
Dissertação de Mestrado

Estudo da Estrutura-Atividade das Esterilquinolinas e Tetraciclinas Através de Descritores Quânticos

Por **Fernando Sato**

Orientado por **Douglas Soares Galvão**

Grupo de Sólidos Orgânicos e Novos Materiais - GSONM

Departamento de Física Aplicada - DFA

Instituto de Física Gleb Wataghin - IFGW

Universidade Estadual de Campinas - Unicamp

Dissertação de Mestrado apresentada no
Instituto de Física Gleb Wataghin da
Universidade Estadual de Campinas para
obtenção do Título de Mestre em Física.

Campinas, 25 de Fevereiro de 2003

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO IFGW - UNICAMP

Sa83e

Sato, Fernando

Estudo da estrutura-atividade das Esterilquinolinas e Tetraciclina através de descritores quânticos / Fernando Sato. -- Campinas, SP : [s.n.], 2003.

Orientador: Douglas Soares Galvão.

Dissertação (mestrado) - Universidade Estadual de Campinas, Instituto de Física "Gleb Wataghin".

1. Relação estrutura-atividade (Bioquímica).
2. Química quântica. 3. Estrutura eletrônica.
- I. Galvão, Douglas Soares. II. Universidade Estadual de Campinas. Instituto de Física "Gleb Wataghin".
- III. Título.



MEMBROS DA COMISSÃO JULGADORA DA TESE DE MESTRADO DE **FERNANDO SATO – RA 007137** APRESENTADA E APROVADA AO INSTITUTO DE FÍSICA “GLEB WATAGHIN”, DA UNIVERSIDADE ESTADUAL DE CAMPINAS, EM 25/02/2003.

COMISSÃO JULGADORA:

Prof. Dr. Douglas Soares Galvão (Orientador do Candidato) –
IFGW/UNICAMP

Profa. Dra. Íris Concepcion Linares de Torriani – IFGW/UNICAMP

Prof. Dr. Paulo Monteiro Vieira Braga Barone – IF/UFJF

Agradecimentos

Primeiramente gostaria de agradecer a CNPq pela minha bolsa no início do mestrado e a FAPESP pelo auxílio financeiro cedido para o desenvolvimento deste trabalho e pelo auxílio das viagens.

Ao IFGW pelas instalações cedidas como o Prédio D e a sala de computação do grupo GSONM no Departamento de Física Aplicada.

Ao prof. Douglas Soares Galvão que me auxiliou durante o desenvolvimento deste trabalho, e por preocupação a pesquisa e com período de duração de minha bolsa. Pessoa de grandes idéias sobre a física e bastante flexível na relação entre orientador e orientando.

Ao pessoal do GSONM (Giro, Vitor, Rosana, Karla, Sérgio B. Legoas, Abdala e Nei) que realmente trabalharam como um grupo de pesquisa, como boas dicas com relação ao trabalho. Em especial para Scheila que me ajudou muito em meus primeiros contatos com o meu projeto de pesquisa.

Aos companheiros extra física (aquele do churrasquinho) Tércio, Zolacir, Juliano, Rogério V. Gelamo*, Sérgio (super), ...

Ao pessoal da APGF: Alex EB, Manoel, Pepe, Júlio Garcia, Walter (sala11), Daniel, Edson (mineiro), ... pelas conversas, xadrez e violão.

Ao pessoal da limpeza Lú, Rosa e Dalva pelo cafezinho e pela conversa no almoço.

Às secretárias do DFA Lúcia, Flávia e Márcia pela orientação com burocracia do dia-a-dia.

À CPG Maria Ignez, Armando e Alessandra pelas informações quando o assunto

era a pós graduação.

Aos professores responsáveis pelas disciplinas que cursei onde pude aprender mais sobre a física (Cabreira, Kleimam, Márcia da Química).

Ao prof. Bernardo Laks pelas boas conversas compartilhadas.

E a todos aqueles(as) que esqueci de colocar nesta lista.

Dedicatória

Dedico este trabalho à minha família, meu pai Masuo, minha Mãe Tamai, meu irmão Carlos e sua esposa Sônia G., Julinha e Netinho, minha tia Tosko e minha prima Lúcia pela torcida e o grande incentivo dado desde o início dos meus estudos. Também para minha companheira de todos os dias e esposa Sônia que nunca mediu esforços para me ajudar e me acompanhar na rotina de todos os dias. Também para o Francisco Carlos Lavarda (Chico) que me ensinou sobre a pesquisa (no meu estágio de três anos de iniciação científica) e sobre a vida acadêmica na qual estou experimentando.

Resumo

Neste trabalho estudamos dois grupos de moléculas na tentativa de correlacionar a estrutura eletrônica com a atividade biológica. Um grupo é formado por inibidores da enzima integrase, necessária para replicação do HIV-1, e o outro grupo por inibidores da *Escherichia Coli*, uma das bactérias causadoras da diarreia.

Em nossas investigações aplicamos a Metodologia dos Índices Eletrônicos (MIE) para verificar a correlação estrutura-atividade dos compostos. Em caráter de comparação e validação dos dados produzidos pela MIE utilizamos também algumas das metodologias mais usuais na área de reconhecimento de padrões: a Análise de Componentes Principais (PCA), Análise Hierárquica de Agrupamentos (HCA) e as Redes Neurais Artificiais. Após o desenvolvimento completo deste trabalho concluímos que a MIE tratou com êxito o problema de reconhecimentos de padrões relacionados à atividade biológica das moléculas estudadas.

Em um passo mais especulativo propomos novos derivados de tetraciclinas e submetemos à previsão pelo modelo proposto no início da segunda parte do trabalho. Nesta segunda parte foi possível sugerir observando os resultados e os parâmetros da MIE alguns compostos, que apresentam características de inibidores da *Escherichia Coli*.

Os resultados deste trabalho somam-se às diferentes classes de outros compostos já estudados, totalizando mais de 2000 compostos em que os mesmos descritores da MIE se mostraram eficientes na classificação biológica.

Abstract

In this work we have studied two classes of molecular groups trying to establish correlations between their electronic structure and correlated biological activity. The first group contains integrase inhibitors, which are necessary to HIV-1 replication; the second group contains Escherichia Coli inhibitors, one bacterium which caused diarrhea.

Electronic Index Methodology (EIM) was used in the Structure-Activity Relationship (SAR) analysis. In order to have an independent statistical validation, the data produced from EIM were contrasted with more standard pattern recognition methodologies: Principal Component Analysis (PCA), Hierarchical Cluster Analysis (HCA), and Artificial Neural Networks (ANN). Our results shown that EIM can be sucessfully use for pattern recognition. As in previous works for other organic families the EIM out performed PCA, HCA and ANN techniques presenting the same predictive power with a significant reduction of computational effort.

In more speculative approach EIM was used to propose new tetraciclines derivatives.

Our results with other class of compounds recently studied sum more them 2000 compounds where EIM have been proved very efficient for biological classification.

Sumário

1	Introdução e Objetivos	8
2	Esterilquinolinas e HIV	11
3	Metodologia	21
3.1	Métodos Semi-empíricos	21
3.2	Metodologia dos Índices Eletrônicos (MIE)	24
3.3	Métodos de Quimiometria	27
3.3.1	Análise de Componentes Principais (PCA)	27
3.3.2	Análise Hierárquica de Agrupamentos (HCA)	30
3.4	Redes Neurais	31
4	Resultados - Esterilquinolinas	39
4.1	Busca conformacional	39
4.2	Cálculo da MIE	42
4.3	PCA e HCA	51
4.4	Redes Neurais	56
5	Tetraciclinas	63
6	Resultados - Tetraciclinas	68
6.1	Busca Conformacional	68

6.2	Cálculo da MIE	71
6.3	PCA e HCA	72
6.4	Novos Compostos	77
7	Sumário dos Resultados	84
7.1	Esterilquinolinas	84
7.2	Tetraciclinas	85
8	Conclusões	87
9	Apêndice A	89

Lista de Figuras

2.1	Esquema da estrutura do HIV-1	12
2.2	Replicação do vírus HIV	13
3.1	Exemplo de um tipo de grau de liberdade do composto C_3H_6 , rotação em torno da ligação que une os dois carbonos.	22
3.2	Exemplo dos graus de liberdade do grupo hidroxila, rotações em torno da ligação que une o carbono e a mobilidade do oxigênio.	23
3.3	Gráfico de Densidade de Estados Total - DOS	26
3.4	Gráfico de Densidade de Estados Local - LDOS, base de um dos índices da MIE	26
3.5	Exemplo de um cálculo com PCA, representação dos compostos em termos das componentes principais (PC's) - Score.	29
3.6	Exemplo de um cálculo com PCA, representação dos descritores em termos das componentes principais (PC's) - Loadings	30
3.7	Exemplo de um cálculo com HCA, cálculo de similaridade dado pela distância entre vizinhos.	31
3.8	Cérebro Humano, contém aproximadamente 10^{11} neurônios altamente interconectados.	33
3.9	Esquema de um neurônio biológico, fonte de inspiração para a construção dos neurônios artificiais, cada neurônio biológico faz aproximadamente 10^4 ligações.	33

3.10	Esquema de um neurônio artificial	34
3.11	Redes Neurais Artificiais - rede neural tipo perceptron, uma das arquiteturas mais simples das redes neurais artificiais	35
4.1	Exemplos dos graus de liberdade, representados pelas ligações simples (traço simples que une dois átomos), de um composto de esterilquinolina.	40
4.2	Um derivado que possui um mínimo de calor de formação	40
4.3	Outro derivado que possui um mínimo de calor de formação	41
4.4	Mapa do Calor de Formação na busca conformacional	41
4.5	Regiões utilizadas para o cálculo da LDOS. Os valores da LDOS são base de um índice da MIE	44
4.6	Análise de Componentes Principais (PCA) com dados MIE/AM1 e demais parâmetros.	52
4.7	Análise de Componentes Principais (PCA) com dados MIE/PM3 e demais parâmetros.	53
4.8	HCA - AM1	54
4.9	HCA - PM3	54
5.1	Figura artística da Bactéria <i>E. coli</i>	64
5.2	Foto de microscopia de uma colônia de <i>E. coli</i>	65
5.3	Estrutura da Tetraciclina na forma zwitteriônica.	65
6.1	Confórmero tipo Extended. O hidrogênio que está ligado ao grupo $N(Me)_2$ está apontando para a parte inferior da página.	69
6.2	Confórmero tipo Twisted. O hidrogênio que está ligado ao grupo $N(Me)_2$ está apontando para a parte superior da página.	69
6.3	Região onde foi realizado o cálculo da LDOS para os índices da MIE.	71
6.4	PCA - AM1.	73
6.5	PCA - PM3	74
6.6	HCA - AM1	75
6.7	HCA - PM3	75
6.8	Estrutura dos novos substituintes - TA	77
6.9	Estrutura dos novos substituintes - TB	77
6.10	Estrutura dos novos substituintes - TC	79

Lista de Tabelas

2.1	A tabela apresenta os compostos derivados da esterilquinolina. A sigla Nd indica que para tais compostos o valor de IC ₅₀ não foi determinado. (A = <i>3'processing</i> e B = <i>integration</i>)	20
3.1	Matriz de dados hipotética de entrada para PCA	28
3.2	Exemplo do tratamento de dados com a rede neural, com o número de descritores utilizados, os padrões de treinamentos e os padrões obtidos pela predição.	38
4.1	Valores do calor de formação para os compostos de mínima energia. A numeração refere-se às estruturas da tabela 2.1.	43
4.2	Átomos que compõem as regiões da figura 4.5	44
4.3	Parâmetros da MIE para os dados dos cálculos realizados com AM1. Os compostos destacados em negrito são inativos e os demais ativos.	46
4.4	Parâmetros da MIE para os dados dos cálculos realizados com PM3. Os compostos destacados em negrito são inativos e os demais ativos.	47

4.5	Porcentagens de acertos com a MIE, para os dados dos cálculos realizados com AM1, baseados nos padrões experimentais. Os destaques em itálico são correspondentes aos índices dos estados ocupados e os demais são correspondentes aos estados desocupados. A última coluna da direita a letra H indica os resultados para os dois últimos orbitais moleculares ocupados (HOMO e HOMO-1) e a letra L indica os resultados para os dois primeiros orbitais moleculares desocupados (LUMO e LUMO+1) .	49
4.6	Porcentagens de acertos com a MIE, para os dados dos cálculos realizados com PM3, baseados nos padrões experimentais. Os destaques em itálico são correspondentes aos índices dos estados ocupados e os demais são correspondentes aos estados desocupados. A última coluna da direita a letra H indica os resultados para os dois últimos orbitais moleculares ocupados (HOMO e HOMO-1) e a letra L indica os resultados para os dois primeiros orbitais moleculares desocupados (LUMO e LUMO+1). .	50
4.7	Dados da variância das componentes principais PC1 e PC2, para os dados dos cálculos realizados com AM1.	51
4.8	Dados da variância das componentes principais PC1 e PC2, para os dados dos cálculos realizados com PM3.	55
4.9	Conjuntos submetidos para as redes neurais: as colunas são os conjuntos de treinamentos (T) e predições (P); as linhas são os compostos (A = ativa e I = inativa). Dados dos cálculos realizados com AM1, com base nos padrões experimentais.	57
4.10	Conjuntos submetidos para as redes neurais: as colunas são os conjuntos de treinamentos (T) e predições (P); as linhas são os compostos (A = ativa e I = inativa). Dados dos cálculos realizados com PM3, com base nos padrões experimentais.	58
4.11	Parâmetros do treinamento da rede neural e erros nos treinamentos e predições. Dados dos cálculos realizados com AM1	59
4.12	Parâmetros do treinamento da rede neural e erros nos treinamentos e predições. Dados dos cálculos realizados com PM3	60

4.13	Conjuntos submetidos para as redes neurais: as colunas são os conjuntos de treinamentos (T) e predições (P); as linhas são os compostos (A = ativa e I = inativa). Baseado nos padrões de classificações da MIE, utilizando somente índices da MIE (AM1 e PM3) como dados de entrada.	61
4.14	Parâmetros de treinamento da rede neural e erros nos treinamentos e nas predições, dados AM1	62
4.15	Parâmetros de treinamento da rede neural e erros nos treinamentos e nas predições, dados PM3	62
4.16	Tabela das porcentagens de acertos das metodologias utilizadas	62
5.1	Substituintes da estrutura da tetraciclina figura 5.3 que compõem os derivados analisados.	66
6.1	Valores do calor de formação para os compostos de mínima energia, referente aos compostos da tabela 5.1	70
6.2	Parâmetros para MIE, para os compostos de tetraciclina da tabela 5.1 e ambos os dados dos cálculos AM1 e PM3	72
6.3	Novos substituintes para formação das novas tetraciclina	78
6.4	Índices da MIE para as nova tetraciclina. Os compostos inativos estão destacados com os caracteres em negrito e os demais compostos são ativos	80
6.5	Resultados da predições das novas tetraciclina pela MIE.	81
6.6	Parâmetros de treinamento da rede neural, baseado na atividade biológica experimental. Foram utilizados os índices eletrônicos da MIE, do cálculos AM1, como entrada para a rede.	81
6.7	Resultados da preidção dos novos derivados de Tetraciclina com as redes neurais.	82
6.8	Resultados da predições das novas tetraciclina pela RN.	82
7.1	Tabela das porcentagens de acertos das metodologias utilizadas para as Esterilquinolinas	85
7.2	Tabela das porcentagens de acertos das metodologias utilizadas para o conjunto das Tetraciclina	86

CAPÍTULO 1

Introdução e Objetivos

A história da humanidade sobre a busca de curas para doenças é muito longa e antiga, repleta de casos muito bem sucedidos como a descoberta da penicilina. Apesar dos sucessos já obtidos temos casos que ainda não estão completamente resolvidos como o câncer e a AIDS.

Talvez uma das mais antigas formas de se realizar a cura estava por conta de algo proveniente da natureza (por exemplo as plantas medicinais) e também do aspecto religioso. Atualmente a busca pela cura de doenças conta com a ajuda da tecnologia interligando as diversas áreas do conhecimento (Física, Química, Biologia, Medicina, etc). Dentro da pesquisa científica um fator muito importante é o da interdisciplinaridade, que passou a ficar mais intensa recentemente. Comunicações entre áreas do conhecimento têm acelerado as pesquisas tanto na área biológica como nas demais áreas.

Essas "curas" se resumem grosseiramente em tratamentos feitos através de remédios, tecnicamente chamados de drogas ou fármacos. O seu desenvolvimento, que inicialmente limitava-se às indústrias farmacêuticas hoje é motivo de muitas pesquisas dentro das universidades de todo o mundo. A inclusão deste assunto dentro das universidades deve-se ao fato da melhor compreensão do funcionamento do corpo humano a nível microscópico (pesquisa de base) e das próprias universidades realizarem estudos de base atrelados à tecnologia.

Na produção de um fármaco existem vários processos entre a elaboração ou "de-

scoberta” do composto molecular até o produto final que chega ao consumidor. O ponto de partida, independente da necessidade humana da droga, ocorre quando é verificado alguma atividade biológica de algum composto e o mesmo é colocado em testes iniciais para verificação de sua potência, que é a medida da sua eficiência de ”cura”. Um próximo passo seria verificar a potência dos compostos derivados do testado inicialmente, tanto experimentalmente como computacionalmente. Durante esses processos intermediários é verificada também a toxicidade do composto em questão pois existe a possibilidade de existir um alto grau de toxicidade. Nesses estudos preliminares o estudo de toxicidade é importante pois a droga pode ser muito eficiente ao que se pretende combater, por outro lado pode comprometer outras partes do organismo. Seguindo com essa descrição simplista o próximo passo seriam os testes *in vitro* e *in vivo* em cobaias, depois em seres humanos, e em seguida o produto final para a grande população. O processo completo pode se estender por até 10 anos de trabalho^[1].

O presente trabalho está relacionado ao processo do desenvolvimento das drogas, mais especificamente nos modelos computacionais. Ele pretende contribuir com uma nova metodologia que relaciona estrutura molecular com atividade biológica (*SAR*, *Structure Activity Relationship*) introduzindo novos descritores quânticos moleculares de caráter provavelmente universal. Esses descritores são frutos do desenvolvimento da Metodologia dos Índices Eletrônicos (MIE), que inicialmente foi proposta para qualificar dentre um conjunto de hidrocarbonetos policíclicos aromáticos quais eram carcinogênicos ou não.

A metodologia dos índices eletrônicos é baseada nos conceitos físicos de densidade local de estados e na diferença de energia dos estados de fronteira, provenientes de cálculos de métodos de mecânica quântica. A MIE utiliza regras booleanas simples a partir desses dois parâmetros para determinar a classificação de um conjunto de compostos.

Neste trabalho aplicamos a MIE para dois conjuntos moleculares totalmente distintos em suas ações biológicas. O primeiro conjunto explorado envolve os derivados da esterilquinolina, potentes inibidores da integrase do HIV-1. O segundo conjunto envolve os derivados da tetraciclina que têm uma grande gama de ação como inibidor biológico, mais especificamente no combate contra a *Escherichia Coli* (um dos causadores da diarreia).

Para validarmos estatisticamente e compararmos a MIE com outras metodologias

muito utilizadas na área de reconhecimentos de padrões, no universo acadêmico e comercial, utilizamos: a Análise de Componentes Principais (PCA - *Principal Component Analysis*), Análise Hierárquica de Agrupamentos (HCA - *Hierarchical Cluster Analysis*) e as Redes Neurais (RN - *Neural Networks*).

Os novos índices propostos são extremamente úteis para a seletividade de compostos orgânicos. Em todos os casos onde foi utilizada a MIE o reconhecimento de padrões de atividade biológica, em média, apresentou-se superior a 80%.

CAPÍTULO 2

Esterilquinolinas e HIV

A Síndrome da Imunodeficiência Adquirida (SIDA), mais conhecida mundialmente pela sigla AIDS, é causada, ou pelo menos se acredita, pelo Vírus da Imunodeficiência Humana (HIV). A AIDS não era conhecida até início da década de oitenta e foi investigada pela primeira vez por Robert Gallo^[2] e Luc Montagnier^[3]. A partir desta data um trabalho intenso tem sido desenvolvido na busca de algum tipo de vacina para conter a proliferação do vírus. Existem dois tipos de HIV: HIV tipo 1 (HIV-1) e tipo 2 (HIV-2). O HIV-2 aparece nas regiões da África e sudeste da Ásia e é similar ao vírus da imunodeficiência dos símios (SIV) e o modo de ação semelhante com o HIV-1. O HIV-1 aparece em todas as partes do mundo, o modo de ação (infecção) se faz através das mucosas ou pela introdução do sangue contaminado.

O vírus apresenta a seguinte estrutura:

1. Camada externa, tecnicamente ***gp120***^{[4],[5]};
2. Camada que faz interface entre o interior e o exterior do vírus, chamada de ***gp41*** (*transmembrane protein*);
3. Camada associada à membrana anterior, ***p17*** (*protein matrix*);
4. Camada que envolve o RNA e outras proteínas, denominada capsídeo ***p24*** CA (*core shell*);

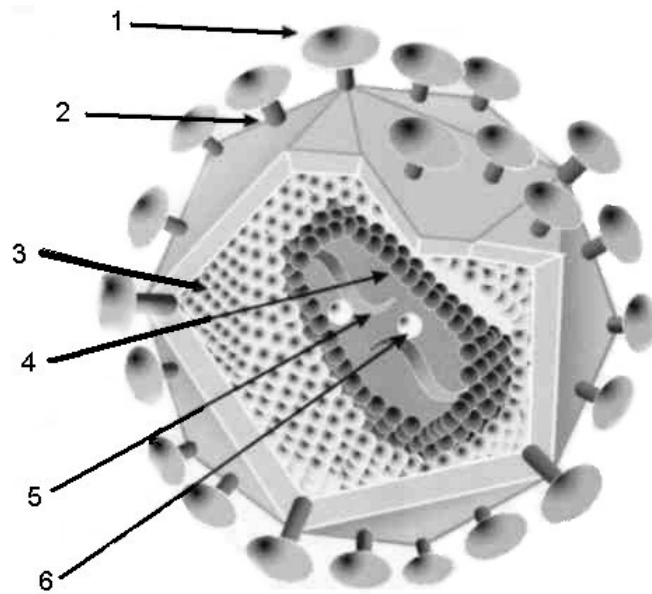


Figura 2.1: Esquema da estrutura do HIV-1

5. RNA, onde toda informação genética está armazenada, que é a característica de um retrovírus. Os terminais do RNA recebem nomes especiais de terminais 3' e 5';
6. As proteínas e também enzimas: RNase H ou transcriptase reversa (TR), p66; protease(PR), p11; integrase (IN), p32.

O vírus após entrar no corpo humano é atraído para células apropriadas, como as células receptoras CD4, às quais se juntam pela fusão de uma membrana celular susceptível ou por endocitose (que é a molécula *gp120* junto com as células T). O vírus então penetra na célula liberando o capsídeo *p24*, RNA e as enzimas. Neste momento o RNA viral é convertido em DNA viral catalizado pela transcriptase reversa. Uma vez convertido em DNA ele precisa ser integrado ao genoma do DNA da célula hospedeira, neste caso propiciado pela enzima integrase. Nesta etapa ocorrem dois processos:

1. **3'-processing** – que é a retirada de dois nucleotídeos de cada terminal 3' do DNA viral;
2. **strand transfer** – que é o processo de integração ocorridas nas extremidades dos terminais 3'e 5', são reações do tipo cortar e colar unindo o DNA viral com o

DNA da célula hospedeira.

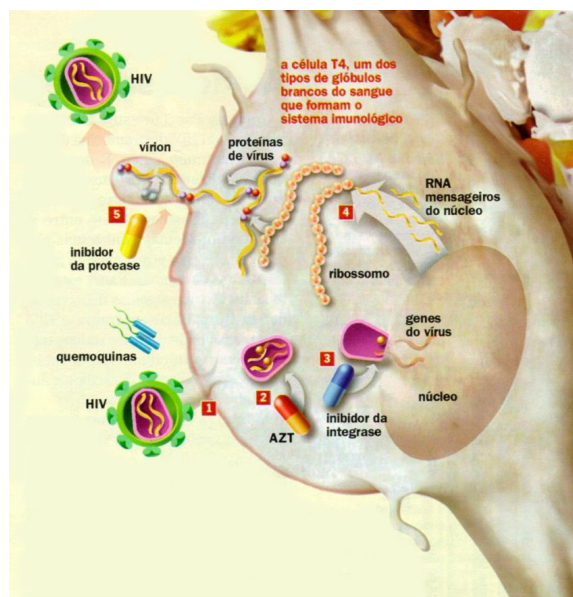


Figura 2.2: Replicação do vírus HIV

Com o DNA viral acoplado ao DNA da célula hospedeira, através da enzima RNA polimerase, ocorre a múltipla transcrição dos RNAs mensageiros (mRNA). Os mRNAs deixam o núcleo em direção ao ribossomo onde ocorre a tradução dos mRNAs em RNA, polipeptídeos e protease, que são as proteínas que possuem em sua estrutura quase todos os componentes para formação de um novo vírus. Nesta fase todos os elementos ditos anteriormente estão próximos da superfície celular. Após a clivagem dos polipetídeos^[6] pela enzima protease eles se encapsulam com a parede celular^[7] de tal forma a "brotarem" no lado externo da célula formando uma nova célula viral, ganhando a possibilidade de atacar outras novas células receptoras disponíveis.

As drogas pesquisadas e disponibilizadas atuam justamente na inibição das enzimas essenciais para a reprodução e a proliferação do vírus HIV-1. As três enzimas virais são: transcriptase reversa, protease e integrase. As duas primeiras já foram exaustivamente estudadas e a última ainda têm sido motivo de muita pesquisa para a verificação de uma boa droga capaz de inibir a enzima integrase. As drogas atuam nas enzimas da seguinte forma:

- transcriptase reversa: como essa enzima é responsável pela transcrição do RNA viral em DNA viral, a droga atua substituindo o substrato que é o RNA viral. A

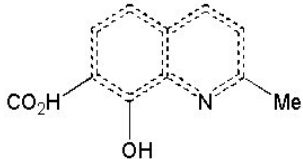
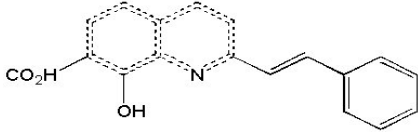
droga comercialmente utilizada e mais conhecida para essa parte do processo é o AZT (3'-azido-3'-deoxythimidina)^[8];

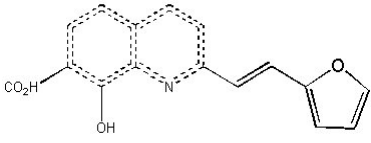
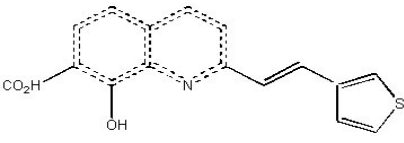
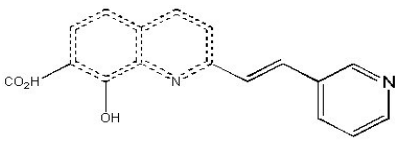
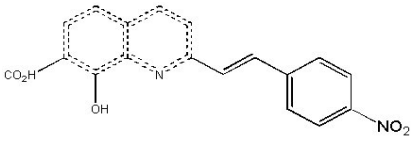
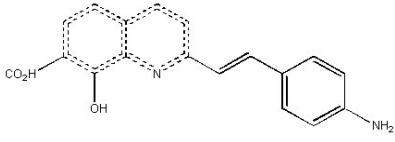
- Protease: nesta parte os inibidores atuam impedindo o processo de maturação do vírus, atacando múltiplas enzimas^[9];
- Integrase: os inibidores da integrase atuam na etapa de acoplamento do DNA viral no DNA da célula hospedeira, impedindo ambas etapas, *3'-processing* e *strand transfer*.

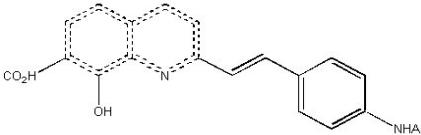
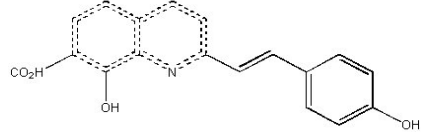
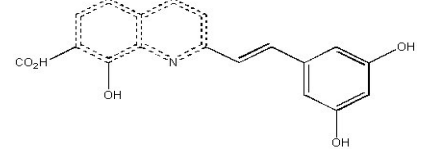
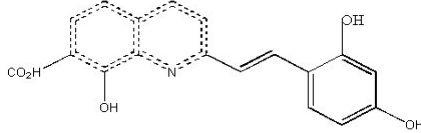
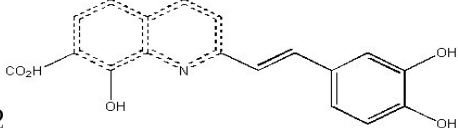
Como dito anteriormente os inibidores de integrase são muito importantes para a contenção da replicação do vírus HIV-1. O trabalho que será descrito no capítulo 4 está justamente desenvolvido para classificação da atividade biológica de alguns derivados das moléculas Esterilquinolinas^{[10], [11]}, potentes inibidores da integrase do HIV-1.

As esterilquinolinas polihidroxiladas são moléculas provenientes (sintetizadas) a partir de uma “matriz” quinolina por vários tipos de reações e são potentes inibidores da terceira enzima viral (integrase), em ambos os processos, “*3'-processing*” e “*strand-transfer*”. Um dos grandes interesses por esses derivados das esterilquinolinas é devido à potência de inibição em concentrações não tóxicas.

Todos os compostos estudados tiveram seus valores da taxa de inibição obtidos pelo índice experimental IC₅₀. Esse índice significa qual é a quantidade necessária da droga (em μM) para inibir 50% da replicação, que neste caso é o vírus do HIV-1. Os valores de IC₅₀ e as moléculas investigadas estão indicadas na tabela 2.1.

	Forma Estrutural	Nome	IC ₅₀ (μM) ^{[10],[11]}	
			A	B
01		8-hydroxy-7-quinaldic-acid	>100	>100
02		(E)-8-hydroxy-2-(2-phenylethenyl)-7-quinolinecarboxylic Acid	5.3	2.1

	Forma Estrutural	Nome	IC ₅₀ (μ M) ^{[10],[11]}	
			A	B
03		(E)-8-hydroxy-2-(2-(2-furyl)ethenyl]-7-quinolinecarboxylic Acid	1.9	5.1
04		(E)-8-hydroxy-2-(2-(3-thienyl)ethenyl]-7-quinolinecarboxylic Acid	3.4	3.0
05		(E)-8-hydroxy-2-(2-(2-pyridinyl)ethenyl]-7-quinolinecarboxylic Acid	4.1	11
06		(E)-8-hydroxy-2-(2-(4-nitrophenyl)ethenyl]-7-quinolinecarboxylic Acid	1.2	1.7
07		(E)-8-hydroxy-2-(2-(4-aminophenyl)ethenyl]-7-quinolinecarboxylic Acid	3.5	2.2

	Forma Estrutural	Nome	IC ₅₀ (μ M) ^{[10],[11]}	
			A	B
08		(E)-8-hydroxy-2-(2-(4-acetamidophenyl)ethenyl)-7-quinolinecarboxylic Acid	1.4	1.2
09		(E)-8-hydroxy-2-(2-(4-hydroxyphenyl)ethenyl)-7-quinolinecarboxylic Acid	1.6	1.6
10		(E)-8-hydroxy-2-(2-(3,5-dihydroxyphenyl)ethenyl)-7-quinolinecarboxylic Acid	3.2	3.2
11		(E)-8-hydroxy-2-(2-(4-dihydroxyphenyl)ethenyl)-7-quinolinecarboxylic Acid	3.7	2.8
12		(E)-8-hydroxy-2-(2-(3,4-dihydroxyphenyl)ethenyl)-7-quinolinecarboxylic Acid	2.4	1.0

	Forma Estrutural	Nome	IC ₅₀ (μ M) ^{[10],[11]}	
			A	B
13		(E)-8-hydroxy-2-(2-(4-hydroxy-3-methoxyphenyl)ethenyl]-7-quinolinecarboxylic Acid	2.8	3.7
14		(E)-8-hydroxy-2-(2-(3-hydroxy-4-methoxyphenyl)ethenyl]-7-quinolinecarboxylic Acid	0.9	3.3
15		3,4-dihydroxy-5-methoxybenzaldehyde	0.3	0.4
16		(E)-8-hydroxy-2-(2-(4-hydroxy-5-methoxyphenyl)ethenyl]-7-quinolinecarboxylic Acid	0.7	1.7

	Forma Estrutural	Nome	IC ₅₀ (μ M) ^{[10],[11]}	
			A	B
17		(E)-8-hydroxy-2-(2-(4-hydroxy-3,5-dimethoxyphenyl)ethenyl]-7-quinolinecarboxylic Acid	4.9	4.5
18		(E)-8-hydroxy-2-(2-(3,5-dibromo-4-hydroxyphenyl)ethenyl]-7-quinolinecarboxylic Acid	1.3	1.2
19		(E)-8-hydroxy-2-(2-(3-iodo-4,5-dihydroxyphenyl)ethenyl]-7-quinolinecarboxylic Acid	4.0	4.9
20			>100	>100

	Forma Estrutural	Nome	IC ₅₀ (μ M) ^{[10],[11]}	
			A	B
21		(E)-5,7-dicloro-8-hidroxy-2-[2-(3,4-dihidroxyfenil)etênil]-7-quinolinecarboxylic Acid	>100	>100
22		8-hidroxy-2-[2-(3,4-dihidroxyfenil)etênil]-7-quinolinecarboxylic Acid	2.3	1.5
23		8-hidroxyquinaldine	>100	>100
24		(E)-quinaldin-8-yl-3,4-dimetoxycinnamate	>100	>100
25		(E)-quinaldin-8-yl-3,4-dihidroxicinnamate	>100	>100
26		(E)-2-(2-fênileetênil)-8-acetoxiquinoline	>100	>100

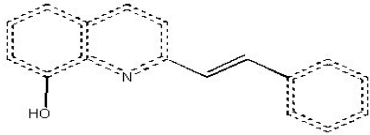
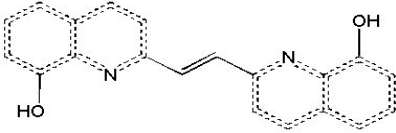
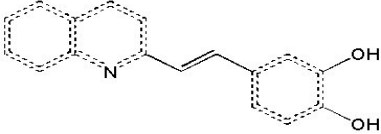
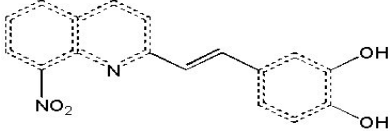
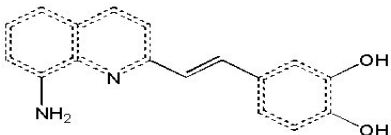
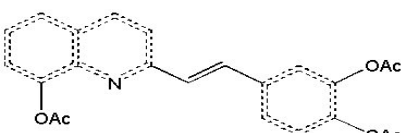
	Forma Estrutural	Nome	IC ₅₀ (μ M) ^{[10],[11]}	
			A	B
27		(E)-2-(2-phenylethenyl)-8-hydroxyquinoline	>100	>100
28		(E)-1,2-bis[2-(8-hydroxyquinolyl)]ethylene	>100	Nd
29		(E)-2-(2-[3,4-dihydroxyphenyl]ethenyl)quinoline	>100	Nd
30		(E)-2-[2-(3,4-dihydroxyphenyl)ethenyl]-8-nitroquinoline	>100	>100
31		(E)-2-[2-(3,4-dihydroxyphenyl)ethenyl]-8-aminoquinoline	>100	>100
32		(E)-2-[2-(3,4-diacetoxyphenyl)ethenyl]-8-acetoxyquinoline	>100	>100

Tabela 2.1: A tabela apresenta os compostos derivados da esterilquinolina. A sigla Nd indica que para tais compostos o valor de IC₅₀ não foi determinado. (A = *3'processing* e B = *integration*)

3.1 Métodos Semi-empíricos

Um método semi-empírico na descrição mais resumida possível é a resolução da Equação de Schrödinger para um arranjo de átomos dentro de várias aproximações. O método semi-empírico recebe este nome justamente porque ele não resolve por completo a equação de Schrödinger, pois utiliza aproximações e alguns parâmetros provenientes de dados empíricos ou cálculos de primeiros princípios (*ab initio*), ao invés de resolver todas as integrais necessárias. Este tipo de cálculo se posiciona entre os cálculos de mecânica molecular (que utilizam somente equações clássicas de movimento, equações de Newton) e cálculos quânticos *ab initio* (cálculos de primeiros princípios que resolvem por completo a equação de Schrödinger, determinando “todas as integrais”).

Todos os métodos de cálculos moleculares possuem características próprias e se adequam a um determinado tipo de aplicação, isto é, não temos uma metodologia única que abranja todo o tratamento de cálculos moleculares. Neste trabalho utilizamos dois métodos semi-empíricos dentre os mais utilizados atualmente. Dentre as várias metodologias semi-empíricas existentes as diferenças principais são, a grosso modo, as formas de aproximação e parametrização estabelecidas. Com o grande número de trabalhos produzidos sobre os métodos de cálculos moleculares, não estarão descritos neste trabalho detalhes dos tratamentos matemáticos dos métodos utilizados. Uma

descrição completa dos métodos poderá ser encontrada nas referências^{[12],[13],[14],[15],[16],[17]}.

Em todo o trabalho desenvolvido utilizamos os métodos semi-empíricos AM1 (*Austin Method One*)^[18] e PM3 (*Parametric Method 3*)^[19] para realização da busca conformacional, obtenção dos dados geométricos, e dos parâmetros da MIE. A opção de usar um cálculo semi-empírico é pelo fato de termos bons resultados e também por estar compatível com os recursos computacionais disponíveis no momento; o tratamento com um método *ab initio* levaria mais tempo que o disponível para a realização deste trabalho. Para os conjuntos de moléculas estudados houve a necessidade de se realizar uma busca conformacional detalhada, pelo fato de serem moléculas que surgiram recentemente e não existirem dados sobre as suas estruturas disponíveis na literatura. Utilizamos os dois métodos semi-empíricos citados acima na busca da melhor estrutura geométrica dos compostos. Esta abordagem nos permitiu comparar os resultados dos métodos e ao mesmo tempo verificar a dependência da análise da MIE.

O processo da busca conformacional consiste em realizar o cálculo molecular para cada conformero gerado a partir de uma dada estrutura inicial. Cada conformero é caracterizado pelas diversas posições que os átomos que constituem a molécula podem ocupar. Como exemplo temos a figura 3.1 e a figura 3.2:

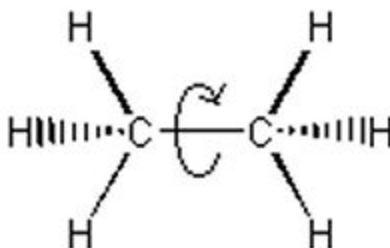


Figura 3.1: Exemplo de um tipo de grau de liberdade do composto C_2H_6 , rotação em torno da ligação que une os dois carbonos.

Na figura 3.1 temos uma molécula dimetila (C_2H_6), a ligação que une os dois carbonos é uma ligação simples (longa) e portanto vemos que pode existir uma rotação de 360° de um grupo CH_3 com relação ao outro grupo CH_3 (ângulo rotacional ou diedral). Na figura 3.2 temos um fragmento de uma molécula representada pelo grupo hidroxila, neste caso o hidrogênio possui dois graus de liberdade, um que varia de 0° a 180° (posição do hidrogênio, ângulo de ligação) e outro que varia de 0° a 360° (ligação entre o carbono e o oxigênio). Em ambos os casos para determinarmos qual a posição atômica

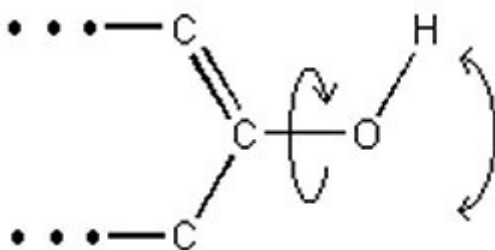


Figura 3.2: Exemplo dos graus de liberdade do grupo hidroxila, rotações em torno da ligação que une o carbono e a mobilidade do oxigênio.

do conjunto que resulta a maior estabilidade, é necessário realizar o cálculo molecular para várias posições atômicas possíveis. Utilizamos o valor do calor de formação molecular para determinar comparativamente qual conformero possui a maior estabilidade.

Esta parte do trabalho foi a que requereu de um maior esforço computacional devido ao grande número de compostos analisados. No total foram 136 compostos entre as esterilquinolinas e as tetraciclinas apresentando uma média de 150 conformeros por composto.

3.2 Metodologia dos Índices Eletrônicos (MIE)

A MIE foi desenvolvida inicialmente com o objetivo principal de distinguir qualitativamente os compostos, Hidrocarbonetos Aromáticos Policíclicos (PAH's)^[20] entre ativos e inativos.

Os PAH's são moléculas compostas por anéis aromáticos dispostos em uma forma linear ou em aglomerados. São indutores do câncer e estão presentes, por exemplo, no cigarro e gases produzidos por automóveis. Para os compostos dos PAH's a MIE aplicada na região molecular de maior ordem de ligação apresentou resultados de $\sim 80\%$ de acerto na preidção de moléculas com ou sem atividade carcinogênica. Depois desta aplicação a MIE foi aplicada para diversos tipos de compostos orgânicos que hoje já somam mais de dois mil compostos analisados^{[21], [22], [23], [24], [25], [26], [27]}.

A MIE é baseada no conceito de densidade local de estados (LDOS, Local Density of States) e na diferença entre as energias dos orbitais de fronteira. A densidade local de estados é a soma dos quadrados dos coeficientes dos orbitais previamente escolhidos. Para hamiltonianos baseados na aproximação da combinação linear dos orbitais atômicos (LCAO) e envolvendo vários orbitais ou regiões moleculares a LDOS é obtida somando-se as contribuições dos orbitais atômicos:

$$LDOS(E_i) = 2 \sum_{m=n_i}^{n_f} |C_{mi}|^2 \quad (3.1)$$

onde:

- E_i é a energia do orbital molecular;
- n_i e n_f são os sítios atômicos iniciais e finais respectivamente;
- C_{mi} é a contribuição do i -ésimo orbital atômico no m -ésimo sítio;
- o fator 2 vem do princípio de exclusão de Pauli (máximo de 2 elétrons por nível eletrônico).

O primeiro parâmetro utilizado pela MIE (η (eta)) é definido como a diferença dos valores das densidades locais de estado, entre dois estados de fronteira que podem ser os dois últimos orbitais ocupados [HOMO (Highest occupied molecular orbital), HOMO-1] ou os dois primeiros estados desocupados [LUMO (Lowest unoccupied molecular orbital), LUMO+1]:

$$\eta = 2 * \sum_{m=n_i}^{n_f} (C_{mn_i1})^2 - (C_{mn_i2})^2 \quad (3.2)$$

Para os últimos orbitais ocupados a equação acima se resume a

$$\eta H = LDOS(E_{HOMO}) - LDOS(E_{HOMO-1}). \quad (3.3)$$

Para os primeiros orbitais desocupados

$$\eta L = LDOS(E_{LUMO+1}) - LDOS(E_{LUMO}). \quad (3.4)$$

O segundo parâmetro da MIE, denominado Δ , é definido como a diferença de energia entre dois orbitais moleculares. A forma geral é dada pela equação abaixo:

$$\Delta = E_{nível1} - E_{nível2} \quad (3.5)$$

Para os dois últimos orbitais moleculares ocupados a equação é dada por

$$\Delta H = E_{HOMO} - E_{HOMO-1} \quad (3.6)$$

e para os dois primeiros orbitais desocupados a equação é dada por

$$\Delta L = E_{LUMO+1} - E_{LUMO}. \quad (3.7)$$

Nos gráficos abaixo podemos visualizar as equações descritas acima e também podemos verificar, para efeito de comparação, os gráficos de DOS (Density of States) e LDOS. Novamente os parâmetros utilizados pela MIE estão representados por Δ e η , sendo um a diferença dos orbitais de fronteira no eixo das abscissas e o outro a diferença na contribuição dos orbitais de fronteira no eixo das ordenadas respectivamente.

Ambos os parâmetros utilizados pela MIE, Δ e η , permitem uma fácil identificação dos padrões correlacionados com a atividade biológica. Em resumo, o índice Δ é o mais simples, sendo a diferença direta entre os orbitais de fronteira; o índice η pode ser encontrado através da LDOS de uma busca sistemática sobre as várias regiões moleculares, ou através das regiões mais relevantes do ponto de vista químico ou biológico.

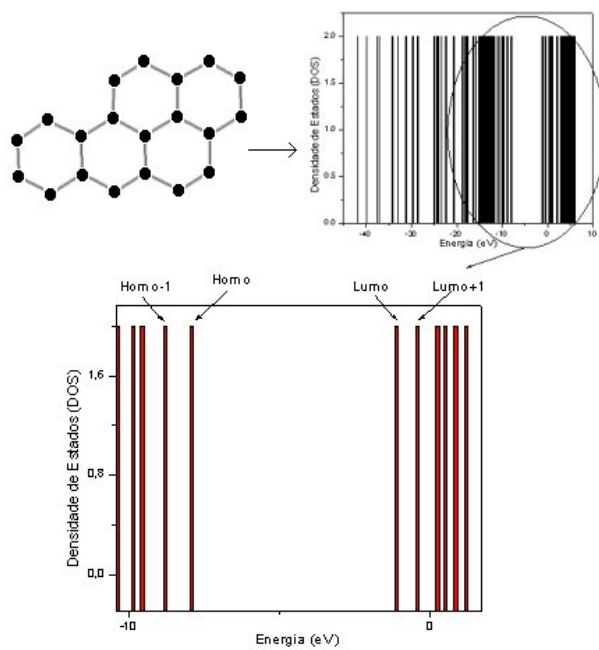


Figura 3.3: Gráfico de Densidade de Estados Total - DOS

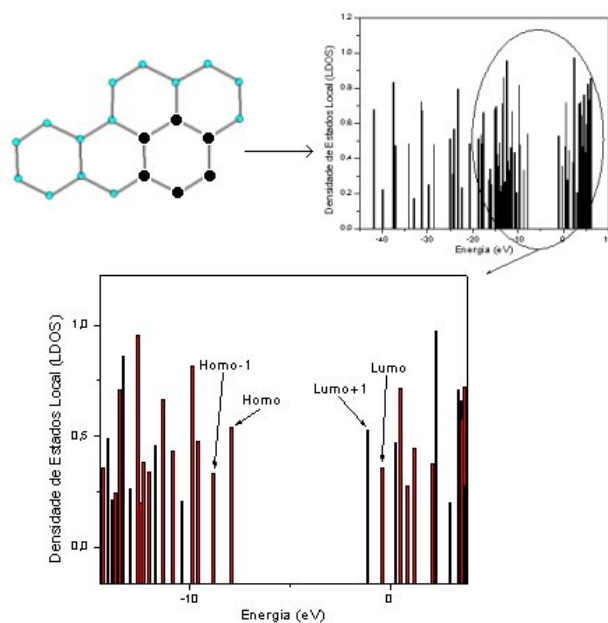


Figura 3.4: Gráfico de Densidade de Estados Local - LDOS, base de um dos índices da MIE

3.3 Métodos de Quimiometria

A quimiometria é uma área que se refere à aplicação de métodos estatísticos e matemáticos, assim como aqueles baseados em lógica matemática, a problemas de origem química.

Com a sofisticação crescente das técnicas instrumentais, impulsionadas pelo desenvolvimento de microprocessadores e microcomputadores no laboratório químico, tornaram-se necessários tratamentos de dados mais complexos do ponto de vista matemático e estatístico. Muitas vezes essas técnicas instrumentais não permitem uma informação direta sobre problema, mas produzem uma grande quantidade de dados que precisam ser tratados para elaboração e interpretação de modelos matemáticos. Nesses sistemas, a conversão da resposta instrumental no dado químico de interesse pode requerer utilização de técnicas de estatística multivariada, álgebra matricial e análise numérica. Essas técnicas se constituem no momento em específicas ferramentas para interpretação de dados e para a extração do máximo de informação sobre os sistemas analisados.

No sentido de ter uma ferramenta para análise de resultados muitas áreas utilizam técnicas de quimiometria. Dentre as técnicas mais utilizadas podemos citar: *Principal Component Analysis* (PCA); *Hierarchy Cluster Analysis* (HCA); *Principal Component Regression* (PCR); *Partial Least Squares* (PLS); *Kth Nearest Neighbor* (KNN); *Soft Independent Modeling Of Class Analogy* (SIMCA). Alguns desses métodos podem ser encontrados no programa computacional Pirouette, comercializado pela empresa Infometrix^[28].

Para análise de nossos resultados utilizamos somente as metodologias: análise de componentes principais (PCA) e análise hierárquica de clusters (HCA), que estão disponibilizados no programa computacional *Einsight*^[29].

3.3.1 Análise de Componentes Principais (PCA)

A PCA utiliza os dados decompondo as informações iniciais em duas matrizes denominadas *loadings* e *scores*, sendo a primeira relacionada com as variáveis dos dados e a segunda com as amostras. A matriz original de dados é organizada de acordo com a tabela 3.1.

Onde as linhas representam as amostras que podem ser, por exemplo, moléculas ou cristais, as colunas representam diferentes qualidades das amostras, no caso de uma análise de espectroscopia será o espectro do material, ou, no caso de uma análise de

Amostras	Var01	Var02	Var03	Var04	Var05	Var06	Var07	Var08	Var09
01	I	II	III	IV	V	VI	VII	VIII	IX
02	I	II	III	IV	V	VI	VII	VIII	IX
03	I	II	III	IV	V	VI	VII	VIII	IX
04	I	II	III	IV	V	VI	VII	VIII	IX

Tabela 3.1: Matriz de dados hipotética de entrada para PCA

SAR ou QSAR serão dados de eletronegatividade, massa, densidade, etc. Essa matriz é denominada matematicamente como matriz X .

A decomposição é feita como descrita pela equação :

$$X = T * L^T \quad (3.8)$$

As colunas de L são as componentes principais, os novos fatores são combinações lineares das variáveis originais. A primeira coluna de L indica então quanto cada variável original contribui para a primeira componente principal, PC1; a segunda coluna para a segunda componente principal, PC2; assim até a coluna n para a PC n . A matriz T é uma projeção das amostras nos eixos definidos pelas variáveis, assim para cada coluna de L as amostras terão novas coordenadas.

Uma forma mais simples de entender o conceito da PCA é imaginando um sistema bidimensional em que os eixos são representados pelos descritores (as variáveis referentes às amostras) e os pontos representados são as amostras. A PCA realiza então uma rotação dos eixos cartesianos até que antigo eixo horizontal (agora representado por PC1) contenha a maior informação possível sobre a distribuição das amostras, a maior variância, e o antigo eixo vertical (agora representado pela PC2) a segunda maior. Em seguida os descritores e as amostras são reescritos, através de uma combinação linear, em termos de PC1 e PC2 gerando assim os gráficos dos loadings e scores. Esta foi uma descrição para um sistema bidimensional, se tivermos um sistema n -dimensional a PCA realizará rotações no espaço n -dimensional até encontrar a maior variância na primeira componente principal PC1.

Após o tratamento matemático da PCA teremos o maior efeito para análise dos dados juntamente com a visualização das matrizes T e L^T , onde é possível verificar em até três dimensões os resultados do PCA. Abaixo exemplificamos resultados PCA^[30], um gráfico bi-dimensional de amostras (scores) (figura 3.5) e um gráfico de variáveis

(loadings) (figura 3.6):

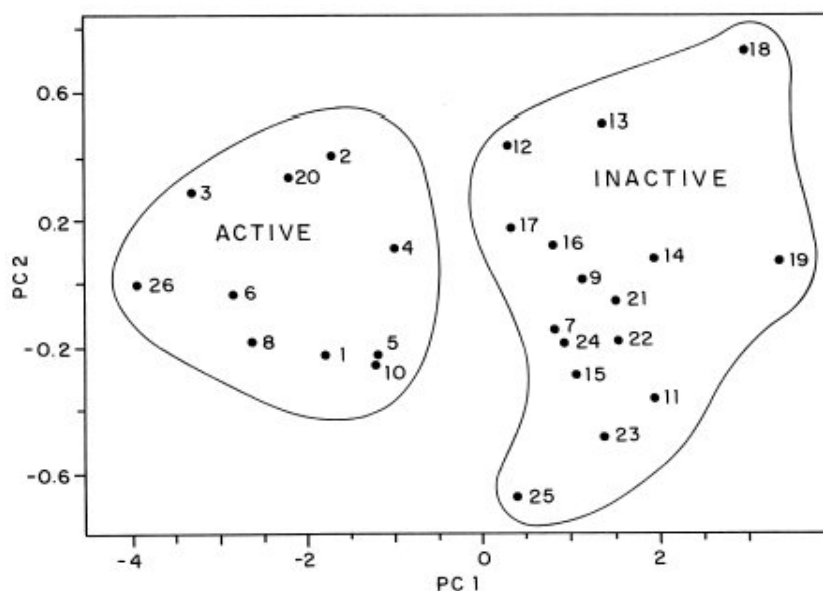


Figura 3.5: Exemplo de um cálculo com PCA, representação dos compostos em termos das componentes principais (PC's) - Score.

Como dito anteriormente o gráfico de variáveis está totalmente correlacionado com o gráfico de amostras, pois a manipulação dos dados em um dos gráficos afetará o outro, e vice-versa. O gráfico de *loadings* mostra as variáveis responsáveis pela separação nas amostras, representada pelo gráfico dos *scores*. Esta ferramenta é muito útil para selecionarmos quais são os descritores mais relevantes para verificação de padrões existentes ou não em um determinado conjunto de amostra.

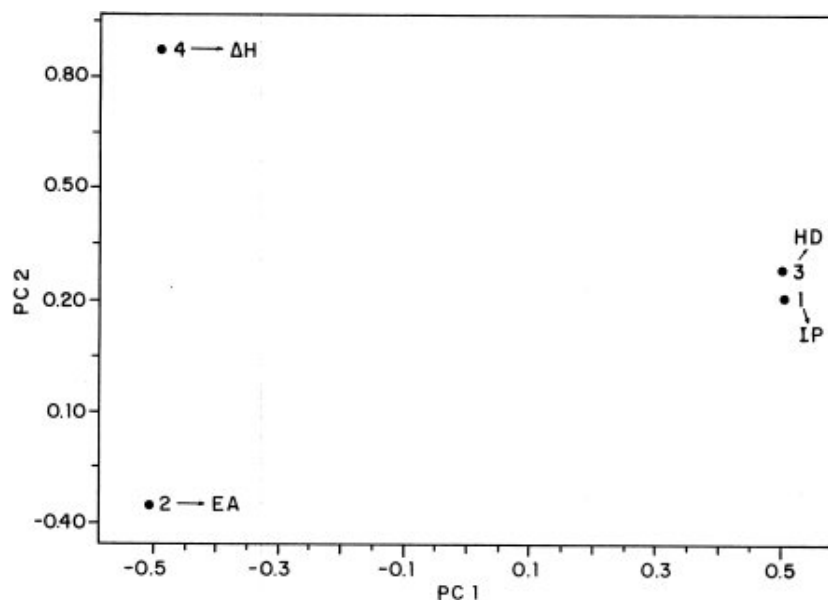


Figura 3.6: Exemplo de um cálculo com PCA, representação dos descritores em termos das componentes principais (PC's) - Loadings

3.3.2 Análise Hierárquica de Agrupamentos (HCA)

A partir dos gráficos obtidos da PCA (*loadings* e *scores*), a HCA verifica a semelhança entre amostras ou variáveis pela distância entre eles no espaço das componentes principais ($PC_x, x = 1, 2, \dots$) entre eles. As distâncias entre pares de amostras são calculadas e comparadas. Quando a distância entre amostras é relativamente pequena elas são consideradas semelhantes, e quando a distância é considerada grande as amostras são consideradas dissimilares. O índice de similaridade é dado pela equação abaixo.

$$similaridade_{ab} = 1 - \frac{d_{ab}}{d_{max}} \quad (3.9)$$

Se tivermos valores próximos de 1 temos amostras muito semelhantes e se tivermos valores próximos de 0 teremos amostras muito diferentes.

O valor d_{max} é a maior distância calculada entre as amostras ou variáveis do sistema. O valor d_{ab} pode ser calculado de várias maneiras dependendo da consideração feita sobre a amostra vizinha, que pode estar isolada ou em um agrupamento já pré-definido.

Os resultados da HCA são apresentados em forma de dendrograma como na figura 3.7:

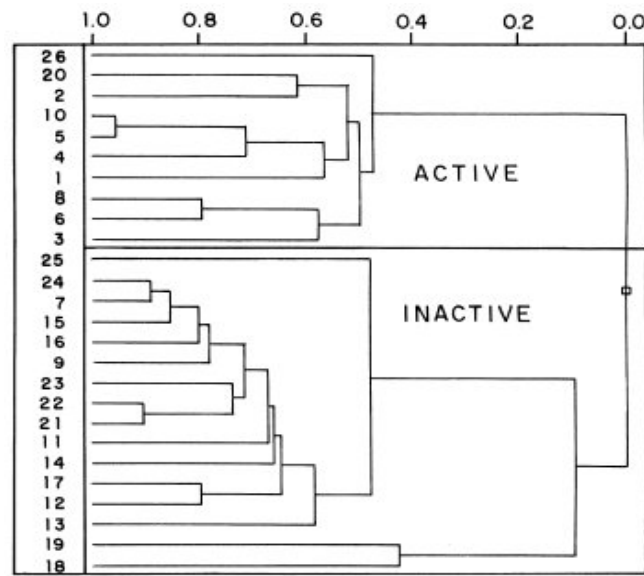


Figura 3.7: Exemplo de um cálculo com HCA, cálculo de similaridade dado pela distância entre vizinhos.

3.4 Redes Neurais

As redes neurais artificiais, ou comumente ditas redes neurais, tiveram como base a idéia das redes dos neurônios do cérebro humano. Os primeiros indícios de trabalhos que abordaram esta idéia ocorreram no final do século XIX com os cientistas Hermann von Helmholtz, Ernst Mach e Ivan Pavlov onde o trabalho enfatizava as teorias, entre as mais comuns, de aprendizagem, da visão e condicionamento. Foi somente no início da década de 40 do século passado que as redes neurais artificiais apresentaram conectividade com algoritmos matemáticos. McCulloch e Walter Pitts^[31] demonstraram que neurônios artificiais poderiam, em princípio, realizar qualquer função aritmética ou lógica. Após muito esforço dos pesquisadores a primeira evidência prática do funcionamento das redes neurais aconteceu no final da década de 50 por Frank Rosenblatt^[32]. Ele demonstrou que uma rede "perceptron" juntamente com regras de aprendizado realizava reconhecimentos de padrões. O trabalho de Rosenblatt gerou muita euforia junto aos pesquisadores da época e embora o esquema proposto por Rosenblatt só resolvesse somente alguns casos particulares.

Um dos grandes interesses da época era também utilizar a arquitetura das redes neurais (de inspiração biológica) para o desenvolvimento de um *hardware* com as mesmas características, levando em conta que o cérebro humano, numa visão tecnológica,

é um grande computador de processamento paralelo, não linear e de extrema complexidade. Em paralelo com o desenvolvimento das pesquisas das redes neurais estava o desenvolvimento do computador digital.

Durante as décadas de 50, 60 e 70 do século XX foi quase totalmente, se não totalmente, abandonado o desenvolvimento de um *hardware* baseado na arquitetura das redes neurais do cérebro humano. O motivo foi que o desenvolvimento do *hardware* baseado na tecnologia digital estava bem à frente daqueles baseados no cérebro humano. Isto foi uma grande desmotivação para que as pesquisas em massa das redes neurais artificiais aumentassem o seu desenvolvimento.

Mesmo com o desinteresse parcial, os trabalhos com as redes neurais continuavam sendo explorados por seus idealizadores Rosenblatt e também por Bernard Widrow e Ted Hoff^[33]. No início da década de 80 com a implementação de modelos de mecânica estatística pelo físico John Hopfield^[34] e do algoritmo *backpropagation* por David Rumelhart^[35] e James McClelland é que as redes neurais artificiais tiveram caráter generalizado para resolver diversos tipos de problemas. Desde então o desenvolvimento das redes neurais foi bastante intenso a ponto de ser implementado em diversas áreas, como^[36],^[37],^[38]: aeroespacial (piloto automático de aviões), automotivo (sistemas de direção automática), bancos (checagem de leitura de documentos), navegação (sonares e radares), eletrônica (visão eletrônica, sintetizador de voz, controle de processamento), produção (diagnóstico de processamento, teste de bebidas, análise da qualidade de papel), medicina (análise de câncer de mama), telecomunicações (automatização de serviços de informação), entre outros.

Antes da descrição de um neurônio artificial é muito interessante ter em mente a fonte de inspiração das redes neurais artificiais. Baseado em estudos do cérebro humano (figura 3.8), as redes neurais compõem um conjunto de aproximadamente 10^{11} elementos altamente interconectados, cada neurônio faz em torno de 10^4 ligações com os demais vizinhos^[36]. Um neurônio biológico (figura 3.9) é constituído basicamente por: membrana celular, citoplasma, núcleo celular (soma), os dendritos e axônio, onde as entradas dos pulsos elétricos são feitos através dos dendritos e os encaminhamentos desses pulsos são feitos pelo axônio. Esse último tem a sua extremidade, aquela oposta ao corpo celular, totalmente ramificada como uma árvore que faz conexão com os dendritos de outros neurônios.

As redes neurais artificiais foram desenvolvidas a partir de neurônios artificiais que



Figura 3.8: Cérebro Humano, contém aproximadamente 10^{11} neurônios altamente interconectados.

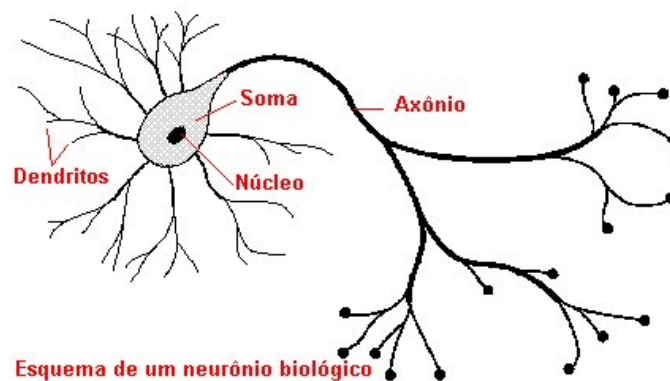


Figura 3.9: Esquema de um neurônio biológico, fonte de inspiração para a construção dos neurônios artificiais, cada neurônio biológico faz aproximadamente 10^4 ligações.

possuem algumas das características dos neurônios biológicos. As partes de um neurônio artificial que fazem analogia com um neurônio biológico são: as entradas (input) são correspondentes aos dendritos, à função de ativação é correspondente ao corpo celular, a saída (output) é correspondente ao axônio e o elemento responsável pela comunicação entre os neurônios artificiais é denominado peso, que são as intensidades das conexões (figura 3.10).

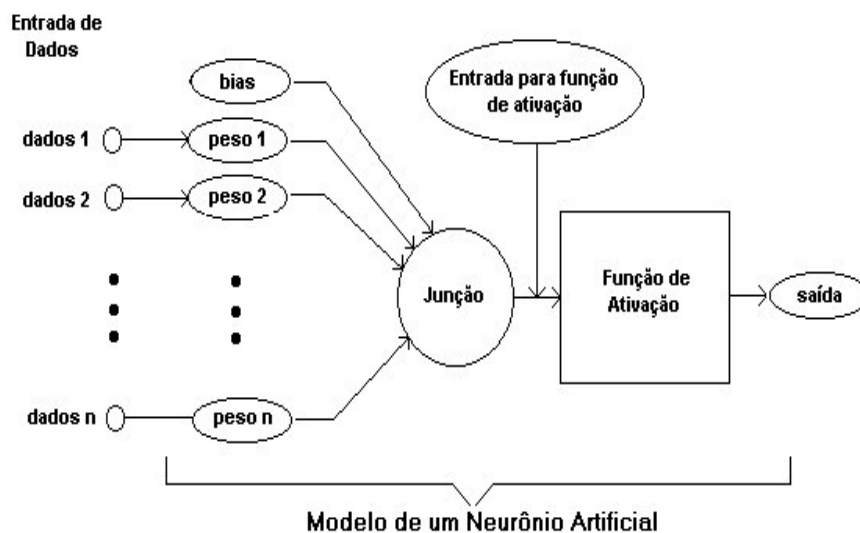


Figura 3.10: Esquema de um neurônio artificial

Uma rede neural artificial pode ser construída com m camadas contendo n neurônios, dependendo do tipo de aplicação a que se designam as redes neurais. Dentre as diversas arquiteturas das redes neurais artificiais, nos restringiremos à rede neural do tipo *perceptron* com três camadas, com aprendizado supervisionado e utilização do algoritmo *backpropagation*. (figura 3.11), implementados no pacote computacional, PSDD (*perceptron simulator for drug design*), que foi elaborado para o desenvolvimento de drogas^[39] e reconhecimento de padrões de moléculas com atividade biológica.

Cada círculo se refere a um neurônio, exemplificado pela figura 3.11. As intensidades das ligações entre eles (representados pelas linhas cheias ligando os neurônios) são dados pelo “**peso**”. Os círculos negros representam a primeira camada, (de entrada) que no caso de QSAR (*quantitative structure activity relationship*) ou SAR (*structure activity relationship*) são os descritores associados à amostra. Os círculos com listras horizontais representam a camada intermediária (*hidden layer*), que é adicionada para melhorar

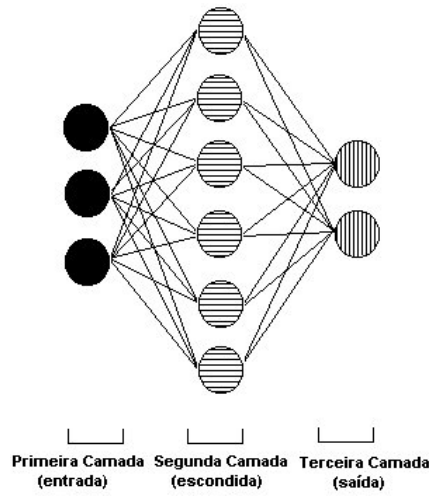


Figura 3.11: Redes Neurais Artificiais - rede neural tipo perceptron, uma das arquiteturas mais simples das redes neurais artificiais

o trabalho de classificação ou quantificação das redes neurais. Os círculos com listras verticais representam a saída de dados após terem passado pela primeira e segunda camada.

Para esta arquitetura os dados de entrada são reescalados para valores entre 0 e 1, pela equação abaixo.

$$\tilde{x} = \frac{x_i - x_{\min} + 0,1}{x_{\max} - x_{\min} + 0,1} \quad (3.10)$$

onde $x_{\min}$ e $x_{\max}$ são os valores mínimos e máximos dos dados. Depois de reescalados veremos que os menores valores não são 0 e sim 0,1, isto é feito pois cada valor de entrada é multiplicado pelo fator “**peso**” conseqüentemente se o valor for zero estaríamos eliminando um dado que poderia ou não ser importante para análise.

A função de ativação utilizada é a do tipo sigmóide, ao verificar os dados de saída sejam esses dados de saída da primeira, segunda ou terceira camada terá a forma dada pela equação abaixo.

$$O_j = \frac{1}{1 + e^{-\alpha \cdot y_j}} \equiv f(y_j) \quad (3.11)$$

e

$$y_j = (\sum W_{ij} x_i) - \theta_j \quad (3.12)$$

x_i é o valor do neurônio na camada $m - 1$; W_{ij} é o elemento da matriz peso (igual ao

fator peso citado anteriormente) e expressa o peso da conexão entre os neurônios i e j ; θ_j é o valor característico do neurônio j (é o parâmetro que expressa a não linearidade do neurônio). Os índices i e j nos dizem com qual neurônio da camada posterior o neurônio de uma camada está se comunicando, por exemplo (vide figura 3.11): o primeiro neurônio (de cima para baixo) da primeira camada pode se comunicar com todos os seis neurônios da segunda camada e portanto o índice j varia de 1 a 6.

Para nos familiarizarmos mais com os nomes no espaço das redes neurais os dados de entrada são chamados de padrão de entrada e os dados de saída são chamados de padrão de saída, ambos são representados como vetores. A dinâmica de inserção dos valores do padrão de entrada que após percorrer as camadas das redes neurais, fornece o padrão de saída, é denominado padrão de treinamento (t_{ij} , valor do neurônio da última camada). O treinamento é realizado de acordo com as equações abaixo.

$$W_{ij} = -d_j x_i \varepsilon \quad (3.13)$$

$$d_j^{(3)} = (O_j - t_j) f'(y_j) \quad (3.14)$$

$$d_j = (W'_{j1} d'_1) f'(y_j) \quad (3.15)$$

ε (parâmetro de ajuste da rede denominado EPSIRON) é o parâmetro que determina o acréscimo para a troca na correção entre os recursivos ciclos. Na equação 3.14 o índice sobrescrito indica que ela só será utilizada na última camada e a equação 3.15 é utilizada nas outras camadas. Após ocorrido o primeiro ciclo os novos valores de W'_{j1} e d'_1 de uma camada m passam a ser os antigos valores de W_{j1} e d_1 da camada $m + 1$. A função $f'(y_j)$ (novo valor do neurônio j) é dada pela equação abaixo.

$$f'(y_j) = f(y_j)[1 - f(y_j)]\alpha \quad (3.16)$$

Aqui ambos os índices ε e α são ajustados independentemente da camada considerada. Com estes recursivos ciclos a minimização do erro é feita até que haja uma pequena diferença entre o padrão de saída gerada pela rede com o padrão de saída desejado inicialmente, o erro é dado pela equação abaixo:

$$E = \sum (O_j - t_j)^2 \quad (3.17)$$

Fazendo uma descrição simplista um ciclo das redes neurais funciona da seguinte maneira: dada a entrada padrão e a saída padrão, os dados de entrada previamente reescalados são multiplicados pelos fatores peso (que são inicialmente gerados aleatoriamente) em seguida discriminados pela função de ativação. Com isso são geradas saídas que servirão de entrada para a segunda camada. Assim como na primeira a situação ela é repetida na segunda e terceira camadas e com as saídas da terceira camada é verificada a compatibilidade com a saída padrão. Se o erro não estiver dentro do estabelecido os fatores pesos (W_{ij}) são modificados de acordo com as equações 3.13, 3.14 e 3.15 recomeçando então o segundo ciclo. Isto ocorre até que o erro seja minimizado dentro da quantidade de ciclo desejado. Esta simples descrição é justamente o algoritmo *backpropagation* com treinamento supervisionado.

Em um exemplo da utilização desta ferramenta os autores Ayoma, Suzuki e Ichikawa^[40], em uma análise de *SAR*, utilizam um conjunto de 38 amostras e 7 descritores. Eles utilizaram 25 amostras no conjunto de treinamento e 13 amostras no conjunto de predição, em ambos conjuntos com amostras do tipo *endo* e tipo *exo*. A saída padrão para amostras do tipo *exo* é 1 0 0 1 e para o tipo *endo* é 0 1 1 0, os resultados do treinamento e da predição podem ser vistos na tabela 3.2.

No resultado do treinamento observamos que alguns valores são absolutos (1,00 ou 0,00), ou são próximos do absoluto (0,99 ou 0,08), isto indica que o treinamento ficou bem próximo da saída padrão, que foi inicialmente desejada. Por exemplo, para a amostra 1 tipo *exo*, a saída padrão é do tipo 1 0 0 1 e o valor obtido pelo treinamento da rede foi 0,97 0,02 0,02 0,97. Isto demonstra um bom resultado do treinamento da rede. No conjunto de predição os resultados obtidos apresentaram-se próximos da saída padrão e o índice de acerto foi estatisticamente muito bom (12/13) $\sim 93\%$ de acerto, apenas a amostra 31 ficou fora do padrão inicialmente estabelecido.

Observamos no trabalho acima que as redes neurais são uma ferramenta muito eficiente e garantem uma boa porcentagem de acerto em relação a outros métodos utilizados na área de reconhecimento de padrões. Esta rede em particular possui uma das arquiteturas mais simples comparadas com as existentes atualmente e os índices de acertos em outros trabalhos que utilizam redes tipo perceptron também são bastante consideráveis^{[41], [42]}.

Treinamento												
No.	endo/exo	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	C ₇	Saída padrão			
1	exo	6.7	6.7	10.1	0.5	0.2	-1.1	-3.7	0.97	0.02	0.02	0.97
2	exo	8.9	25.3	12.4	-0.4	-1.2	-3.1	-4.4	1.00	0.00	0.00	1.00
3	exo	7.7	44.3	12.3	-1.0	-1.3	-5.2	-4.4	1.00	0.00	0.00	1.00
4	exo	4.6	16.7	4.4	-0.2	-0.3	-1.0	-1.8	0.94	0.06	0.06	0.94
5	exo	1.8	15.1	4.4	-0.2	0.2	-0.7	-3.3	0.99	0.01	0.01	0.94
6	exo	5.7	3.2	2.6	-0.5	-0.4	0.7	-3.5	0.99	0.01	0.01	0.94
7	exo	6.1	5.9	10.6	0.6	0.2	0.2	-3.7	0.99	0.01	0.01	0.99
8	exo	6.5	6.3	10.4	0.3	-0.8	-0.1	-3.5	0.99	0.01	0.01	0.99
9	exo	6.5	7.5	9.5	0.5	1.7	0.7	-3.8	0.97	0.03	0.03	0.97
10	exo	7.8	47.0	11.7	-1.3	3.9	-2.7	-3.2	0.99	0.01	0.01	0.99
11	exo	6.9	6.4	10.1	0.7	-1.2	0.1	-3.9	1.00	0.00	0.00	1.00
12	exo	5.6	4.9	7.0	0.2	-1.1	0.2	-3.9	1.00	0.00	0.00	1.00
13	exo	2.5	42.5	11.9	-0.8	-1.1	-2.4	1.4	0.95	0.05	0.05	0.95
14	endo	5.4	4.5	10.6	1.4	0.5	-7.7	0.2	0.00	1.00	1.00	0.00
15	endo	6.8	23.3	10.5	1.2	0.6	-9.5	0.3	0.00	1.00	1.00	0.00
16	endo	6.3	42.4	9.5	0.9	0.2	-9.7	-0.9	0.03	0.96	0.96	0.03
17	endo	4.2	16.2	2.1	0.9	-0.6	-4.8	1.9	0.00	1.00	1.00	0.00
18	endo	1.7	12.8	4.0	0.4	0.2	-7.2	1.4	0.00	1.00	1.00	0.00
19	endo	4.7	3.1	2.2	0.3	1.3	-6.5	-0.6	0.00	1.00	1.00	0.00
20	endo	4.7	5.3	9.2	1.3	-0.4	-6.5	1.4	0.00	1.00	1.00	0.00
21	endo	4.6	11.5	8.9	-0.1	0.8	0.4	1.8	0.08	0.93	0.93	0.08
22	endo	5.6	7.5	8.7	1.4	1.7	-3.0	1.7	0.00	1.00	1.00	0.00
23	endo	7.1	47.8	13.3	2.2	3.6	-3.4	0.3	0.01	0.99	0.99	0.01
24	endo	4.1	4.2	7.0	0.7	0.5	-7.4	0.0	0.00	1.00	1.00	0.00
25	endo	3.2	40.2	10.4	-0.5	0.0	-10.3	3.1	0.00	1.00	1.00	0.00

Predição													
No.	endo/exo	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	C ₇	Saída padrão				Decisão
26	exo	5.5	1.0	6.3	-0.3	-1.5	-1.6	-1.3	0.83	0.18	0.18	0.83	exo
27	endo	3.4	0.1	5.5	0.2	-0.7	-4.9	0.0	0.02	0.98	0.98	0.02	endo
28	exo	5.1	16.4	4.2	-0.4	-1.1	-1.4	-2.1	0.98	0.02	0.02	0.98	exo
29	endo	4.0	15.9	2.2	0.7	-0.7	-5.0	1.7	0.00	1.00	1.00	0.00	endo
30	exo	6.6	7.0	10.1	0.2	-1.2	-0.5	-3.7	1.00	0.00	0.00	1.00	exo
31	endo	6.0	8.4	11.2	-0.1	0.7	-1.5	-1.6	0.74	0.24	0.24	0.74	exo
32	exo	6.3	7.2	9.8	0.7	-0.1	0.8	-3.5	0.99	0.01	0.01	0.99	exo
33	endo	5.1	4.8	8.4	1.1	-0.1	-7.3	1.6	0.00	1.00	1.00	0.00	endo
34	exo	1.9	17.1	5.2	-0.1	0.9	0.9	-3.4	0.99	0.01	0.01	0.99	exo
35	endo	2.3	18.3	5.0	0.3	1.3	-2.9	-1.4	0.01	0.99	0.99	0.01	endo
36	endo	5.1	4.0	8.4	1.1	0.2	-7.7	1.6	0.00	1.00	1.00	0.00	endo
37	exo	2.9	30.3	13.4	-0.5	-2.1	-0.7	2.0	0.94	0.06	0.06	0.94	exo
38	endo	3.7	29.8	10.8	-1.6	-1.1	-9.0	2.2	0.04	0.96	0.96	0.04	endo

Tabela 3.2: Exemplo do tratamento de dados com a rede neural, com o número de descritores utilizados, os padrões de treinamentos e os padrões obtidos pela predição.

Resultados - Esterilquinolinas

Todos os derivados de esterilquinolinas indicadas na tabela 2.1 foram submetidos a uma busca conformacional detalhada, verificando possíveis variações em seus comprimentos e ângulos de ligações e ângulos diedrais. Utilizamos os métodos semi-empíricos, como dito anteriormente, AM1 e PM3 para realizar as otimizações juntamente com os programas Chem2Pac (desenvolvido pelo nosso grupo de pesquisa), Spartan^[43] (programa comercial), CAChe^[44] (programa comercial), todos eles contendo os dois métodos. A opção de utilizar esses diferentes programas foi para obter uma boa estrutura geométrica final sobre efeito de comparação das moléculas estudadas. Utilizamos também o programa HyperChem^[45] que fornece boa interface gráfica para construção das moléculas.

4.1 Busca conformacional

Com o programa Chem2Pac realizamos a busca conformacional dos diedros aos pares com passos de 10°, já com os demais programas foi possível realizar a busca conformacional com o mesmo passo e com todos os diedros que fornecessem possíveis rotações simultaneamente.

Na figura 4.1 (molécula no.18) podemos ver os vários graus de liberdade das moléculas de esterilquinolinas. Todas as ligações simples (ligações longas e proporcionam rotações em torno do seu próprio eixo, ao contrário das ligações duplas que não permitem

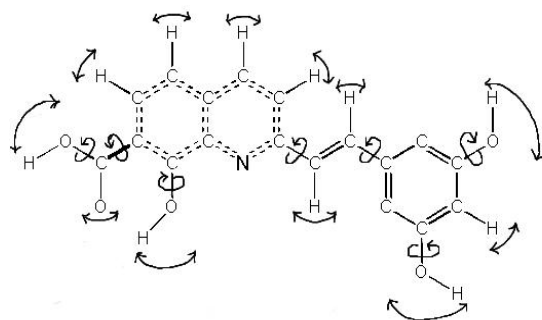


Figura 4.1: Exemplos dos graus de liberdade, representados pelas ligações simples (traço simples que une dois átomos), de um composto de esterilquinolina.

rotações em torno do seu próprio eixo) dos grupos laterais foram sujeitas à análise, o que demandou um grande custo computacional para o grande número de conformémeros gerados.

Com a busca conformacional realizada obtivemos duas conformações geométricas que forneceram mínimos próximos valor de calor de formação. As geometrias das figuras 4.2 e 4.3 apresentaram estas conformações.

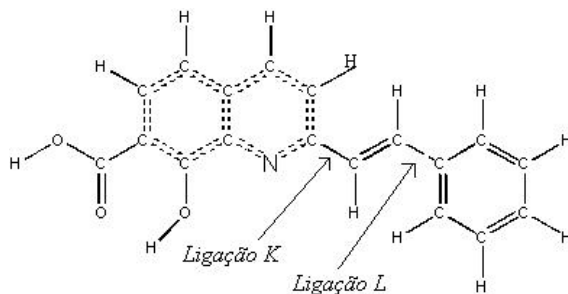


Figura 4.2: Um derivado que possui um mínimo de calor de formação

As moléculas se diferenciam apenas nas rotações em torno das ligações K e L . No conformémero B, o radical que está à direita da ligação K está rotacionado de 180° , em relação ao conformémero A. Isto aconteceu para todas as moléculas, com exceção das moléculas números 01, 23, 24 e 25 porque não possuem a estrutura das demais moléculas. Dentre as duas moléculas utilizamos o conformémero tipo A. As diferenças de valores no calor de formação entres os conformémeros tipo A e B são muito pequenas (milésimos de kcal/mol). Para ilustrar esta escolha fizemos um mapa do calor de formação versus variação do diedro (figura 4.3):

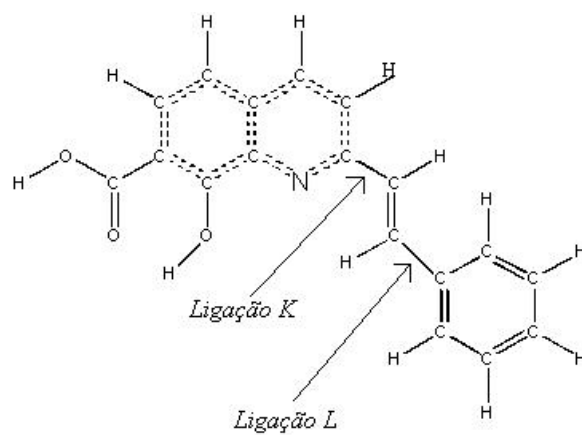


Figura 4.3: Outro derivado que possui um mínimo de calor de formação

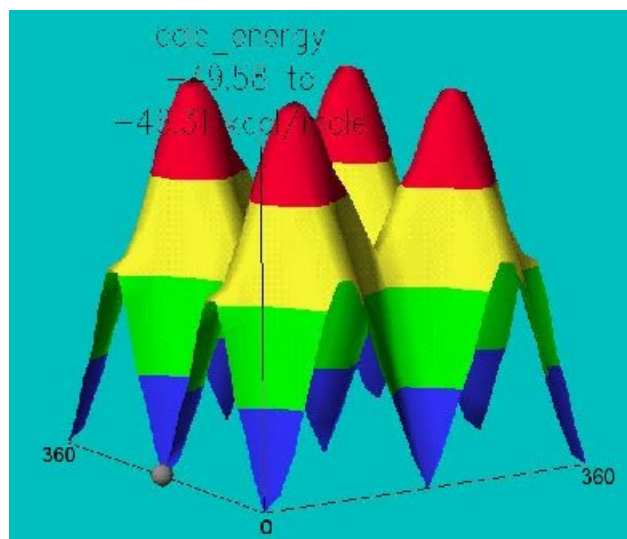


Figura 4.4: Mapa do Calor de Formação na busca conformacional

Em 0° e 360° a forma geométrica é a do confôrmere A e em 180° temos a forma geométrica do confôrmere B. O azul no gráfico (parte inferior) indica o local de menor calor de formação e a região vermelha indica o maior calor de formação (parte superior).

Na tabela 4.1 indicamos os valores de calor de formação para os resultados de otimização de geometria dos compostos de esterilquinolinas. A tabela 4.1 apresenta os resultados somente para os confôrmeros de menor calor de formação de cada derivado de esterilquinolina. Todas as moléculas que apresentaram os valores de energia acima possuem uma estrutura quase planar, ou totalmente planar se desconsiderarmos diferença de 30° entre planos formados pelos dois anéis principais e os radicais associados a eles. As únicas exceções foram os compostos de números 22, 24 e 25. O composto 22 apresenta três ligações simples entre o anel que contém o nitrogênio da estrutura principal da esterilquinolina e o radical que é um anel aromático, neste caso temos três ângulos diedrais que estão completamente livres para rotação. Os compostos 24 e 25 possuem um radical grande ligado ao anel aromático da estrutura principal da esterilquinolina com três ligações simples em sequência, estes compostos tendem a sair do plano devido a repulsão entre os átomos de hidrogênio do radical e da estrutura base da esterilquinolina. Em se tratando de cálculos com moléculas no vácuo as ligações simples são totalmente flexíveis, o que permite a torção, enquanto para a ligação dupla isto não ocorre.

4.2 Cálculo da MIE

Com a busca conformacional realizada e com os dados da estrutura eletrônica (energias e autoestados) já coletados para o mínimo global realizamos então o cálculo dos parâmetros utilizados na metodologia dos índices eletrônicos. Como já previsto, uma dificuldade desta metodologia é localizar a região, ou as regiões, que podem nos fornecer as melhores regras (o melhor padrão) de identificação de atividade biológica com os parâmetros Δ e η . Em trabalhos anteriores utilizando a MIE os autores indicam alguns caminhos para iniciarmos a busca da região mais adequada. Essas regiões podem ser, por exemplo, a região de maior densidade de carga, da maior localização dos orbitais moleculares HOMO ou LUMO, da região de maior ordem de ligação, etc. Independente de ser qualquer uma das regiões acima citadas foram feitas coletas dos dados para várias as regiões comuns para todos os compostos, indicadas na figura 4.5:

No. Da molécula	AM1	PM3
01	-87.138172	-96.726.323
02	-39.120788	-49.321794
03	-56.879630	-76.396565
04	-34.791991	-41.968191
05	-25.922028	-42.516718
06	-34.934088	-58.137689
07	-41.231342	-51.775053
08	-66.286299	-94.134997
09	-83.437021	-94.663413
10	-125.877834	-139.842356
11	-120.832660	-137.229328
12	-126.027665	-137.989034
13	-119.869233	-130.999484
14	-80.407464	-130.811570
15	-171.572960	-181.605239
16	-165.082812	-174.462615
17	-132.304180	-165.408768
18	-68.638130	-76.492224
19	-108.933879	-111.225695
20	-120.458480	-66.882780
21	5.310934	-59.495623
22	-153.230732	-161.463849
23	0.801032	-7.421687
24	-53.057025	-63.371043
25	-65.913317	-81.000529
26	-18.934410	6.222460
27	-52.538650	41.849704
28	-36.068553	19.578798
29	6.354784	-0.710653
30	12.928263	-9.611411
31	1.750220	-4.652928
32	-130.789573	-154.084388

Tabela 4.1: Valores do calor de formação para os compostos de mínima energia. A numeração refere-se às estruturas da tabela 2.1.

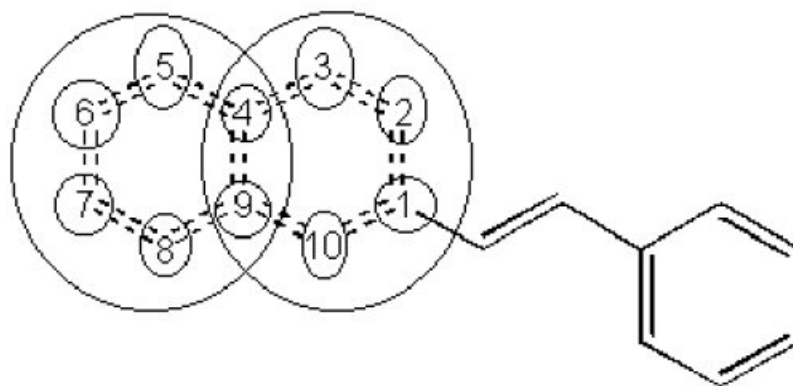


Figura 4.5: Regiões utilizadas para o cálculo da LDOS. Os valores da LDOS são base de um índice da MIE

Região	Átomos
A	C4, C5, C6, C7, C8 e C9
B	C1, C2, C3, C4, C9 e N10
C	C1
D	C2
E	C3
F	C4
G	C5
H	C6
I	C7
J	C8
K	C9
L	N10

Tabela 4.2: Átomos que compõem as regiões da figura 4.5

Na tabela 4.2, os símbolos na coluna dos átomos indicam o símbolo do átomo e a sua respectiva numeração na molécula. Por exemplo, C1 significa carbono da posição 1, N10 significa nitrogênio da posição 10.

Os valores dos parâmetros ΔL e ηL utilizados pela MIE para algumas regiões destacadas na tabela 4.2 estão nas tabelas 4.3 e 4.4.

Para os dados dos compostos obtidos com as otimizações pelo método semi-empírico AM1 as melhores regras foram obtidas para as regiões B, J e H e para o método semi-empírico PM3 as melhores regras foram para as regiões A, B, L, J, G e E.

Regras para AM1

Regra tipo “e” para região B:

Se $\Delta L \leq 0,400$ e $\eta L \leq 0,298$ o composto será ativo, caso não satisfaça essa condição simultaneamente o composto será inativo. Para esta regra os compostos de números 14 e 28 são classificados erroneamente.

Regra tipo “ou” para região H:

Se $\Delta L \leq 0,206$ ou $\eta L \geq 0,414$ o composto será ativo, caso não satisfaça algumas dessas condições o composto será inativo. Para esta regra as mesmas moléculas da regra anterior são classificadas erroneamente.

Regras para PM3

Regra tipo “e” para a região A:

Se $\Delta L \leq 0,122$ e $\eta L \leq 0,855$ o composto será ativo, caso não satisfaça essa condição simultaneamente o composto será inativo. Para esta regra os compostos de no. 28 e 30 são classificadas erroneamente.

Regra tipo “ou” para região E:

Se $\Delta L \leq 0,115$ ou $\eta L \geq 0,800$ o composto será ativo, caso não satisfaça algumas dessas condições o composto será inativo. Para esta regra as moléculas de números 30 e 32 são classificadas erroneamente.

Essas regras podem ser conferidas nas tabelas 4.3 e 4.4, em que os compostos discriminados com os caracteres na forma *itálico* são biologicamente ativos.

Estas regras acima, tipo “e” e “ou”, são somente para algumas das melhores regiões

AM1					
IC ₅₀ (μ M)	Região B		Região J	Região H	No.
	ΔL	ηL	ηL	ηL	
0.7	0,05085	-0,0281	-0,31103	0,21122	16
0.9	0,06099	0,02047	0,36522	-0,20747	14
2.8	0,02023	-0,06439	-0,18579	0,16424	13
2.4	0,08299	-0,22376	0,48152	-0,22506	12
0.3	0,0709	0,01074	-0,32178	0,22871	15
4.9	0,09236	0,04184	-0,32035	0,24026	17
2.3	0,39269	0,2980	-0,27137	0,41414	22
4.0	0,01535	-0,2008	-0,11357	0,10851	19
1.3	0,01157	-0,05545	0,46694	-0,21352	18
3.7	0,05093	0,0466	0,37978	-0,18842	11
3.2	0,04841	0,18315	-0,08049	-0,04752	10
3.4	0,09287	0,02287	-0,3025	0,23102	04
1.9	0,02655	-0,02309	-0,25151	0,19899	03
5.3	0,08858	0,03008	-0,28839	0,23224	02
4.1	0,05202	0,07742	0,3172	-0,19043	05
1.6	0,07611	0,01634	-0,28414	0,22555	09
1.4	0,02529	-0,1367	-0,22811	0,16121	08
3.5	0,1185	0,06002	-0,30074	0,24505	07
1.2	0,33196	0,55868	0,3646	-0,06412	06
>100	0,46496	0,13343	0,09231	-0,05163	28
>100	0,32321	0,32753	-0,03319	-0,20166	27
>100	0,49606	0,41355	0,06378	-0,15706	31
>100	0,53722	0,2647	0,05685	-0,16231	29
>100	0,51428	0,32046	-0,28142	-0,07686	30
>100	0,41538	0,19502	0,30356	-0,18426	26
>100	0,21133	0,84244	0,32194	0,14769	25
>100	0,40456	0,28723	-0,25458	0,41279	01
>100	0,02513	-0,04081	-0,31206	0,2164	20
>100	0,37678	0,30699	-0,01745	-0,20887	21
>100	0,25459	0,37116	-0,31731	0,08987	23
>100	0,25102	0,7978	0,28691	0,14626	24
>100	0,41094	0,14106	-0,10218	-0,17776	32

Tabela 4.3: Parâmetros da MIE para os dados dos cálculos realizados com AM1. Os compostos destacados em negrito são inativos e os demais ativos.

PM3								
IC ₅₀ (μ M)	Região A		Região B		Região L		Região J	
	ΔL	ηL	ηL	ηL	ηL	ηL	ηL	No.
0.7	0,02077	0,51094	0,07867	-0,09082	0,27662	0,04251	0,00777	16
0.9	0,04276	0,49625	0,09939	-0,03924	0,31978	0,06222	0,04728	14
2.8	0,00787	0,47596	0,06219	-0,06288	0,30031	0,05055	0,02463	13
2.4	0,01072	0,53249	0,08314	-0,13711	0,2292	0,02391	-0,02698	12
0.3	0,02723	0,51423	0,08378	-0,2873	0,27745	0,04415	0,01069	15
4.9	0,02542	0,48618	0,08102	-0,05092	0,30986	0,05683	0,03675	17
2.3	0,40548	0,04506	0,26315	0,0094	-0,27742	0,02685	0,06753	22
4.0	0,0498	0,50033	0,1107	-0,0313	0,32526	0,06547	0,05424	27
1.3	0,09224	0,54683	0,17263	-0,02722	0,32361	0,0709	0,06641	18
3.7	0,01315	0,5793	0,12639	-0,20159	0,14602	-0,00309	-0,07569	11
3.2	0,03284	0,49294	0,07469	-0,06997	0,29599	0,05209	0,02662	10
3.4	0,04213	0,51099	0,11939	-0,0411	0,31516	0,06102	0,04583	04
1.9	0,0256	0,38087	0,01019	0,00829	0,35726	0,07753	0,07823	03
5.3	0,00838	0,39446	0,02514	0,00631	0,35518	0,07381	0,07172	02
4.1	0,08608	0,53837	0,15341	-0,03784	0,31708	0,06796	0,05941	05
1.6	0,00604	-0,43278	-0,03429	0,03268	-0,32629	-0,0589	-0,04129	09
1.4	0,0681	0,53074	0,14846	-0,03239	0,32056	0,06701	0,05791	08
3.5	0,0126	-0,46956	-0,0476	0,07189	-0,29308	-0,04294	-0,01064	07
1.2	0,40783	0,8055	0,56512	0,04228	0,33871	0,09606	0,13158	06
>100	0,53415	0,10166	0,00712	0,13989	0,05174	0,09947	-0,07713	28
>100	0,49627	0,49664	0,39583	-0,20664	-0,01408	0,09615	0,09505	27
>100	0,49584	0,47878	0,2929	-0,20654	0,01855	0,11685	0,08972	31
>100	0,45903	0,46281	0,22912	0,02498	-0,01545	0,07516	-0,23214	29
>100	0,52115	0,24959	0,18384	-0,25943	-0,21985	-0,17936	-0,19918	30
>100	0,47074	0,44649	0,19977	-0,24831	-0,04518	0,05067	-0,01622	26
>100	0,12549	-0,94471	-0,91982	-0,26111	-0,33761	-0,32581	-0,34848	25
>100	0,41349	0,04634	0,25744	0,0133	-0,27145	0,03421	0,07401	01
>100	0,25895	0,48597	0,1737	0,02551	0,39386	0,11595	0,09895	20
>100	0,40709	0,5387	0,2852	-0,24886	-0,04076	0,0479	-0,04153	21
>100	0,35234	0,62677	0,24909	-0,35017	-0,26871	-0,24606	-0,45587	23
>100	0,07195	0,87337	0,92888	0,26525	0,30827	0,302	0,35414	24
>100	0,47479	0,43732	0,17435	-0,23216	-0,03764	0,05728	-0,01331	32

Tabela 4.4: Parâmetros da MIE para os dados dos cálculos realizados com PM3. Os compostos destacados em negrito são inativos e os demais ativos.

para ambos os métodos semi-empíricos. Para as outras regiões analisadas indicamos somente a sua performance no reconhecimento de padrão em porcentagens de acertos:

Nas tabelas 4.6 e 4.7 a última coluna da direita a letra H indica os resultados para os dois últimos orbitais moleculares ocupados (HOMO e HOMO-1). A letra L indica os resultados para os dois primeiros orbitais moleculares desocupados (LUMO e LUMO+1). Com os resultados acima podemos observar que os padrões obtidos pelos dois primeiros orbitais moleculares desocupados são melhores que os padrões fornecidos pelos dois últimos orbitais moleculares ocupados.

Verificamos que MIE se mostrou extremamente eficiente (até 94% de acerto) para classificar os compostos ativos e inativos de acordo com os padrões experimentais.

	AM1 – Porcentagem de acerto								
	Regras Tipo “e”				Regras Tipo “ou”				
Região	$\Delta \geq$ e $\eta \geq$	$\Delta \geq$ e $\eta \leq$	$\Delta \leq$ e $\eta \geq$	$\Delta \leq$ e $\eta \leq$	$\Delta \geq$ ou $\eta \geq$	$\Delta \geq$ ou $\eta \leq$	$\Delta \leq$ ou $\eta \geq$	$\Delta \leq$ ou $\eta \leq$	
A	71.88%	78.12%	68.75%	71.88%	75.00%	75.00%	65.62%	68.75%	H
	59.38%	75.00%	90.62%	90.62%	59.38%	71.88%	90.62%	90.62%	L
B	78.12%	71.88%	68.75%	65.62%	75.00%	75.00%	68.75%	62.50%	H
	59.38%	87.50%	90.62%	93.75%	59.38%	84.38%	90.62%	90.62%	L
C	75.00%	71.88%	65.62%	65.62%	75.00%	75.00%	68.75%	65.62%	H
	68.75%	62.50%	90.62%	90.62%	68.75%	59.38%	90.62%	90.62%	L
D	78.12%	78.12%	68.75%	75.00%	75.00%	71.88%	68.75%	65.62%	H
	59.38%	75.00%	90.62%	90.62%	59.38%	75.00%	87.50%	90.62%	L
E	71.88%	84.38%	68.75%	78.12%	75.00%	78.12%	62.50%	75.00%	H
	81.25%	65.62%	90.62%	90.62%	78.12%	65.62%	87.50%	87.50%	L
F	87.50%	71.88%	84.38%	65.62%	84.38%	71.88%	81.25%	62.50%	H
	59.38%	75.00%	90.62%	90.62%	59.38%	75.00%	90.62%	90.62%	L
G	71.88%	84.38%	65.62%	75.00%	75.00%	75.00%	65.62%	71.88%	H
	65.62%	68.75%	90.62%	90.62%	65.62%	68.75%	87.50%	87.50%	L
H	71.88%	75.00%	65.62%	68.75%	75.00%	75.00%	65.62%	68.75%	H
	68.75%	59.38%	90.62%	90.62%	65.62%	62.50%	93.75%	90.62%	L
I	71.88%	84.38%	68.75%	78.12%	75.00%	75.00%	65.62%	71.88%	H
	59.38%	75.00%	90.62%	90.62%	59.38%	71.88%	87.50%	90.62%	L
J	75.00%	81.25%	75.00%	68.75%	75.00%	68.75%	68.75%	68.75%	H
	59.38%	65.62%	90.62%	90.62%	62.50%	62.50%	93.75%	90.62%	L
K	78.12%	75.00%	65.62%	68.75%	75.00%	75.00%	65.62%	62.50%	H
	59.38%	78.12%	90.62%	90.62%	59.38%	78.12%	87.50%	90.62%	L
L	75.00%	75.00%	71.88%	68.75%	75.00%	71.88%	65.62%	68.75%	H
	78.12%	62.50%	90.62%	90.62%	78.12%	62.50%	90.62%	87.50%	L

Tabela 4.5: Porcentagens de acertos com a MIE, para os dados dos cálculos realizados com AM1, baseados nos padrões experimentais. Os destaques em itálico são correspondentes aos índices dos estados ocupados e os demais são correspondentes ao estados desocupados. A última coluna da direita a letra H indica os resultados para os dois últimos orbitais moleculares ocupados (HOMO e HOMO-1) e a letra L indica os resultados para os dois primeiros orbitais moleculares desocupados (LUMO e LUMO+1)

	PM3 – Porcentagem de acerto								
	Regras Tipo “e”				Regras Tipo “ou”				
Região	$\Delta \geq$ e $\eta \geq$	$\Delta \geq$ e $\eta \leq$	$\Delta \leq$ e $\eta \geq$	$\Delta \leq$ e $\eta \leq$	$\Delta \geq$ ou $\eta \geq$	$\Delta \geq$ ou $\eta \leq$	$\Delta \leq$ ou $\eta \geq$	$\Delta \leq$ ou $\eta \leq$	
A	<i>68.75%</i>	<i>75.00%</i>	<i>78.12%</i>	<i>75.00%</i>	<i>71.88%</i>	<i>68.75%</i>	<i>75.00%</i>	<i>75.00%</i>	<i>H</i>
	65.62%	62.50%	90.62%	93.75%	62.50%	65.62%	93.75%	87.50%	L
B	<i>78.12%</i>	<i>68.75%</i>	<i>81.25%</i>	<i>71.88%</i>	<i>75.00%</i>	<i>68.75%</i>	<i>75.00%</i>	<i>71.88%</i>	<i>H</i>
	62.50%	84.38%	90.62%	93.75%	62.50%	84.38%	93.75%	87.50%	L
C	<i>78.12%</i>	<i>68.75%</i>	<i>78.12%</i>	<i>75.00%</i>	<i>71.88%</i>	<i>71.88%</i>	<i>75.00%</i>	<i>71.88%</i>	<i>H</i>
	59.38%	75.00%	90.62%	90.62%	59.38%	71.88%	87.50%	90.62%	L
D	<i>81.25%</i>	<i>71.88%</i>	<i>81.25%</i>	<i>78.12%</i>	<i>81.25%</i>	<i>68.75%</i>	<i>75.00%</i>	<i>71.88%</i>	<i>H</i>
	68.75%	75.00%	90.62%	90.62%	65.62%	75.00%	87.50%	90.62%	L
E	<i>75.00%</i>	<i>75.00%</i>	<i>81.25%</i>	<i>78.12%</i>	<i>68.75%</i>	<i>68.75%</i>	<i>71.88%</i>	<i>75.00%</i>	<i>H</i>
	71.88%	68.75%	90.62%	93.75%	73.88%	68.75%	93.75%	87.50%	L
F	<i>84.38%</i>	<i>71.88%</i>	<i>90.62%</i>	<i>75.00%</i>	<i>81.25%</i>	<i>68.75%</i>	<i>84.38%</i>	<i>71.88%</i>	<i>H</i>
	65.62%	84.25%	90.62%	90.62%	62.50%	81.25%	87.50%	87.50%	L
G	<i>68.75%</i>	<i>78.12%</i>	<i>78.12%</i>	<i>78.12%</i>	<i>68.75%</i>	<i>71.88%</i>	<i>75.00%</i>	<i>75.00%</i>	<i>H</i>
	68.75%	71.88%	90.62%	93.75%	68.75%	68.75%	90.62%	87.50%	L
H	<i>75.00%</i>	<i>68.75%</i>	<i>78.12%</i>	<i>75.00%</i>	<i>71.88%</i>	<i>71.88%</i>	<i>75.00%</i>	<i>75.00%</i>	<i>H</i>
	59.38%	78.12%	90.62%	90.62%	59.38%	75.00%	90.62%	90.62%	L
I	<i>75.00%</i>	<i>84.38%</i>	<i>78.12%</i>	<i>87.50%</i>	<i>71.88%</i>	<i>81.25%</i>	<i>75.00%</i>	<i>81.25%</i>	<i>H</i>
	68.75%	81.25%	90.62%	90.62%	65.62%	81.25%	90.62%	90.62%	L
J	<i>75.00%</i>	<i>81.25%</i>	<i>81.25%</i>	<i>81.25%</i>	<i>68.75%</i>	<i>68.75%</i>	<i>71.88%</i>	<i>75.00%</i>	<i>H</i>
	84.38%	62.50%	90.62%	90.62%	81.25%	62.50%	93.75%	90.62%	L
K	<i>75.00%</i>	<i>71.88%</i>	<i>78.12%</i>	<i>78.12%</i>	<i>68.75%</i>	<i>71.88%</i>	<i>75.00%</i>	<i>71.88%</i>	<i>H</i>
	59.38%	78.12%	90.62%	90.62%	59.38%	78.12%	87.50%	90.62%	L
L	<i>71.88%</i>	<i>71.88%</i>	<i>81.25%</i>	<i>78.12%</i>	<i>71.88%</i>	<i>68.75%</i>	<i>75.00%</i>	<i>75.00%</i>	<i>H</i>
	81.25%	65.62%	90.62%	93.75%	81.25%	62.50%	90.62%	87.50%	L

Tabela 4.6: Porcentagens de acertos com a MIE, para os dados dos cálculos realizados com PM3, baseados nos padrões experimentais. Os destaques em itálico são correspondentes aos índices dos estados ocupados e os demais são correspondentes aos estados desocupados. A última coluna da direita a letra H indica os resultados para os dois últimos orbitais moleculares ocupados (HOMO e HOMO-1) e a letra L indica os resultados para os dois primeiros orbitais moleculares desocupados (LUMO e LUMO+1).

4.3 PCA e HCA

O próximo estudo foi realizado com a Análise de Componentes Principais (PCA) e Análise Hierárquica de Agrupamentos (HCA). Para a análise utilizamos os parâmetros da MIE (ΔH , ηH , ΔL , ηL) de todas as regiões analisadas e, E_{HOMO} , E_{HOMO-1} , E_{LUMO} , E_{LUMO+1} , valores das densidades locais de estados dos orbitais moleculares de fronteira, calor de formação, momento dipolar, potencial de ionização, dureza ($(E_{HOMO} - E_{LUMO})/2$), área da superfície aproximada, volume, energia de hidratação, Log P, refratividade, polarizabilidade, massa, energia eletrônica, repulsão nuclear, totalizando um total de 93 variáveis para análise. Esses índices foram obtidos com o programa Chem2Pac, arquivos de saída do programa mopac, e também com o programa HyperChem que possui uma ferramenta denominada como *QSAR properties*. Nossa matriz inicial de dados foi composta por 32 (amostras) x 93 (variáveis). Primeiramente foi realizada uma seleção de variáveis que melhor separassem as classes de moléculas em ativas e inativas. Essa seleção foi realizada visualizando variável contra variável em um gráfico bi-dimensional, uma outra forma de se realizar esta análise seria verificar a distribuição dos compostos para um determinado descritor no sentido de verificar se existe uma separação entre compostos ativos e inativos. A seleção de variáveis é feita até chegarmos a um número mínimo de descritores capazes de distinguir entre compostos ativos e inativos. O resultado ideal seria o caso em que todos os compostos se apresentassem em dois grupos (“clusters”) distintos se tivéssemos duas classes de compostos, mas um resultado pode ser considerado bom desde que tenha uma separação visual entre as classes estudadas e também seus compostos estejam em sua respectiva classe. Os resultados podem ser vistos nas figuras 4.6 e 4.7. Para a análise de PCA utilizamos o pré-processamento de autoescalamento dos dados com duas componentes principais. Para a análise de HCA utilizamos o mesmo tipo de pré-processamento do PCA com a técnica de agrupamento tipo “*incremental*”.

PCA – AM1			
	Variância	Porcentagem	Acumulativo
PC1	50.2090	53.9882	53.9882
PC2	32.0640	34.4774	88.4656

Tabela 4.7: Dados da variância das componentes principais PC1 e PC2, para os dados dos cálculos realizados com AM1.

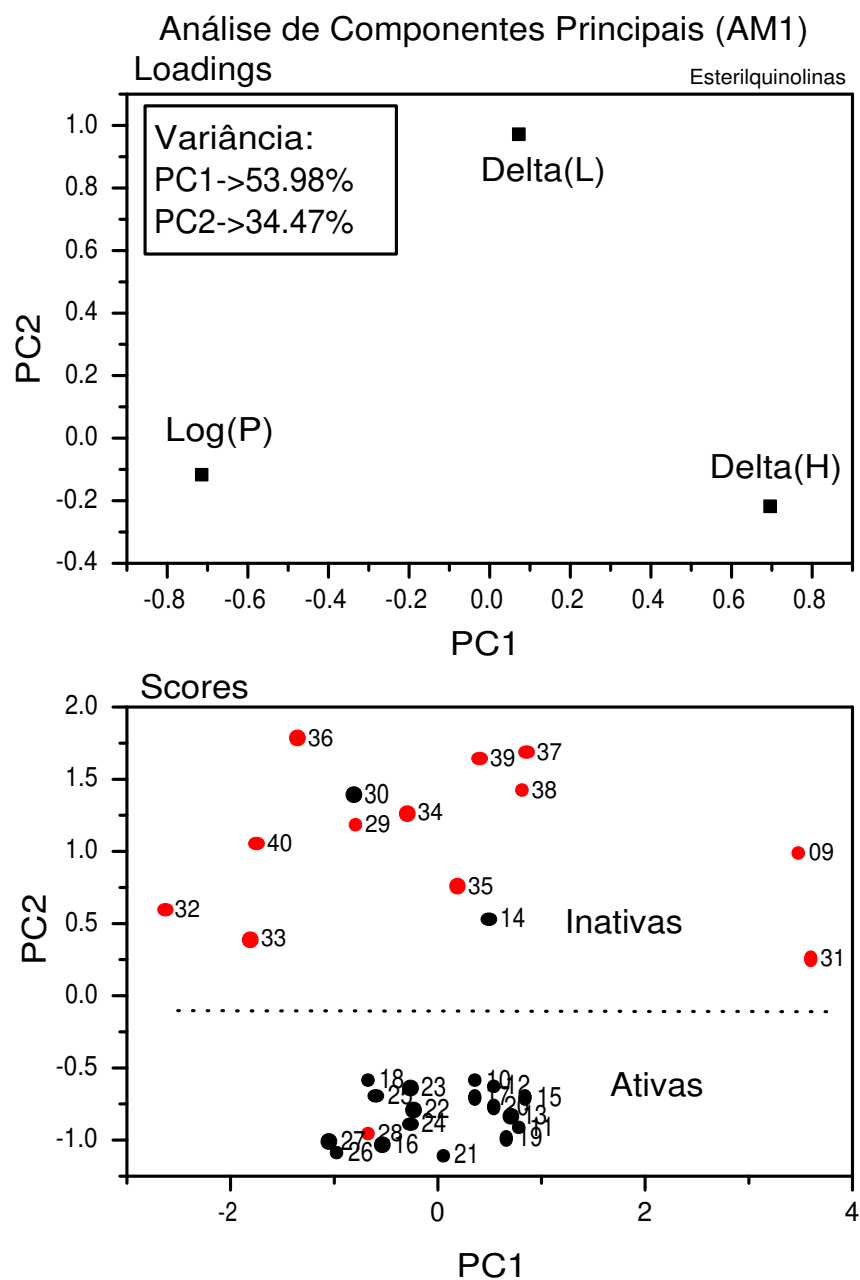


Figura 4.6: Análise de Componentes Principais (PCA) com dados MIE/AM1 e demais parâmetros.

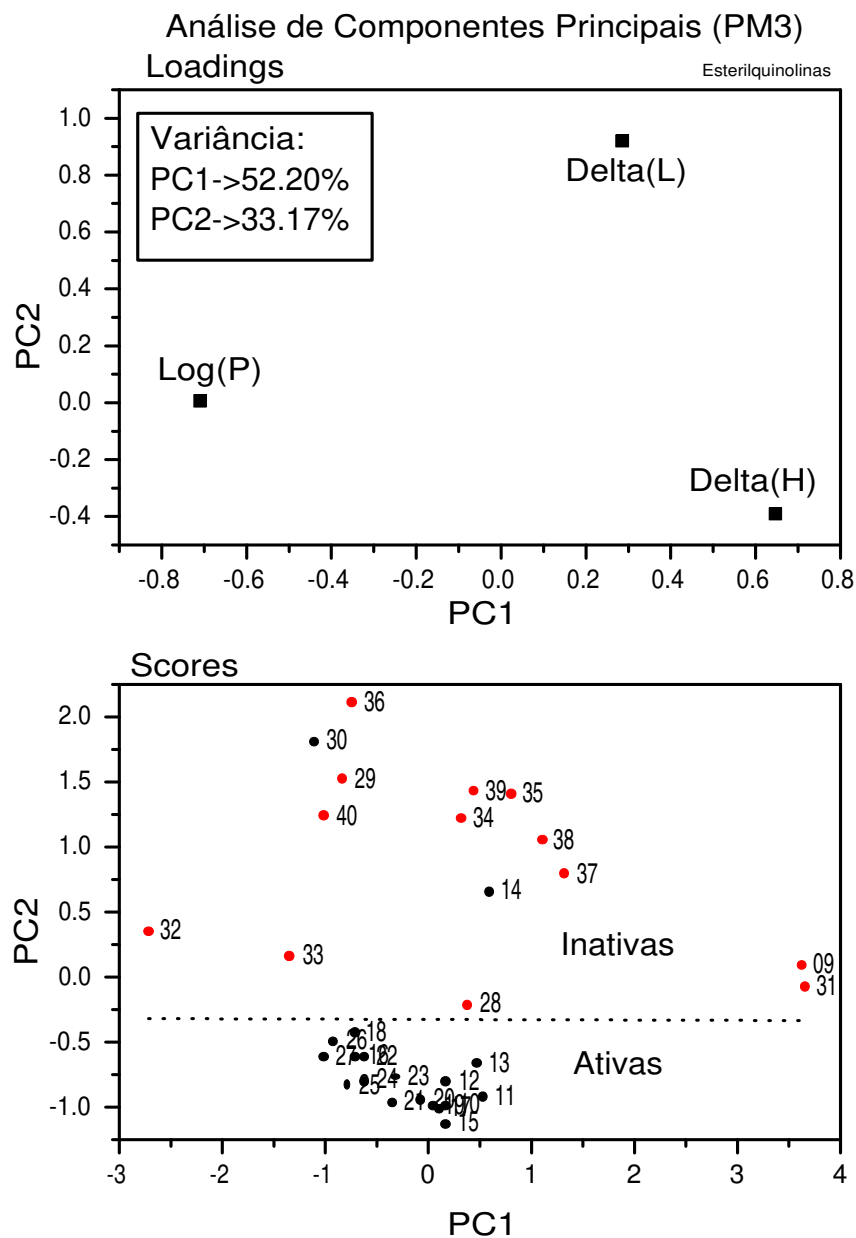


Figura 4.7: Análise de Componentes Principais (PCA) com dados MIE/PM3 e demais parâmetros.

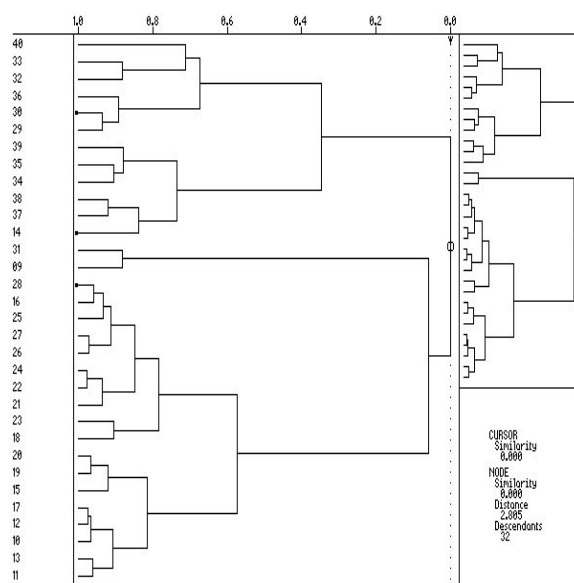


Figura 4.8: HCA - AM1

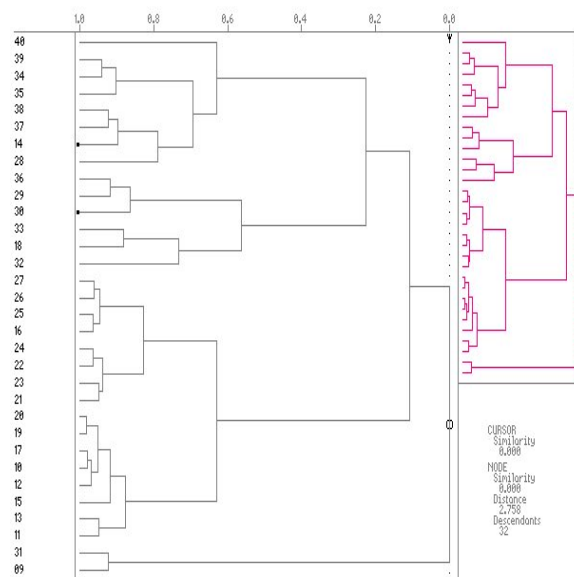


Figura 4.9: HCA - PM3

Equações para as variáveis (loadings) PCA – AM1:

$$PC1=0.69857(\Delta H)+0.07355(\Delta L)-0.7117(\text{Log P})$$

$$PC2=-0.2171(\Delta H)+0.9696(\Delta L)-0.1128(\text{Log P})$$

PCA – PM3			
	Variância	Porcentagem	Acumulativo
PC1	48.550	52.2096	52.2096
PC2	30.8562	33.1787	85.3884

Tabela 4.8: Dados da variância das componentes principais PC1 e PC2, para os dados dos cálculos realizados com PM3.

Equações para as variáveis (loadings) PCA – PM3:

$$PC1=0.6968(\Delta H)+0.28284(\Delta L)-0.7056(\text{Log P})$$

$$PC2=-0.3919(\Delta H)+0.91996(\Delta L)-0.00790(\text{Log P})$$

Nas figuras 4.6 e 4.7 podemos verificar os padrões fornecidos pelas análises de PCA. As variáveis ΔH , ΔL e Log P foram suficientes para a separação das classes de atividades distintas com grandes contribuições nas componentes principais como pode ser visto pelas equações acima. As componentes principais PC1 e PC2 foram suficientes para a separação das classes, não sendo necessário utilizar outras componentes principais. Outro motivo para utilização de duas componentes principais foi o fato de termos somente duas classes de moléculas uma inativa e outra ativa. Essa não é uma regra definida, mas podemos nos utilizar desse recurso para melhor desenvolvimento da análise exploratória.

A análise de PCA com os dados das otimizações com AM1 previu somente três moléculas fora das suas classes de origem. As moléculas classificadas incorretamente foram as de números 14, 28 e 38. Os resultados de HCA nos mostram que as moléculas classificadas incorretamente são as mesmas dos resultados da análise feita com o PCA. Na análise feita com o PCA para os dados das otimizações com PM3 temos somente duas moléculas classificadas erroneamente. Os resultados de PCA e HCA, agora para o caso dos dados provenientes das otimizações feitas com PM3, tiveram os mesmos acertos. As

moléculas classificadas incorretamente são as mesmas para ambos os métodos, moléculas 14 e 38.

Os resultados de HCA podem ser vistos pelas figuras 4.8 e 4.9 onde as moléculas que foram previstas incorretamente estão com um círculo negro no início do lado esquerdo do dendrograma. Estas são incorretamente agrupadas bem no início da análise. Esse resultados repete os resultados obtidos com a análise de PCA.

Em seqüência ao desenvolvimento do trabalho, o próximo estudo realizado foi utilizando as redes neurais artificiais. Neste estudo utilizamos como dados de entrada para a rede os descritores que foram utilizados para a análise de PCA, já mencionados. Não utilizamos todos os 93 descritores, somente aqueles que foram relevantes para o PCA (ΔL , ΔH e $\text{Log}(P)$) e os parâmetros da MIE (ΔL e ηL) da região B que melhor descreveu o padrão experimental.

Todos os dados referentes às saídas fornecidas pelas redes neurais (padrões de treinamentos e predições) se encontram em forma de tabelas no apêndice A.

4.4 Redes Neurais

O estudo das redes neurais foi desenvolvido com o propósito de verificar se a rede consegue “enxergar” os padrões estabelecidos pelos dados experimentais de atividade biológica (esta parte consiste no treinamento da rede) e também realizar a predição. Para proceder o estudo com as redes neurais utilizamos 50% dos compostos (ativos e inativos) como conjunto de treinamento e os outros 50% como conjunto de predição. Para verificar a dependência do conjunto tanto do treinamento como da predição, selecionamos 10 conjuntos, aleatoriamente, para o treinamento e para predição, que podem ser vistos nas tabelas 4.9 e 4.10.

Os resultados dos treinamentos e das predições estão explicitados nas tabelas 4.11 e 4.12. Os valores de ALPHA e HALPHA são os parâmetros de não linearidade da função sigmoidal da segunda e terceira camada, HDNLAYER é o número de neurônios da segunda camada, EPSIRON e HEPSIRON é o fator de troca dos pesos da segunda e terceira camada e a bias é o termo independente da função sigmoidal da primeira camada, como visto na secção das metodologias. As duas últimas linhas das tabelas contêm os números de erros no treinamento da rede e na predição. Os conjuntos 3, 6 e 7 foram os que apresentaram os menores erros para os dados do cálculo AM1 e os

		01		02		03		04		05		06		07		08		09		10	
		T	P	T	P	T	P	T	P	T	P	T	P	T	P	T	P	T	P	T	P
I	01		X	X			X		X		X		X	X		X		X		X	
A	02		X		X	X		X		X		X		X	X		X		X		
A	03	X		X			X		X		X		X	X			X	X		X	
A	04		X		X		X	X			X		X		X		X		X	X	
A	05		X		X		X		X	X		X			X	X		X		X	
A	06	X		X			X	X			X	X		X		X			X	X	
A	07		X	X		X		X			X		X	X			X		X		X
A	08	X		X			X	X			X		X		X	X		X			X
A	09	X			X		X	X		X		X			X		X	X			X
A	10	X			X		X		X	X		X		X			X	X			X
A	11		X		X		X	X		X		X		X		X			X	X	
A	12		X	X		X			X		X	X		X		X			X	X	
A	13	X		X		X			X		X		X	X			X		X		X
A	14	X			X	X			X	X			X	X		X			X	X	
A	15		X	X		X			X	X		X		X		X		X		X	
A	16		X	X			X	X		X			X		X		X		X		X
A	17		X	X		X		X		X		X			X		X	X		X	
A	18	X			X	X			X		X		X		X	X			X		X
A	19	X		X		X			X		X		X	X			X	X			X
I	20	X		X		X		X			X	X		X		X			X		X
I	21		X	X		X			X	X		X			X	X			X		X
A	22	X			X	X		X		X		X			X	X		X			X
I	23	X			X	X			X		X		X		X		X	X		X	
I	24		X	X			X	X		X			X	X		X		X			X
I	25	X			X	X			X	X		X		X		X			X	X	
I	26		X	X			X		X		X		X		X		X	X			X
I	27		X	X		X		X		X			X	X			X		X		X
I	28		X	X			X	X		X		X			X		X		X	X	
I	29	X			X	X		X		X		X		X			X	X		X	
I	30	X			X		X	X			X	X			X		X	X		X	
I	31	X			X		X		X	X		X			X	X			X	X	
I	32	X			X	X		X			X		X	X		X		X			X

Tabela 4.9: Conjuntos submetidos para as redes neurais: as colunas são os conjuntos de treinamentos (T) e predições (P); as linhas são os compostos (A = ativa e I = inativa). Dados dos cálculos realizados com AM1, com base nos padrões experimentais.

		01		02		03		04		05		06		07		08		09		10	
		T	P	T	P	T	P	T	P	T	P	T	P	T	P	T	P	T	P	T	P
I	01		X	X			X		X		X		X	X		X		X		X	
A	02		X		X	X		X		X		X		X	X		X		X		
A	03	X		X			X		X		X		X	X			X	X		X	
A	04		X		X		X	X			X		X		X		X		X	X	
A	05		X		X		X		X	X		X			X	X		X		X	
A	06	X		X			X	X			X	X		X		X			X	X	
A	07		X	X		X		X			X		X	X			X		X		X
A	08	X		X			X	X			X		X		X	X		X			X
A	09	X			X		X	X		X		X			X		X	X			X
A	10	X			X		X		X	X		X		X			X	X			X
A	11		X		X		X	X		X		X		X		X			X	X	
A	12		X	X		X			X		X	X		X		X			X		
A	13	X		X		X			X		X		X	X			X		X		X
A	14	X			X	X			X	X			X	X		X			X	X	
A	15		X	X		X			X	X		X		X		X		X		X	
A	16		X	X			X	X		X			X		X		X		X		X
A	17		X	X		X		X		X		X			X		X	X		X	
A	18	X			X	X			X		X		X		X	X			X		X
A	19	X		X		X			X		X		X	X			X	X			X
I	20	X		X		X		X			X	X		X		X			X		X
I	21		X	X		X			X	X		X			X	X			X		X
A	22	X			X	X		X		X		X			X	X		X			X
I	23	X			X	X			X		X		X		X		X	X		X	
I	24		X	X			X	X		X			X	X		X		X			X
I	25	X			X	X			X	X		X		X		X			X	X	
I	26		X	X			X		X		X		X		X		X	X			X
I	27		X	X		X		X		X			X	X			X		X		X
I	28		X	X			X	X		X		X			X		X		X	X	
I	29	X			X	X		X		X		X		X			X	X		X	
I	30	X			X		X	X			X	X			X		X	X		X	
I	31	X			X		X		X	X		X			X	X			X	X	
I	32	X			X	X		X			X		X	X		X		X			X

Tabela 4.10: Conjuntos submetidos para as redes neurais: as colunas são os conjuntos de treinamentos (T) e predições (P); as linhas são os compostos (A = ativa e I = inativa). Dados dos cálculos realizados com PM3, com base nos padrões experimentais.

conjuntos 1, 3, 7 e 8 para os dados do cálculos PM3. Esses conjuntos apresentam ter um caráter mais universal dentro do conjunto dos compostos estudados.

Conjunto	1	2	3	4	5	6	7	8	9	10
ALPHA	4,0	4,0	3,5	4,0	2,5	3,0	4,5	4,0	4,0	4,0
HALPHA	4,0	4,0	3,5	4,0	2,5	3,0	4,5	4,0	4,0	4,0
HDNLAYER	6	9	6	6	6	6	6	10	10	6
EPSIRON	0,45	0,25	0,80	0,95	2,50	1,50	0,70	0,15	0,30	0,25
HEPSIRON	0,45	0,25	0,80	0,95	2,50	1,50	0,70	0,15	0,30	0,25
BIAS	sim	sim	sim	sim	sim	sim	sim	sim	não	sim
No. De In- terações	30.000	30.000	30.000	30.000	30.000	30.000	30.000	30.000	30.000	30.000
Precisão	10^{-5}	10^{-5}	10^{-5}	10^{-5}	10^{-5}	10^{-5}	10^{-5}	10^{-4}	10^{-5}	10^{-5}
Erros no Treina- mento	2	0	1	1	1	1	1	0	0	0
Erros na Predição	1	2	1	4	3	1	1	4	3	3

Tabela 4.11: Parâmetros do treinamento da rede neural e erros nos treinamentos e predições. Dados dos cálculos realizados com AM1

Complementando o estudo com as redes neurais, utilizamos agora somente os parâmetros da MIE para treinar e prever os compostos, de acordo com os padrões fornecidos pela própria MIE. De maneira semelhante a usada no caso anterior, foram utilizados 50% dos compostos ativos e inativos para o treinamento e 50% para predição com conjuntos gerados aleatoriamente.

Nas tabelas 4.14 e 4.15 (duas últimas linhas) temos a performance da rede neural utilizando os parâmetros da MIE. Nestes casos é possível verificar que independente do conjunto utilizado os resultados possuem a mesma margem de erro, para ambos os casos dos dados AM1 e PM3, com exceção dos grupos 3 e 4 da tabela 4.15. Os resultados finais desta análise podem ser resumidos na tabela 4.16.

Os resultados finais desta análise podem ser resumidos na tabela 4.16:

Os melhores resultados foram obtidos com dados dos cálculos realizados com o método semi-empírico PM3. Com a metodologia dos índices eletrônicos, em ambos os métodos semi-empíricos, os índices de acertos foram relativamente superiores e somente dois parâmetros foram usados contra três dos outros métodos.

Conjunto	1	2	3	4	5	6	7	8	9	10
ALPHA	4,0	4,5	2,7	4,0	4,5	2,0	2,0	2,0	2,4	3,9
HALPHA	4,0	4,5	2,7	4,0	4,5	2,0	2,0	2,0	2,4	3,9
HDNLAYER	6	6	6	6	6	6	6	6	6	6
EPSIRON	0,80	0,70	0,10	0,10	0,95	0,95	0,82	0,67	0,70	0,91
HEPSIRON	0,90	0,70	0,10	0,10	0,95	0,10	0,82	0,67	0,70	0,91
BIAS	sim	sim	sim	sim	sim	sim	sim	sim	sim	sim
No. De In- terações	30.000	30.000	30.000	30.000	30.000	30.000	30.000	30.000	30.000	30.000
Precisão	10^{-5}	10^{-5}	10^{-5}	10^{-5}	10^{-5}	10^{-5}	10^{-5}	10^{-5}	10^{-5}	10^{-5}
Erros no Treina- mento	1	0	0	1	0	0	0	0	0	0
Erros na Predição	0	2	1	1	3	2	1	0	3	2

Tabela 4.12: Parâmetros do treinamento da rede neural e erros nos treinamentos e predições. Dados dos cálculos realizados com PM3

Na análise dos compostos de esterilquinolinas, a MIE mostrou não somente um grande acerto classificatório, como também superou o acerto de técnicas mais tradicionais como PCA e HCA. Vale a pena ressaltar que se os índices eletrônicos da MIE não são utilizados na PCA e HCA os padrões fornecidos são diferentes e o índice de acerto diminui significativamente, o que confirma a importância estatística dos parâmetros da MIE. Mais especificamente a análise de estrutura-atividade sem os parâmetros da MIE do total de parâmetros utilizados, no caso das esterilquinolinas, traz poucas informações sobre o padrão comparado ao padrão experimental assumido como base.

		AM1												PM3									
		01		02		03		04		05				01		02		03		04		05	
		T	P	T	P	T	P	T	P	T	P			T	P	T	P	T	P	T	P	T	P
I	01		X	X			X		X		X	I	01		X	X			X		X		X
A	02		X		X	X		X		X		A	02		X		X	X		X		X	
A	03	X		X			X		X		X	A	03	X		X			X		X		X
A	04		X		X		X	X			X	A	04		X		X		X	X			X
A	05		X		X		X		X	X		A	05		X		X		X		X	X	
I	06	X		X		X		X			X	A	06	X		X			X	X			X
A	07		X	X		X		X			X	A	07		X	X		X		X			X
A	08	X		X			X	X			X	A	08	X		X			X	X			X
A	09	X			X		X	X		X		A	09	X			X		X	X		X	
A	10	X			X		X		X	X		A	10	X			X		X		X	X	
A	11		X		X		X	X		X		A	11		X		X		X	X		X	
A	12		X	X		X			X		X	A	12		X	X		X			X		X
A	13	X		X		X			X		X	A	13	X		X		X			X		X
A	14	X			X	X			X	X		A	14	X			X	X			X	X	
A	15		X	X		X			X	X		A	15		X	X		X			X	X	
A	16		X	X			X	X		X		A	16		X	X			X	X		X	
A	17		X	X		X		X		X		A	17		X	X		X		X		X	
A	18	X			X	X			X		X	A	18	X			X	X			X		X
A	19	X		X		X			X		X	A	19	X		X		X			X		X
A	20	X		X		X		X		X		A	20	X			X	X		X		X	
I	21		X	X		X			X	X		I	21		X	X		X			X		X
A	22	X			X		X	X			X	I	22	X		X		X		X		X	
I	23	X			X	X			X		X	I	23	X			X	X			X		X
I	24		X	X			X	X		X		I	24		X	X			X	X		X	
I	25	X			X	X			X	X		I	25	X			X	X			X	X	
I	26		X	X			X		X		X	I	26		X	X			X		X		X
I	27		X	X		X		X		X		I	27		X	X		X		X		X	
I	28		X	X			X	X		X		I	28		X	X			X	X		X	
I	29	X			X	X		X		X		I	29	X			X	X		X		X	
I	30	X			X		X	X			X	I	30	X			X		X	X			X
I	31	X			X		X		X	X		I	31	X			X		X		X	X	
I	32	X			X	X		X			X	I	32	X			X	X		X			X

Tabela 4.13: Conjuntos submetidos para as redes neurais: as colunas são os conjuntos de treinamentos (T) e predições (P); as linhas são os compostos (A = ativa e I = inativa). Baseado nos padrões de classificações da MIE, utilizando somente índices da MIE (AM1 e PM3) como dados de entrada.

Conjunto	1	2	3	4	5
ALPHA	4,5	1,0	4,0	4,5	4,0
HALPHA	4,5	1,0	4,0	4,5	4,0
HDNLAYER	12	4	4	10	10
EPSIRON	0,71	0,50	0,90	0,74	0,50
HEPSIRON	0,74	0,50	0,90	0,74	0,50
BIAS	sim	sim	sim	sim	sim
No. De Interações	50.000	30.000	30.000	30.000	30.000
Precisão	10^{-5}	10^{-5}	10^{-5}	10^{-5}	10^{-5}
Erros no Treinamento	1	0	0	1	0
Erros na Predição	0	1	1	0	1

Tabela 4.14: Parâmetros de treinamento da rede neural e erros nos treinamentos e nas predições, dados AM1

Conjunto	1	2	3	4	5
ALPHA	2,0	4,0	2,0	2,0	1,0
HALPHA	2,0	4,0	2,0	2,0	1,0
HDNLAYER	4	4	10	4	4
EPSIRON	0,35	0,10	0,09	0,10	1,90
HEPSIRON	0,35	0,10	0,09	0,10	1,90
BIAS	sim	sim	sim	sim	sim
No. De Interações	30.000	30.000	30.000	30.000	30.000
Precisão	10^{-5}	10^{-5}	10^{-5}	10^{-5}	10^{-5}
Erros no Treinamento	0	0	0	0	1
Erros na Predição	1	1	2	2	1

Tabela 4.15: Parâmetros de treinamento da rede neural e erros nos treinamentos e nas predições, dados PM3

	AM1	PM3
MIE	93.75%	93.75%
PCA	90.62%	93.75%
HCA	90.62%	93.75%
RN/PCA	93,75%	100%
RN/MIE	93,75%	93,75%

Tabela 4.16: Tabela das porcentagens de acertos das metodologias utilizadas

CAPÍTULO 5

Tetraciclinas

Nesta segunda parte do trabalho aplicamos a MIE para uma outra classe de importantes antibióticos que são as tetraciclinas. Utilizamos também os outros métodos de reconhecimento de padrões com o mesmo objetivo dos propostos para o conjunto de esterilquinolinas. Essa parte do trabalho foi realizada em colaboração com o prof. Hélio Ferreira do Santos, do Departamento de Química da Universidade Federal de Juiz de Fora – MG.

A primeira tetraciclina foi isolada em 1947 e desde então outros novos compostos de tetraciclinas foram obtidos. Dentre a grande diversidade de atuação das tetraciclinas podemos citar a ação contra bactérias aeróbicas e anaeróbicas Gram-positivas e Gram-negativas, parede celular de micoplasmas livres, “*chlamydae*”, “*mycobacterium*”, ricketsia, *Helicobacter*, *Listeria* e parasitas protozoários tais como *Entamoeba histolytica*, *Giardia lamblia* e *Plasmodium falciparum*^[48]. As tetraciclinas (inibidores da *E. coli* figuras 5.1 e 5.2) é o primeiro grupo de antibióticos que, devido ao seu grande espectro de atividade, relativa segurança e baixo custo, são muito utilizadas mundialmente e são a segunda após a penicilina em quantidade de toneladas utilizadas por ano.

As tetraciclinas atravessam a parede da célula bacteriana por difusão passiva pelos poros hidrofílicos externos da membrana celular, e depois, através da membrana citoplasmática interna por transporte ativo dependente de energia^{[46],[47]}. “Típicas tetraciclinas”, entre as quais se incluem a tetraciclina, clorotetraciclina, minociclina e deox-

iciclina, impedem o crescimento da bactéria pelas ligações reversíveis nos ribossomos, então inibindo a síntese protéica; eles são agentes bacteriostáticos que têm alta potência de inibição tanto *in vitro* quanto *in vivo*.

As tetraciclinas agem inibindo a interação do aminioacil-tRNA com o sítio “A” da sub-unidade menor 30S do ribossomo 70S durante a etapa de translação do processo de síntese protéica em bactérias. O preciso mecanismo, devido aos aspectos moleculares relacionados à natureza das interações fármacos e receptor, ainda não está bem entendido, porém é possível que as tetraciclinas se liguem ao ribossomo, causando uma distorção tri-dimensional na estrutura do ribossomo ganhando uma vantagem para troca de acessibilidade das bases dentro da subunidade 16s rRNA^[48].

O desenvolvimento de novas drogas tem um processo de alto custo e demanda muito tempo, desde a síntese até os testes biológicos, dificultando a colocação de novos produtos no mercado e também o fácil acesso para a maioria da população. Dentro desse “espírito” aproximações teóricas podem *a priori* selecionar bons candidatos que podem ser muito úteis para reduzir tempo e custo requerido.

A bactéria *E. Coli* é a responsável pela diarréia devido a ingestão de alimentos com mais de 2000 nmp/100ml (unidades de *E. Coli* por 100 ml)^[50]. Muito se tem feito para controlar a proliferação da bactéria *E. Coli*. Nos dias atuais são observadas variedades resistentes da bactéria que já foram encontradas no Brasil, o que é mais um estímulo para o desenvolvimento para drogas inibidoras da *E. Coli*.



Figura 5.1: Figura artística da Bactéria *E. coli*.

As moléculas derivadas das tetraciclinas, apresentadas na tabela 5.1, foram muito utilizadas para inibição do fator-R Resistente da *Escherichia coli*. O índice biológico K_{sens} é o índice que representa a constante da velocidade de inibição e é baseada na lei

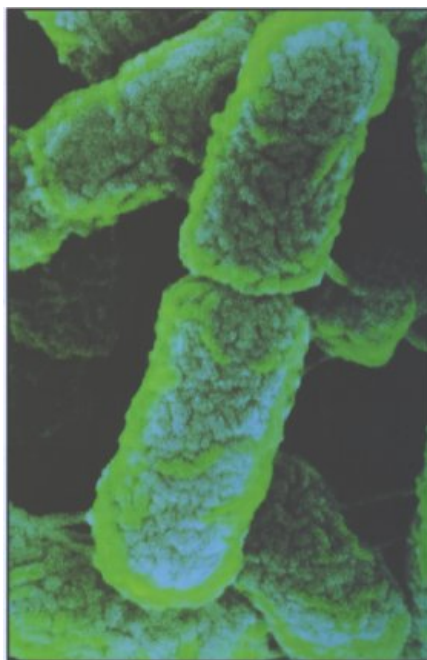


Figura 5.2: Foto de microscopia de uma colônia de *E. coli*.

de velocidade do crescimento celular de primeira ordem^[51].

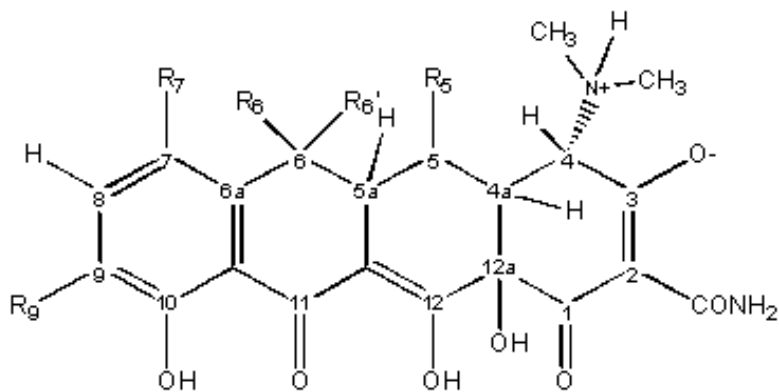


Figura 5.3: Estrutura da Tetraciclina na forma zwitteriônica.

A primeira coluna indica o número do composto, a segunda coluna indica o nome do composto, as colunas de 3 a 7 são os substituintes da figura 5.3, e a última coluna indica o valor utilizado para análise de atividade K_{sens} .

Os compostos abaixo apresentam uma pequena diferença com relação à estrutura básica:

- ♣ —O composto 12 possui um hidrogênio na posição 12a e não um grupo hidroxila;

		Substituintes da Figura 5.3					
Número	Composto	R ₅	R ₆	R _{6'}	R ₇	R ₉	K_{sens}
01	Tetracycline	H	OH	Me	H	H	600,00
02	7-NH₂-6-Demethyl-6-deoxytetracycline	H	H	H	NH ₂	H	190,00
03	5-Hydroxytetracycline	OH	OH	Me	H	H	650,00
04	9-NO₂-6-Demethyl-6-deoxytetracycline	H	H	H	H	NO ₂	145,00
05	9-NH₂-6-Demethyl-6-deoxytetracycline	H	H	H	H	NH ₂	350,00
06	7-Cl-6-Demethyltetracycline	H	OH	H	Cl	H	1100,00
07	7-Cl-Tetracycline	H	OH	Me	Cl	H	950,00
08	7-NH₂-6-Demethyl-6-deoxytetracycline	H	H	H	NO ₂	H	1450,00
09	9-N(CH₃)₂-6-Demethyldoxytetracycline	H	H	H	H	N(Me) ₂	76,00
10	6-Demethyl-6-deoxytetracycline	H	H	H	H	H	250,00
11	6-Demethylamino-6-demethyldeoxytetracycline	H	H	H	N(Me) ₂	H	250,00
♣ 12	5a(6)-Anhydrotetracycline	H	Me	H	H	H	120,00
♠ 13	12a-Deoxytetracycline	H	OH	Me	H	H	8,50
14	7-Br-6-Demethyl-6-deoxytetracycline	H	H	H	Br	H	70,00

Tabela 5.1: Substituintes da estrutura da tetraciclina figura 5.3 que compõem os derivados analisados.

♠ –O composto 13 na posição 11 possui um grupo hidroxila (ao invés de um oxigênio), na posição 12 um oxigênio (ao invés de um grupo hidroxila) e na posição 5a não possui o hidrogênio.

Todos os derivados de tetraciclina aqui estudados estão na forma zwitteriônica. A forma zwitteriônica, neste caso, apresenta porções moleculares com cargas opostas permanecendo a carga total zero na molécula. Neste caso o cátion é o nitrogênio que está ligado ao carbono 4 e o ânion é o oxigênio que está ligado ao carbono 3. As tetraciclinas quando estão em meio aquoso apresentam-se nessa forma^[52].

6.1 Busca Conformacional

Como no caso das esterilquinolinas, as tetraciclinas foram submetidas à busca conformacional detalhada, utilizando os mesmos softwares e os métodos semi-empíricos descritos anteriormente. Realizada a busca conformacional detalhada a estrutura geral das tetraciclinas de menor calor de formação foram as do tipo *extended*. Este confômero recebe este nome pela posição do hidrogênio que está ligado no grupo dimetilamino $[\text{NH}(\text{CH}_3)_2]$ (figura 6.1).

Os confômeros *extended* e *twisted* se diferenciam pela posição do hidrogênio ligado ao nitrogênio, o primeiro apontando para a parte superior da página e o outro apontando para a parte inferior da página. As formas desses confômeros são importantes para o reconhecimento quanto a atividades biológica e a nomenclatura dos compostos, um confômero tipo *extended* ou *twisted*, ambos zwitteriônicos, é a forma em que se encontram em ambiente aquoso e é um inibidor das tetraciclinas tipo resistentes.

Os valores do calor de formação para os compostos mais estáveis estão indicados na tabela 6.1.

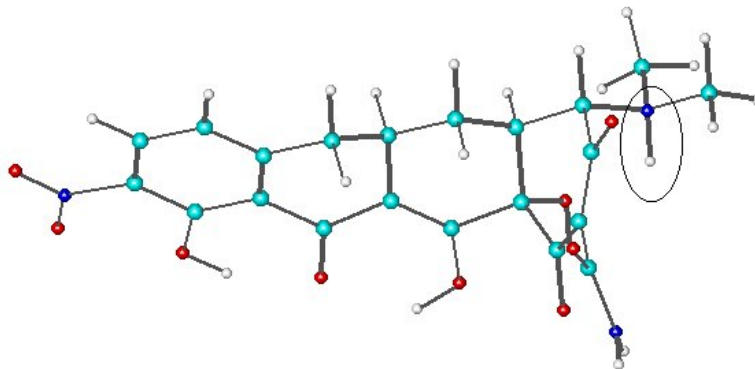


Figura 6.1: Confórmero tipo Extended. O hidrogênio que está ligado ao grupo $N(Me)_2$ está apontando para a parte inferior da página.

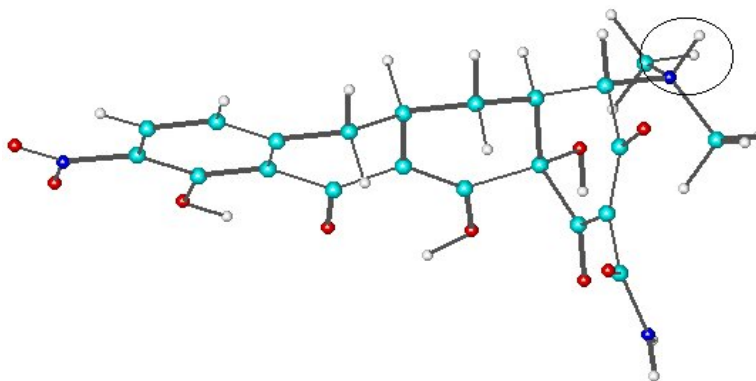


Figura 6.2: Confórmero tipo Twisted. O hidrogênio que está ligado ao grupo $N(Me)_2$ está apontando para a parte superior da página.

Calor de Formação (Kcal/Mol)		
No. Da Molécula	AM1	PM3
01	-244.71	-254.58
02	-201.19	-205.07
03	-286.40	-297.02
04	-194.17	-216.04
05	-201.85	-209.28
06	-251.06	-252.47
07	-240.79	-249.71
08	-198.99	-218.33
09	-194.30	-212.92
10	-201.63	-209.29
11	-193.54	-213.30
12	-190.45	-206.14
13	-198.30	-210.75
14	-198.72	-206.08

Tabela 6.1: Valores do calor de formação para os compostos de mínima energia, referente aos compostos da tabela 5.1

6.2 Cálculo da MIE

Em seguida realizamos o estudo da MIE para diversas regiões da molécula de tetraciclina. A região que melhor propiciou uma regra para os padrões entre moléculas ativas e inativas está indicada na figura 6.3.

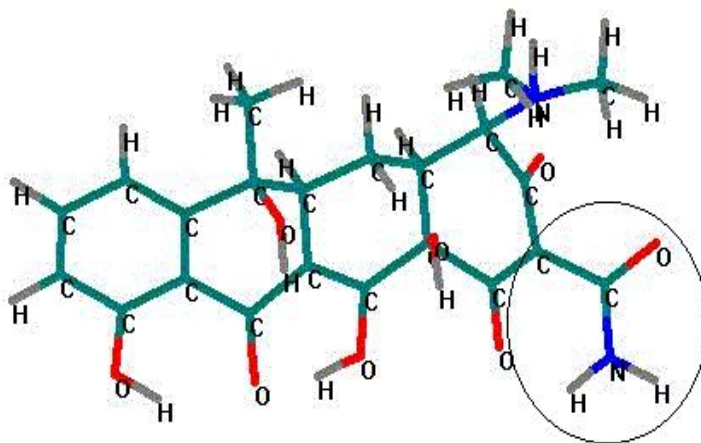


Figura 6.3: Região onde foi realizado o cálculo da LDOS para os índices da MIE.

A região possui os átomos C-CNOH₂. Os parâmetros da MIE para essa região estão na tabela 6.2, para ambos dados os cálculos AM1 e PM3.

Para estabelecer um parâmetro de corte utilizamos valores de K_{sens} abaixo de 100 para moléculas inativas e acima para as moléculas ativas. Para ambos os métodos semi-empíricos a melhor região foi à mesma. Os parâmetros de corte para as regras da MIE (Δ e η) foram:

Regras para AM1

Regra tipo “e”:

Se $\Delta H \geq 0,70$ e $\eta H \geq 1,39$ o composto será inativo, caso não satisfaça alguma dessas condições o composto será ativo. Para esta regra todas as moléculas são classificadas corretamente de acordo com o índice K_{sens} .

Regras para PM3

Regra tipo “ou”:

Se $\Delta H \leq 0,58$ ou $\eta H \geq -0,53$ o composto será ativo, caso não satisfaça alguma dessas condições o composto será inativo. Para esta regra a molécula de n^o 11 é classificada

No. da Molécula	AM1		PM3		K_{sens}	
	ΔH	ηH	ΔH	ηH		
14	0,89087	1,39134	0.92557	-0.55894	70.00	Inativas
9	0,76369	1,39352	0.58841	-0.62668	76.00	
13	0,98351	1,4187	0.63619	-0.56416	8.50	
10	0,86306	1,3317	0.50062	-0.71141	250.00	<i>Ativas</i>
11	0,38482	1,3937	0.59861	-0.62682	250.00	
12	0,182	1,25667	0.28154	1.23213	120.00	
8	1,17464	-0,25789	0.55445	-0.62782	1450.00	
7	0,81229	1,32955	0.57544	-0.62679	950.00	
2	0,10518	-1,37719	0.19437	-1.27861	190.00	
3	1,00269	1,3861	0.89320	-0.52663	650.00	
4	0,9409	-0,5899	0.46468	-0.71502	145.00	
5	0,11706	1,33444	0.31112	1.28549	350.00	
6	0,93911	1,34164	0.46537	-0.70580	1100.00	
1	0,98587	1,38572	0.58068	-0.62704	600.00	

Tabela 6.2: Parâmetros para MIE, para os compostos de tetraciclinas da tabela 5.1 e ambos os dados dos cálculos AM1 e PM3

erroneamente.

6.3 PCA e HCA

Nos cálculos quimiométricos (PCA e HCA), o cálculo de PCA apresentou um acerto correspondente à MIE. O cálculo de HCA apresentou resultados que ficaram fora destes padrões.

Na figura 6.4 o gráfico das moléculas (scores) nos mostra uma total separação entre moléculas ativas e inativas. Este padrão para os resultados obtidos com o AM1 foi obtido depois de uma exaustiva seleção das variáveis, em que concluímos que as de maior importância foram: ΔH , ηH e Log P. Na figura 6.5, que é para os dados das otimizações realizadas com PM3, a separação entre as classes ativas e inativas é mais evidente que com os dados do AM1. Nesta análise a única molécula classificada incorretamente de acordo com o índice K_{sens} foi a molécula de número 11. As variáveis desta análise também foram outras: Calor de Formação, energia do HOMO, ΔH e ηH .

Os resultados de HCA não foram como esperados porque não são observados agrupamentos de moléculas ativas e inativas, como mostrado pelos dendrogramas das figuras

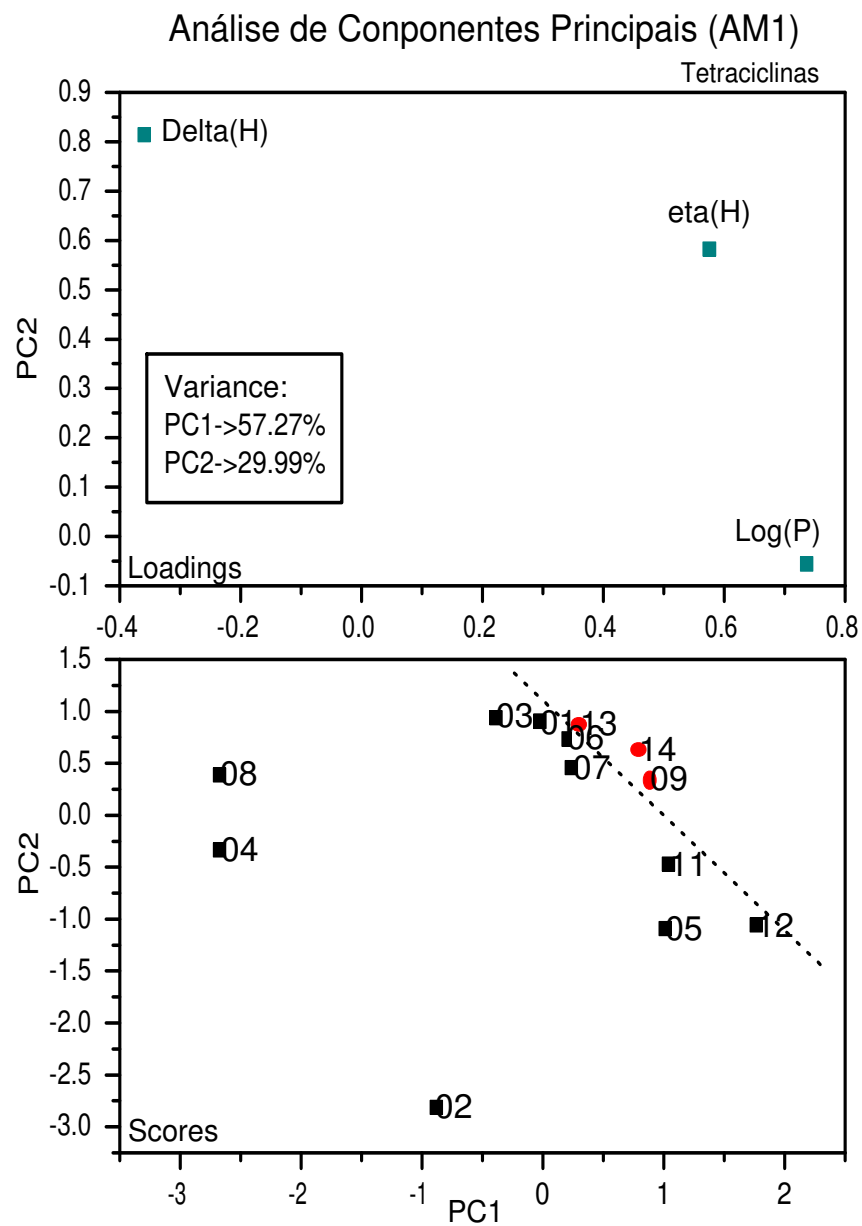


Figura 6.4: PCA - AM1.

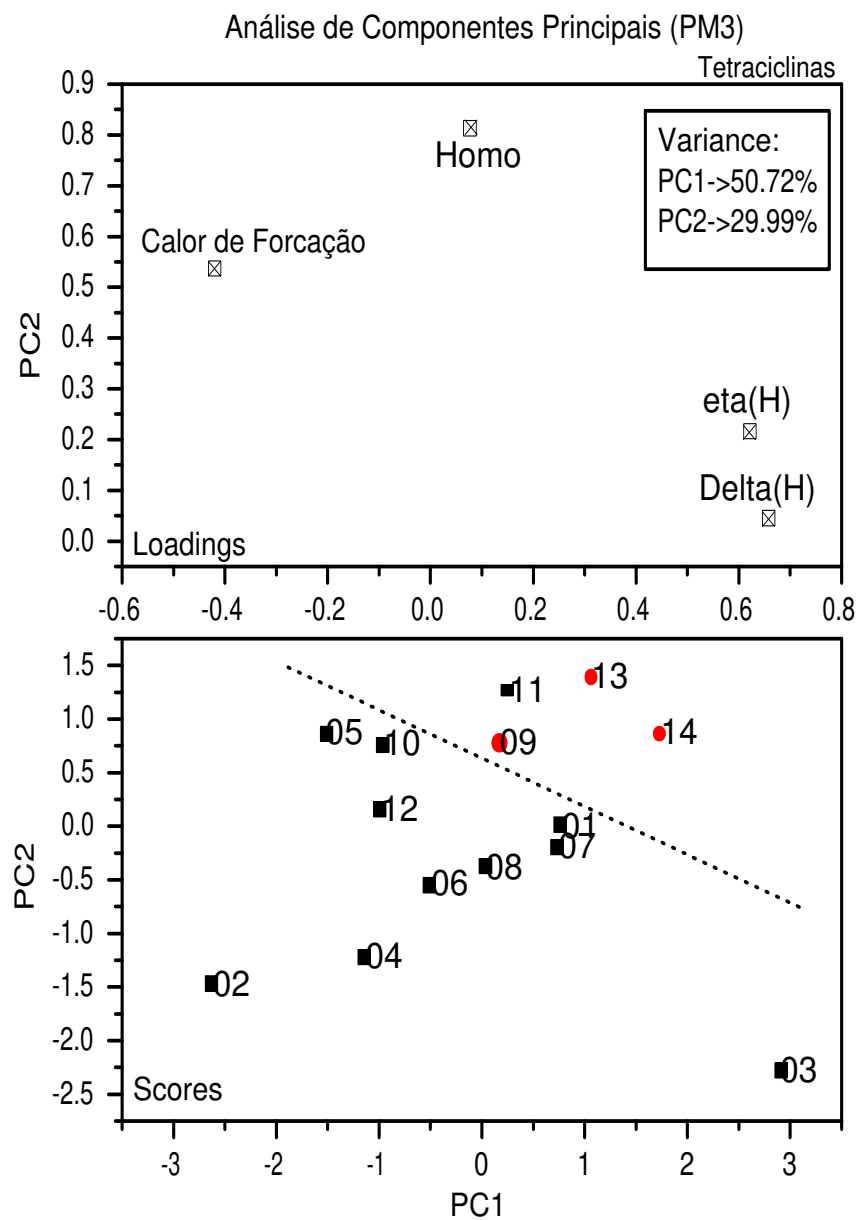


Figura 6.5: PCA - PM3

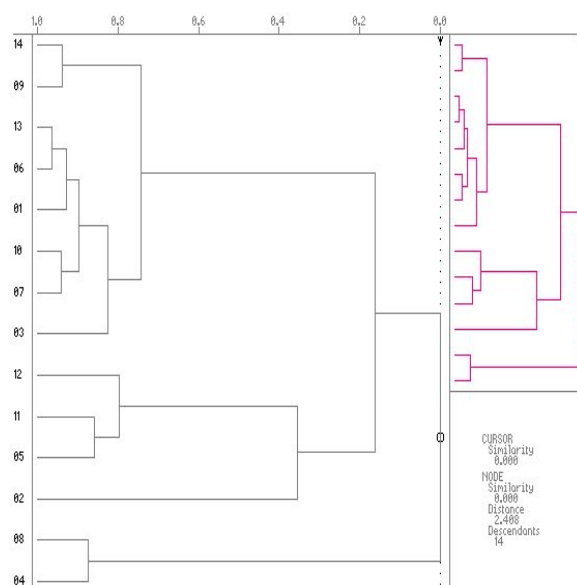


Figura 6.6: HCA - AM1

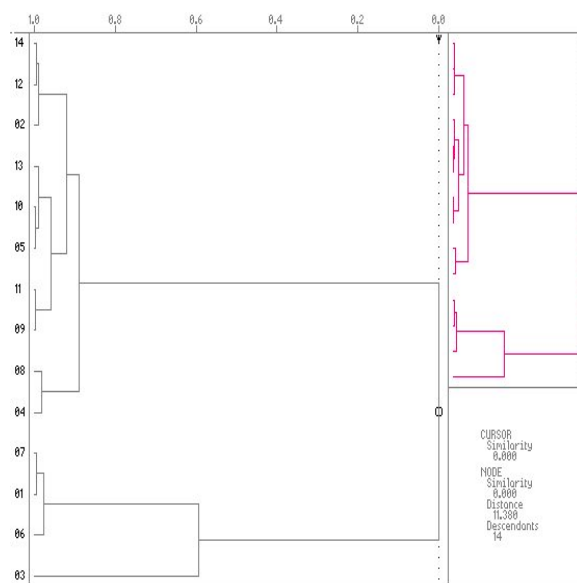


Figura 6.7: HCA - PM3

6.6 e 6.7. As moléculas 09, 13 e 14 que são inativas, deveriam estar em um só agrupamento distinto das outras moléculas.

Os resultados de HCA são consequência dos resultados de PCA, pois a análise de HCA utiliza os resultados do PCA para calcular as distâncias métricas entre as amostras. Os resultados de PCA para as tetraciclinas não tiveram uma faixa de separação suficiente para que o HCA pudesse fazer a distinção entre os grupos de moléculas, apesar dos bons resultados. Nos resultados de PCA para as moléculas de esterilquinolinas as separações entre as classes são bem mais evidentes (têm um intervalo de separação) e as moléculas estão mais agrupadas. É por esse motivo os melhores resultados de HCA para as moléculas de esterilquinolinas. Mesmo neste caso onde o HCA não apresentou bons resultados a MIE apresenta 100% de acerto.

6.4 Novos Compostos

Desenvolvendo a etapa seguinte, após a verificação de padrões de compostos já testados experimentalmente propomos uma série de derivados de tetraciclina que foram classificados (ativos e inativos) a partir do modelo proposto anteriormente.

Baseado nos padrões encontrados e modificando os substituintes dos compostos de tetraciclina, elaboramos um conjunto de 90 novas estruturas de tetraciclina. Primeiro verificamos a possibilidade de permutação dos radicais OH, Me, NO₂, NH₂ e N(Me)₂ nas posições R₅, R₆, R_{6'}, R₇ e R₉. Nestas mesmas posições utilizamos mais alguns elementos da tabela periódica da coluna VIIA para a substituição, que são os elementos F (Flúor) e I (Iodo), os elementos Br (Bromo) e Cl (Cloro) já tinham sido usados anteriormente. Em seguida, com esses possíveis substituintes utilizamos as outras estruturas da tetraciclina semelhantes às estruturas dos compostos 12, 13 e dos demais compostos para a formação das novas tetraciclina.

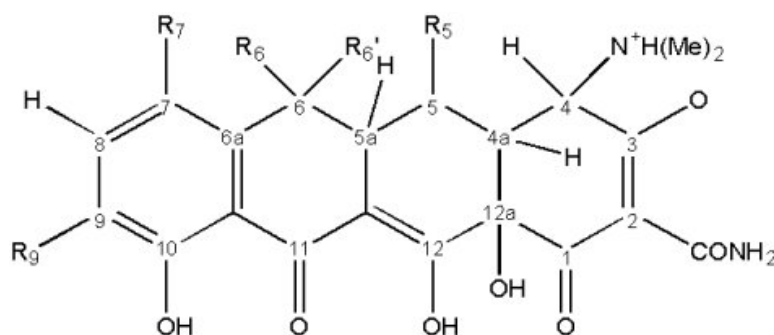


Figura 6.8: Estrutura dos novos substituintes - TA

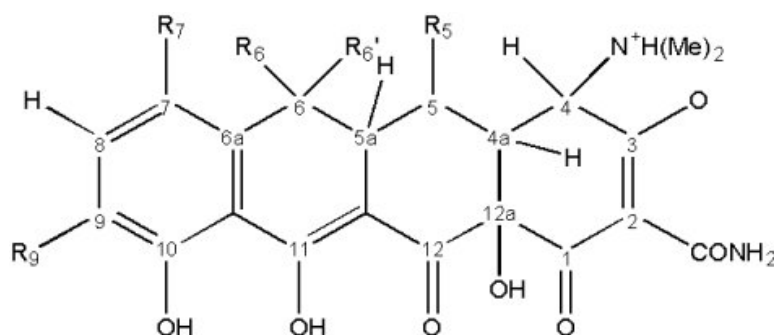


Figura 6.9: Estrutura dos novos substituintes - TB

	R ₅	R ₆	R _{6'}	R ₇	R ₉
01	Br	H	H	H	H
02	Br	H	H	H	Br
03	H	H	H	H	Br
04	H	H	Br	H	H
05	H	H	H	Cl	H
06	Cl	H	H	H	H
07	Cl	H	H	H	Cl
08	H	H	Cl	H	H
09	H	OH	H	Br	H
10	H	OH	Me	Br	H
11	H	OH	H	I	H
12	H	OH	Me	I	H
13	H	OH	H	F	H
14	H	OH	Me	F	H
15	I	H	H	H	H
16	I	H	H	H	I
17	H	H	H	H	I
18	H	H	H	I	H
19	H	H	I	H	H
20	H	H	H	F	H
21	H	OH	Me	NH ₂	H
22	H	OH	Me	NO ₂	H
23	H	OH	Me	H	NH ₂
24	H	OH	Me	H	NO ₂
25	OH	OH	OH	OH	OH
26	OH	H	H	H	N(Me) ₂
27	H	OH	Me	H	N(Me) ₂
28	H	H	H	OH	N(Me) ₂
29	OH	H	H	H	NH ₂
30	OH	H	H	H	NO ₂

Tabela 6.3: Novos substituintes para formação das novas tetraciclinas

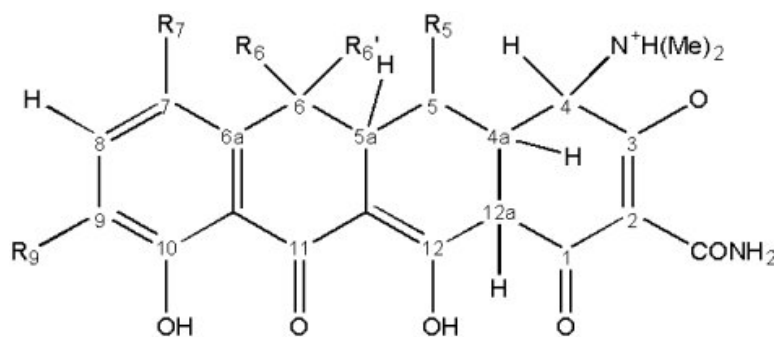


Figura 6.10: Estrutura dos novos substituintes - TC

Na figura TA corresponde à estrutura da tetraciclina inicial, igual à estrutura inicialmente estudada para elaboração do modelo. A estrutura TB possui um grupo OH na posição 11 ao invés de um O, e um O na posição 12 ao invés de um grupo OH. A estrutura TC possui um H ao invés de um grupo OH na posição 12a.

Como o número de compostos para análise é relativamente alto, devido ao tempo disponível realizamos apenas a busca conformacional detalhada para o método semi-empírico AM1, cujos dados produziram a melhor classificação das moléculas ativas e inativas para o conjunto inicialmente estudado. Para verificação dos padrões dessas “novas tetraciclinas” os parâmetros da MIE (ΔH e ηH) estão na tabela 6.4.

As moléculas de 1 a 30 são os compostos dos substituintes da tabela 6.3 com a estrutura TA, de 31 a 60 são os compostos dos substituintes da tabela 6.3 com a estrutura TB, e de 61 a 90 são os compostos dos substituintes da tabela 6.3 com a estrutura TC.

Dentro da regra para os dados do método AM1 classificamos as moléculas (tabela 6.4) em ativas e inativas. As moléculas números: **01, 02, 03, 04, 05, 06, 07, 08, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 26, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 52, 55, 56, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 85 e 86** foram classificadas como inativas, estão destacadas em negrito. As moléculas números: 09, 21, 22, 23, 24, 27, 28, 29, 30, 51, 53, 54, 57, 58, 59, 60, 81, 82, 83, 84, 87, 88, 89 e 90 foram classificadas como ativas.

O próximo estudo realizado com as tetraciclinas foi a aplicação das redes neurais para o conjunto já conhecido (os 14 compostos de tetraciclinas estudados inicialmente) e os novos compostos (os 90 compostos sugeridos). Utilizamos o conjunto com atividade

ΔH	ηH	No.	ΔH	ηH	No.	ΔH	ηH	No.
0.81866	1.38598	1	0.79833	1.38745	31	0.83554	1.41992	61
0.80444	1.38578	2	0.90265	1.38693	32	0.82775	1.42014	62
0.857	1.39171	3	0.92911	1.39352	33	0.86963	1.42355	63
0.93227	1.39044	4	0.86917	1.39272	34	0.94946	1.42225	64
0.82747	1.39082	5	0.83862	1.39356	35	0.8379	1.4219	65
0.81805	1.38653	6	0.80458	1.38813	36	0.83461	1.42026	66
0.7567	1.38645	7	0.88564	1.38765	37	0.77851	1.42103	67
0.95641	1.38996	8	0.893	1.39304	38	0.96989	1.42192	68
1.12838	1.22781	9	1.04288	1.39225	39	0.99903	1.41681	69
0.92687	1.38737	10	0.90953	1.39316	40	0.93523	1.42042	70
1.04079	1.38039	11	1.03177	1.39284	41	1.04845	1.41199	71
0.96759	1.38555	12	0.93654	1.39281	42	0.97592	1.41827	72
0.90679	1.38793	13	0.93194	1.39303	43	0.914	1.42064	73
0.84067	1.38964	14	0.83704	1.39254	44	0.85898	1.42367	74
0.81879	1.38556	15	0.78934	1.38727	45	0.83534	1.4192	75
0.85629	1.38548	16	0.89949	1.38678	46	0.87845	1.41986	76
0.90824	1.39158	17	0.92908	1.39351	47	0.92102	1.42296	77
0.94217	1.38916	18	0.90764	1.39342	48	0.95276	1.41917	78
0.89823	1.39038	19	0.84866	1.3926	49	0.91465	1.42168	79
0.78269	1.39117	20	0.81154	1.39338	50	0.79225	1.42243	80
0.04766	1.39047	21	0.16325	1.39415	51	0.07619	1.42449	81
1.18383	-0.25949	22	1.02315	1.38713	52	1.2335	0.27641	82
0.26312	1.39128	23	0.42919	1.39326	53	0.27454	1.42464	83
1.16637	-0.49712	24	1.19231	-0.55167	54	1.22174	-0.44071	84
0.68029	1.38489	25	0.82486	1.38848	55	0.72508	1.42496	85
0.85328	1.3898	26	0.85905	1.39142	56	0.85231	1.42413	86
0.02959	1.39048	27	0.18808	1.39476	57	0.24952	1.42483	87
0.13244	-1.39217	28	0.03192	1.3938	58	0.04051	1.42429	88
0.17058	1.39045	29	0.33831	1.39157	59	0.20886	1.42474	89
1.15873	-0.37796	30	1.17236	1.3269	60	1.21997	-0.19147	90

Tabela 6.4: Índices da MIE para as nova tetraciclina. Os compostos inativos estão destacados com os caracteres em negrito e os demais compostos são ativos

Quantidade de Moléculas	Classe
24	Ativas
66	Inativas

Tabela 6.5: Resultados da predições das novas tetraciclinas pela MIE.

biológica experimental já conhecida para treinarmos as redes neurais cujos parâmetros da rede e os erros no treinamento estão indicados na tabela 6.6.

Conjunto	1
ALPHA	3,0
HALPHA	3,0
HDNLAYER	20
EPSIRON	0,20
HEPSIRON	0,20
BIAS	sim
No. De Interações	100.000
Precisão	10^{-5}
Erros no Treinamento	0

Tabela 6.6: Parâmetros de treinamento da rede neural, baseado na atividade biológica experimental. Foram utilizados os índices eletrônicos da MIE, do cálculos AM1, como entrada para a rede.

Observamos que a rede neural verifica o padrão estabelecido pelos índices experimentais biológicos e obteve o mesmo índice de acerto que a MIE, não errando nenhum composto na fase de treinamento. Utilizando deste padrão de treinamento realizamos a predição das novas tetraciclinas, como pode ser visto na tabela 6.7.

A tabela 6.7 nos mostra os resultados de predição da MIE e também da rede neural. Pode-se verificar que os resultados obtidos pelos dois métodos são semelhantes e pode ser conferidos observando a classe em que cada método selecionou cada um dos compostos. A classe tipo 1 é para o composto ativo e a classe tipo 2 e para o composto inativo, tendo base nos índices experimentais biológicos. Apenas os compostos 11, 12, 39, 41, 52 e 71 apresentaram estar em desacordo com ambas metodologias. A classificação dos novos compostos pela rede neural apresentou estar em acordo com a MIE. Contudo podemos agora separar as novas tetraciclinas, entre aquelas que tem atividade biológica

RN		MIE	RN		MIE	RN		MIE
No.	Pred.	Class.	No.	Pred.	Class.	No.	Pred.	Class.
1	0.002	0.998	2	2	31	0.001	0.999	2
2	0.001	0.998	2	2	32	0.017	0.983	2
3	0.001	0.999	2	2	33	0.015	0.986	2
4	0.044	0.959	2	2	34	0.001	0.999	2
5	0.001	0.999	2	2	35	0.000	0.999	2
6	0.002	0.998	2	2	36	0.001	0.999	2
7	0.001	0.999	2	2	37	0.007	0.993	2
8	0.383	0.633	2	2	38	0.002	0.998	2
9	1.000	0.000	1	1	39	1.000	0.000	1
10	0.066	0.938	2	2	40	0.005	0.995	2
11	1.000	0.000	1	2	41	1.000	0.000	1
12	0.924	0.080	1	2	42	0.032	0.970	2
13	0.016	0.984	2	2	43	0.021	0.980	2
14	0.001	0.999	2	2	44	0.001	0.999	2
15	0.002	0.998	2	2	45	0.001	0.999	2
16	0.005	0.995	2	2	46	0.015	0.985	2
17	0.007	0.993	2	2	47	0.015	0.986	2
18	0.138	0.870	2	2	48	0.004	0.996	2
19	0.006	0.994	2	2	49	0.001	0.999	2
20	0.000	1.000	2	2	50	0.000	1.000	2
21	1.000	0.000	1	1	51	1.000	0.000	1
22	1.000	0.000	1	1	52	1.000	0.000	1
23	1.000	0.000	1	1	53	1.000	0.000	1
24	1.000	0.000	1	1	54	1.000	0.000	1
25	0.004	0.995	2	2	55	0.001	0.999	2
26	0.002	0.998	2	2	56	0.001	0.999	2
27	1.000	0.000	1	1	57	1.000	0.000	1
28	1.000	0.000	1	1	58	1.000	0.000	1
29	1.000	0.000	1	1	59	1.000	0.000	1
30	1.000	0.000	1	1	60	1.000	0.000	1

Tabela 6.7: Resultados da predição dos novos derivados de Tetraciclina com as redes neurais.

Quantidade de Moléculas	Classe
30	Ativas
60	Inativas

Tabela 6.8: Resultados da predições das novas tetraciclinas pela RN.

inibitória e as que não tem atividade. Para os compostos que tiveram discordância entre os métodos existe a necessidade de se realizar os testes experimentais para apurar a atividade biológica.

CAPÍTULO 7

Sumário dos Resultados

7.1 Esterilquinolinas

Com os derivados de esterilquinolinas analisamos um total de 32 compostos sendo 19 compostos ativos e 13 inativos, em que a classificação inicial foi dada pelo índice experimental IC_{50} . Primeiramente foi realizada uma busca conformacional detalhada de todos os compostos com os métodos semi-empíricos AM1 e PM3. Em seguida utilizamos a nova metodologia MIE para obter o reconhecimento do padrão inicial (experimental). Para efeito de comparação e validação dos dados da MIE utilizamos outras metodologias de amplo reconhecimento acadêmico e comercial que foram: a PCA, a HCA e as Redes Neurais Artificiais. Na tabela 7.1 podemos verificar o quanto cada metodologia consegue chegar ao padrão proposto inicialmente pelos dados experimentais. Na análise de PCA e HCA, iniciamos com 93 descritores e chegamos, depois de uma grande otimização, a somente 3 descritores em que 2 são descritores da MIE. Com a rede neural utilizamos os descritores sugeridos pelo cálculo da PCA como dados de entrada e depois somente os descritores da MIE como dados de entrada.

	AM1	PM3
MIE	93.75%	93.75%
PCA	90.62%	93.75%
HCA	90.62%	93.75%
RN/PCA	93,75%	100%
RN/MIE	93,75%	93,75%

Tabela 7.1: Tabela das porcentagens de acertos das metodologias utilizadas para as Esterilquinolinas

7.2 Tetraciclinas

Com os derivados de tetraciclinas analisamos um total de 14 compostos sendo 11 compostos ativos e 3 inativos, em que a classificação inicial foi dada pelo índice experimental K_{sens} . Primeiramente foi realizada uma busca conformacional detalhada de todos os compostos com os métodos semi-empíricos AM1 e PM3. Em seguida utilizamos a nova metodologia MIE para obter o reconhecimento do padrão inicial (experimental). Para efeito de comparação e validação dos dados da MIE utilizamos outras metodologias de amplo reconhecimento acadêmico e comercial que foram: a PCA, a HCA e as Redes Neurais Artificiais. Na tabela 7.2 podemos verificar quanto que cada metodologia consegue chegar ao padrão proposto inicialmente pelos dados experimentais. Na análise de PCA e HCA, iniciamos com 93 descritores e chegamos, depois de uma grande otimização dos parâmetros, a somente 3/4 descritores para os dados de AM1/PM3 em que 2/2 são descritores da MIE. A HCA não conseguiu averiguar o padrão experimental, isto porque a separação entre as moléculas ativas e inativas não é tão evidente como no caso das esterilquinolinas. Em seguida fizemos a proposta de 90 novos derivados de tetraciclinas entre compostos ativos e inativos. Nesta parte do trabalho utilizamos as redes neurais da seguinte forma: utilizamos o conjunto dos 14 compostos para realizar o treinamento da rede neural e depois submetemos os novos compostos para predição. A MIE previu 24 compostos ativos e 66 inativos e a rede neural previu 30 compostos ativos e 60 inativos, dentre os 90 compostos submetidos a previsão somente 6 apresentaram estar em desacordo com as MIE e a rede neural.

	AM1	PM3
MIE	100.00%	92.85%
PCA	100.00%	92.85%
HCA	—	—
RN Trein.	100.00%	—

Tabela 7.2: Tabela das porcentagens de acertos das metodologias utilizadas para o conjunto das Tetraciclinas

CAPÍTULO 8

Conclusões

Neste trabalho foi possível verificar a grande importância dos índices eletrônicos da MIE. Nas metodologias PCA, HCA e Redes Neurais foi necessária utilização dos parâmetros da MIE para conseguirmos ter os padrões previamente estabelecidos experimentalmente. Na PCA e HCA a dependência dos parâmetros da MIE é mais evidente para ambos os conjuntos moleculares estudados.

A aplicação da rede neural demonstrou ser muito eficiente para verificação de padrões. Durante o processo de treinamento e predição foi encontrada uma grande dificuldade do método que é o ajuste dos parâmetros relacionados aos pesos (conectividade) entre os neurônios. Neste tratamento foram utilizadas seis variáveis; dentre elas, cinco que podem estar dentro de espectro de 0,01-4,00, o que torna o método lento, necessitando de muitas horas de cpu e trabalho humano. Nos métodos PCA e HCA o tempo de trabalho para se chegar em um resultado satisfatório também é grande. Inicialmente entramos com um número grande de descritores, depois da tentativa sistemática de combinação dos descritores é que obtivemos resultados satisfatórios. Superficialmente temos que realizar a combinação de 93 descritores e selecionarmos os melhores, os que indicam algum padrão, para otimização do modelo a ser proposto.

A MIE apresentou um grande desempenho na verificação dos padrões experimentais, o possibilita uma exploração rápida e eficaz de padrões de compostos biologicamente ativos. Talvez, uma parte que impossibilitaria o desenvolvimento rápido da MIE seria

a determinação de uma região para o cálculo da LDOS, referente ao índice η , que nos fornecesse um bom padrão. Tendo em vista esta dificuldade foi implementado dentro do programa Chem2Pac um algoritmo para realizar o cálculo da LDOS da região molecular desejada. Concluimos de fato que a MIE é muito eficiente na verificação de padrões e se torna muito mais rápida que os outros métodos aqui aplicados.

Um grande passo dado na análise destes compostos foi para o conjunto dos derivados das tetraciclinas; pois quase não se encontram trabalhos teóricos na literatura devido ao grau de dificuldade encontrado para estabelecer os padrões de forma teórica de atividade biológica, visto que os trabalhos iniciais são da década de 40. Propomos dentre 90 compostos 27 que poderiam ser ativos. Nenhum outro trabalho do gênero foi publicado nesses anos de pesquisa sobre as tetraciclinas.

Em resumo, a metodologia dos Índices Eletrônicos (MIE) quando aplicada às esterilquinolinas e tetraciclinas, se mostrou altamente eficiente na classificação/separação de compostos ativos e inativos.

Esses resultados se somam às diferentes classes de outros compostos já estudados ^[22], ^[23], ^[24], ^[25], ^[26], ^[27], ^[30] totalizando mais de 2000 compostos onde os mesmos descritores da MIE se mostraram eficientes na classificação biológica. Essa aparente “classe de universalidade” dos descritores é certamente intrigante. Como é possível que com apenas dois parâmetros classes completamente distintas de compostos (carcinogênicos, antibióticos, hormônios, inibidores de proteínas, antitumorais, etc) possam ser enquadrados mecanismos bioquímicos?

Essa “universalidade” dos parâmetros da MIE sugere que talvez alguma lei de escala (fractal/caótica) esteja presente. Estudos mais detalhados neste sentido se fazem necessários. Caso isto possa ser provado abririam-se novas e interessantes perspectivas para o *design* de novos fármacos com significativa redução de tempo e custos operacionais.

CAPÍTULO 9

Apêndice A

Resultados - Redes Neurais - Esterilquinolinas

Esterilquinolinas - AM1

Conjunto de Treinamento 01					Conjunto de Treinamento 02				
Molécula	Treinamento Padrão		Resultado do Treinamento		Molécula	Treinamento Padrão		Resultado do Treinamento	
3	1	0	1,000	0,000	3	1	0	1,000	0,000
6	1	0	0,000	1,000	6	1	0	0,946	0,055
8	1	0	0,997	0,003	7	1	0	1,000	0,000
9	1	0	0,998	0,002	8	1	0	0,757	0,246
10	1	0	0,994	0,006	12	1	0	1,000	0,000
13	1	0	0,997	0,003	13	1	0	1,000	0,000
14	1	0	0,990	0,010	15	1	0	0,963	0,039
18	1	0	0,998	0,002	16	1	0	1,000	0,000
19	1	0	0,989	0,011	17	1	0	1,000	0,000
22	1	0	0,000	1,000	19	1	0	0,733	0,273
20	0	1	0,015	0,985	1	0	1	0,001	0,999
23	0	1	0,002	0,998	20	0	1	0,211	0,796
25	0	1	0,002	0,998	21	0	1	0,000	1,000
29	0	1	0,000	1,000	24	0	1	0,000	1,000
30	0	1	0,000	1,000	26	0	1	0,000	1,000
31	0	1	0,000	1,000	27	0	1	0,001	0,999
32	0	1	0,000	1,000	28	0	1	0,000	1,000
Conjunto de Treinamento 01					Conjunto de Predição 02				
Molécula	Padrão da Predição		Resultado da Predição		Molécula	Padrão da Predição		Resultado da Predição	
2	1	0	0,998	0,002	2	1	0	0,999	0,001
4	1	0	0,996	0,004	4	1	0	1,000	0,000
5	1	0	0,999	0,001	5	1	0	1,000	0,000
7	1	0	0,933	0,066	9	1	0	1,000	0,000
11	1	0	0,997	0,003	10	1	0	0,000	1,000
12	1	0	0,996	0,004	11	1	0	1,000	0,000
15	1	0	0,014	0,986	14	1	0	1,000	0,000
16	1	0	0,997	0,003	18	1	0	0,744	0,260
17	1	0	0,997	0,003	22	1	0	0,000	1,000
1	0	1	0,000	1,000	23	0	1	0,003	0,997
21	0	1	0,000	1,000	25	0	1	0,000	1,000
24	0	1	0,010	0,989	29	0	1	0,001	0,999
26	0	1	0,000	1,000	30	0	1	0,035	0,963
27	0	1	0,000	1,000	31	0	1	0,000	1,000
28	0	1	0,000	1,000	32	0	1	0,071	0,925

Esterilquinolinas – PM3

Conjunto de Treinamento 03					Conjunto de Treinamento 04				
Molécula	Treinamento Padrão		Resultado do Treinamento		Molécula	Treinamento Padrão		Resultado do Treinamento	
2	1	0	1,000	0,000	3	1	0	1,000	0,000
7	1	0	1,000	0,000	4	1	0	1,000	0,000
12	1	0	1,000	0,000	6	1	0	1,000	0,000
13	1	0	1,000	0,000	7	1	0	1,000	0,000
15	1	0	1,000	0,000	8	1	0	1,000	0,000
14	1	0	1,000	0,000	9	1	0	1,000	0,000
17	1	0	1,000	0,000	11	1	0	1,000	0,000
18	1	0	1,000	0,000	16	1	0	1,000	0,000
19	1	0	1,000	0,000	17	1	0	1,000	0,000
22	1	0	0,995	0,005	22	1	0	1,000	0,000
20	0	1	1,000	0,000	20	0	1	1,000	0,000
21	0	1	0,003	0,997	24	0	1	0,000	1,000
23	0	1	0,000	1,000	27	0	1	0,000	1,000
25	0	1	0,001	0,999	28	0	1	0,000	1,000
27	0	1	0,000	1,000	29	0	1	0,000	1,000
29	0	1	0,000	1,000	30	0	1	0,000	1,000
32	0	1	0,001	0,999	32	0	1	0,000	1,000
Conjunto de Treinamento 03					Conjunto de Predição 04				
Molécula	Padrão da Predição		Resultado da Predição		Molécula	Padrão da Predição		Resultado da Predição	
3	1	0	1,000	0,000	3	1	0	1,000	0,000
4	1	0	1,000	0,000	5	1	0	1,000	0,000
5	1	0	1,000	0,000	10	1	0	0,176	0,846
6	1	0	0,000	1,000	12	1	0	1,000	0,000
8	1	0	1,000	0,000	13	1	0	1,000	0,000
9	1	0	1,000	0,000	14	1	0	1,000	0,000
10	1	0	1,000	0,000	15	1	0	1,000	0,000
11	1	0	1,000	0,000	18	1	0	1,000	0,000
16	1	0	1,000	0,000	19	1	0	1,000	0,007
1	0	1	0,000	1,000	1	0	1	1,000	0,000
24	0	1	0,002	0,998	21	0	1	0,999	0,000
26	0	1	0,000	1,000	23	0	1	1,000	0,000
28	0	1	0,000	1,000	25	0	1	0,001	1,000
30	0	1	0,000	1,000	26	0	1	0,048	0,766
31	0	1	0,000	1,000	31	0	1	0,001	0,995

Conjunto de Treinamento 05					Conjunto de Treinamento 06				
Molécula	Treinamento Padrão		Resultado do Treinamento		Molécula	Treinamento Padrão		Resultado do Treinamento	
2	1	0	1,000	0,000	2	1	0	1,000	0,000
5	1	0	1,000	0,000	5	1	0	1,000	0,000
9	1	0	1,000	0,000	6	1	0	1,000	0,000
10	1	0	1,000	0,000	9	1	0	1,000	0,000
11	1	0	0,999	0,001	10	1	0	1,000	0,000
14	1	0	0,999	0,001	11	1	0	1,000	0,000
15	1	0	0,999	0,002	12	1	0	1,000	0,000
16	1	0	0,999	0,001	15	1	0	1,000	0,000
17	1	0	0,998	0,002	17	1	0	1,000	0,000
22	1	0	0,995	0,005	22	1	0	0,997	0,003
21	0	1	0,002	0,998	20	1	0	1,000	0,000
24	0	1	0,000	1,000	21	1	0	0,001	0,999
25	0	1	0,002	0,998	25	1	0	0,000	1,000
27	0	1	0,003	0,997	28	1	0	0,000	1,000
28	0	1	0,003	0,997	29	1	0	0,001	0,999
29	0	1	0,000	1,000	30	1	0	0,000	1,000
31	0	1	0,002	0,998	31	1	0	0,002	0,998
Conjunto de Treinamento 05					Conjunto de Predição 06				
Molécula	Padrão da Predição		Resultado da Predição		Molécula	Padrão da Predição		Resultado da Predição	
3	1	0	1,000	0,000	3	1	0	0,987	0,014
4	1	0	1,000	0,000	4	1	0	1,000	0,000
6	1	0	0,000	1,000	7	1	0	1,000	0,000
7	1	0	0,999	0,001	8	1	0	1,000	0,000
8	1	0	0,999	0,001	13	1	0	1,000	0,000
12	1	0	0,999	0,001	14	1	0	1,000	0,000
13	1	0	0,999	0,001	16	1	0	1,000	0,000
18	1	0	0,999	0,001	18	1	0	0,998	0,002
19	1	0	1,000	0,000	19	1	0	1,000	0,000
1	0	1	0,069	0,951	1	0	1	0,220	0,784
20	0	1	0,999	0,001	23	0	1	0,026	0,981
23	0	1	1,000	0,000	24	0	1	0,000	1,000
26	0	1	0,000	1,000	26	0	1	0,351	0,649
30	0	1	0,000	1,000	27	0	1	1,000	0,000
32	0	1	0,000	1,000	32	0	1	0,000	1,000

Conjunto de Treinamento 07					Conjunto de Treinamento 08				
Molécula	Treinamento Padrão		Resultado do Treinamento		Molécula	Treinamento Padrão		Resultado do Treinamento	
3	1	0	1,000	0,000	2	1	0	1,000	0,000
6	1	0	0,998	0,000	5	1	0	1,000	0,000
7	1	0	0,999	0,000	6	1	0	1,000	0,000
10	1	0	0,999	0,003	8	1	0	0,995	0,005
11	1	0	0,999	0,000	11	1	0	1,000	0,000
12	1	0	0,999	0,000	12	1	0	1,000	0,000
13	1	0	1,000	0,000	14	1	0	0,995	0,005
14	1	0	0,999	0,000	15	1	0	0,996	0,004
15	1	0	0,999	0,000	18	1	0	0,993	0,008
19	1	0	1,000	0,000	22	1	0	0,993	0,007
1	0	1	0,001	0,998	1	0	1	0,001	0,998
20	0	1	1,000	0,000	20	0	1	0,010	0,990
24	0	1	0,001	1,000	21	0	1	0,007	0,993
25	0	1	0,000	1,000	24	0	1	0,000	1,000
27	0	1	0,003	0,998	25	0	1	0,000	1,000
29	0	1	0,001	0,999	31	0	1	0,001	0,999
32	0	1	0,000	1,000	32	0	1	0,000	1,000
Conjunto de Treinamento 07					Conjunto de Predição 08				
Molécula	Padrão da Predição		Resultado da Predição		Molécula	Padrão da Predição		Resultado da Predição	
2	1	0	0,999	0,000	3	1	0	1,000	0,000
4	1	0	0,999	0,000	4	1	0	1,000	0,000
5	1	0	0,999	0,000	7	1	0	1,000	0,000
8	1	0	1,000	0,000	9	1	0	1,000	0,000
9	1	0	0,999	0,000	10	1	0	0,969	0,027
16	1	0	1,000	0,000	13	1	0	1,000	0,000
17	1	0	1,000	0,000	16	1	0	0,997	0,003
18	1	0	1,000	0,000	17	1	0	0,898	0,116
22	1	0	0,000	1,000	19	1	0	0,049	0,950
21	0	1	0,002	1,000	23	0	1	0,000	0,999
23	0	1	0,001	0,998	26	0	1	0,960	0,040
26	0	1	0,002	0,999	27	0	1	1,000	0,000
28	0	1	0,000	1,000	28	0	1	0,000	1,000
30	0	1	0,002	0,999	29	0	1	0,000	1,000
31	0	1	0,002	0,999	30	0	1	0,015	0,986

Conjunto de Treinamento 09					Conjunto de Treinamento 10				
Molécula	Treinamento Padrão		Resultado do Treinamento		Molécula	Treinamento Padrão		Resultado do Treinamento	
2	1	0	0,9980	0,000	2	1	0	1,000	0,000
3	1	0	1,000	0,000	3	1	0	1,000	0,000
5	1	0	1,000	0,000	4	1	0	1,000	0,000
8	1	0	1,000	0,000	5	1	0	1,000	0,000
9	1	0	0,999	0,001	6	1	0	0,990	0,010
10	1	0	0,999	0,001	11	1	0	1,000	0,000
15	1	0	1,000	0,000	12	1	0	1,000	0,000
17	1	0	1,000	0,000	14	1	0	1,000	0,000
19	1	0	1,000	0,000	15	1	0	1,000	0,000
22	1	0	0,997	0,003	17	1	0	1,000	0,000
1	0	1	0,000	1,000	1	0	1	0,000	1,000
23	0	1	0,001	0,999	23	0	1	0,006	0,995
24	0	1	0,002	0,998	25	0	1	0,007	0,993
26	0	1	0,003	0,997	28	0	1	0,000	1,000
29	0	1	0,000	1,000	29	0	1	0,000	1,000
30	0	1	0,000	1,000	30	0	1	0,007	0,992
32	0	1	0,000	1,000	31	0	1	0,000	1,000
Conjunto de Treinamento 09					Conjunto de Predição 10				
Molécula	Padrão da Predição		Resultado da Predição		Molécula	Padrão da Predição		Resultado da Predição	
4	1	0	0,998	0,002	7	1	0	1,000	0,000
6	1	0	0,000	1,000	8	1	0	1,000	0,000
7	1	0	0,999	0,001	9	1	0	1,000	0,000
11	1	0	1,000	0,000	10	1	0	0,999	0,001
12	1	0	0,999	0,001	13	1	0	1,000	0,000
13	1	0	1,000	0,000	16	1	0	1,000	0,000
14	1	0	1,000	0,000	18	1	0	1,000	0,000
16	1	0	1,000	0,000	19	1	0	1,000	0,000
18	1	0	1,000	0,000	22	1	0	0,000	1,000
20	0	1	1,000	0,000	20	0	1	1,000	0,000
21	0	1	0,099	0,900	21	0	1	0,000	1,000
25	0	1	0,003	0,997	24	0	1	0,007	0,993
27	0	1	0,000	1,000	26	0	1	0,000	1,000
28	0	1	1,000	0,000	27	0	1	0,001	0,999
31	0	1	0,047	0,953	32	0	1	0,649	0,336

Conjunto de Treinamento 01					Conjunto de Treinamento 02				
Molécula	Treinamento Padrão		Resultado do Treinamento		Molécula	Treinamento Padrão		Resultado do Treinamento	
3	1	0	0,999	0,000	3	1	0	1,000	0,000
6	1	0	0,005	0,992	6	1	0	0,997	0,003
8	1	0	0,997	0,002	7	1	0	1,000	0,000
9	1	0	0,999	0,000	8	1	0	0,998	0,002
10	1	0	0,990	0,006	12	1	0	1,000	0,000
13	1	0	0,999	0,000	13	1	0	1,000	0,000
14	1	0	0,998	0,001	15	1	0	0,999	0,001
18	1	0	0,971	0,016	16	1	0	1,000	0,000
19	1	0	0,999	0,001	17	1	0	1,000	0,000
22	1	0	0,992	0,009	19	1	0	0,999	0,001
20	0	1	0,000	1,000	1	0	1	0,002	0,998
23	0	1	0,000	1,000	20	0	1	0,003	0,997
25	0	1	0,000	1,000	21	0	1	0,000	1,000
29	0	1	0,001	1,000	24	0	1	0,001	0,999
30	0	1	0,001	1,000	26	0	1	0,002	0,998
31	0	1	0,001	1,000	27	0	1	0,001	0,999
32	0	1	0,001	1,000	28	0	1	0,000	1,000
Conjunto de Treinamento 01					Conjunto de Predição 02				
Molécula	Padrão da Predição		Resultado da Predição		Molécula	Padrão da Predição		Resultado da Predição	
2	1	0	0,999	0,000	2	1	0	1,000	0,000
4	1	0	0,999	0,001	4	1	0	0,999	0,001
5	1	0	0,961	0,022	5	1	0	0,996	0,004
7	1	0	1,000	0,000	9	1	0	1,000	0,000
11	1	0	0,999	0,000	10	1	0	0,999	0,001
12	1	0	0,999	0,000	11	1	0	1,000	0,000
15	1	0	0,999	0,000	14	1	0	0,999	0,001
16	1	0	0,999	0,000	18	1	0	0,988	0,012
17	1	0	0,999	0,000	22	1	0	0,000	1,000
1	0	1	0,001	0,999	23	0	1	0,001	0,999
21	0	1	0,003	0,999	25	0	1	0,033	0,967
24	0	1	0,000	0,999	29	0	1	0,151	0,848
26	0	1	0,001	1,000	30	0	1	0,034	0,965
27	0	1	0,001	1,000	31	0	1	0,001	0,999
28	0	1	0,023	0,992	32	0	1	0,997	0,003

Esterilquinolina – AM1 - MIE

Conjunto de Treinamento 03					Conjunto de Treinamento 04				
Molécula	Treinamento Padrão		Resultado do Treinamento		Molécula	Treinamento Padrão		Resultado do Treinamento	
2	1	0	0,999	0,001	2	1	0	0,999	0,001
7	1	0	0,998	0,002	4	1	0	1,000	0,000
12	1	0	0,999	0,001	6	1	0	0,990	0,010
13	1	0	0,999	0,001	7	1	0	0,993	0,007
14	1	0	0,997	0,003	8	1	0	1,000	0,000
15	1	0	0,999	0,002	9	1	0	0,999	0,001
17	1	0	0,999	0,001	11	1	0	0,999	0,001
18	1	0	0,992	0,008	16	1	0	1,000	0,000
19	1	0	0,997	0,003	17	1	0	1,000	0,000
22	1	0	0,992	0,008	22	0	1	0,994	0,006
20	0	1	0,006	0,995	20	0	1	0,008	0,992
21	0	1	0,007	0,993	24	0	1	0,004	0,996
23	0	1	0,000	1,000	27	0	1	0,007	0,993
25	0	1	0,009	0,991	28	0	1	0,005	0,994
27	0	1	0,000	1,000	29	0	1	0,002	0,998
29	0	1	0,000	1,000	30	0	1	0,007	0,993
32	0	1	0,000	1,000	32	0	1	0,000	1,000
Conjunto de Treinamento 03					Conjunto de Predição 04				
Molécula	Padrão da Predição		Resultado da Predição		Molécula	Padrão da Predição		Resultado da Predição	
3	1	0	0,998	0,002	3	1	0	0,995	0,005
4	1	0	0,998	0,002	5	1	0	0,995	0,005
5	1	0	0,993	0,008	10	1	0	1,000	0,000
6	1	0	0,000	1,000	12	1	0	1,000	0,000
8	1	0	0,997	0,004	13	1	0	1,000	0,000
9	1	0	0,999	0,001	14	1	0	1,000	0,000
10	1	0	0,973	0,030	15	1	0	1,000	0,000
11	1	0	0,999	0,001	18	1	0	1,000	0,000
16	1	0	0,999	0,001	19	1	0	1,000	0,000
1	0	1	0,000	1,000	1	0	1	0,002	0,998
24	0	1	0,009	0,991	21	0	1	0,006	0,994
26	0	1	0,000	1,000	23	0	1	0,082	0,918
28	0	1	0,035	0,972	25	0	1	0,999	0,001
30	0	1	0,000	1,000	26	0	1	0,006	0,994
31	0	1	0,000	1,000	31	0	1	0,005	0,996

Conjunto de Treinamento 05					Conjunto de Treinamento 06				
Molécula	Treinamento Padrão		Resultado do Treinamento		Molécula	Treinamento Padrão		Resultado do Treinamento	
2	1	0	0,999	0,001	2	1	0	1,000	0,000
5	1	0	0,999	0,001	5	1	0	1,000	0,000
9	1	0	0,999	0,001	6	1	0	0,992	0,008
10	1	0	0,999	0,001	9	1	0	1,000	0,000
11	1	0	0,999	0,001	10	1	0	0,998	0,002
14	1	0	0,999	0,001	11	1	0	1,000	0,000
15	1	0	0,999	0,001	12	1	0	1,000	0,000
16	1	0	0,999	0,001	15	1	0	0,999	0,001
17	1	0	0,998	0,002	17	1	0	0,995	0,005
22	1	0	0,997	0,003	22	1	0	0,995	0,005
21	0	1	0,002	0,998	20	0	1	0,005	0,995
24	0	1	0,001	0,999	21	0	1	0,003	0,997
25	0	1	0,002	0,998	25	0	1	0,005	0,995
27	0	1	0,001	0,999	28	0	1	0,003	0,997
28	0	1	0,003	0,997	29	0	1	0,000	1,000
29	0	1	0,001	0,999	30	0	1	0,006	0,994
31	0	1	0,001	0,999	31	0	1	0,000	1,000
Conjunto de Treinamento 05					Conjunto de Predição 06				
Molécula	Padrão da Predição		Resultado da Predição		Molécula	Padrão da Predição		Resultado da Predição	
3	1	0	0,999	0,001	3	1	0	1,000	0,000
4	1	0	0,999	0,001	4	1	0	1,000	0,000
6	1	0	0,001	0,999	7	1	0	1,000	0,000
7	1	0	0,999	0,001	8	1	0	0,830	0,170
8	1	0	0,996	0,004	13	1	0	0,999	0,001
12	1	0	0,999	0,001	14	1	0	0,991	0,009
13	1	0	0,999	0,001	16	1	0	0,998	0,002
18	1	0	0,792	0,204	18	1	0	0,136	0,864
19	1	0	0,990	0,009	19	1	0	0,891	0,109
1	0	1	0,999	0,001	1	0	1	0,002	0,998
20	0	1	0,064	0,935	23	0	1	0,961	0,036
23	0	1	0,999	0,001	24	0	1	0,002	0,998
26	0	1	0,001	0,999	26	0	1	0,001	0,999
30	0	1	0,001	0,999	27	0	1	0,001	0,999
32	0	1	0,000	1,000	32	0	1	0,000	1,000

Conjunto de Treinamento 07					Conjunto de Treinamento 08				
Molécula	Treinamento Padrão		Resultado do Treinamento		Molécula	Treinamento Padrão		Resultado do Treinamento	
3	1	0	0,999	0,001	2	1	0	1,000	0,000
6	1	0	0,993	0,007	5	1	0	0,999	0,001
7	1	0	0,999	0,001	6	1	0	0,993	0,007
10	1	0	0,998	0,002	8	1	0	0,999	0,001
11	1	0	0,999	0,001	11	1	0	1,000	0,000
12	1	0	0,999	0,001	12	1	0	1,000	0,000
13	1	0	0,999	0,001	14	1	0	0,999	0,001
14	1	0	0,998	0,002	15	1	0	1,000	0,000
15	1	0	0,999	0,001	18	1	0	0,996	0,004
19	1	0	0,997	0,003	22	1	0	0,995	0,005
1	0	1	0,001	0,999	1	0	1	0,005	0,995
20	0	1	0,005	0,995	20	0	1	0,003	0,997
24	0	1	0,002	0,998	21	0	1	0,005	0,995
25	0	1	0,000	1,000	24	0	1	0,005	0,995
27	0	1	0,000	1,000	25	0	1	0,002	0,998
29	0	1	0,005	0,995	31	0	1	0,005	0,995
32	0	1	0,005	0,995	32	0	1	0,003	0,997
Conjunto de Treinamento 07					Conjunto de Predição 08				
Molécula	Padrão da Predição		Resultado da Predição		Molécula	Padrão da Predição		Resultado da Predição	
2	1	0	0,999	0,001	3	1	0	1,000	0,000
4	1	0	0,998	0,002	4	1	0	1,000	0,000
5	1	0	0,975	0,025	7	1	0	1,000	0,000
8	1	0	0,994	0,006	9	1	0	1,000	0,000
9	1	0	0,999	0,001	10	1	0	0,999	0,001
16	1	0	0,999	0,001	13	1	0	1,000	0,000
17	1	0	0,999	0,001	16	1	0	1,000	0,000
18	1	0	0,967	0,033	17	1	0	1,000	0,000
22	1	0	0,001	0,999	19	1	0	0,999	0,001
21	0	1	0,002	0,998	23	0	1	0,008	0,992
23	0	1	0,001	0,999	26	0	1	0,099	0,901
26	0	1	0,101	0,899	27	0	1	0,003	0,997
28	0	1	0,000	1,000	28	0	1	0,466	0,534
30	0	1	0,013	0,987	29	0	1	0,235	0,765
31	0	1	0,000	1,000	30	0	1	0,058	0,942

Conjunto de Treinamento 09					Conjunto de Treinamento 10				
Molécula	Treinamento Padrão		Resultado do Treinamento		Molécula	Treinamento Padrão		Resultado do Treinamento	
2	1	0	0,999	0,001	2	1	0	1,000	0,001
3	1	0	0,999	0,001	3	1	0	0,999	0,001
5	1	0	0,999	0,001	4	1	0	0,999	0,001
8	1	0	0,999	0,001	5	1	0	0,999	0,001
9	1	0	0,999	0,001	6	1	0	0,997	0,003
10	1	0	0,998	0,002	11	1	0	0,999	0,001
15	1	0	0,999	0,001	12	1	0	0,999	0,001
17	1	0	0,999	0,001	14	1	0	0,999	0,001
19	1	0	0,998	0,002	15	1	0	0,999	0,001
22	1	0	0,997	0,003	17	1	0	0,999	0,001
1	0	1	0,000	1,000	1	0	1	0,002	0,998
23	0	1	0,001	0,999	23	0	1	0,002	0,998
24	0	1	0,003	0,997	25	0	1	0,002	0,998
26	0	1	0,002	0,998	28	0	1	0,000	1,000
29	0	1	0,001	0,999	29	0	1	0,002	0,998
30	0	1	0,000	1,000	30	0	1	0,001	0,999
32	0	1	0,002	0,998	31	0	1	0,000	1,000
Conjunto de Treinamento 09					Conjunto de Predição 10				
Molécula	Padrão da Predição		Resultado da Predição		Molécula	Padrão da Predição		Resultado da Predição	
4	1	0	0,999	0,001	7	1	0	1,000	0,001
6	1	0	0,001	0,999	8	1	0	0,999	0,001
7	1	0	0,999	0,001	9	1	0	1,000	0,001
11	1	0	0,999	0,001	10	1	0	0,927	0,079
12	1	0	0,999	0,001	13	1	0	0,999	0,001
13	1	0	0,999	0,001	16	1	0	0,999	0,001
14	1	0	0,999	0,001	18	1	0	0,999	0,001
16	1	0	0,999	0,001	19	1	0	0,999	0,001
18	1	0	0,998	0,002	22	1	0	0,000	1,000
20	0	1	0,020	0,981	20	0	1	0,999	0,001
21	0	1	0,978	0,021	21	0	1	0,000	1,000
25	0	1	0,993	0,007	24	0	1	0,000	1,000
27	0	1	0,002	0,998	26	0	1	0,005	0,996
28	0	1	0,155	0,841	27	0	1	0,000	1,000
31	0	1	0,002	0,998	32	0	1	0,077	0,922

Conjunto de Treinamento 01					Conjunto de Treinamento 02				
Molécula	Treinamento Padrão		Resultado do Treinamento		Molécula	Treinamento Padrão		Resultado do Treinamento	
3	1	0	1,000	0,000	3	1	0	1,000	0,000
8	1	0	1,000	0,000	7	1	0	0,994	006
9	1	0	0,998	0,001	8	1	0	1,000	0,000
10	1	0	0,952	0,047	12	1	0	1,000	0,000
13	1	0	1,000	0,000	13	1	0	1,000	0,000
14	1	0	0,999	0,001	15	1	0	0,999	0,001
18	1	0	1,000	0,000	16	1	0	1,000	0,000
19	1	0	1,000	0,000	17	1	0	0,998	0,002
20	1	0	1,000	0,000	19	1	0	1,000	0,000
22	1	0	0,257	0,731	20	1	0	1,000	0,000
6	0	1	0,005	0,995	1	0	1	0,001	0,999
23	0	1	0,024	0,978	6	0	1	0,002	0,998
25	0	1	0,000	1,000	21	0	1	0,002	0,998
29	0	1	0,001	0,999	24	0	1	0,002	0,998
30	0	1	0,002	0,998	26	0	1	0,002	0,998
31	0	1	0,006	0,994	27	0	1	0,006	0,994
32	0	1	0,094	0,901	28	0	1	0,001	0,999
Conjunto de Treinamento 01					Conjunto de Predição 02				
Molécula	Padrão da Predição		Resultado da Predição		Molécula	Padrão da Predição		Resultado da Predição	
2	1	0	0,997	0,003	2	1	0	0,998	0,002
4	1	0	0,998	0,002	4	1	0	0,998	0,002
5	1	0	0,994	0,006	5	1	0	0,999	0,001
7	1	0	0,985	0,014	9	1	0	0,999	0,001
11	1	0	0,997	0,003	10	1	0	0,998	0,002
12	1	0	1,000	0,000	11	1	0	0,999	0,001
15	1	0	0,999	0,001	14	1	0	0,999	0,001
16	1	0	1,000	0,000	18	1	0	1,000	0,000
17	1	0	0,995	0,004	22	1	0	0,002	0,998
1	0	1	0,204	0,783	23	0	1	0,032	0,968
21	0	1	0,301	0,690	25	0	1	0,005	0,996
24	0	1	0,000	0,999	29	0	1	0,000	1,000
26	0	1	0,104	0,889	30	0	1	0,000	1,000
27	0	1	0,225	0,778	31	0	1	0,000	1,000
28	0	1	0,004	0,996	32	0	1	0,002	0,998

Esterilquinolina – PM3 - MIE

Conjunto de Treinamento 03					Conjunto de Treinamento 04				
Molécula	Treinamento Padrão		Resultado do Treinamento		Molécula	Treinamento Padrão		Resultado do Treinamento	
2	1	0	0,999	0,001	2	1	0	0,998	0,002
7	1	0	0,997	0,003	4	1	0	0,998	0,002
12	1	0	1,000	0,000	7	1	0	0,989	0,016
13	1	0	1,000	0,000	8	1	0	1,000	0,000
14	1	0	1,000	0,000	9	1	0	0,999	0,001
15	1	0	1,000	0,000	11	1	0	0,999	0,001
17	1	0	0,999	0,001	16	1	0	0,999	0,000
18	1	0	1,000	0,000	17	1	0	0,997	0,003
19	1	0	1,000	0,000	20	1	0	1,000	0,000
20	1	0	1,000	0,000	22	1	0	0,001	0,999
6	0	1	0,000	1,000	6	0	1	0,001	1,000
21	0	1	0,000	1,000	24	0	1	0,001	1,000
23	0	1	0,003	0,997	27	0	1	0,001	0,999
25	0	1	0,000	1,000	28	0	1	0,000	1,000
27	0	1	0,001	0,999	29	0	1	0,000	1,000
29	0	1	0,000	1,000	30	0	1	0,000	1,000
32	0	1	0,001	0,999	32	0	1	0,001	0,999
Conjunto de Treinamento 03					Conjunto de Predição 04				
Molécula	Padrão da Predição		Resultado da Predição		Molécula	Padrão da Predição		Resultado da Predição	
3	1	0	1,000	0,000	3	1	0	1,000	0,000
4	1	0	0,999	0,001	5	1	0	0,998	0,002
5	1	0	1,000	0,000	10	1	0	0,993	0,012
8	1	0	1,000	0,000	12	1	0	1,000	0,000
9	1	0	1,000	0,000	13	1	0	1,000	0,000
10	1	0	0,999	0,001	14	1	0	0,999	0,001
11	1	0	1,000	0,000	15	1	0	0,999	0,001
16	1	0	1,000	0,000	18	1	0	1,000	0,000
22	1	0	0,000	1,000	19	1	0	1,000	0,000
1	0	1	0,000	1,000	1	0	1	0,001	0,999
24	0	1	0,000	1,000	21	0	1	0,001	0,999
26	0	1	0,000	1,000	23	0	1	0,001	0,999
28	0	1	0,000	1,000	25	0	1	0,001	1,000
30	0	1	0,000	1,000	26	0	1	0,001	0,999
31	0	1	0,000	1,000	31	0	1	0,001	1,000

Conjunto de Treinamento 05				
Molécula	Treinamento Padrão		Resultado do Treinamento	
2	1	0	0,998	0,002
5	1	0	0,999	0,001
9	1	0	0,999	0,001
10	1	0	0,998	0,002
11	1	0	0,999	0,001
14	1	0	0,999	0,001
15	1	0	0,999	0,001
16	1	0	1,000	0,000
17	1	0	0,998	0,002
20	1	0	1,000	0,000
21	0	1	0,001	0,999
24	0	1	0,001	0,999
25	0	1	0,001	0,999
27	0	1	0,003	0,997
28	0	1	0,001	0,999
29	0	1	0,000	1,000
31	0	1	0,000	1,000
Conjunto de Treinamento 05				
Molécula	Padrão da Predição		Resultado da Predição	
3	1	0	1,000	0,000
4	1	0	0,998	0,002
7	1	0	0,993	0,008
8	1	0	1,000	0,000
12	1	0	1,000	0,000
13	1	0	1,000	0,000
18	1	0	1,000	0,000
19	1	0	1,000	0,000
22	1	0	0,001	0,999
1	0	1	0,001	0,999
6	0	1	0,001	0,999
23	0	1	0,013	0,986
26	0	1	0,001	0,999
30	0	1	0,000	1,000
32	0	1	0,002	0,998

Conjunto de Treinamento 01					Conjunto de Treinamento 02				
Molécula	Treinamento Padrão		Resultado do Treinamento		Molécula	Treinamento Padrão		Resultado do Treinamento	
3	1	0	1,000	0,000	3	1	0	0,966	0,034
6	1	0	0,991	0,009	6	1	0	0,983	0,016
8	1	0	1,000	0,000	7	1	0	0,987	0,013
9	1	0	1,000	0,000	8	1	0	0,865	0,133
10	1	0	1,000	0,000	12	1	0	0,927	0,072
13	1	0	1,000	0,000	13	1	0	0,955	0,044
14	1	0	1,000	0,000	15	1	0	0,886	0,113
18	1	0	1,000	0,000	16	1	0	0,907	0,092
19	1	0	1,000	0,000	17	1	0	0,912	0,087
20	1	0	0,992	0,008	19	1	0	0,874	0,124
22	0	1	0,000	1,000	1	0	1	0,000	1,000
23	0	1	0,011	0,989	21	0	1	0,014	0,986
25	0	1	0,003	0,997	22	0	1	0,000	1,000
29	0	1	0,000	1,000	24	0	1	0,167	0,835
30	0	1	0,000	1,000	26	0	1	0,000	1,000
31	0	1	0,004	0,997	27	0	1	0,000	1,000
32	0	1	0,001	1,000	28	0	1	0,000	1,000
Conjunto de Treinamento 01					Conjunto de Predição 02				
Molécula	Padrão da Predição		Resultado da Predição		Molécula	Padrão da Predição		Resultado da Predição	
2	1	0	1,000	0,000	2	1	0	0,975	0,024
4	1	0	1,000	0,000	4	1	0	0,862	0,137
5	1	0	1,000	0,000	5	1	0	0,913	0,085
7	1	0	1,000	0,000	9	1	0	0,998	0,002
11	1	0	1,000	0,000	10	1	0	0,892	0,106
12	1	0	1,000	0,000	11	1	0	0,892	0,107
15	1	0	1,000	0,000	14	1	0	0,879	0,120
16	1	0	1,000	0,000	18	1	0	0,925	0,074
17	1	0	1,000	0,000	20	1	0	0,996	0,004
1	0	1	0,000	1,000	23	0	1	0,974	0,025
21	0	1	0,000	1,000	25	0	1	0,001	0,999
24	0	1	1,000	0,000	29	0	1	0,000	1,000
26	0	1	0,001	1,000	30	0	1	0,000	1,000
27	0	1	0,005	0,996	31	0	1	0,000	1,000
28	0	1	0,000	1,000	32	0	1	0,000	1,000

Conjunto de Treinamento 03					Conjunto de Treinamento 04				
Molécula	Treinamento Padrão		Resultado do Treinamento		Molécula	Treinamento Padrão		Resultado do Treinamento	
2	1	0	1,000	0,000	2	1	0	1,000	0,000
7	1	0	1,000	0,000	4	1	0	0,999	0,001
12	1	0	1,000	0,000	6	1	0	0,985	0,015
13	1	0	1,000	0,000	7	1	0	1,000	0,000
14	1	0	1,000	0,000	8	1	0	0,999	0,001
15	1	0	1,000	0,000	9	1	0	1,000	0,000
17	1	0	1,000	0,000	11	1	0	0,989	0,011
18	1	0	1,000	0,000	16	1	0	0,999	0,001
19	1	0	1,000	0,000	17	1	0	1,000	0,000
20	1	0	0,987	0,013	20	1	0	0,997	0,003
21	0	1	0,000	1,000	22	0	1	0,002	0,998
22	0	1	0,000	1,000	24	0	1	0,015	0,985
23	0	1	0,013	0,987	27	0	1	0,003	0,997
25	0	1	0,007	0,993	28	0	1	0,000	1,000
27	0	1	0,000	1,000	29	0	1	0,011	0,989
29	0	1	0,000	1,000	30	0	1	0,000	1,000
32	0	1	0,000	1,000	32	0	1	0,003	0,997
Conjunto de Treinamento 03					Conjunto de Predição 04				
Molécula	Padrão da Predição		Resultado da Predição		Molécula	Padrão da Predição		Resultado da Predição	
3	1	0	1,000	0,000	3	1	0	1,000	0,000
4	1	0	1,000	0,000	5	1	0	0,999	0,001
5	1	0	1,000	0,000	10	1	0	1,000	0,000
6	1	0	0,001	0,999	12	1	0	0,998	0,002
8	1	0	1,000	0,000	13	1	0	1,000	0,000
9	1	0	1,000	0,000	14	1	0	1,000	0,000
10	1	0	1,000	0,000	15	1	0	0,999	0,001
11	1	0	1,000	0,000	18	1	0	0,999	0,001
16	1	0	1,000	0,000	19	1	0	1,000	0,000
1	0	1	0,000	1,000	1	0	1	0,002	0,998
24	0	1	1,000	0,000	21	0	1	0,620	0,380
26	0	1	0,000	1,000	23	0	1	0,985	0,015
28	0	1	0,000	1,000	25	0	1	0,362	0,635
30	0	1	0,000	1,000	26	0	1	0,004	0,996
31	0	1	0,000	1,000	31	0	1	0,002	0,998

Conjunto de Treinamento 05				
Molécula	Treinamento Padrão		Resultado do Treinamento	
2	1	0	1,000	0,0000
5	1	0	0,997	0,003
9	1	0	0,999	0,001
10	1	0	0,999	0,001
11	1	0	0,994	0,006
14	1	0	0,999	0,001
15	1	0	0,998	0,002
16	1	0	0,999	0,001
17	1	0	0,999	0,001
20	1	0	0,995	0,005
22	1	0	0,004	0,996
24	0	1	0,007	0,993
25	0	1	0,004	0,996
27	0	1	0,000	1,000
28	0	1	0,002	0,998
29	0	1	0,003	0,997
31	0	1	0,001	0,999
Conjunto de Treinamento 05				
Molécula	Padrão da Predição		Resultado da Predição	
3	1	0	1,000	0,000
4	1	0	0,998	0,002
6	1	0	0,001	0,999
7	1	0	0,998	0,002
8	1	0	0,998	0,002
12	1	0	0,998	0,002
13	1	0	0,999	0,001
18	1	0	0,996	0,004
19	1	0	0,999	0,001
1	0	1	0,004	0,996
21	0	1	0,053	0,947
23	0	1	0,347	0,653
26	0	1	0,002	0,998
30	0	1	0,002	0,998
32	0	1	0,002	0,998

Referências Bibliográficas

- [1] Revista Veja, Super-Remédios, Editora Abril, edição 1757, ano 35, no. 25, 26 de junho de 2002, páginas 94 à 101;
- [2] Gallo, R.C., Sarin, P.S., Galmann, E.P., Robert-Guroff, M., et al, *Science* 1983, 220, 865-867;
- [3] Montagnier, L., A history of HIV Discovery, *Science*, 298, 1727(2002);
- [4] **gp** significa glicoproteína e **p** proteína, os números que os seguem dão o tamanho aproximado da molécula em kilodaltons;
- [5] Volker, E.J., *J. Chem. Ed.* 1993, 70, 3;
- [6] Essas enzimas são designadas pela sigla **Pr160**;
- [7] Nesta parte o novo vírus é um víron que significa um vírus imturo sendo catalizado essencialmente pela protease;
- [8] Hiebl, J., Zbiral, E., Balzarini, J. e De Clercq, E., *J. Med. Chem.*, 1992, 35, 3016;
- [9] Ho, D.D., et al, *Nature* 1995, 373, 123-126 e referências citadas;
- [10] Zouhiri, F., et al, *J. Med. Chem.*, 2000, 43, 1533-1540;
- [11] Mekouar, K., et al, *J. Med. Chem.*, 1998, 41, 2846-2857;

- [12] Santo, L.L.E., Aplicação de Métodos Semi-empíricos ao Estudo da Estrutura Eletrônica de Compostos Bioativos, *Tese de Doutorado, IFGW-UNICAMP*, Campinas, 2001;
- [13] Giro, R., Estudo de Estrutura Eletrônica de Copolímeros de Polianilina, *Dissertação de mestrado, IFGW-UNICAMP*, Campinas, Março de 2000;
- [14] Cyrillo, M., Estudo Estrutur-atividade de Alguns Inibidores da Enzima Integrase do Vírus HIV-1, *Dissertação de mestrado, IFGW-UNICAMP*, Campinas, Junho de 1999;
- [15] Pople, J.A., and Beveridge, D. L., Approximate Molecular Orbital Theory, *Mc Graw-Hill*, 1970;
- [16] Levine, I.A., *Quantum Chemistry*, Fifth Edition, *Prentice Hall*, ;
- [17] Szabo, A., and Ostlund, N.S., Modern Quantum Chemistry - Introduction to Advanced Electronic Structure Theory, *Dover Publications*, New York, 1996;
- [18] Dewar, M.J.S., Zoebish, E.G., Healy, E.F., and Stewart, J.J.P., *J. Amer. Soc.*, 107, 3902(1985);
- [19] Stewart, J.J.P., *J. Comp. Chem.*, 209, 10(1991);
- [20] Braga, R.S., Barone, P.M.V.B., Galvão, D.G., *J. Mol. Struc.*, 464, 257(1999);
- [21] Barone, P.M.V.B., Camilo Jr, A. e Galvão, D.S., *J. Mol. Struc. (THEOCHEM)*, 505, 55(2000);
- [22] Vendrame, R., Braga, R.S., Takahata, Y. e Galvão, D.S., *J. Mol. Struc. (THEOCHEM)*, 539, 253(2001);
- [23] Vendrame, R., Braga, R.S., Takahata, Y. e Galvão, D.S., *J. Chem. Inf. Comp. Sci.*, 39, 1094(1999);
- [24] Santo, L.L.E. e Galvão, D.S., *J. Mol. Struc. (THEOCHEM)* 464, 273(1999);
- [25] Mazzali, L.L. e Galvão, D.S., *J. Chem. Phys. B* – submetido;
- [26] Braga, R.S., Vendrame, R. e Galvão, D.S., *J. Chem. Inf. Comp. Sci.* 40, 1377(2000);

- [27] Cyrillo, M. e Galvão, D.S., J. Mol. Struc. (*THEOCHEM*), 464, 267(1999);
- [28] Sítio da empresa Infometrix: <http://www.infometrix.com>;
- [29] Einsigh Program, Infometrix, Seattle, USA;
- [30] Vendrame, R., Braga, R.S., Takahata, Y., Galvão, D.S., *J. Mol. Struct* (THEOCHEM), 529, 253(2001);
- [31] McCulloch, W., and Pitts, W., "A logical calculus of ideas immanent in nervous activity, "Bulletin of Mathematical Biophysics, vol, 5, pp. 115-133, 1943;
- [32] Rosenblatt, F., "The perceptron: A probabilistic model for information storage and organization in the brain," Psychological Review, vol 65, pp. 386-408, 1958;
- [33] Widrow, B., and Hoff, M.E., "Adaptive switching circuits", 1960 IRE WESCON Convention Record, New York: IRE Part 4, pp. 96-104, 1960;
- [34] Hopfield, J.J., "Neural Network and physical systems with emergent collective computational properties, " Proceedings of the National Academy of Sciences, vol. 79, pp. 2554-2558, 1982;
- [35] Rumelhart, D.E., and McClelland, J.L., eds., Parallel Distributed Processing: Exploration in the Microstructure of Cognition, vol. 1, Cambridge, MA: MIT Press, 1986;
- [36] Hagan, M.T., Demuth, M.B. e Beale, M., Neural Network Design, *PWS Publishing Company*, Boston, 1996;
- [37] Haykin, S., Neural Network: A Comprehensive Foundation, Second Edition, *Prentice-Hall*, New Jersey, 1999;
- [38] Bishop, C.M., *Neural Network for Pattern Recognition*, Oxford University Press Inc, New York, 1995;
- [39] Ichikawa, H., PSDD: Perceptron-type Neural Network Simulator, QCPE 615, Indiana, USA;
- [40] Ayoama, T., Suzuki, Y. e Ichikawa, H., Neural Network Applied to Pharmaceutical Problems: I. Method and Application to Decision Making, Communication to the editor, *Chem. Pharm. Bull* 37(9), 2558-2560(1989);

- [41] H. Oinuma, K. Miyake, *et al*, Neural Network Applied to Structure-Activity Relationship, *J. Med. Chem.*, 33, 905-908(1990);
- [42] T. Aoyama, Y. Suzuki, and H. Ichikawa, Neural Network Applied to Quantitative Structure-Activity Relationship Analysis, *J. Med. Chem.*, 33, 2583-2590(1990);
- [43] PC Spartan Pro vs 1.0, <http://www.wavefun.com>;
- [44] CAChe vs 5.0 for windows, Fujitsu, <http://www.cachefun.com>;
- [45] HyperChem; <http://www.hyper.com>;
- [46] Chopra, I., Hawkey, P.M., and Hinton, M., *J.Antimicrob. Chemother*, 29, 245(1992);
- [47] Levy, S.B., *Antimicrob. Agents Chemother*, 36, 695(1992);
- [48] Roberts, M.C., *FEMS Microbiology Rev.*, 19, 1(1996);
- [49] Revista Scientific Americam Brasil, no.1, Jun, 35(2002);
- [50] Revista Pesquisa FAPESP, no.76, Jun, 45(2002);
- [51] Miller, G.H., Smith, H.L, and Heldberg, S., *J.Pharm Sci.*, 66, 88(1977);
- [52] Dos Santos, H.F., Xavier É.S., Zerner, M.C., De Almeida, W.B., *J. Mol. Struc. (THEOCHEM)*, 527, 193(2000);