressão, de forma semelhante aos AG, bem como o PLS como processo de comparação e exploração do número de variáveis em estudo.

Finalmente, a selecção final do modelo preditivo baseou-se nos seguintes critérios:

- O domínio de aplicação obtido pelo estudo dos gráficos de desvio padrão *versus leverage*, possibilitando assim a detecção de *outliers* e a identificação de moléculas que influenciam a preditividade dos modelos construídos.
- A capacidade preditiva que é estimada pelo valor de  $Q^2$  obtido, através da validação cruzada interna (*Leave-one-out internal cross-validation*, LOO-CV; *Leave-Multiple-out internal cross-validation*, LMO-CV; e BOO-CV, *Bootstrap cross-validation*), pelos coeficientes de determinação ( $R^2$  e  $R^2$  ajustado) e, bem como, pelos valores significantes de  $p$  em cada um dos modelos possíveis.

É importante referir que todos os critérios, anteriormente referidos, foram obtidos pelo programa *STATISCA* e utilizados no estudo de cada um dos *clusters*, quer no receptor 5-HT<sub>1A</sub> quer no receptor 5-HT<sub>2A</sub>.

## 5 – Resultados e discussão

A aplicação dos diversos métodos estatísticos anteriores proporcionou a escolha da melhor metodologia para o nosso estudo computacional. Deste modo, foram elaboradas várias combinações dos procedimentos disponíveis, tendo-se começado pelo estudo dos métodos vinculados à simplificação de variáveis, mais especificamente o PCA e o PLS.

Posteriormente, a eliminação dos descritores com elevada intercolinearidade (> 90%) ou um grande número de valores nulos permitiu a obtenção de 572 e 573 descritores *e-Dragon*, a partir dos 1666 descritores iniciais, para os receptores 5-HT<sub>1A</sub> e 5-HT<sub>2A</sub>, respectivamente. Enquanto que, no caso dos 94 descritores iniciais do *VolSurf*, foram somente seleccionados 63 e 64 descritores para o estudo QSAR dos receptores 5-HT<sub>1A</sub> e 5HT<sub>2A</sub>, respectivamente. Uma vez que o número de descritores *VolSurf* é relativamente escasso, logo torna-se desnecessário realizar uma análise dos componentes principais (PCA), ao contrário dos descritores *e-Dragon*. Os resultados do PCA demonstraram a indispensabilidade de um grande número de componentes principais (PC) para atingir uma elevada descrição da variância inicial, como se pode observar na figura 15. Deste modo, foi seleccionado um conjunto de 90 PC, a partir dos descritores *e-Dragon*, e que foram seguidamente aplicados aos GA juntamente com os descritores *VolSurf*.

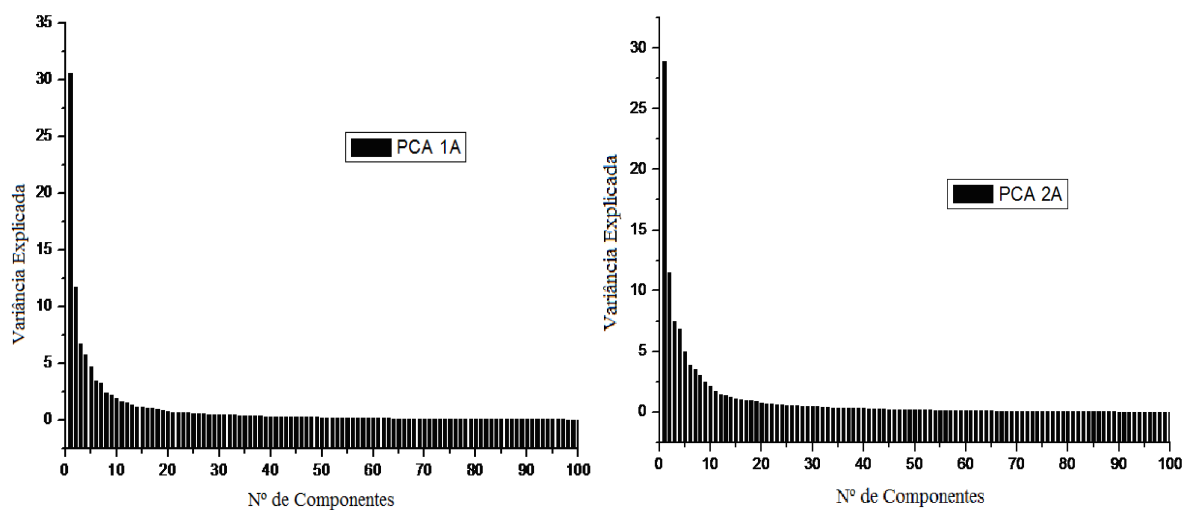


Figura 15 - Representação gráfica da análise de componentes principais (PCA) na série em estudo para os receptores 5-HT<sub>1A</sub> (à esquerda) e 5-HT<sub>2A</sub> (à direita).

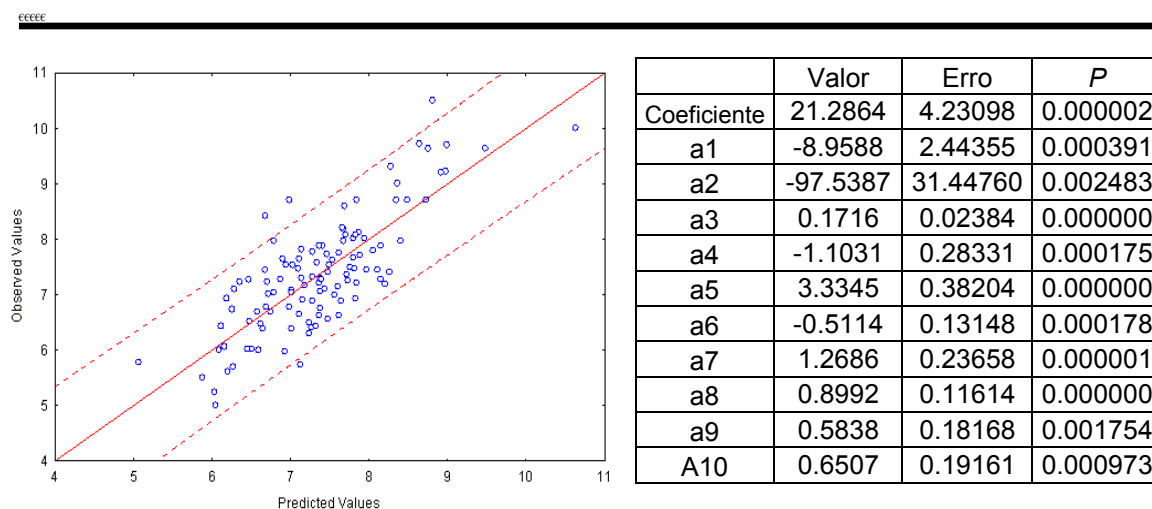
Os noventa componentes principais originaram, por intermédio do GA, um modelo de cinco variáveis com baixo valor de preditividade ( $Q^2 < 0.30$ ) para os dois receptores, sendo que, portanto, a combinação do PCA com o GA não possibilita a obtenção de modelos produtivos. No caso dos modelos construídos, a partir dos descritores do *VolSurf*, revelaram também uma fraca preditividade, mais especificamente de  $Q^2 = 0.11-0.13$  e de  $Q^2 = 0.30-0.32$  para os receptores 5-HT<sub>1A</sub> e 5-HT<sub>2A</sub>, respectivamente. Ambos os modelos preditivos possuem cinco variáveis de natureza diferente, ou seja, o modelo 5-HT<sub>1A</sub> foi construído a partir dos descritores *D3\_DRY*, *R\_OH2*, *D13\_OH2*, *ID8\_DRY* e *HB3\_NI*; enquanto que, o modelo 5-HT<sub>2A</sub> foi elaborado através dos descritores *V\_OH2*, *W6\_OH2*, *Cw2\_OH2*, *Cw5\_OH2* e *HB7\_NI*.

Por fim, podemos concluir que a metodologia por PCA e por descritores *VolSurf* é inadequada para o nosso estudo, devido a uma fraca qualidade preditiva e à incapacidade de recolher o comportamento global da série em estudo (figura 15). Assim, a aplicação do PLS e a obtenção dos modelos finais foi restritamente realizada com os descritores *e-Dragon* e sem recorrer aos respectivos componentes principais.

### 5.1 – Análise dos modelos preditivos para os receptores 5-HT

O método estatístico alternativo, em relação ao PCA, para o estudo proposto consiste na selecção dos melhores descritores por GA, enquanto que a correlação dos  $pK_i$  experimentais com os descritores moleculares seleccionados foi realizada por MLR e por PLS. O MLR, ao contrário do PLS, facilita a interpretação das propriedades físico-químicas inerentes à afinidade com os receptores 5-HT, mas, por outro lado limita a capacidade de exploração de um grande número de variáveis.

A figura 16 ilustra a aplicação da metodologia adoptada anteriormente para a construção de um modelo preditivo geral para os receptores 5-HT<sub>1A</sub>, onde o gráfico  $pK_i$  teóricos *versus*  $pK_i$  experimentais (à esquerda) mostra a qualidade de ajuste do modelo preditivo. Por seu lado, a tabela (à direita) apresenta o intervalo do erro e o valor de  $p$  associados a cada coeficiente inerente à equação do modelo construído.



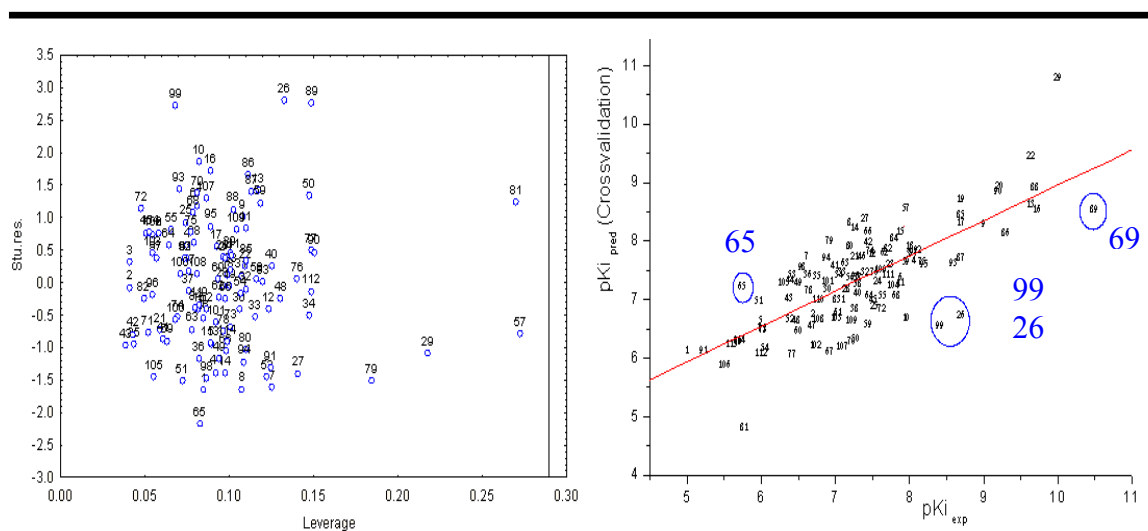
**Figura 16 – Representação gráfica por MLR (à esquerda) dos valores de  $pK_i$  teóricos *versus* experimentais e tabela (à direita) dos coeficientes do modelo preditivo geral para os receptores 5-HT<sub>1A</sub>.**

Verifica-se, então, que o modelo construído para o receptor 5-HT<sub>1A</sub> ilustra um ajuste de preditividade razoável, pois apresenta um coeficiente de preditividade ( $Q^2$ ) igual a 0.5538, e um  $p$  sempre inferior a 0.01 ( $p < 0.01$ ). No entanto, é de referir que o erro associado a alguns coeficientes de cada descritor utilizado não é o ideal e que  $Q^2$  pode ser optimizado.

Embora o estudo do ajuste do modelo construído, bem como o erro e o  $p$  associado a cada coeficiente sejam necessários para a validação do modelo preditivo,

estes parâmetros não são suficientes para comprovar a qualidade do modelo construído. Daí que se tenha procedido a um estudo mais pormenorizado (figura 17), mais especificamente no que diz respeito ao domínio de aplicabilidade e à validação do modelo construído, em particular a validação interna através da validação cruzada.

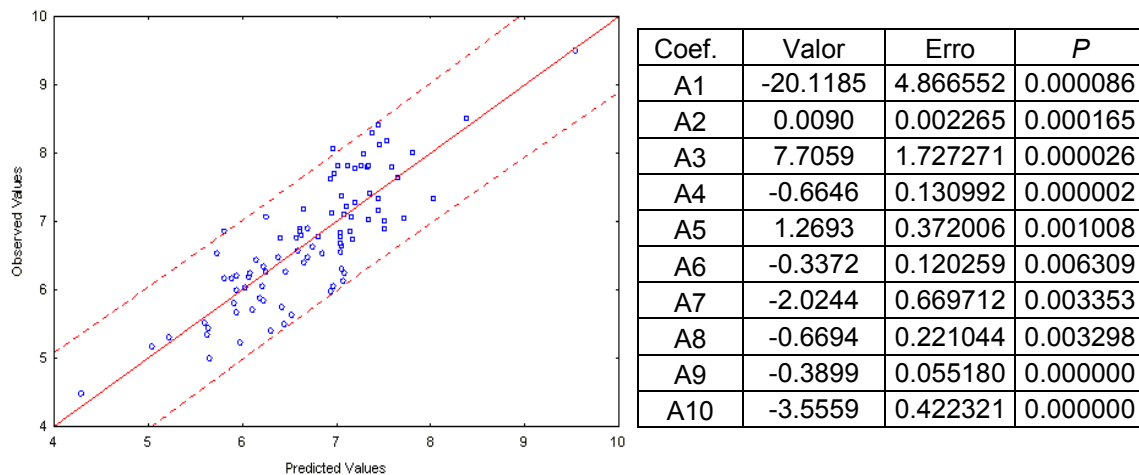
Na figura 17 observa-se o domínio de aplicabilidade e a validação interna do modelo construído para a série em estudo.



**Figura 17 - Representação gráfica (à esquerda) dos valores de *leverage* ( $h^* = 0.29$ ) versus desvio padrão das moléculas da série em estudo e representação gráfica (à direita) da validação cruzada para o modelo preditivo geral dos receptores 5-HT<sub>1A</sub>. Os compostos com círculos azuis apresentam um desvio padrão fora do intervalo normalizado ( $> \pm 2S$ ).**

Como se pode observar, o gráfico *leverage versus* desvio padrão (à esquerda) não apresenta nenhum composto com um *leverage* superior ao valor crítico ( $h^* > 0.29$ ). No entanto, evidencia quatro compostos que apresentam um desvio padrão fora do intervalo normalizado ( $> \pm 2S$ ), mais especificamente as moléculas 26, 65, 89 e 99 (ver as tabelas I-VII em anexo). Por sua vez, a validação cruzada interna ( $Q^2_{LOO}$ ; gráfico a direita) comprova, de facto, que estes compostos estão realmente afastados da recta traçada. Embora a eliminação das moléculas com um desvio padrão fora do intervalo normalizado (*oulier*) permita um aumento do valor de  $Q^2_{LOO}$ , bem como, das qualidades do ajuste, é preferível explorar outras alternativas que facilitem a diminuição do número de variáveis e *ouliers*.

O mesmo estudo metodológico foi realizado para os receptores 5-HT<sub>2A</sub>, onde a figura 18 demonstra a qualidade do ajuste, através do gráfico  $pK_i$  teórico *versus*  $pK_i$  experimental (à esquerda), bem como o intervalo do erro e o valor de  $p$  associados a cada coeficiente inerente à equação do modelo preditivo, como se pode observar na tabela (à direita).

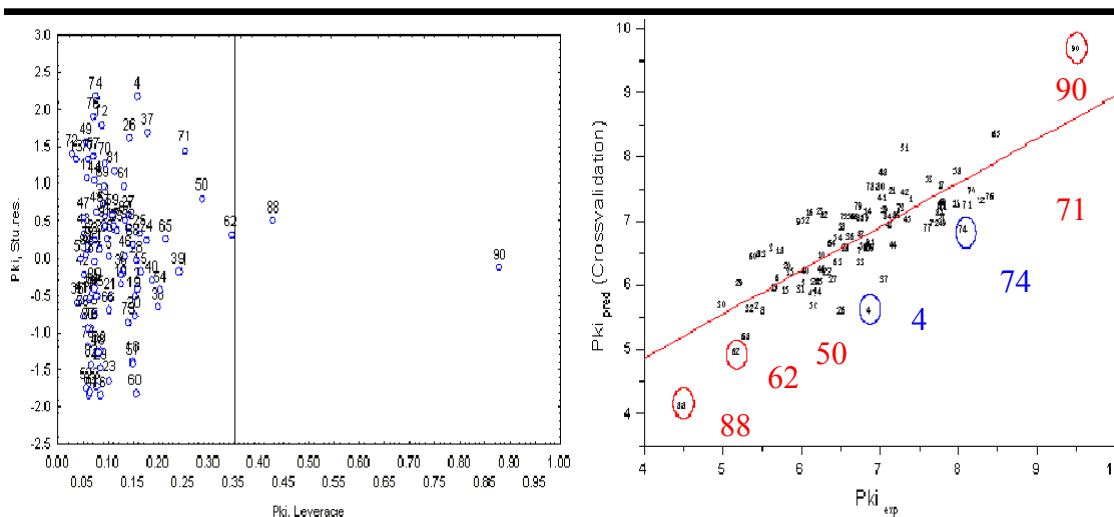


**Figura 18** - Representação gráfica por MLR (à esquerda) dos valores de  $pK_i$  teóricos *versus* experimentais e tabela (à direita) dos coeficientes do modelo preditivo geral para os receptores 5-HT<sub>2A</sub>.

Pode-se, então, constatar que o modelo elaborado apresenta um ajuste de qualidade ligeiramente superior ao obtido anteriormente dado que apresenta um coeficiente de preditividade de 0.6313, enquanto que no modelo do receptor 5-HT<sub>1A</sub> é somente de 0.5538. Apesar disto, o erro associado ao coeficiente de cada descritor seleccionado continua longe do ideal e o  $Q^2$  pode ser optimizado, embora, ambos os modelos demonstrem um  $p$  sempre inferior a 0.01 ( $p < 0.01$ ).

De forma semelhante ao verificado anteriormente, estes parâmetros não são suficientes para comprovar a qualidade do modelo construído. Por isso, realizou-se novamente um estudo minucioso ao domínio de aplicabilidade e à validação cruzada do modelo preditivo (figura 19).

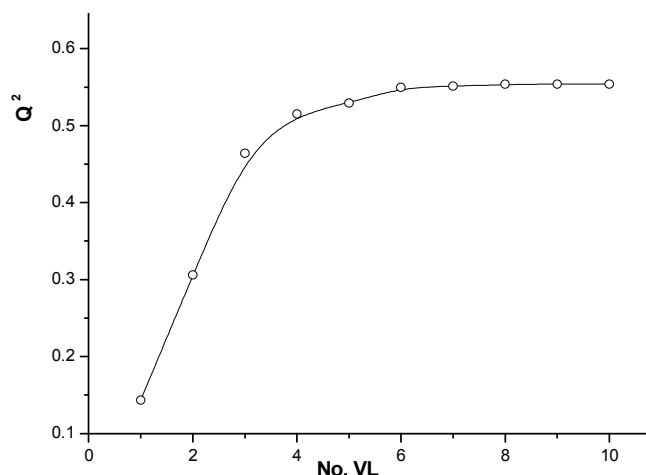




**Figura 19 - Representação gráfica (à esquerda) dos valores de *leverage* ( $h^* = 0.36$ ) versus desvio padrão das moléculas da série em estudo e representação gráfica (à direita) da validação cruzada para o modelo preditivo geral dos receptores 5-HT<sub>2A</sub>. Os compostos com círculos vermelhos apresentam um valor de  $h > h^*$ , enquanto que os compostos com círculos azuis apresentam um desvio padrão fora do intervalo normalizado ( $> \pm 2S$ ).**

O gráfico *leverage versus* desvio padrão (à esquerda) adverte para três moléculas, mais propriamente os compostos 62, 88 e 90 (ver as tabelas I-VII em anexo), que se encontram além do domínio de aplicação. Quanto ao desvio padrão, apenas os compostos 4 e 74 revelam valores superiores ao intervalo normalizado ( $> \pm 2S$ ). Por conseguinte, o gráfico da validação cruzada ( $Q^2_{LOO}$ ; gráfico à direita) confirma, de facto, que todos estes compostos estão afastados da recta delineada pelos valores de  $pK_i$  experimentais e teóricos, prejudicando assim o modelo preditivo.

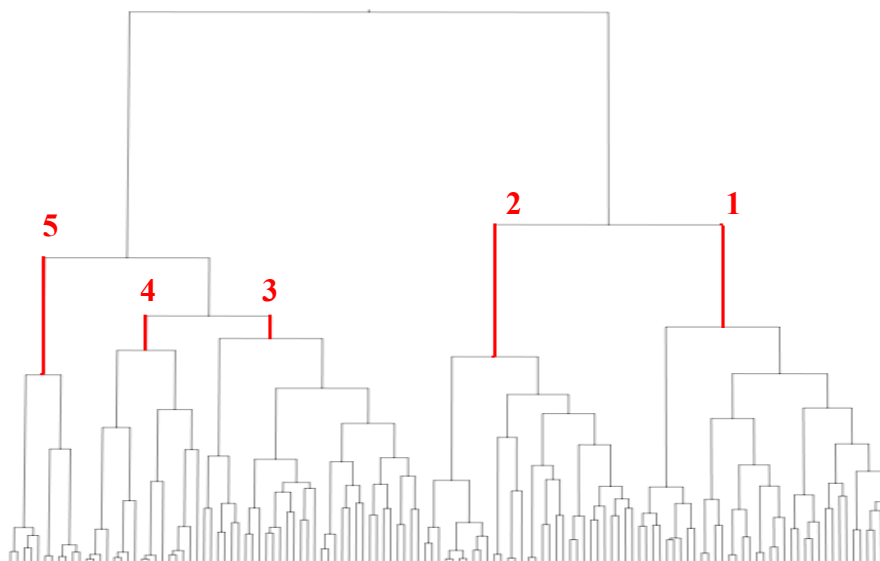
Por fim, é importante referir que foi utilizado o número fixo de dez descritores para a construção dos modelos preditivos gerais, quer para os receptores 5-HT<sub>1A</sub> quer para os receptores 5-HT<sub>2A</sub>. No entanto, não se pode afirmar que este seja o número ideal de descritores, pois a utilização de um número excessivo de descritores incrementa a probabilidade de *overfitting*, estabelecendo assim um modelo potencialmente inapropriado. Deste modo, o próximo passo consistiu no estudo do número de variáveis, neste caso de descritores moleculares, necessários para a obtenção de um modelo preditivo adequado ao estudo proposto (figura 20).



**Figura 20 - Representação gráfica do número de variáveis latentes *versus* o coeficiente de preditividade ( $Q^2$ ) obtido no modelo preditivo construído para os receptores 5-HT<sub>1A</sub>.**

A figura 20 exibe o estudo realizado do número de variáveis latentes, isto é, os descritores convertidos pelo método estatístico PLS *versus* o coeficiente de preditividade ( $Q^2$ ) máximo obtido, através dos GA, no modelo preditivo para os receptores 5-HT<sub>1A</sub>. Desta forma, observa-se que a recta do gráfico possui um comportamento típico de uma curva de saturação, concluindo-se então que somente três a cinco variáveis são necessárias para atingir o patamar de saturação na preditividade e formular assim um modelo com o mesmo grau de preditividade. No entanto, os modelos apresentam ainda assim um nível de preditividade relativamente baixo.

Após este estudo aos modelos preditivos iniciais concluiu-se que existem duas falhas que devem ser resolvidas ou optimizadas na metodologia: a preditividade e o domínio de aplicação. Verifica-se, de facto, que a preditividade dos dois modelos estava longe do valor mínimo permitido ( $Q^2 \geq 0.70$ ), pois apresentavam somente um coeficiente  $Q^2$  máximo de 0.5538 para 5-HT<sub>1A</sub> e 0.6313 para o 5-HT<sub>2A</sub>. Quanto ao domínio de aplicação, os modelos 5-HT<sub>1A</sub> e 5-HT<sub>2A</sub> apresentavam alguns compostos com  $h > h^*$ , bem como, quatro e dois compostos fora do intervalo normalizado para o desvio padrão, respectivamente. Estas complicações procedimentais derivam do grande número de piperazinas da série em estudo, conjugando uma enorme variedade estrutural implícita nos substituintes associados à piperazina. Daí ter-se optado pela subdivisão das moléculas em conjuntos com similaridade estrutural, ou seja, realizou-se uma análise de *clusters* (figura 21). Este passo procedimental permite diminuir o ruído de fundo na informação dos descritores, aumentar a preditividade do modelo para cada *cluster*, pois torna-se mais específico, e, conseqüentemente, cria um domínio de aplicação mais apropriado para cada subconjunto de piperazinas.



**Figura 21 - Ilustração do conjunto de *clusters* obtidos pelo programa *Moloc* por similaridade estrutural a partir das moléculas pertencentes à série em estudo. Os conjuntos seleccionados (1 a 5) para este estudo estão assinalados a vermelho.**

A figura 21 ilustra a divisão das 129 piperazinas da série em estudo em diversos *clusters*, conforme a similaridade estrutural partilhada. Como se pode observar, o número de *clusters* da série aumenta proporcionalmente com o grau de similaridade partilhada pelas moléculas do subconjunto. Os *clusters* seleccionados para o estudo computacional estão marcados a vermelho na figura acima, os quais se acredita possuírem o número de moléculas ideais para a obtenção de um modelo com maior preditividade e um melhor domínio de aplicação, dado o perfil estrutural de cada subconjunto ser mais refinado.

Por último, deve ter-se em mente que os mesmos cinco *clusters* são utilizados na construção de modelos preditivos para os receptores 5-HT<sub>1A</sub> e 5-HT<sub>2A</sub>, embora nem todos os compostos de cada *cluster* possuam ambos os valores de pK<sub>i</sub> para os dois receptores serotoninérgicos (tabela 4).

**Tabela 4 - O número de piperazinas que constituem cada um dos *clusters* seleccionados para a construção de modelos preditivos mais específicos para os receptores 5-HT<sub>1A</sub> e 5-HT<sub>2A</sub>.**

|   | <i>Cluster 1</i> | <i>Cluster 2</i> | <i>Cluster 3</i> | <i>Cluster 4</i> | <i>Cluster 5</i> | Total |
|---|------------------|------------------|------------------|------------------|------------------|-------|
| Nº de compostos para o 5-HT <sub>1A</sub> | 32               | 22               | 32               | 17               | 11               | 114   |
| Nº de compostos para o 5-HT <sub>2A</sub> | 16               | 23               | 24               | 16               | 11               | 90    |

## 5.2 – Análise dos clusters para os receptores 5-HT<sub>1A</sub>

Nesta componente experimental optou-se pela divisão da série de piperazinas analisadas em cinco *clusters* conforme a similaridade estrutural partilhada. Posteriormente, em cada um dos *clusters*, procedeu-se à construção de um modelo preditivo para cada um dos tipos de receptor (5-HT<sub>1A</sub> e 5-HT<sub>2A</sub>), como se pode observar nas tabelas 5 e 6.

**Tabela 5 - Os modelos preditivos obtidos para cada um dos cinco *clusters* por regressão múltipla a partir dos descritores seleccionados e os parâmetros estatísticos mais relevantes para os receptores 5-HT<sub>1A</sub>.**

|                                      |                  |  |                      |              |          |                                    |                                    |                                    |
|--------------------------------------|------------------|--|----------------------|--------------|----------|------------------------------------|------------------------------------|------------------------------------|
| <b>Modelos<br/>5-HT<sub>1A</sub></b> | <b>Cluster 1</b> | $pK_{1A} = 11.437(1.80) - 4.254(0.73) GATS4m + 0.148(0.04) Mor02m - 53.062(9.77) HATS7m$   |                      |              |          |                                    |                                    |                                    |
|                                      |                  | <b>R</b>   | <b>R<sup>2</sup></b> | <b>F</b>     | <b>P</b> | <b>Q<sup>2</sup><sub>LOO</sub></b> | <b>Q<sup>2</sup><sub>LMO</sub></b> | <b>Q<sup>2</sup><sub>BOO</sub></b> |
|                                      |                  | 0.845  | 0.715                | 23.387       | <0.01    | 0.621                              | 0.320                              | 0.550                              |
|                                      | <b>Cluster 2</b> | $pK_{1A} = 7.166(0.175) + 0.688(0.150) nCs - 1.094(0.157) nRNR2 + 0.453(0.125) C-025$      |                      |              |          |                                    |                                    |                                    |
|                                      |                  | <b>R</b>   | <b>R<sup>2</sup></b> | <b>F</b>     | <b>P</b> | <b>Q<sup>2</sup><sub>LOO</sub></b> | <b>Q<sup>2</sup><sub>LMO</sub></b> | <b>Q<sup>2</sup><sub>BOO</sub></b> |
|                                      |                  | 0.912  | 0.832                | 29.794       | <0.01    | 0.745                              | 0.610                              | 0.690                              |
|                                      | <b>Cluster 3</b> | $pK_{1A} = 5.708(0.117) RCI - 0.203(0.046) RDF140m - 3.422(0.588) Mor28v$                  |                      |              |          |                                    |                                    |                                    |
|                                      |                  | <b>R</b>   | <b>R<sup>2</sup></b> | <b>F</b>     | <b>P</b> | <b>Q<sup>2</sup><sub>LOO</sub></b> | <b>Q<sup>2</sup><sub>LMO</sub></b> | <b>Q<sup>2</sup><sub>BOO</sub></b> |
|                                      |                  | 0.997  | 0.994                | 1613.5<br>29 | <0.01    | 0.642                              | 0.400                              | 0.380                              |
|                                      | <b>Cluster 4</b> | $pK_{1A} = 19.211(1.781) - 0.410(0.100) MAXDN - 12.660(1.606) MATS2v - 72.323(11.768) Gle$ |                      |              |          |                                    |                                    |                                    |
|                                      |                  | <b>R</b>   | <b>R<sup>2</sup></b> | <b>F</b>     | <b>P</b> | <b>Q<sup>2</sup><sub>LOO</sub></b> | <b>Q<sup>2</sup><sub>LMO</sub></b> | <b>Q<sup>2</sup><sub>BOO</sub></b> |
|                                      |                  | 0.949  | 0.900                | 39.087       | <0.01    | 0.829                              | 0.813                              | 0.670                              |
|                                      | <b>Cluster 5</b> | $pK_{1A} = -3.667(0.951) + 2.990(0.361) PJI3 - 0.021(0.007) Mor02u + 20.283(1.301) Du$     |                      |              |          |                                    |                                    |                                    |
|                                      |                  | <b>R</b>   | <b>R<sup>2</sup></b> | <b>F</b>     | <b>P</b> | <b>Q<sup>2</sup><sub>LOO</sub></b> | <b>Q<sup>2</sup><sub>LMO</sub></b> | <b>Q<sup>2</sup><sub>BOO</sub></b> |
|                                      |                  | 0.990  | 0.980                | 112.744      | <0.01    | 0.960                              | 0.898                              | 0.895                              |

Os modelos empíricos para o receptor 5-HT<sub>1A</sub> demonstraram relativamente a cada um dos *clusters* uma enorme melhoria, quer na preditividade, quer na validação estatística, em comparação com o modelo geral (ver *secção 5.1*).

O modelo geral para o receptor 5-HT<sub>1A</sub> apresenta somente um coeficiente de preditividade (Q<sup>2</sup><sub>LOO</sub>) de 0.55, enquanto que o Q<sup>2</sup><sub>LOO</sub> dos modelos preditivos dos *clusters* varia entre 0.62 e 0.96, por sua vez, o cálculo dos coeficientes de preditividade Q<sup>2</sup><sub>LMO</sub> e Q<sup>2</sup><sub>BOO</sub> demonstra intervalos de 0.32-0.90 e de 0.38-0.90, respectivamente. Assim, observa-se, estatisticamente, a coerência da viabilidade dos modelos preditivos para cada *cluster*.

Na análise estatística dos modelos foram utilizados mais três parâmetros de avaliação. Na tabela 5 pode constatar-se que o intervalo de confiança é sempre superior a 99% ( $p < 0.01$ ), enquanto que o coeficiente de correlação e o  $F$  se encontram nos intervalos 0.72-0.98 e 23.39-1613.53, respectivamente. Pode-se então chegar à conclusão que a construção dos *clusters* e seus respectivos modelos permitem uma melhor preditividade e validação estatística.

No entanto, o ajuste dos modelos anteriores foi comprovado pelo estudo do domínio de aplicação das moléculas de cada *cluster* através do respectivo gráfico *leverage versus* desvio padrão, como se pode observar na figura 22.

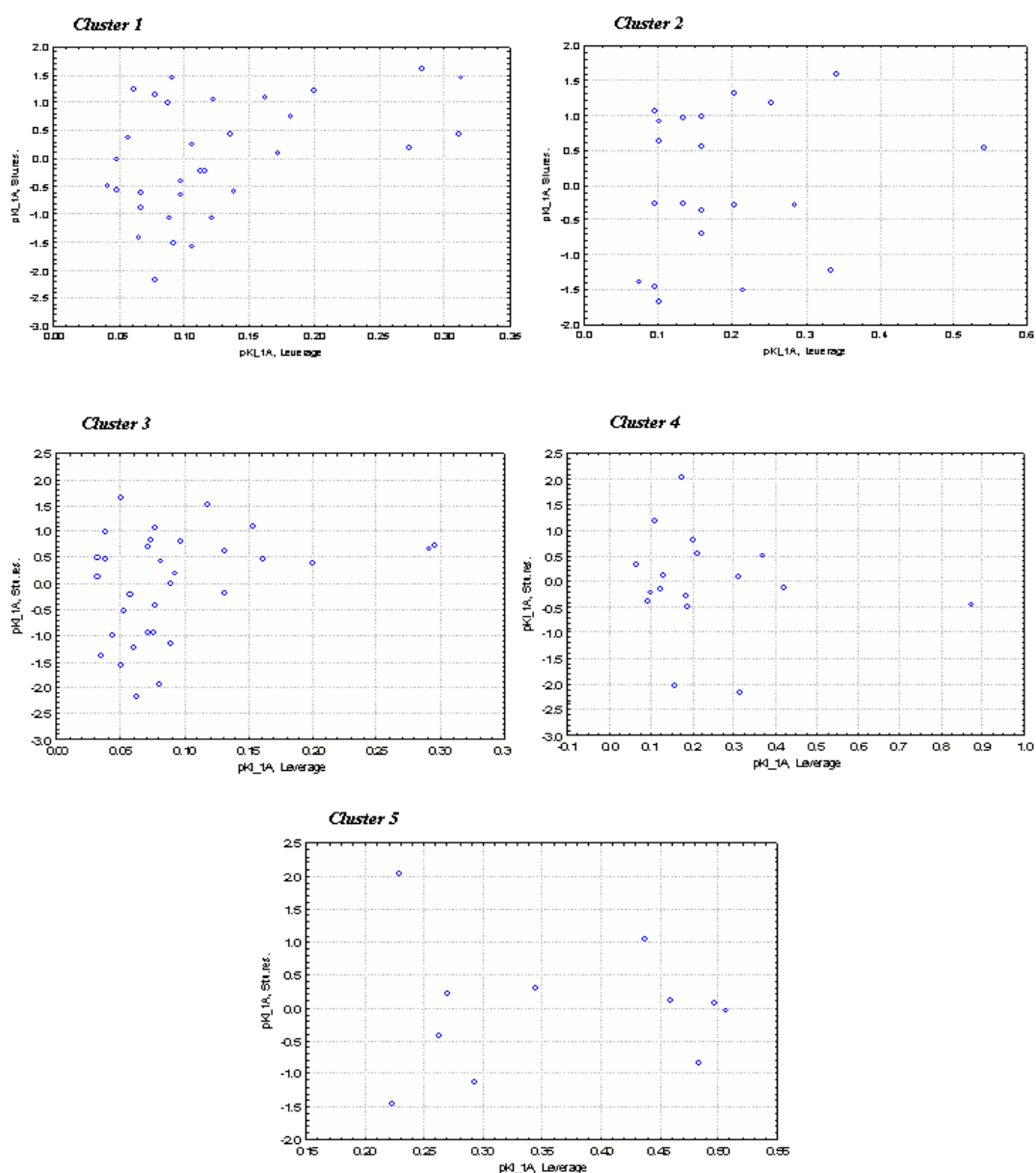


Figura 22 - Ilustração do conjunto de gráficos *leverage versus* desvio padrão para cada um dos *clusters* dos modelos preditivos relativamente ao receptor 5-HT<sub>1A</sub>. Os *leverages* críticos ( $h^*$ ) por *clusters* (1 a 5) são, respectivamente, os seguintes: 0.38, 0.54, 0.38, 0.71 e 1.09.

Os gráficos demonstram, de forma geral, um bom domínio de aplicação em cada um dos cinco *clusters*, apresentando apenas sete moléculas fora do desvio padrão ( $\pm 2s$ ) e três moléculas acima do *leverage* crítico ( $h > h_c$ ).

Os compostos fora do desvio padrão, mais propriamente as piperazinas 32<sup>11</sup>, 46<sup>11</sup>, 49<sup>12</sup>, 65<sup>13</sup>, 78<sup>14</sup>, 97<sup>12</sup> e 103<sup>12</sup> (ver os compostos na tabela em anexo), encontram-se no limiar do intervalo  $\pm 2s$ , de modo que não devem ser consideradas como sendo inaceitáveis, mas sim como pontos influentes que requerem uma maior atenção por parte do investigador. De igual modo, os compostos acima do *leverage* crítico, mais precisamente as piperazinas 29<sup>13</sup>, 69<sup>13</sup> e 106<sup>12</sup>, não representam qualquer risco para a autenticidade dos modelos construídos e, por isso, devem também ser considerados como pontos influentes dos respectivos *clusters*.

No modelo geral, nenhum composto se encontrava fora do domínio de aplicação (ver secção 5.1); quanto ao *leverage*, no entanto, o mesmo não se verifica nos *clusters*. Mais especificamente, os compostos 29, 69 e 106 surgem agora como moléculas acima do *leverage* crítico ( $h > h_c$ ), quando inicialmente apareciam como pertencentes ao domínio de aplicação. Por outro lado, o mesmo desfecho se observa no estudo do domínio de aplicação quanto ao desvio padrão, onde inicialmente os compostos 26, 65, 89 e 99 se apresentavam fora do intervalo  $\pm 2s$  e, actualmente, apenas o composto 65 se mantém fora do intervalo do desvio padrão, verificando-se também o surgimento de outros compostos, mais precisamente 46, 64, 78, 97 e 103, com um desvio superior a  $\pm 2s$ . Estas mudanças devem-se, simplesmente, ao facto de a série inicial ter sofrido a divisão em *clusters* (ver gráfico 21), isto é, as 129 piperazinas em estudo foram subdivididas em cinco conjuntos diferentes segundo critérios de similaridade estrutural, alterando assim os descritores seleccionados para o modelo preditivo em cada *cluster* (ver a tabela 5). Por conseguinte, o domínio de aplicação deixa de ser igual para toda a série e varia consoante o *cluster* que se esteja a estudar.

Embora se verifique uma melhoria bastante significativa na preditividade dos modelos preditivos de cada um dos cinco *clusters* em relação ao modelo geral, note-se que os *clusters* 1 e 3 apresentam uma preditividade inferior aos restantes conjuntos. Isto deve-se ao facto de estes dois conjuntos apresentarem um número elevado de compostos tendo, por isso, uma maior diversidade estrutural. Logo, uma hipótese de melhorar a

---

<sup>11</sup> Cluster 5

<sup>12</sup> Cluster 4

<sup>13</sup> Cluster 3

<sup>14</sup> Cluster 1

preditividade dos modelos seria a re-fragmentação destes dois *clusters* ou a eliminação dos respectivos *outliers*.

Com o propósito de comprovar, uma vez mais, a capacidade dos descritores *VolSurf*, estes foram empregados no grupo 1 (Cluster 1) com a mesma metodologia utilizada anteriormente nos descritores *e-Dragon*. Os resultados finais demonstraram, os quais não estão aqui representados, uma qualidade muito inferior, mais exactamente de  $Q^2_{\text{LOO}} = 0.34-0.45$ . Por esse motivo, neste trabalho computacional foram somente utilizados os descritores do *e-Dragon*.

Por fim, pode-se concluir que os modelos preditivos para o receptor 5-HT<sub>1A</sub> dos cinco *clusters* se apresentam aptos, quer em ajuste quer em validação estatística, para a previsão da relação estrutura-actividade de novas piperazinas.

### 5.3 – Análise dos clusters para os receptores 5-HT<sub>2A</sub>

Os modelos empíricos para o receptor 5-HT<sub>2A</sub>, obtidos a partir dos mesmos *clusters*, evidenciam melhores resultados que o modelo geral inicial e os modelos preditivos para o receptor 5-HT<sub>1A</sub>. Assim, o modelo 5-HT<sub>2A</sub> inicial apresenta um coeficiente de preditividade de 0.63, enquanto que os modelos 5-HT<sub>2A</sub> dos *clusters* ostentam um coeficiente  $Q^2_{LOO}$  no intervalo entre 0.83 e 1.00 (tabela 6), bem como coeficientes  $Q^2_{LMO}$  e  $Q^2_{BOO}$  com intervalos de 0.57-0.98 e de 0.67-0.94, respectivamente, os quais são também superiores aos intervalos dos modelos 5-HT<sub>1A</sub> (ver secção 5.2).

**Tabela 6 - Os modelos preditivos obtidos para cada um dos cinco *clusters* por regressão múltipla a partir dos descritores seleccionados e os parâmetros estatísticos mais relevantes para os receptores 5-HT<sub>2A</sub>.**

|                                      |                  |   |                      |              |          |                                    |                                    |                                    |
|--------------------------------------|------------------|---|----------------------|--------------|----------|------------------------------------|------------------------------------|------------------------------------|
| <b>Modelos<br/>5-HT<sub>2A</sub></b> | <b>Cluster 1</b> | $pK_{2A} = 12.301(0.625) + 14.119(1.248) MATS1v - 1.489(0.176) GATS8m - 0.113(0.024) RDF045m$ |                      |              |          |                                    |                                    |                                    |
|                                      |                  | <b>R</b>  | <b>R<sup>2</sup></b> | <b>F</b>     | <b>P</b> | <b>Q<sup>2</sup><sub>LOO</sub></b> | <b>Q<sup>2</sup><sub>LMO</sub></b> | <b>Q<sup>2</sup><sub>BOO</sub></b> |
|                                      |                  | 0.971   | 0.942                | 64.974       | <0.01    | 0.928                              | 0.870                              | 0.870                              |
|                                      | <b>Cluster 2</b> | $pK_{2A} = 7.761(0.573) - 1.162(0.182) Mor15m + 5.072(1.350) E3u + 0.713(0.139) C-025$        |                      |              |          |                                    |                                    |                                    |
|                                      |                  | <b>R</b>  | <b>R<sup>2</sup></b> | <b>F</b>     | <b>P</b> | <b>Q<sup>2</sup><sub>LOO</sub></b> | <b>Q<sup>2</sup><sub>LMO</sub></b> | <b>Q<sup>2</sup><sub>BOO</sub></b> |
|                                      |                  | 0.909   | 0.827                | 30.268       | <0.01    | 0.776                              | 0.740                              | 0.670                              |
|                                      | <b>Cluster 3</b> | $pK_{2A} = 3.328(0.423) + 0.110(0.013) RDF065m + 0.144(0.044) Mor03m + 162.926(23.488) R4v+$  |                      |              |          |                                    |                                    |                                    |
|                                      |                  | <b>R</b>  | <b>R<sup>2</sup></b> | <b>F</b>     | <b>P</b> | <b>Q<sup>2</sup><sub>LOO</sub></b> | <b>Q<sup>2</sup><sub>LMO</sub></b> | <b>Q<sup>2</sup><sub>BOO</sub></b> |
|                                      |                  | 0.907   | 0.822                | 30.806       | <0.01    | 0.751                              | 0.570                              | 0.710                              |
|                                      | <b>Cluster 4</b> | $pK_{2A} = 24.826(1.930) - 24.947(3.033) P2u - 31.617(3.391) Du - 1.451(0.301) nHDon$         |                      |              |          |                                    |                                    |                                    |
|                                      |                  | <b>R</b>  | <b>R<sup>2</sup></b> | <b>F</b>     | <b>P</b> | <b>Q<sup>2</sup><sub>LOO</sub></b> | <b>Q<sup>2</sup><sub>LMO</sub></b> | <b>Q<sup>2</sup><sub>BOO</sub></b> |
|                                      |                  | 0.952   | 0.906                | 38.608       | <0.01    | 0.848                              | 0.870                              | 0.800                              |
|                                      | <b>Cluster 5</b> | $pK_{2A} = 6.219(0.479) DISPe + 20.623(1.163) Glu + 0.492(0.034) nCb-$                        |                      |              |          |                                    |                                    |                                    |
|                                      |                  | <b>R</b>  | <b>R<sup>2</sup></b> | <b>F</b>     | <b>P</b> | <b>Q<sup>2</sup><sub>LOO</sub></b> | <b>Q<sup>2</sup><sub>LMO</sub></b> | <b>Q<sup>2</sup><sub>BOO</sub></b> |
|                                      |                  | 1.000   | 1.000                | 12893.5<br>4 | <0.01    | 0.975                              | 0.980                              | 0.940                              |

A análise estatística para estes modelos dependeu, novamente, dos parâmetros estatísticos apresentados na tabela acima; mais precisamente do coeficiente de correlação ( $R^2$ ), do teste estatístico  $F$  e do intervalo de confiança ( $p$ ). Na tabela 6, os modelos apresentam um intervalo de confiança sempre superior a 99% ( $p < 0.01$ ), um coeficiente de correlação e um  $F$  nos intervalos situados entre 0.83 até 1.00 e 30.27 até



12893.54, respectivamente. Assim, os modelos preditivos dos *clusters* permitem, uma vez mais, um aumento da preditividade e da validação estatística.

Entretanto, o ajuste dos modelos preditivos foi estudado através do domínio de aplicação das moléculas de cada um dos *clusters* por intermédio dos respectivos gráficos *leverage versus* desvio padrão (figura 23).

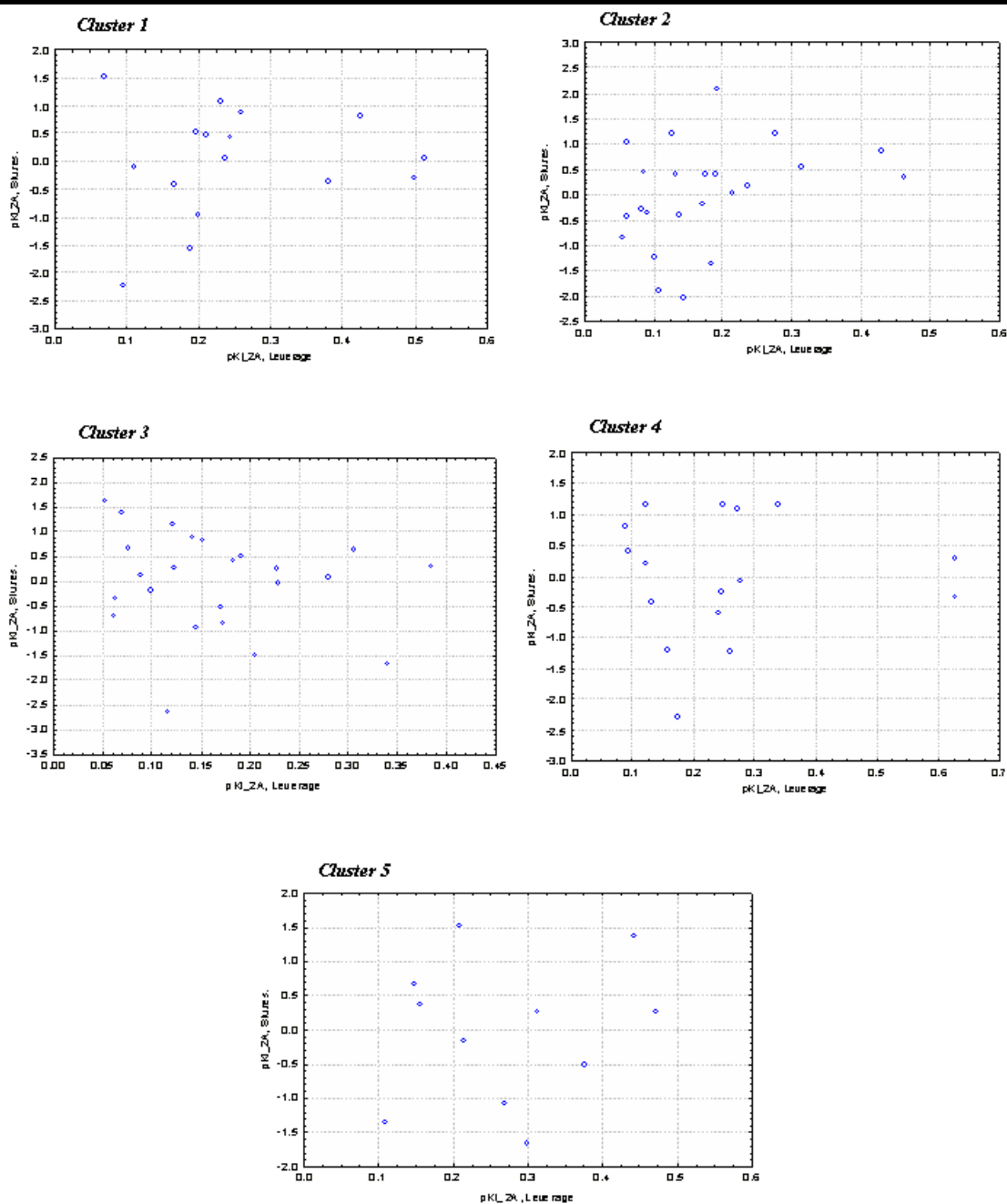


Figura 23 - Ilustração do conjunto de gráficos *leverage versus* desvio padrão para cada um dos *clusters* dos modelos preditivos para o receptor 5-HT<sub>2A</sub>. Os *leverages* críticos ( $h^*$ ) por *clusters* (1 a 5) são, respectivamente, os seguintes: 0.75, 0.52, 0.50, 0.75 e 1.09.

A figura 23 comprova um bom domínio de aplicação para cada um dos cinco modelos preditivos 5-HT<sub>2A</sub>, os quais apresentam um ajuste superior comparativamente aos modelos 5-HT<sub>1A</sub>. Verifica-se que nenhuma molécula possuía um *leverage* superior ao valor crítico ( $h > h_c$ ) e que apenas cinco moléculas evidenciavam um desvio padrão ligeiramente fora do intervalo ( $\pm 2s$ ), mais exactamente os compostos 2<sup>15</sup>, 4<sup>16</sup>, 15<sup>16</sup>, 19<sup>17</sup> e 35<sup>18</sup> (ver os compostos na tabela em anexo). Embora existam cinco piperazinas com desvio padrão superior ao intervalo estabelecido, estas devem ser consideradas nos respectivos modelos como pontos influentes e não como pontos conflituosos, sendo uma mais valia na interpretação das propriedades moleculares inerentes à relação estrutura-afinidade com os receptores em questão. De referir que um determinado composto apenas deve ser excluído da série em estudo quando apresentar simultaneamente um *leverage* ( $h > h_c$ ) e um desvio padrão ( $> \pm 2s$ ) fora do domínio de aplicação estabelecido.

No modelo 5-HT<sub>2A</sub> geral, observou-se que três compostos, mais especificamente 62, 88 e 90, possuem um *leverage* superior ao valor crítico ( $h > h_c$ ); enquanto que nenhum modelo preditivo dos *clusters* para o mesmo receptor apresenta um caso semelhante. No entanto, o modelo geral apresenta somente duas piperazinas com um desvio padrão fora do intervalo ( $\pm 2s$ ), mais propriamente as moléculas 4 e 74. Embora o composto 74 já não se apresente fora do intervalo do desvio padrão, o composto 4 e mais quatro piperazinas (2, 15, 19 e 35) encontram-se na situação oposta. Esta mudança no domínio de aplicação, tal como se verificou anteriormente (ver secção 5.2), deve-se à divisão da série inicial (ver gráfico 21) por cinco *clusters* (ver a tabela 4), alterando assim os descritores seleccionados em cada modelo preditivo. Consequentemente, os critérios do domínio de aplicação para cada *cluster* são também modificados, de forma a maximizar a preditividade e o ajuste das piperazinas em cada conjunto segundo a similaridade estrutural partilhada por estas.

Deste modo, pode-se concluir que os modelos preditivos 5-HT<sub>2A</sub> obtidos, a partir dos cinco *clusters*, são adequados, quer estatisticamente quer pelo ajuste, para a previsão da relação estrutura-actividade de novas piperazinas.

---

<sup>15</sup> Cluster 1

<sup>16</sup> Cluster 2

<sup>17</sup> Cluster 3

<sup>18</sup> Cluster 4

#### 5.4 – Afinidade das piperazinas para os receptores 5-HT<sub>1A</sub> e 5-HT<sub>2A</sub>

Nesta última componente computacional são discutidos os parâmetros moleculares, como as propriedades topológicas e as propriedades químico-físicas, inerentes às piperazinas da série em estudo que hipoteticamente elucidam a sua afinidade para os receptores serotoninérgicos 5-HT<sub>1A</sub> e 5-HT<sub>2A</sub>.

A selecção e a obtenção de descritores moleculares, através de programas como o *Dragon*, serviram para a construção de modelos preditivos para cada receptor, mas também para o estudo das propriedades mais relevantes na afinidade das piperazinas para com os receptores. É importante referir que a interpretação das propriedades moleculares não foi exclusivamente baseada nos descritores pertencentes aos modelos preditivos, os quais, na sua maioria, não continham uma informação viável ou de fácil interpretação química e topológica. Para isso, foi elaborada uma matriz de correlação entre todos os descritores finais e aqueles de fácil interpretação, de modo a esclarecer a informação estrutural e molecular partilhada por estes. Assim, a interpretação molecular dos descritores pertencentes aos modelos preditivos tornou-se mais acessível e simples de “traduzir”.

O estudo da matriz permitiu observar a informação compartilhada e dissociada pelos dois tipos de receptores serotoninérgicos. Note-se que, apesar do número de *clusters* existente, algumas piperazinas têm simultaneamente afinidade para os receptores 5-HT<sub>1A</sub> e 5-HT<sub>2A</sub>, os quais pertencem à mesma família (5-HT), partilhando assim determinado tipo de características intrínsecas no centro de activo. Além do grupo funcional principal, ambos os receptores partilham uma afinidade com grupos aromáticos com substituintes fortemente electrotopológicos ou electronegativos, bem como, uma certa conformidade para a presença de ligações RCONHR na molécula da piperazina semelhantes às ligações peptídicas presentes nas proteínas. Estes receptores compartilham também uma empatia para o carbono *alfa* (*sp*<sup>3</sup>) ligado a certos heteroátomos (O e N) e para as aminas alifáticas, embora de natureza diferente como se verá a seguir. Além disso, observou-se que ambos apresentam uma desarmonia com o

elemento Bromo (Br) e com determinados grupos funcionais, como por exemplo, o -N(CO)<sub>2</sub> e -CX<sub>3</sub> (onde cada X equivale a um determinado halogéneo ou calcogénio<sup>19</sup>).

Quanto à especificidade, verificou-se, de facto, que os receptores 5-HT<sub>1A</sub> demonstravam uma maior afinidade electrónica para piperazinas com calcogéneos, mais especificamente com os elementos de Oxigénio (O) e de Enxofre (S), enquanto que os receptores 5-HT<sub>2A</sub> elegem piperazinas com halogéneos, mais exactamente os elementos de Flúor (F) e de Cloro (Cl). Além disso, ambos os receptores revelam uma afinidade acrescida para os respectivos grupos (calcogéneos e halogénios) quando presentes em estruturas aromáticas, principalmente nas posições *orto*- e *meta*-.

Embora cada *cluster* apresente no seu grupo de piperazinas um peso molecular e um grau de saturação semelhante, os resultados apontam para a existência de uma correlação entre estes factores e os receptores 5-HT, onde uma maior ou menor taxa destes aumenta a selectividade para o 5-HT<sub>1A</sub> ou o 5-HT<sub>2A</sub>, respectivamente. É importante referir ainda que o número de benzenos com substituintes, que permitem ligações rotativas, é uma mais valia para os receptores 5-HT<sub>2A</sub>.

Por último, ambos os receptores beneficiam estruturas com os elementos de azoto e de oxigénio, que promovam a electronegatividade do composto, mas sem a possibilidade de ceder ou captar iões de hidrogénio, ou seja, sem a capacidade de exercer efectivamente uma função ácida ou básica que, na maioria dos casos estudados, era alcançada por impedimento estérico ou conformacional dos substituintes.

Na construção dos modelos farmacóforos para os receptores serotoninérgicos (5-HT<sub>1A</sub> e 5-HT<sub>2A</sub>) optou-se pela selecção de um conjunto de moléculas que apresentassem um elevado grau de afinidade para um dos receptores e, simultaneamente, um reduzido valor para o outro receptor. Deste modo, as estruturas seleccionadas apresentam as propriedades mais relevantes para o receptor em estudo, distinguindo-se facilmente as características físico-químicas inerentes a cada um dos receptores.

Seguidamente, cada um dos conjuntos anteriores é submetido a um alinhamento flexível por intermédio de um algoritmo de sobreposição presente no *software Moloc* (Gerber, 2001), possibilitando assim uma maior sobreposição das moléculas através da rotação das suas ligações simples. Deste modo, as propriedades em comuns são

---

<sup>19</sup> Cada grupo da tabela periódica constitui uma família de elementos, às quais se atribui, por vezes, uma designação própria. Neste caso, os elementos do grupo 16 denominam-se por calcogéneos, enquanto que os compostos do grupo 17 intitulam-se como halogéneos.

sobrevalorizadas na construção do respectivo modelo farmacóforo, como se pode observar na figura 24.

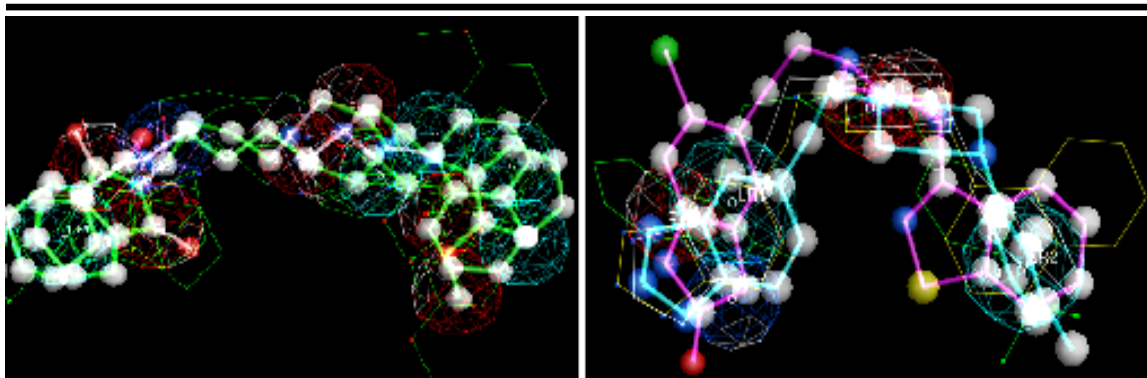


Figura 24 - Representação dos farmacóforos para os receptores 5-HT<sub>1A</sub> e 5-HT<sub>2A</sub> encontram-se nas imagens à esquerda e à direita, respectivamente, obtidas através do *software Moloc*.

**Legenda** – Nas figuras acima as esferas representam certas características da conformação tridimensional na respectiva zona espacial, mais especificamente:

Esfera azul escura – Dador de átomos de hidrogénio (ácido);

Esfera azul clara – Aromaticidade;

Esfera vermelha – Aceitador de átomos de hidrogénio (base).

Na figura 24 pode observar-se a presença de um centro básico em ambos os modelos, mais especificamente um aceitador de protões na zona do átomo N4 do grupo funcional piperazina (marcado com a esfera vermelha); enquanto que a esfera azul claro representa o efeito estérico devido ao predomínio de estruturas aromáticas que circundam o respectivo átomo N4. Note-se que o modelo 2A exibe uma “blindagem” estérica mais cerrada, mas sem que o átomo N4 perca o seu carácter básico. Além disso, a presença de grupos com densidade electrónica de carga negativa (C=O) possibilitam também uma certa basicidade, sendo essa circunstância mais evidente no modelo 1A. Por sua vez, os grupos ácidos (esfera azul escura) aparentam uma maior importância para a interacção com o receptor 2A, em comparação ao receptor 1A, devido as suas posições em relação aos centros hidrofóbicos nos compostos em estudo, facto que está de acordo com a literatura (Chidester *et al.*, 1993). No entanto, é importante ter em mente que, um farmacóforo é apenas um modelo geral das propriedades comuns de um certo número de moléculas que compartilham determinada afinidade, a qual também depende dos aspectos químicos locais que interactivam com as respectivas estruturas privilegiadas, como por exemplo, a basicidade do N4 ou do C=O e a densidade pi ( $\pi$ ) dos anéis aromáticos presentes.

No caso do receptor 2A verificou-se, de facto, uma maior complexidade nas características do farmacóforo devido à presença de um grupo C=O entre o átomo N4 e

o grupo aromático em algumas moléculas (compostos 78, 92, 128 e 129; ver na secção anexo) que tende a modificar o próprio farmacóforo, mas preservando uma alta afinidade. No entanto, a análise destas influências na interacção com o respectivo receptor necessita de um estudo por *docking* molecular, com o intuito de averiguar o contorno molecular do receptor inerente às modificações da benzilpiperazina.

Pode-se, então, concluir que os receptores 5-HT partilham propriedades químicas e topológicas na relação estrutura-actividade, mas também divergem nas propriedades inerentes ao ligando na sua afinidade para cada um dos subtipos 1A e 2A. Este trabalho computacional comprova a possibilidade de caracterizar *in silico* a afinidade para os receptores 5-HT<sub>1A</sub> e 5-HT<sub>2A</sub>, bem como, a sua previsão para outros compostos agonistas. Futuramente, esta investigação pode ser continuada, mais precisamente no estudo conformacional (por métodos de dinâmica molecular) da série em estudo para averiguar com uma maior precisão as propriedades inerentes ao farmacóforo. É de referir que ao longo deste trabalho computacional foi construído um modelo que permite elucidar e modificar a selectividade das piperazinas para um dos dois receptores em estudo. Este estudo encontra-se submetido para publicação.

## 6 – Conclusões

O trabalho computacional desenvolvido permitiu delinear as seguintes conclusões:

- 1) A conjugação da divisão de *clusters* com o GA possibilitou uma melhoria na preditividade dos modelos construídos.
- 2) A caracterização estrutural da série em estudo por descritores do *e-Dragon* alcançou melhores resultados comparativamente aos descritores de *VolSurf*.
- 3) A análise estrutural e os descritores moleculares da família das piperazinas validam a presença do centro básico na zona do átomo N4 do grupo funcional piperazina e o seu efeito estérico, devido ao predomínio de estruturas aromáticas que circundam o respectivo átomo N4, sendo esta estericidade mais significativa para o receptor 5-HT<sub>2A</sub>.
- 4) A existência de um equilíbrio estérico-electrónico nas piperazinas da série em estudo influencia a afinidade dos compostos para um dos dois receptores serotoninérgicos (5-HT<sub>1A</sub> e 5-HT<sub>2A</sub>), mais especificamente, o tipo de substituintes (elementos dos calcogéneos ou dos halogéneos) presentes nos centros hidrofóbicos (estruturas aromáticas, como por exemplo, o benzeno).

## 7 – *Perspectivas futuras*

Ao longo desta dissertação foram obtidos resultados bastante coerentes, o que sugere que o trabalho computacional deva ser continuado. No entanto, a futura investigação necessitaria de um estudo computacional mais profundo, especificamente um estudo dinâmico da ligação das piperazinas aos respectivos receptores, como por exemplo o *docking* molecular. O *docking* molecular poderia ainda analisar detalhadamente a influência da presença e da ausência de grupos básicos na vizinhança conformacional do N4 das piperazinas para ambos os receptores serotoninérgicos 1A e 2A.

Por outro lado, a validação cruzada (*cross-validation*) para a avaliação dos modelos construídos deveria ser não só interna, conforme realizado, mas também externa. Além disto, seria aconselhável experimentar outros métodos de validação estatística para otimizar os cálculos do modelo preditivo.

Por fim, é importante referir que seria também recomendável, além da validação estatística, a execução de ensaios experimentais *in vitro* para comprovar os resultados teóricos alcançados.



## *Bibliografia*

**Bibliografia**

1. Bajorath Jürgen. (2004). *Chemoinformatics Concepts, Methods, and Tools for Drug Discovery*. Humana Press Totowa, New Jersey.
2. Barnett Gene, Trsic Milan and Willette Robert E. (1978). *QuaSAR Quantitative Structure Activity Relationships of Analgesics, Narcotic Antagonists, and Hallucinogens*. NIDA Research Monograph 22.
3. Basak S. C., Gute B. D. and Ghatak S. (1999). *Prediction of complement inhibitory activity of benzamidines using topological and geometrical parameters*. J. Chem. Inf. Comput. Sci.; 39: 255-260.
4. Basak Subhash C., Gute Brian D., Grunwald Gregory D., Opitz David W. and Balasubramanian Krishnan (1999a). *Use of Statistical and Neural Net Methods in Predicting Toxicity of Chemicals: A Hierarchical QSAR Approach*. AAAI Spring Symposium on Predictive Toxicology of Chemicals: Experiences and Impact of AI Tool.
5. Baumann K. (2002). *An alignment-independent versatile structure descriptor for QSAR and QSPR based on the distribution of molecular features*. Journal of Chemical Information and Computer Sciences; 42(1): 26-35.
6. Baumann K. (2002a). *Distance Profiles (DiP): A translationally and rotationally invariant 3D structure descriptor capturing steric properties of molecules*. Quantitative Structure-Activity Relationships; 21 (5): 507-519.
7. Bitetti-Putzer R., Joseph-McCarthy D., Hogle J. M. and Karplus M. (2001). *Functional group placement in protein binding sites: a comparison of GRID and MCSS*. J Comput Aided Mol Des; 15 (10): 935-60.
8. Bonchev D., Balaban A. T. and Mekenyan O. (1980). *Generalization of the Graph Center Concept, and Derived Topological Centric Indexes*. J. Chem. InJ Comput. Sci.; 20: 106-113.
9. Bonchev Danail. (2000). *Overall Connectivities/Topological Complexities: A New Powerful Tool for QSPR/QSAR*. J. Chem. Inf. Comput. Sci.; 40: 934-941.
10. Broto P., Moreau G. and Vanduycke C. (1984). *Molecular Structures – Perception, Auto-Correlation Descriptor and SAR Studies – Auto-Correlation Descriptor*. European Journal of Medicinal Chemistry; 19 (1): 66-70.
11. Bursi R., Dao T., van Wijk T., de Gooyer M., Kellenbach E. and Verwer P. (1999). *Comparative spectra analysis (CoSA): spectra as three-dimensional molecular descriptors for the prediction of biological activities*. J Chem Inf Comput Sci; 39 (5): 861-7.
12. Capela J., Ruscher K., Lautenschlager M., Freyer D., Dirnagl U., Gaio A., Bastos M., Meisel A. Carvalho F. (2006). *Ecstasy-induced cell death in cortical*

- neuronal cultures is serotonin 2A-receptor-dependent and potentiated under hyperthermia.* Neuroscience 139: 1069-1081.
13. Carmo Helena (2007). *Estudo da Influência do Metabolismo na Toxicidade de Derivados Anfetamínicos: 4-MTA, 2C-B e MDMA.* Tese de Doutoramento na Faculdade de Farmacia da Universidade do Porto.
  14. Chidester, C. G., Lin, C., Lahti, R. A., Haadsma-Svensson, S. R., and Smith, M. W. (1993). *Comparison of 5-HT<sub>1A</sub> and Dopamine D<sub>2</sub> Pharmacophores. X-ray Structures and Affinities of Conformationally Constrained Ligands.* Journal of Medicinal Chemistry, 36, 10.
  15. Connie G. Chidester, Chiu-Hong Lin, Robert A. Lahti, Susanne R. Haadsma-Svensson, Martin W. Smith. (1993). *Comparison of 5-HT<sub>1A</sub> and Dopamine D<sub>2</sub> Pharmacophores. X-ray Structures and Affinities of Conformationally Constrained Ligands.* J. Med. Chem; 36 (10): 1301-1315.
  16. Cruciani C., Crivori P., Carrupt P. A. and Testa B. (2000a). *Molecular fields in quantitative structure-permeation relationships: the VolSurf approach.* Journal of Molecular Structure-Theochem; 503 (1-2): 17-30.
  17. Cruciani G., Pastor M. and Guba W. (2000b). *VolSurf: a new tool for the pharmacokinetic optimization of lead compounds.* Eur J Pharm Sci; 11 Suppl 2: S29-39.
  18. Cruciani G., Pastor M. and Mannhold R. (2002). *Suitability of molecular descriptors for database mining. A comparative analysis.* J Med Chem; 45 (13): 2685-94.
  19. Ducki S. D., Mackenzie G., Lawrence N. J. and Snyder J. P. (2005). *Quantitative Structure-Activity Relationship (5D-QSAR) Study of Combretastatin-like Analogues as Inhibitors of Tubulin Assembly.* J. Med. Chem; 48: 457-465.
  20. Ekins S., Bravi G., Binkley S., Gillespie J. S., Ring B. J., Wikel J. H. And Wrighton S. A. (2000). *Three- And Four-Dimensional-Quantitative Structure Activity Relationship (3d/4d-Qsar) Analyses Of Cyp2c9 Inhibitors.* Drug Metabolism and Disposition; Vol. 28, No. 8.
  21. Ekins S., Durst G. L., Stratford R. E., Thorner D. A., Lewis R., Loncharich R. J. (2001). *Three-dimensional quantitative structure-permeability relationship analysis for a series of inhibitors of rhinovirus replication.* J Chem Inf Comput Sci; 41 (6): 1578-86.
  22. Esbensen Kim H. (2002). *Multivariate data Analysis in Practice. 5<sup>th</sup> Edition.* CAMO Process AS.
  23. Estrada Emesto (1995). *Graph Theoretical Invariant of Randić Revisited.* J. Chem. Inf. Comput. Sci; 35: 1022-1025.

24. Fechner U., Frankle L., Renner S., Schneider P. and Schneider G. (2003). *Comparison of correlation vector methods for ligand-based similarity searching*. J Comput Aided Mol Des; 17 (10): 687-98.
25. Ferguson A. M., Heritage T., Jonathon P., Pack S. E., Philips L., Rogan J. (1997). *EVA: a new theoretically based molecular descriptor for use in QSAR/QSPR analysis*. J Comput Aided Mol Des; 11 (2): 143-52.
26. Fichera M., Cruciani G., Bianchi A., Musumarra G. (2000). *A 3D-QSAR Study on the Structural Requirements for Binding to CB1 and CB2 Cannabinoid Receptors*. J. Med. Chem; 43 (12): 2300-2309.
27. Gaillard P, Carrupt P.-A, Testa B, Schambel P. (1996). *Binding of Arylpiperazines, (Aryloxy)propanolamines, and Tetrahydropyridylindoles to the 5-HT1A Receptor: Contribution of the Molecular Lipophilicity Potential to Three-Dimensional Quantitative Structure-Affinity Relationship Models*. J. Med. Chem; 39: 126-134.
28. Garthwaite P.H. (1994). *An Interpretation of Partial Least Square*. Journal of the American Statistical Association; Vol. 89, No.425: pp. 122-127.
29. Gaster, L. M., Blaney, F. E., Davies, S., Duckworth, D. M., Ham, P., Jenkins, S., Jennings, A. J., Joiner, G. F., King, F. D., Mulholland, K. R., Wyman, P. A., Hagan, J. J., Hatcher, J., Jones, B. J., Middlemiss, D. N., Price, G. W., Riley, G., Roberts, C., Routledge, C., Selkirk, J. and Slade, P. D. (1998) *The selective 5-HT1B receptor inverse agonist 1'-methyl-5-[[2'-methyl-4'-(5-methyl-1,2,4-oxadiazol-3-yl)biphenyl-4-yl]carbonyl]-2,3,6,7-tetrahydro-spiro[furo[2,3-f]indole-3,4'-piperidine] (SB224289) potently blocks terminal 5-HT autoreceptor function both in vitro and in vivo*. J Med Chem, 41, 1218 - 1235.
30. Gee P, Richardson S, Woltersdorf W, Moore G (2005). *Toxic effects of BZP-based herbal party pills in humans: a prospective study in Christchurch, New Zealand*. N Z Med J; 118 (1227): U1784.
31. Gerber P R (2001). *Topological Pharmacophore Description of Chemical Structures using MAB-Force-Field-Derived Data and Corresponding Similarity Measures*. In *Fundamental of Molecular Similarity*, eds. Carbó-Dorca R, Giromés X. & Merzey P., Kluwer Academic/Plenum Publishers, New York, 2001; pp. 67-82.
32. Gustafsson M. G. (2001). *A Probabilistic Derivation of the Partial Least-Squares Algorithm*. J. Chem. Inf. Comput. Sci; 41 (2): 288-294.
33. Gute B. D. and Basak S. C. (1997). *Predicting acute toxicity of benzene derivatives using theoretical molecular descriptors: a hierarchical QSAR approach*. SAR QSAR Environ. Res; 7: 117-131.
34. Gutman Ivan, Vidović Dušica and Popović Ljiljana. (1998). *Graph representation of organic molecules Cayley's plerograms vs. his kenograms*. J. Chem. Soc., Faraday T rans; 94 (7): 857-860.

35. Hagadone, T. R. (1992). *Molecular substructure similarity searching: Efficient retrieval in two-dimensional structure databases*. J. Chem. Inf. Comput. Sci; 32: 515–521.
36. Hall A, Henry J (2006). *Acute toxic effects of 'Ecstasy' (MDMA) and related compounds: overview of pathophysiology and clinical management*. Br J Anaesth; 96 (6): 678-85.
37. Hammett L. P. *Physical Organic Chemistry*, 2nd ed., McGraw-Hill, New York, 1970.
38. Hibert M. F, McDermott I, Middlemiss D. N, Mir A. K, Fozard J. R. (1989). *Eur. J. Med. Chem*; 24: 31.
39. Hoyer Daniel, Hannon Jason P., Martin Graeme R. (2002). *Molecular, pharmacological and functional diversity of 5-HT receptors*. Pharmacology, Biochemistry and Behavior; 71: 533–554.
40. <http://awstats.sourceforge.net>
41. <http://openbabel.sourceforge.net/>
42. <http://www.cambridgesoft.com/software/ChemOffice/>
43. [http:// www.drugabuse.gov](http://www.drugabuse.gov)
44. <http://www.mathworks.com/>
45. [http://www.moldiscovery.com/soft\\_volsurf.php](http://www.moldiscovery.com/soft_volsurf.php)
46. <http://www.moloc.ch/>
47. Hurst, T. and Heritage, T. (1997) *HQSAR—A highly predictive QSAR technique based on molecular holograms*. 213th ACS Natl. Meeting, San Francisco, CA, CINF 019.
48. Jensen F. (1999). *Introduction to Computational Chemistry*. John Wiley & Sons, New York.
49. Jorand-Lebrun, C., Pauwels, P. J., Palmier, C., Moret, C., Chopin, P., Perez, M., Marien, M. and Halazy, S. (1997) *5-HT<sub>1B</sub> receptor antagonist properties of novel arylpiperazide derivatives of 1-naphthylpiperazine*. J Med Chem, 40, 3974 - 3978.
50. Jürgen Drews. (2000). *Drug Discovery: A Historical Perspective*. 1960; Science: 287.
51. Kamphausen Stefan, Höltge Nils, Wirsching Frank, Morys-Wortmann Corinna, Riester Daniel, Goetz Ruediger, Thürk Marcel and Schwienhorst Andreas.

- (2002). *Genetic algorithm for the design of molecules with desired properties*. Journal of Computer-Aided Molecular Design; 16: 551–567.
52. Kier L. B., Hall L. H., Murray W. J. and Randic M. (1975). *Molecular connectivity. I: Relationship to nonspecific local anesthesia*. J Pharm Sci; 64 (12): 1971-4.
53. Kier L. B., Murray W. J. and Hall L. H. (1975a). *Molecular connectivity. 4. Relationships to biological activities*. J Med Chem; 18 (12): 1272-4.
54. Klein C. T., Kaiser D. and Ecker G. (2004). *Topological distance based 3D descriptors for use in QSAR and diversity analysis*. J Chem Inf Comput Sci; 44 (1): 200-9.
55. Knight, A. R., Misra, A., Quirk, K., Benwell, K., Revell, D., Kennett, G. and Bickerdike, M. (2004) *Pharmacological characterisation of the agonist radioligand binding site of 5-HT(2A), 5-HT(2B) and 5-HT(2C) receptors*. Naunyn Schmiedebergs Arch Pharmacol, 370, 114-123.
56. Kovalishyn V. V., Kholodovych, V., Tetko, I. V. and Welsh, W. J. (2007). *Volume learning algorithm significantly improved PLS model for predicting the estrogenic activity of xenoestrogens*. J. Mol. Graph. Model.
57. Kubinyi H., Folkers G. and Martin Y. C. (2002). *3D QSAR in Drug Design Volume 2, Ligand-Protein Interactions and Molecular Similarity*. Kluwer Academic Publishers.
58. Kuipers, W, Van Wijngaarden I, Kruse C. G, Ter Horst-Van Amstel M, Tulp, M. Th. M, IJzerman A.P. (1995). *N4-unsubstituted N1-Arylpiperazines as High-Affinity 5-HT1A Receptor Ligands*. J. Med.Chem. 38: 1942-1954.
59. Langer Thierry, Hoffmann Rémy D. (2006). *Pharmacophores and Pharmacophore Searches, Volume 32. Methods and Principles in Medicinal Chemistry* WILEY-VCH.
60. Lawrence Neil. (2005). *Probabilistic Non-linear Principal Component Analysis with Gaussian Process Latent Variable Models*. Journal of Machine Learning Research; 6: 1783–1816.
61. Learch A. R. (2001). *Molecular Modelling. Principles and applications. Second ed.* Harlow: Pearson Education Limited.
62. Lemmen C. and Lengauer T. (2000). *Computational methods for the structural alignment of molecules*. Journal of Computer-Aided Molecular Design; 14: 215–232.
63. Livingstone David J. (2000). *The Characterization of Chemical Structures Using Molecular Properties. A Survey*. J. Chem. Inf. Comput. Sci; 40: 195-209.

64. López-Rodríguez M. L., Ayala D., Benhamú B., Morcillo M. J. and Viso A. (2002). *Arylpiperazine derivatives acting at 5-HT<sub>1A</sub> receptors*. *Curr. Med. Chem*; 9: 443-469.
65. Michael H Baumann, Robert D Clark, Allison G Budzynski, John S Partilla, Bruce E Blough and Richard B Rothman (2005). *N-Substituted Piperazines Abused by Humans. Mimic the Molecular Mechanism of 3,4-Methylenedioxymethamphetamine (MDMA, or 'Ecstasy')*. *Neuropsychopharmacology* ; 30: 550-560.
66. Michael R. Braden, Jason C. Parrish, John C. Naylor, and David E. Nichols. (2006). *Molecular Interaction of Serotonin 5-HT<sub>2A</sub> Receptor Residues Phe339 (6.51) and Phe340 (6.52) with Superpotent N-Benzyl Phenethylamine Agonists*. *Mol. Pharmacol*; 70: 6.
67. Miller Michael D. (1998). *Molecular Superposition*. Encyclopedia of Computational Chemistry, John Wiley & Sons, Ltd.
68. Miranker A. and Karplus M. (1991). *Functionality maps of binding sites: a multiple copy simultaneous search method*. *Proteins*; 11 (1): 29-34.
69. Moon T., Chi M. H., Kim D. H., Yoon C. N. and Choi Y. S. (2000). *Quantitative structure-activity relationships (QSAR) study of flavonoid derivatives for inhibition of cytochrome P450 1A2*. *Quant. Struct.-Act. Relat*; 19: 257-263.
70. Newman-Tancredi, A., Gavaudan, S., Conte, C., Chaput, C., Touzard, M., Verrielle, L., Audinot, V. and Millan, M. J. (1998) *Agonist and antagonist actions of antipsychotic agents at 5-HT<sub>1A</sub> receptors: a [35S]GTPgammaS binding study*. *Eur J Pharmacol*, 355, 245 - 256.
71. Newman-Tancredi, A., Verrielle, L., Chaput, C. and Millan, M. J. (1998a) *Labelling of recombinant human and native rat serotonin 5-HT<sub>1A</sub> receptors by a novel, selective radioligand, [3H]-S 15535: definition of its binding profile using agonists, antagonists and inverse agonists*. *Naunyn Schmiedebergs Arch Pharmacol*, 357, 205 - 217.
72. Nikolova Nina and Jaworska Joanna. (2003). *Approaches to Measure Chemical Similarity ± a Review*. *QSAR Comb. Sci*. 22.
73. Novič M. and Roncaglioni A. (2004). *Application of Artificial Neural Networks to the QSPR Study – Automated Classification of Endocrine Disrupting Chemicals*. *Kem. Ind*; 53 (7-8): 323-331.
74. Philip E. Bourne, Helge Weissing. (2003). *Structural Bioinformatics*. John Wiley & Sons, Chichester, UK.
75. Poda, G. I.; Tetko, I. V.; Rohrer, D. C. (2005). *Towards predictive ADME profiling of drug candidates: Lipophilicity and solubility*, In 229th American

- Chemical Society National Meeting & Exposition; ACS; San Diego, CA: MEDI 514.
76. Popp Herbert, Männer Reinhard, Smith Jeremy. (2002). *Protein Structure Prediction: Improving and Automating Knowledge-based Approaches*. Tag der mündlichen Prüfung.
  77. Price, G. W., Burton, M. J., Collin, L. J., Duckworth, M., Gaster, L., Gothert, M., Jones, B. J., Roberts, C., Watson, J. M. and Middlemiss, D. N. (1999) *SB-216641 and BRL-15572 compounds to pharmacologically discriminate h5-HT1B and h5-HT1D receptors*. Naunyn Schmiedebergs Arch. Pharmacol., 356, 312 - 320.
  78. Ragno Rino, Simeoni Silvia, Valente Sergio, Massa Silvio and Mai Antonello. (2006). *3-D QSAR Studies on Histone Deacetylase Inhibitors. A GOLPE/GRID Approach on Different Series of Compounds*. J. Chem. Inf. Model; 46: 1420-1430.
  79. Randic M. and Basak S. C. (2001). *On use of the variable connectivity index  $1\chi$  in QSAR: Toxicity of aliphatic ethers*. J. Chem. Inf. Comput. Sci; 41: 614-618.
  80. Randic M., Basak N. and Plavsic D. (2004). *Novel Graphical Matrix and Distance-based Molecular Descriptors*. Croatica Chemica Acta CCACAA; 77 (1-2): 251-257.
  81. Randic Milan. (2000). *On Characterization of Pharmacophore*. Acta Chim. Slov; 47: 143-151.
  82. Rapaport D.C. (2004). *The Art of Molecular Dynamics Simulation, 2<sup>nd</sup> Edition*, Cambridge University Press.
  83. Robinson D. D, Winn P. J, Lyne P. D and Richards W. G. (1999). *Self-Organizing Molecular Field Analysis: A Tool for Structure-Activity Studies*. Journal of Medicinal Chemistry; 42 (4): 573-583.
  84. Rowley Michael, Bristow Linda J., Hutson Peter H. (2001). *Current and Novel Approaches to the Drug Treatment of Schizophrenia*. J. Med. Chem; 44: 4.
  85. Sanchez, C; J. Hyttel (1999). "Comparison of the Effects of Antidepressants and Their Metabolites on Reuptake of Biogenic Amines and on Receptor Binding". Cellular and Molecular Neurobiology 19 (4): 467-89.
  86. Santos-Filho Osvaldo A., Mishra Rama K. & Hopfinger A. J. (2001). *Free energy force field (FEFF) 3D-QSAR analysis of a set of Plasmodium falciparum dihydrofolate reductase inhibitors*. Journal of Computer-Aided Molecular Design; 15: 787-810.
  87. Schuur J.H., Selzer P. and Gasteiger J. (1996). *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; 36 (2): 334-344.
88. Selassie C.D. (2003). *The History of Quantitative Structure Activity Relationships.* In, *Burger's Medicinal Chemistry and Drug Discovery, Volume 1, 6<sup>th</sup> Edition*, D.J. Abraham (ed.), John Wiley and Sons Publishers, New York, New York.
  89. Selassie C.D., Garg R. and Mekapati S.B. (2003). *A Mechanism based QSAR Approach to the Study of the Toxicity of Endocrine Active Substances.* Pure and Applied Chemistry; 75: 2363-2373.
  90. Selassie C.D., Mekapati S.B., Verma R.P. (2002). *QSAR: Then and Now.* Current Topics in Medicinal Chemistry; 2: 1357-1379.
  91. Shapiro, D. A., Renock, S., Arrington, E., Chiodo, L. A., Liu, L. X., Sibley, D. R., Roth, B. L. and Mailman, R. (2003) *Aripiprazole, a novel atypical antipsychotic drug with a unique and robust pharmacology.* Neuropsychopharmacology, 28, 1400 - 1411.
  92. Silverman B. D., Platt D. E., Pitman M. and Rigoutsos I. (1998). *Comparative molecular moment analysis (CoMMA).* Perspectives in Drug Discovery and Design; 12: 183-196.
  93. Stewart J. J. P. (1993). *MOPAC 93.00 Manual.* Fujitsu Limited, Tokyo, Japan.
  94. Tammy L, Amanda L. B, Andrew M, Caroline Q, Smita P, Kerry Ch, Angus M. M. (2006). *A new class of selective, non-basic 5-HT<sub>2A</sub> receptor antagonists.* Bioorg. Med. Chem. Lett; 16: 3201-3204.
  95. Terfloth Lothar and Gasteiger Johann. (2001). *Neural networks and genetic algorithms in drug design.* Drug Discovery Technology (DDT); Vol. 6: No. 15.
  96. Tetko I. V. (2002). *Associative neural network.* Neural Processing Letters; 16: 187-199.
  97. Tetko I. V. (2005). *Computing chemistry on the web.* Drug Discov. Today; 10: 1497-500.
  98. Tetko I. V. and Tanchuk V. Y. (2002). *Application of associative neural networks for prediction of lipophilicity in ALOGPS 2.1 program.* J. Chem. Inf. Comput. Sci; 42: 1136-45.
  99. Tetko I. V., Gasteiger J., Todeschini R., Mauri A., Livingstone D., Ertl P., Palyulin V. A., Radchenko E. V., Zefirov N. S., Makarenko A. S., Tanchuk V. Y., Prokopenko V. V. (2005). *Virtual computational chemistry laboratory - design and description.* J. Comput. Aid. Mol. Des; 19: 453-63.

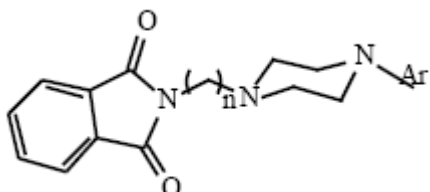
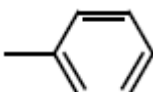
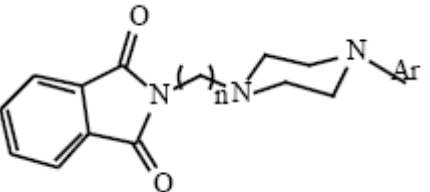
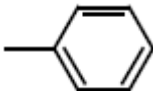
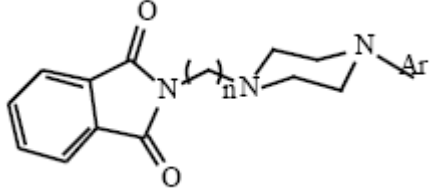
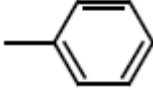
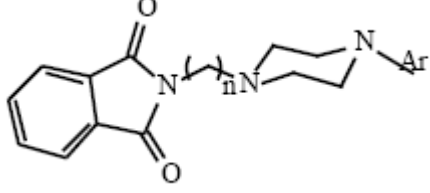
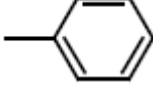
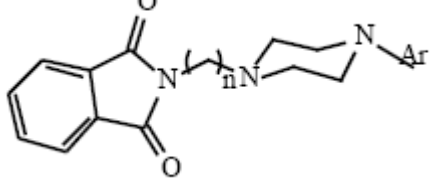
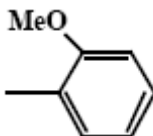
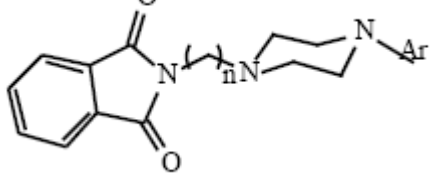
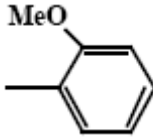
100. Tetko I. V., Kovalishyn V. V., Livingstone D. J. (2001). *Volume learning algorithm artificial neural networks for 3D QSAR studies*. J. Med. Chem; 44: 2411-20.
101. Tetko I. V., Luik A. I., Poda G. I. (1993). *Applications of neural networks in structure-activity relationships of a small number of molecules*. J. Med. Chem; 36: 811-4.
102. Tetko I. V., Tanchuk V. Yu., Villa A. E. P. (2001). *Prediction of n-Octanol/Water Partition Coefficients from PHYSPROP Database Using Artificial Neural Networks and E-State Indices*. J. Chem. Inf. Comput. Sci; 41 (5): 1407-1421.
103. Thilagavathi Ramasamy and Chakraborti Asit K. (2005). *Importance of Alignment in Developing 3-D QSAR Models of 1,5-Diaryl Pyrazoles for the Prediction of COX-2 Inhibitory Activity*. Internet Electronic Journal of Molecular Design; 4: 603-612.
104. Thorner D. A., Willett P., Wright P. M. and Taylor R. (1997). *Similarity searching in files of three-dimensional chemical structures: Representation and searching of molecular electrostatic potentials using field-graphs*. J. Comput.-Aided Mol. Design; 11: 163-174.
105. Tipping M. E. and Bishop Ch. M. (1999). *Probabilistic Principal Component Analysis*. Journal of the Royal Statistical Society; Series B. 61 (Part 3): pp. 611-622.
106. Todeschini R. and Lasagni M. (1994). *New Molecular Descriptors for 2d and 3d Structures- Theory*. Journal of Chemometrics; 8 (4): 263-272.
107. Trinajstić Nenad and Gutman Ivan. (2002). *Mathematical Chemistry*. Croatica Chemica Acta CCACAA; 75 (2): 329-356.
108. Tudor I. Opera. (2005). *Chemoinformatics in Drugs Discovery*. Weinheim; Wiley: Vol. XXIII.
109. Tuppurainen K., Viisas M., Laakitainen R. and Perakyla M. (2002). *Evaluation of a novel electronic eigenvalue (EEVA) molecular descriptor for QSAR/QSPR studies: validation using a benchmark steroid data set*. J Chem Inf Comput Sci; 42 (3): 607-13.
110. Vilca J. C. Vega. (2004). *Generalizaciones de mínimos cuadrados parciales con aplicación en clasificación supervisada*. Tesis del Universidad de Puerto Rico.
111. Vukičević Damir and Trinajstić Nenad. (2003). *Modified Zagreb M2 Index – Comparison with the Randić Connectivity Index for Benzenoid Systems*. Croatica Chemica Acta CCACAA; 76 (2): 183-187.
112. Wagener M., Sadowski J. and Gasteiger J. (1995). *Autocorrelation of Molecular-Surface Properties for Modeling Corticosteroid-Binding Globulin*

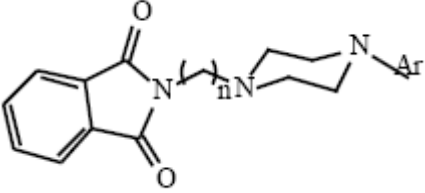
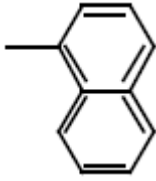
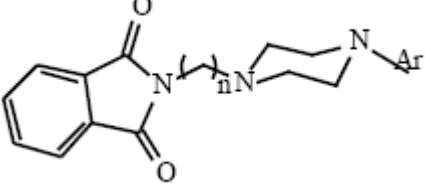
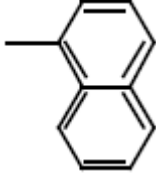
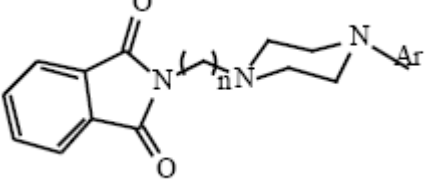
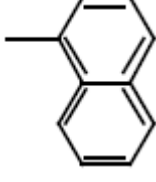
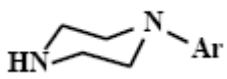
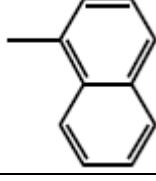

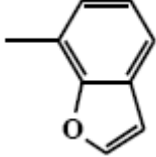
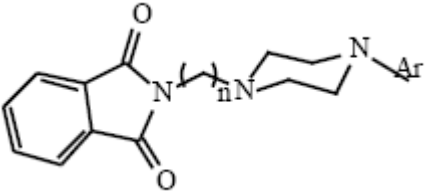
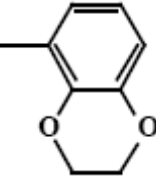
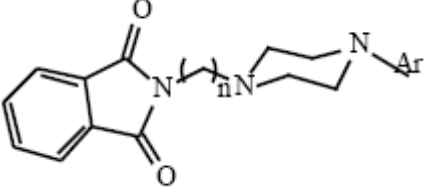
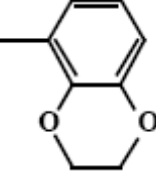
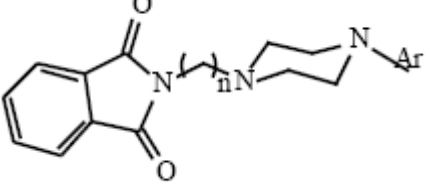
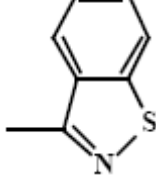
- and Cytosolic Ah Receptor Activity by Neural Networks*. Journal of the American Chemical Society; 117 (29): 7769-7775.
113. Wang C. D, Gallaher T. K, Shih J. C. (1993). *Site-directed mutagenesis of the serotonin 5-hydroxytryptamine<sub>2</sub> receptor: Identification of amino acids necessary for ligand binding and receptor activation*. Mol. Pharmacol; 43: 931-940.
114. Young David C. (2001). *Computational Chemistry: A Practical Guide for Applying Techniques to Real-World Problems*. John Wiley & Sons, Inc.
115. Zamora I., Oprea T., Cruciani G., Pastor M. and Ungell A. L. (2003). *Surface descriptors for protein-ligand affinity prediction*. J Med Chem; 46 (1): 25-33.

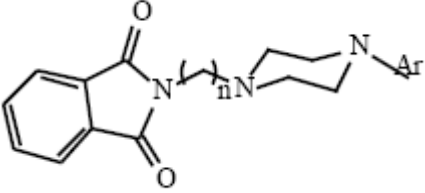
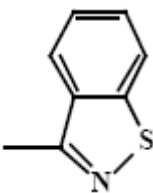
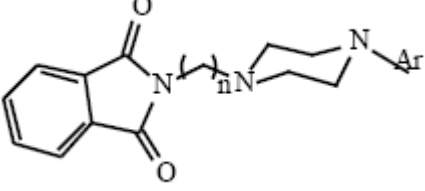
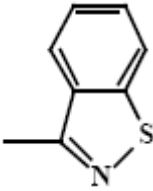
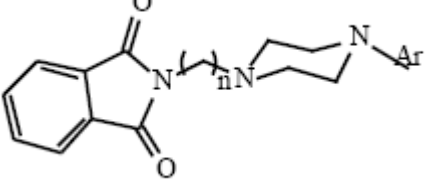
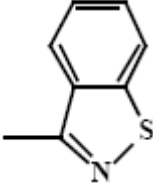
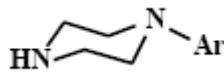
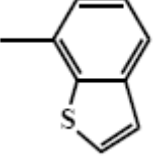
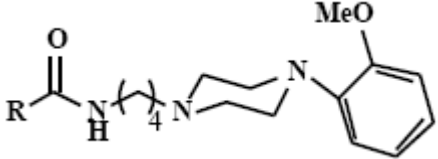
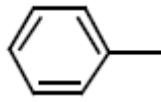
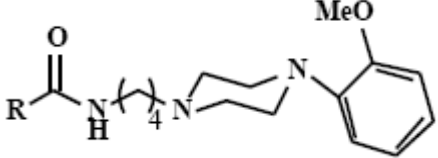
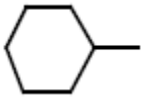
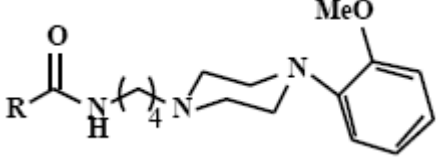

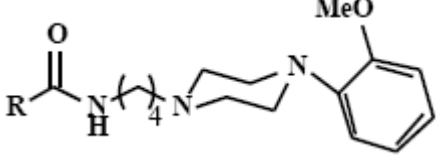

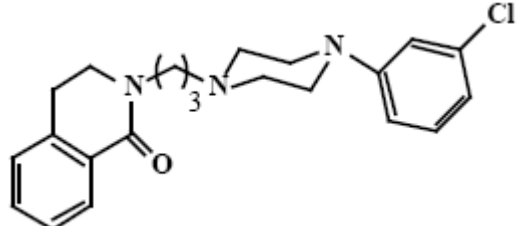
*Anexos*

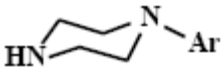
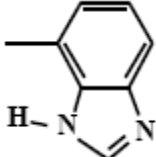
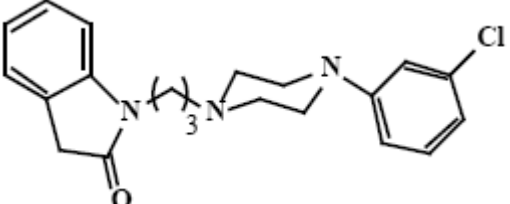
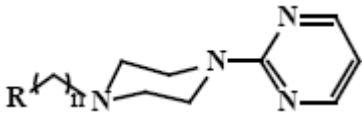
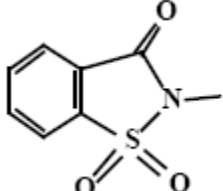

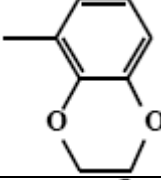
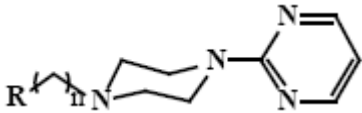
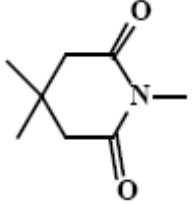
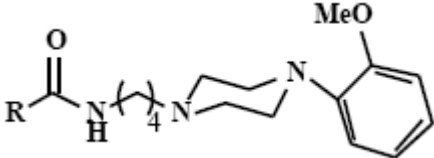

## Anexos

Tabela I - Recolha dos dados experimentais necessários da série de piperazinas (compostos 1 a 29) para a elaboração do estudo da afinidade para o receptor 5-HT<sub>1A</sub> e 5-HT<sub>2A</sub>.

| <i>Composto</i> | <i>Estrutura</i>  | <i>- R</i>   | <i>N</i> | <i>pK<sub>i</sub></i><br><i>5-HT<sub>1A</sub></i> | <i>pK<sub>i</sub></i><br><i>5-HT<sub>2A</sub></i> | <i>Ref.</i> |
|-----------------|---|--|----------|---|---|-------------|
| 1               |    |    | 2        | 5   | -   | a)          |
| 2               |    |    | 3        | 6.69897   | -   | a)          |
| 3               |   |  | 4        | 8   | -   | a)          |
| 4               |  |  | 5        | 8.070581  | -   | a)          |
| 5               |  |  | 2        | 6   | -   | a)          |
| 6               |  |  | 3        | 7.886057  | -   | a)          |

|    |   |  |   |          |        |    |
|----|---|--|---|----------|--------|----|
| 7  |    |    | 2 | 6.619789 | -      | a) |
| 8  |    |    | 3 | 7.187087 | -      | a) |
| 9  |    |    | 4 | 9        | -      | a) |
| 10 |   |   | - | 7.958607 | 7.4085 | a) |
| 11 |  |  | - | 7.886057 | -      | a) |
| 12 |  |  | 2 | 7.49485  | -      | a) |
| 13 |  |  | 4 | 9.638272 | -      | a) |
| 14 |  |  | 2 | 7.267606 | -      | a) |

|    |   |  |   |          |   |    |
|----|---|--|---|----------|---|----|
| 15 |    |    | 3 | 7.886057 | - | a) |
| 16 |    |    | 4 | 9.721246 | - | a) |
| 17 |    |    | 5 | 8.69897  | - | a) |
| 18 |   |   | - | 8.004365 | - | a) |
| 19 |  |  | - | 8.69897  | - | a) |
| 20 |  |  | - | 9.221849 | - | a) |
| 21 |  |  | - | 7.259637 | - | a) |
| 22 |  |  | - | 9.638272 | - | a) |
| 23 |  | -  | - | 7.744727 | - | a) |

|    |   |  |   |          |        |    |
|----|---|--|---|----------|--------|----|
| 24 |    |    | - | 7.568636 | -      | a) |
| 25 |    | -  | - | 7.537602 | -      | a) |
| 26 |    |    | 3 | 8.69897  | -      | a) |
| 27 |    |    | - | 7.39794  | -      | a) |
| 28 |  |   | 4 | 7.154902 | 5.4401 | a) |
| 29 |  |  | - | 10       | -      | a) |

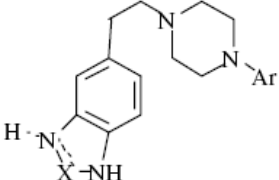
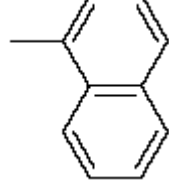
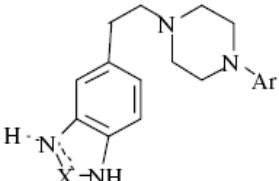
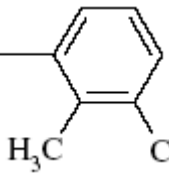
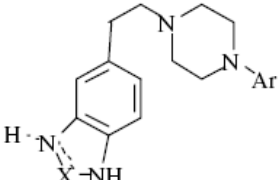
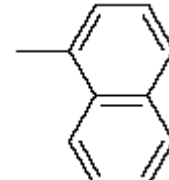
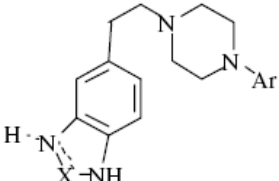
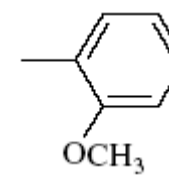
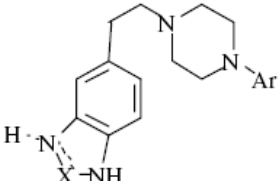
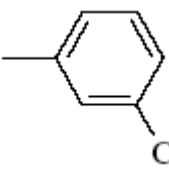
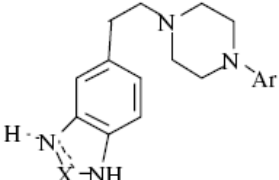
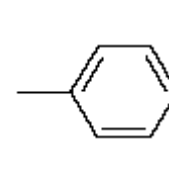
**Legenda:**

a) López-Rodríguez *et al.*, 2002;  
 b) Tomić *et al.*, 2004;  
 c) Bojarski *et al.*, 2002;  
 d) Huang *et al.*, 2005 ;  
 e) Jurczyk *et al.*, 2004 ;  
 f) Grundt *et al.*, 2007 ;  
 g) González-Gómez *et al.*, 2003;  
 h) Sanchez *et al.*, 1999 ;  
 i) Price *et al.*, 1999;

j) Gaster *et al.*, 1998;  
 k) Knight *et al.*, 2004;  
 l) Newman-Tancredi *et al.*, 1998;  
 m) Newman-Tancredi *et al.*, 1998a;  
 n) Jorand-Lebrun *et al.*, 1997;  
 o) Shapiro *et al.*, 2003;  
 p) [www.bindingdb.org](http://www.bindingdb.org);  
 q) [www.iuphar-db.org](http://www.iuphar-db.org);  
 r) Protein ki Database (2008)  
<http://kidb.cwru.edu/pdsp.php>;



Tabela II - Recolha dos dados experimentais necessários da série de piperazinas (compostos 30 a 35) para a elaboração do estudo da afinidade para o receptor 5-HT<sub>1A</sub> e 5-HT<sub>2A</sub>.

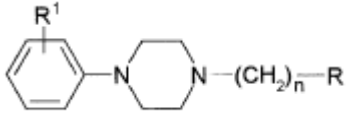
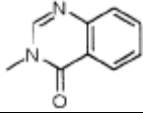
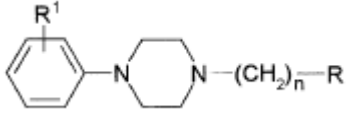
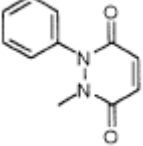
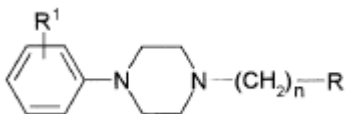
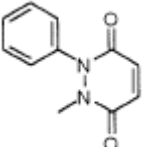
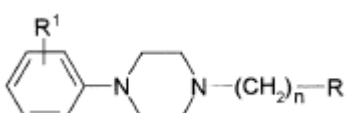
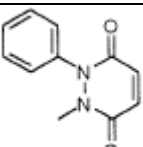
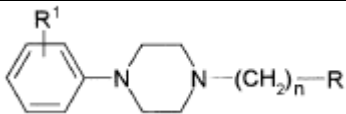
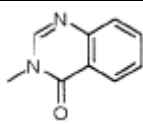
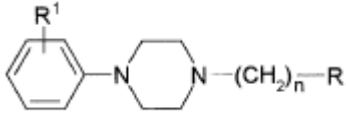
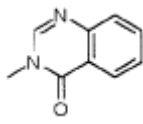
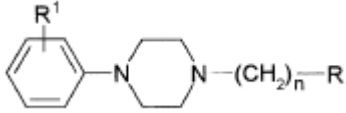
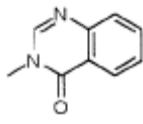
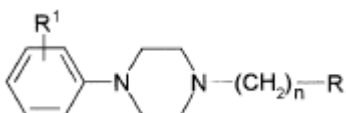
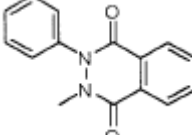
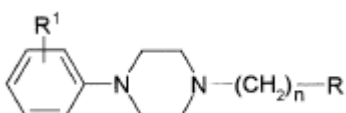
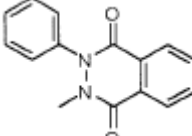
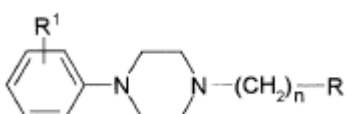
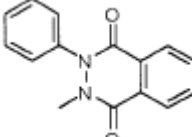
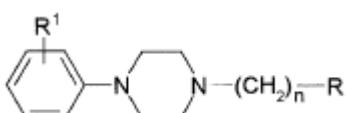
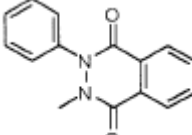
| <i>Composto</i> | <i>Estrutura</i>  | <i>X</i>                        | <i>- R</i>   | <i>pK<sub>i</sub></i><br><i>5-HT<sub>1A</sub></i> | <i>pK<sub>i</sub></i><br><i>5HT-2A</i> | <i>Referência</i> |
|-----------------|---|---------------------------------|--|---|--|-------------------|
| 30              |    | CH                              |    | 6.9   | 7.82                                   | b)                |
| 31              |    | N                               |    | 7.08  | 8.29                                   | b)                |
| 32              |    | N                               |    | 7.41  | 7.98                                   | b)                |
| 33              |   | N                               |   | 7.1   | 6.85                                   | b)                |
| 34              |  | CO-CO                           |  | 7.06  | 5.81                                   | b)                |
| 35              |  | C-C <sub>6</sub> H <sub>5</sub> |  | 6.76  | 6.12                                   | b)                |

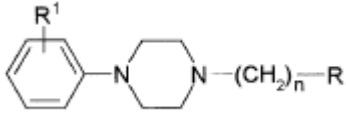
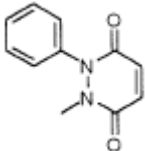
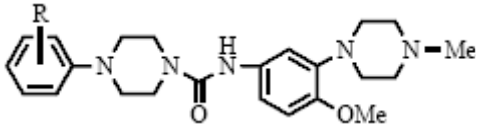
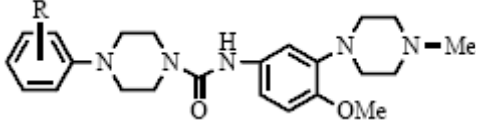
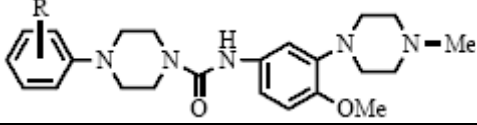
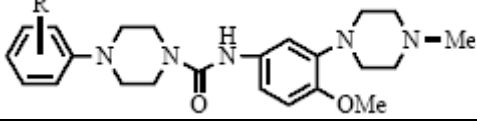
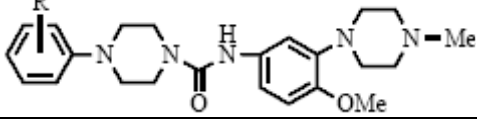
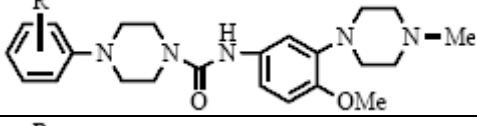
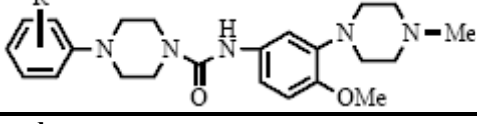
**Legenda:**

a) López-Rodríguez *et al.*, 2002;  
 b) Tomić *et al.*, 2004;  
 c) Bojarski *et al.*, 2002;  
 d) Huang *et al.*, 2005 ;  
 e) Jurczyk *et al.*, 2004 ;  
 f) Grundt *et al.*, 2007 ;  
 g) González-Gómez *et al.*, 2003;  
 h) Sanchez *et al.*, 1999 ;  
 i) Price *et al.*, 1999;

j) Gaster *et al.*, 1998;  
 k) Knight *et al.*, 2004;  
 l) Newman-Tancredi *et al.*, 1998;  
 m) Newman-Tancredi *et al.*, 1998a;  
 n) Jorand-Lebrun *et al.*, 1997;  
 o) Shapiro *et al.*, 2003;  
 p) www.bindingdb.org;  
 q) www.iuphar-db.org;  
 r) Protein ki Database (2008)  
<http://kidb.cwru.edu/pdsp.php>

Tabela III - Recolha dos dados experimentais necessários da série de piperazinas (compostos 36 a 54) para a elaboração do estudo da afinidade para o receptor 5-HT<sub>1A</sub> e 5-HT<sub>2A</sub>.

| <i>Composto</i> | <i>Estrutura</i>  | <i>n</i> | <i>-R1</i>                 | <i>-R</i>  | <i>pK<sub>i</sub></i><br><i>5-HT<sub>1A</sub></i> | <i>pK<sub>i</sub></i><br><i>5HT-2A</i> | <i>Ref.</i> |
|-----------------|---|----------|----------------------------|--|---|--|-------------|
| 36              |    | 3        | <i>m</i> -Cl               |    | 6.628932  | 7.80                                   | c)          |
| 37              |    | 4        | <i>m</i> -Cl               |    | 7.30103   | 5.74                                   | c)          |
| 38              |    | 3        | <i>o</i> -OCH <sub>3</sub> |    | 7.267606  | 5.67                                   | c)          |
| 39              |    | 4        | <i>o</i> -OCH <sub>3</sub> |    | 7.958607  | 5.84                                   | c)          |
| 40              |   | 4        | <i>m</i> -Cl               |   | 7.30103   | 7.17                                   | c)          |
| 41              |  | 3        | <i>o</i> -OCH <sub>3</sub> |  | 7   | 6.34                                   | c)          |
| 42              |  | 4        | <i>o</i> -OCH <sub>3</sub> |  | 7.443697  | 6.25                                   | c)          |
| 43              |  | 3        | <i>m</i> -Cl               |  | 6.381952  | 6.18                                   | c)          |
| 44              |  | 4        | <i>m</i> -Cl               |  | 6.39794   | 6.24                                   | c)          |
| 45              |  | 3        | <i>o</i> -OCH <sub>3</sub> |  | 7.522879  | 6.52                                   | c)          |
| 46              |  | 4        | <i>o</i> -OCH <sub>3</sub> |  | 7.366532  | 6.43                                   | c)          |

|    |   |   |              |  |          |      |    |
|----|---|---|--------------|--|----------|------|----|
| 47 |    | 3 | <i>m</i> -Cl |  | 6.686133 | 6.57 | c) |
| 48 |    | - | -            | 2-Cl   | 6.474955 | 5.22 | d) |
| 49 |    | - | -            | 2-F  | 6.501689 | 5.00 | d) |
| 50 |    | - | -            | 4-Cl   | 7.279841 | 6.00 | d) |
| 51 |    | - | -            | 4-F  | 5.966979 | 5.35 | d) |
| 52 |   | - | -            | 4-Ph   | 6.392545 | 6.77 | d) |
| 53 |  | - | -            | 4-CF <sub>3</sub>  | 6.417937 | 6.21 | d) |
| 54 |  | - | -            | 4-SMe  | 6.054531 | 5.87 | d) |

**Legenda:**

a) López-Rodríguez *et al.*, 2002;  
 b) Tomić *et al.*, 2004;  
 c) Bojarski *et al.*, 2002;  
 d) Huang *et al.*, 2005 ;  
 e) Jurczyk *et al.*, 2004 ;  
 f) Grundt *et al.*, 2007 ;  
 g) González-Gómez *et al.*, 2003;  
 h) Sanchez *et al.*, 1999 ;  
 i) Price *et al.*, 1999;

j) Gaster *et al.*, 1998;  
 k) Knight *et al.*, 2004;  
 l) Newman-Tancredi *et al.*, 1998;  
 m) Newman-Tancredi *et al.*, 1998a;  
 n) Jorand-Lebrun *et al.*, 1997;  
 o) Shapiro *et al.*, 2003;  
 p) www.bindingdb.org;  
 q) www.iuphar-db.org;  
 r) Protein ki Database (2008)  
<http://kidb.cwru.edu/pdsp.php>;

Tabela IV - Recolha dos dados experimentais necessários da série de piperazinas (compostos 55 a 59) para a elaboração do estudo da afinidade para o receptor 5-HT<sub>1A</sub> e 5-HT<sub>2A</sub>.

| Composto | Estrutura | n | X  | -R                  | $pK_i$             | $pK_i$             | Ref. |
|----------|-----------|---|----|---------------------|--------------------|--------------------|------|
|          |           |   |    |                     | 5-HT <sub>1A</sub> | 5-HT <sub>2A</sub> |      |
| 55       |           | 2 | H  | 3'-Cl               | 7.638272           | 6.63               | e)   |
| 56       |           | 2 | Br | 2'-OCH <sub>3</sub> | 7.207608           | 7.07               | e)   |
| 57       |           | 3 | Br | H                   | 7.958607           | 6.53               | e)   |
| 58       |           | 3 | Br | 3'-Cl               | 7.309804           | 7.07               | e)   |
| 59       |           | 3 | Br | 2'-OCH <sub>3</sub> | 7.443697           | 6.06               | e)   |

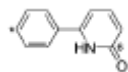
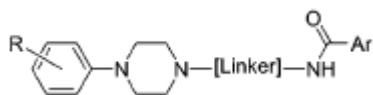
Legenda:

a) López-Rodríguez *et al.*, 2002;  
 b) Tomić *et al.*, 2004;  
 c) Bojarski *et al.*, 2002;  
 d) Huang *et al.*, 2005 ;  
 e) Jurczyk *et al.*, 2004 ;  
 f) Grundt *et al.*, 2007 ;  
 g) González-Gómez *et al.*, 2003;  
 h) Sanchez *et al.*, 1999 ;  
 i) Price *et al.*, 1999;

j) Gaster *et al.*, 1998;  
 k) Knight *et al.*, 2004;  
 l) Newman-Tancredi *et al.*, 1998;  
 m) Newman-Tancredi *et al.*, 1998a;  
 n) Jorand-Lebrun *et al.*, 1997;  
 o) Shapiro *et al.*, 2003;  
 p) www.bindingdb.org;  
 q) www.iuphar-db.org;  
 r) Protein ki Database (2008)  
<http://kidb.cwru.edu/pdsp.php>;

Tabela V - Recolha dos dados experimentais necessários da série de piperazinas (compostos 60 a 76) para a elaboração do estudo da afinidade para o receptor 5-HT<sub>1A</sub> e 5-HT<sub>2A</sub>.

| <i>Composto</i> | <i>Estrutura</i> | <i>Ar</i> | <i>[Linker]</i> | <i>-R</i>          | <i>pK<sub>i</sub></i><br><i>5-HT<sub>1A</sub></i> | <i>pK<sub>i</sub></i><br><i>5-HT<sub>2A</sub></i> | <i>Ref.</i> |
|-----------------|------------------|-----------|-----------------|--------------------|---|---|-------------|
| 60              |                  |           |                 | 2,3-diCl           | 6.510042  | 7.03  | f)          |
| 61              |                  |           |                 | 2,3-diCl           | 7.036212  | 7.33  | f)          |
| 62              |                  |           |                 | 2-OCH <sub>3</sub> | 7.66354   | 7.12  | f)          |
| 63              |                  |           |                 | 2-OCH <sub>3</sub> | 7.143876  | 7.18  | f)          |
| 64              |                  |           |                 | 2,3-diCl           | 7.464706  | 7.37  | f)          |
| 65              |                  |           |                 | 2,3-diCl           | 5.742321  | 6.26  | f)          |
| 66              |                  |           |                 | 2-OCH <sub>3</sub> | 7.441291  | 6.16  | f)          |
| 67              |                  |           |                 | 2,3-diCl           | 6.931814  | 7.05  | f)          |
| 68              |                  |           |                 | 2,3-diCl           | 7.821023  | 7.81  | f)          |
| 69              |                  |           |                 | 2,3-diCl           | 5.69037   | 6.17  | f)          |
| 70              |                  |           |                 | 2,3-diCl           | 7.229148  | 7.33  | f)          |
| 71              |                  |           |                 | 2,3-diCl           | 6.007889  | 6.05  | f)          |
| 72              |                  |           |                 | 2,3-diCl           | 7.630784  | 7.64  | f)          |
| 73              |                  |           |                 | 2,3-diCl           | 6.02641   | 6.48  | f)          |
| 74              |                  |           |                 | 2-OCH <sub>3</sub> | 7.468521  | 6.81  | f)          |
| 75              |                  |           |                 | 2-OCH <sub>3</sub> | 8.173277  | 6.64  | f)          |

2,3-  
diCl

7.527244

7.81

f)

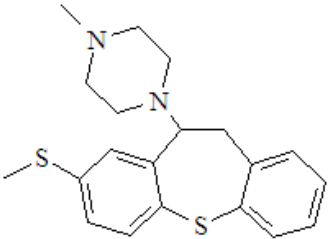
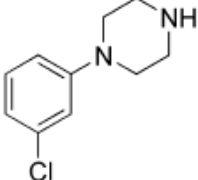
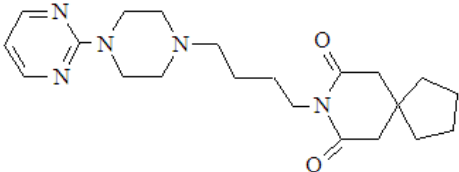
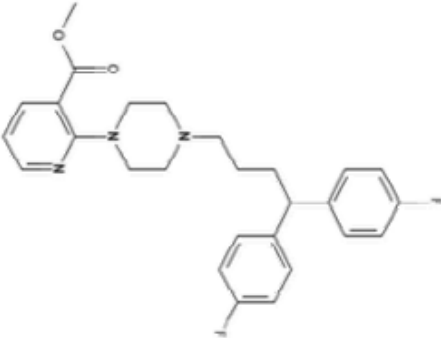
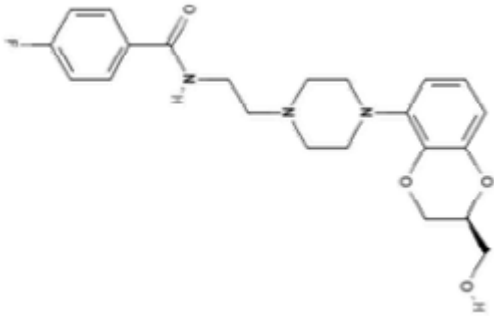
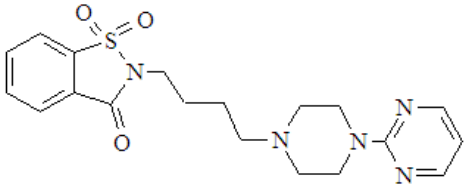
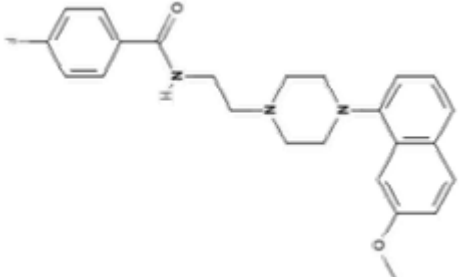
## Legenda:

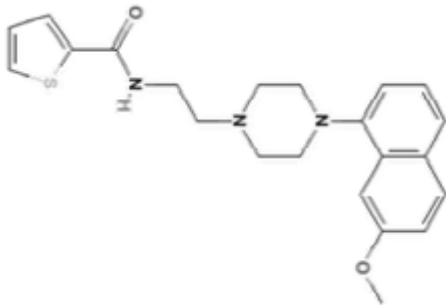
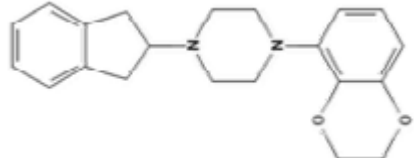
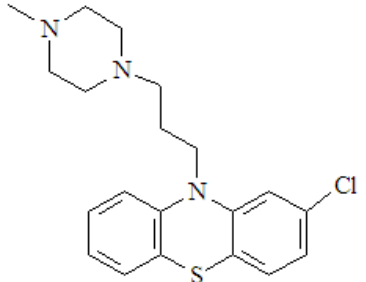
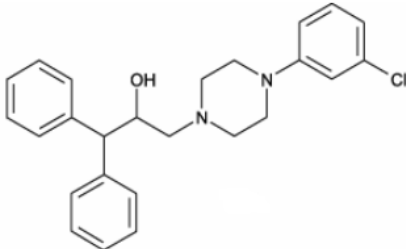
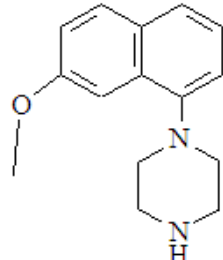
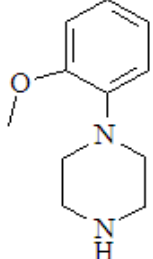
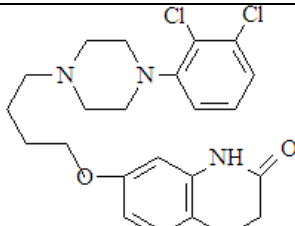
- a) López-Rodríguez *et al.*, 2002;  
 b) Tomić *et al.*, 2004;  
 c) Bojarski *et al.*, 2002;  
 d) Huang *et al.*, 2005 ;  
 e) Jurczyk *et al.*, 2004 ;  
 f) Grundt *et al.*, 2007 ;  
 g) González-Gómez *et al.*, 2003;  
 h) Sanchez *et al.*, 1999 ;  
 i) Price *et al.*, 1999;

- j) Gaster *et al.*, 1998;  
 k) Knight *et al.*, 2004;  
 l) Newman-Tancredi *et al.*, 1998;  
 m) Newman-Tancredi *et al.*, 1998a;  
 n) Jorand-Lebrun *et al.*, 1997;  
 o) Shapiro *et al.*, 2003;  
 p) [www.bindingdb.org](http://www.bindingdb.org);  
 q) [www.iuphar-db.org](http://www.iuphar-db.org);  
 r) Protein ki Database (2008)  
<http://kidb.cwru.edu/pdsp.php>;

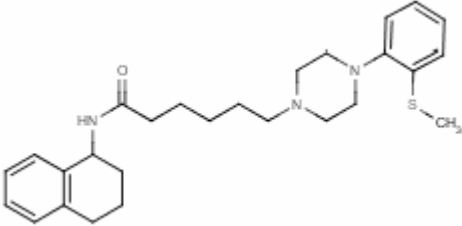
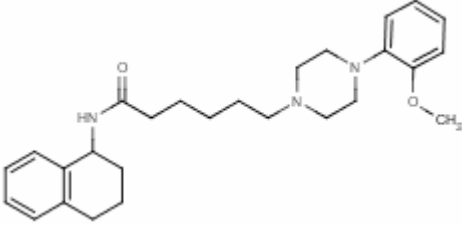
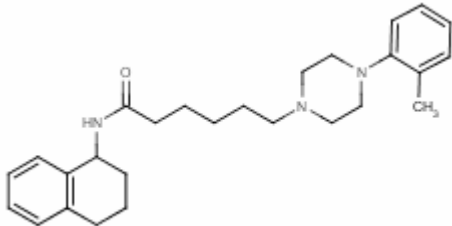
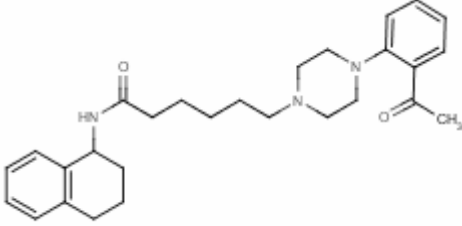
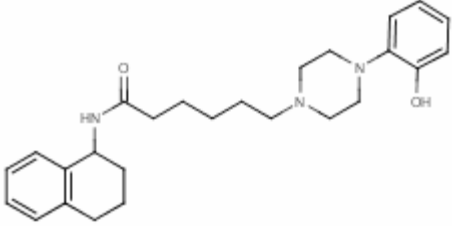
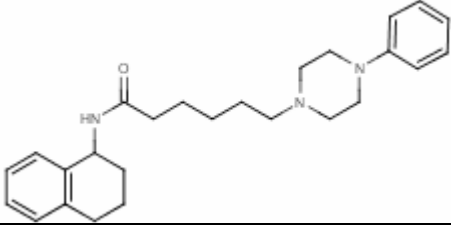
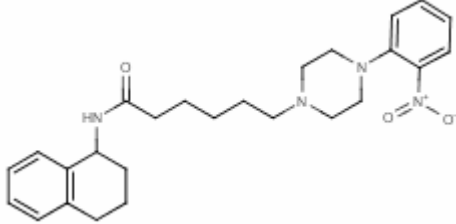
Tabela VI - Recolha dos dados experimentais necessários da série de piperazinas (compostos 77 a 114) para a elaboração do estudo da afinidade para o receptor 5-HT<sub>1A</sub> e 5-HT<sub>2A</sub>.

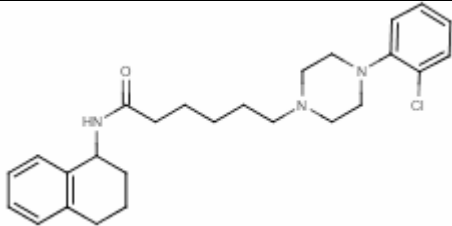
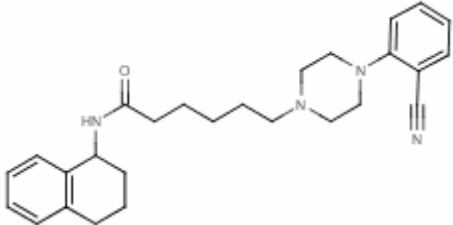
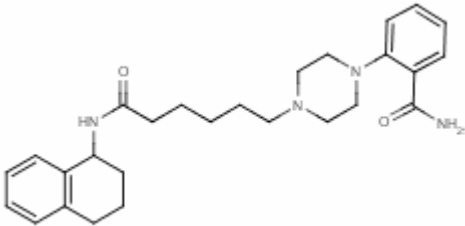
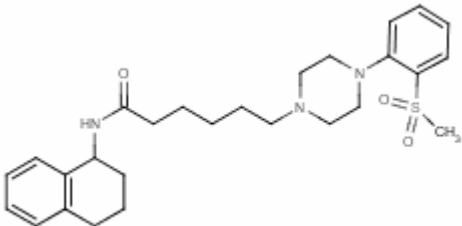
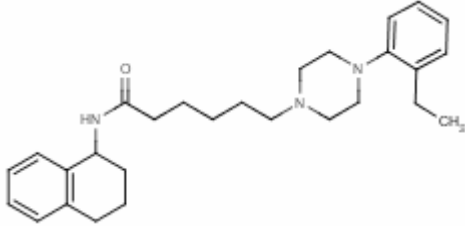
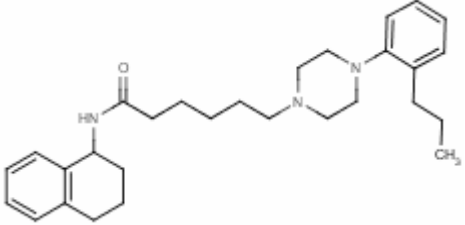
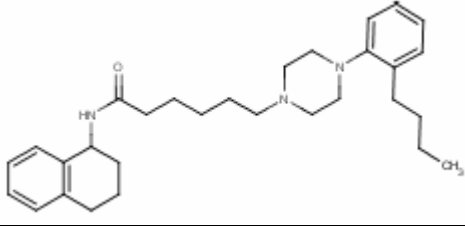
| Composto | Estrutura | $pK_i$             | $pK_i$             | Ref.    |
|----------|-----------|--------------------|--------------------|---------|
|          |           | 5-HT <sub>1A</sub> | 5-HT <sub>2A</sub> |         |
| 77       |           | 6.420216           | -                  | r)      |
| 78       |           | 6.641998           | 8.00               | r) e h) |
| 79       |           | 6.922652           | 6.90               | r)      |
| 80       |           | 7.20               | 7.80               | i) e j) |
| 81       |           | 5.770318           | 6.48               | r)      |

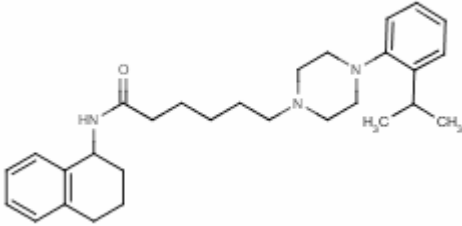
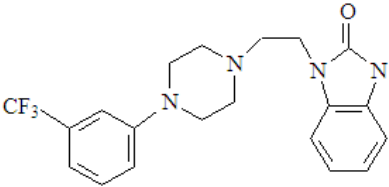
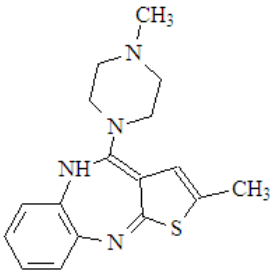
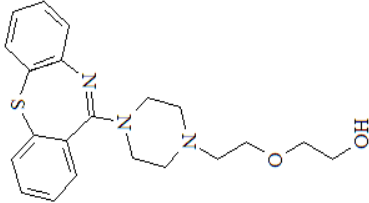
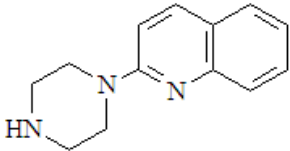
|    |   |          |      |             |
|----|---|----------|------|-------------|
| 82 |    | 7.713971 | 8.50 | r) e k)     |
| 83 |    | 7.026626 | 6.69 | r)          |
| 84 |    | 7.801125 | 6.86 | r)          |
| 85 |   | 8.70     | -    | q) e l)     |
| 86 |  | 9.30     | -    | q) e l)     |
| 87 |  | 8.70     | 5.30 | r); l) e m) |
| 88 |  | 9.70     | -    | l) e m)     |

|    |   |          |      |         |
|----|---|----------|------|---------|
| 89 |    | 10.50    | -    | l) e m) |
| 90 |    | 9.20     | -    | m)      |
| 91 |    | 5.229148 | 7.82 | r)      |
| 92 |   | 8.113509 | 8.18 | r)      |
| 93 |  | 8.60     | -    | n)      |
| 94 |  | 6.884822 | -    | r)      |
| 95 |  | 8.20     | 7.70 | o)      |



|     |   |          |      |    |
|-----|---|----------|------|----|
| 96  |    | 7.278189 | 6.49 | p) |
| 97  |    | 8.065502 | 5.90 | p) |
| 98  |    | 6.554396 | 6.58 | p) |
| 99  |   | 8.414539 | 4.91 | p) |
| 100 |  | 7.619789 | 5.47 | p) |
| 101 |  | 6.89279  | 7.11 | p) |
| 102 |  | 6.737549 | 6.55 | p) |

|     |   |          |      |    |
|-----|---|----------|------|----|
| 103 |    | 7.017729 | 6.52 | p) |
| 104 |    | 7.779892 | 6.15 | p) |
| 105 |    | 6.306273 | 5.40 | p) |
| 106 |   | 5.505289 | 5.40 | p) |
| 107 |  | 7.101275 | 5.47 | p) |
| 108 |  | 6.774691 | 5.50 | p) |
| 109 |  | 7.221849 | -    | p) |

|     |   |          |      |    |
|-----|---|----------|------|----|
| 110 |    | 6.777284 | 5.32 | p) |
| 111 |    | 7.72     | -    | r) |
| 112 |    | 6.00     | -    | r) |
| 113 |   | 5.610834 | -    | r) |
| 114 |  | 5.69897  | -    | r) |

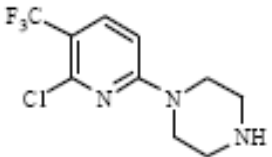
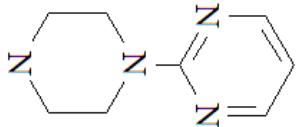
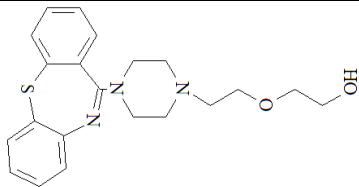
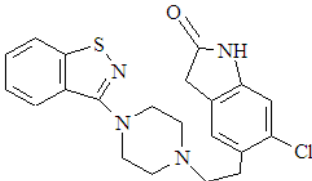
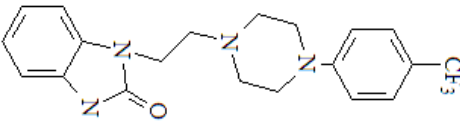
**Legenda:**

a) López-Rodríguez *et al.*, 2002;  
 b) Tomić *et al.*, 2004;  
 c) Bojarski *et al.*, 2002;  
 d) Huang *et al.*, 2005 ;  
 e) Jurczyk *et al.*, 2004 ;  
 f) Grundt *et al.*, 2007 ;  
 g) González-Gómez *et al.*, 2003;  
 h) Sanchez *et al.*, 1999 ;  
 i) Price *et al.*, 1999;

j) Gaster *et al.*, 1998;  
 k) Knight *et al.*, 2004;  
 l) Newman-Tancredi *et al.*, 1998;  
 m) Newman-Tancredi *et al.*, 1998a;  
 n) Jorand-Lebrun *et al.*, 1997;  
 o) Shapiro *et al.*, 2003;  
 p) [www.bindingdb.org](http://www.bindingdb.org);  
 q) [www.iuphar-db.org](http://www.iuphar-db.org);  
 r) Protein ki Database (2008)  
<http://kidb.cwru.edu/pdsp.php>;

Tabela VII - Recolha dos dados experimentais necessários da série de piperazinas (compostos 115 a 129) para a elaboração do estudo da afinidade para o receptor 5-HT<sub>1A</sub> e 5-HT<sub>2A</sub>.

| Composto | Estrutura | Ar | -R               | $pK_i$             | $pK_i$             | Ref. |
|----------|-----------|----|------------------|--------------------|--------------------|------|
|          |           |    |                  | 5-HT <sub>1A</sub> | 5-HT <sub>2A</sub> |      |
| 115      |           |    | CH <sub>3</sub>  | -                  | 5.63               | g)   |
| 116      |           |    | OCH <sub>3</sub> | -                  | 6.86               | g)   |
| 117      |           |    | CH <sub>3</sub>  | -                  | 6.03               | g)   |
| 118      |           |    | OCH <sub>3</sub> | -                  | 5.71               | g)   |
| 119      |           |    | CH <sub>3</sub>  | -                  | 6.76               | g)   |
| 120      |           |    | OCH <sub>3</sub> | -                  | 5.51               | g)   |
| 121      |           |    | CH <sub>3</sub>  | -                  | 5.97               | g)   |
| 122      |           |    | OCH <sub>3</sub> | -                  | 6.27               | g)   |
| 123      |           |    |                  | -                  | 5.40               | k)   |
| 124      |           |    |                  | -                  | 5.17               | r)   |

|     |  |   |      |    |
|-----|--|---|------|----|
| 125 |   | - | 8.12 | r) |
| 126 |   | - | 4.48 | r) |
| 127 |   | - | 6.83 | r) |
| 128 |   | - | 9.51 | r) |
| 129 |  | - | 6.90 | r) |

**Legenda:**

- a) López-Rodríguez *et al.*, 2002;  
 b) Tomić *et al.*, 2004;  
 c) Bojarski *et al.*, 2002;  
 d) Huang *et al.*, 2005 ;  
 e) Jurczyk *et al.*, 2004 ;  
 f) Grundt *et al.*, 2007 ;  
 g) González-Gómez *et al.*, 2003;  
 h) Sanchez *et al.*, 1999 ;  
 i) Price *et al.*, 1999;

- j) Gaster *et al.*, 1998;  
 k) Knight *et al.*, 2004;  
 l) Newman-Tancredi *et al.*, 1998;  
 m) Newman-Tancredi *et al.*, 1998a;  
 n) Jorand-Lebrun *et al.*, 1997;  
 o) Shapiro *et al.*, 2003;  
 p) [www.bindingdb.org](http://www.bindingdb.org);  
 q) [www.iuphar-db.org](http://www.iuphar-db.org);  
 r) Protein ki Database (2008)  
<http://kidb.cwru.edu/pdsp.php>

