

**UNIVERSIDADE ESTADUAL PAULISTA
PÓS-GRADUAÇÃO EM CIÊNCIA E TECNOLOGIA DE MATERIAIS
CAMPUS DE BAURU**

NÉLIO HENRIQUE NICOLETI

**ESTRUTURA ELETRÔNICA DE MATERIAIS ORGÂNICOS: MOLÉCULAS
ANTIMALARIAIS DE SULFONAMIDAS E ANILINOQUINOLINAS**

Bauru

2007

NÉLIO HENRIQUE NICOLETI

**ESTRUTURA ELETRÔNICA DE MATERIAIS ORGÂNICOS: MOLÉCULAS
ANTIMALARIAIS DE SULFONAMIDAS E ANILINOQUINOLINAS**

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Ciência e Tecnologia dos Materiais do Programa de Pós-Graduação em Ciência e Tecnologia de Materiais da Universidade Estadual Paulista “Júlio de Mesquita Filho” - UNESP.

Orientador: Prof. Dr. Francisco Carlos Lavarda.

Bauru

2007

NÉLIO HENRIQUE NICOLETI

**ESTRUTURA ELETRÔNICA DE MATERIAIS ORGÂNICOS: MOLÉCULAS
ANTIMALARIAIS DE SULFONAMIDAS E ANILINOQUINOLINAS**

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Ciência e Tecnologia dos Materiais do Programa de Pós-Graduação em Ciência e Tecnologia de Materiais da Universidade Estadual Paulista “Júlio de Mesquita Filho” - UNESP.

BANCA EXAMINADORA

Prof. Dr. Francisco Carlos Lavarda – Orientador
Faculdade de Ciências – UNESP

Prof^a. Dr^a. Ignez Caracelli
Faculdade de Ciências – UNESP

Prof. Dr. Alexandre Camilo Junior
Departamento de Física – UEPG

Bauru, 11 de maio de 2007

DEDICATÓRIA

Dedico este trabalho a minha família, em especial aos meus pais Nelson e Fátima pelo incentivo e apoio dado em toda minha vida. Também para Valéria, pessoa maravilhosa sem a qual nada disso seria possível, por todo seu amor e paciência. Também para Francisco Carlos Lavarda, que tanto me ensinou e ajudou, pelas horas de trabalho e conversas.

AGRADECIMENTOS

Gostaria de agradecer a todos os amigos da POSMAT, em especial aos companheiros de sempre Tiago, Douglas e Renata.

Aos companheiros de trabalho Prof. Francisco Carlos Lavarda, Pedro, Neto, Éderson, Daniel e Rafael.

Ao departamento de Física pelas instalações cedidas: Laboratório de Física Teórica e Computacional – LFTC, em especial ao Prof. Dr. André Luiz Malvezzi e ao Prof. Dr. Pablo Antonio Venegas Urenda.

Aos professores responsáveis pelas disciplinas que cursei, nas quais aprendi muito (Aguinaldo Robinson de Souza, André Luiz Malvezzi, Ariovaldo de Oliveira Florentino, Carlos Roberto Grandini, Júlio Ricardo Sambrano, Pablo Antonio Venegas Urenda, Paulo Noronha Lisboa Filho).

Ao pessoal da secretária pela atenção dispensada e a eficiência de sempre, em especial a Djanira pelas orientações referente à pós-graduação.

Ao grande amigo Fabiano Paixão, pelo apoio e idéias compartilhadas.

E a todos que de alguma forma me auxiliaram nesta jornada. Obrigado.

RESUMO

Neste trabalho estudamos dois grupos de moléculas: as anilinoquinolinas e as sulfonamidas, inibidores do *Plasmodium* causador da malária, com o objetivo de correlacionar a estrutura eletrônica com a atividade antimalarial. Em nossas buscas utilizamos métodos empíricos e semi-empíricos para o estudo conformacional e obtenção dos descritores eletrônicos. Também aplicamos vários métodos estatísticos como: Regressão Linear Simples e Múltipla, Análise de Componentes Principais (PCA) e Análise Discriminante Linear (LDA), para verificar uma possível correlação estrutura-atividade dessas moléculas. Os resultados apontaram os descritores eletrônicos mais relevantes na classificação das moléculas antimalariais.

Palavras-chave: estrutura eletrônica, antimalariais, sulfonamidas, anilinoquinolinas.

ABSTRACT

In this work we study two groups of antimalarial compounds: the anilinoquinolines and sulfonamides, aiming the correlation of the electronic structure with the antimalarial activity. In our studies we employ empirical and semi empirical quantum chemistry methods for the geometry optimization and calculation of the electronic descriptors. Also we employed the statistical methods Simple and Multiple Linear Regression, Principal Component Analysis (PCA) and Linear Discriminating Analysis (LDA), to verify the existence of a possible structure-activity correlation for these compounds. The results of this work have pointed out the best electronic descriptors in the classification of the active compounds.

Key Words: electronic structure, antimalarials, sulfonamides, anilinoquinolines.

LISTA DE FIGURAS

Figura 2.1: Representação qualitativa de correlações lineares observadas.

Figura 2.2: Regressão linear simples para pontos não igualmente espaçados.

Figura 2.3: Representação gráfica do resultado obtido através da LDA.

Figura 3.1: Representação esquemática da estrutura básica: *2,4-diamino-6-quinazolines sulfonamides*.

Figura 3.2: Análise dos Componentes Principais PC1 *versus* PC2.

Figura 3.3: Gráfico dos escores Z e o parâmetro de corte m .

Figura 3.4: Representação esquemática da estrutura básica da anilinoquinolina.

Figura 3.5: Análise dos componentes principais PC1 *versus* PC2, para o grupo 1.

Figura 3.6: Análise dos componentes principais PC1 *versus* PC2, para o grupo 2.

Figura 3.7: Gráfico dos *scores* de Z e o parâmetro de corte m , para o grupo 1.

Figura 3.8: Gráfico dos *scores* de Z e o parâmetro de corte m , para o grupo 2.

Figura 3.9: Gráfico dos *scores* de Z e o parâmetro de corte m , para o grupo 3.

Figura 4.1: Representação da estrutura básica das sulfonamidas, com a indicação da região de maior significância.

Figura 4.2: Representação da estrutura básica da anilinoquinolina, com a indicação da região de maior significância.

LISTA DE TABELAS

Tabela 2.1: Descritores eletrônicos calculados para cada molécula.

Tabela 3.1: Substituintes da *2,4-diamino-6-quinazolines sulfonamides* e as respectivas propriedades.

Tabela 3.2: Representação esquemática dos 16 compostos derivados da *2,4-diamino-6-quinazolines sulfonamides* (Figura 3.1).

Tabela 3.3: Resultado das regressões lineares simples e múltiplas considerando todas as moléculas.

Tabela 3.4: Dados da variância dos componentes principais PC1 e PC2.

Tabela 3.5: Teste de igualdade das médias entre os grupos.

Tabela 3.6: Valor da função discriminante no centróide dos grupos e parâmetro de corte.

Tabela 3.7: Resultados da classificação para as *2,4-diamino-6-quinazolines sulfonamides*, utilizando LDA com o modelo original e de validação cruzada.

Tabela 3.8: Resultados da classificação, utilizando LDA com o modelo original e de validação cruzada, somente com o descritor E_{ELET} .

Tabela 3.9: Representação esquemática dos 3 grupos derivados das anilinoquinolinas, substituintes e índice antimalarial.

Tabela 3.10: Resultado das regressões lineares simples e múltiplas considerando todas as moléculas.

Tabela 3.11: Resultados referentes aos grupos dos derivados de anilinoquinolinas.

Tabela 3.12: Dados da variância dos componentes principais PC1 e PC2.

Tabela 3.13: Dados da variância dos componentes principais PC1 e PC2.

Tabela 3.14: Teste de igualdade das médias entre as moléculas de maior e menor atividade antimalarial para o grupo 1.

Tabela 3.15: Valor da função discriminante no centróide dos agrupamentos das moléculas mais ativas (A), menos ativas (B) e parâmetro de corte (m).

Tabela 3.16: Resultados da classificação, utilizando LDA com o modelo original e de validação cruzada.

Tabela 3.17: Teste de igualdade das médias entre as moléculas de maior e menor atividade antimalarial para o grupo 2.

Tabela 3.18: Valor da função discriminante no centróide dos grupos e parâmetro de corte.

Tabela 3.19: Resultados da classificação utilizando LDA com o modelo original e de validação cruzada.

Tabela 3.20: Teste de igualdade das médias entre as moléculas de maior e menor atividade para o grupo 3.

Tabela 3.21: Valor da função discriminante no centróide dos grupos e parâmetro de corte.

Tabela 3.22: Resultados da classificação, utilizando LDA com o modelo original e de validação cruzada.

LISTA DE ABREVIATURAS

AM1: Austin Model 1.

BIOMAT: Grupo de Física Teórica de Biomoléculas e Materiais.

CAR: carga líquida.

CL: comprimento de ligação.

DFT: Density Functional Theory.

E_{ELET} : energia referente aos elétrons de valência da molécula.

E_H : energia do último orbital molecular ocupado.

EHF: Equação de Hartree-Fock.

E_L : energia do primeiro orbital molecular desocupado.

E_{L1} : energia do segundo orbital molecular desocupado.

EM : Eletronegatividade de Mulliken $[(E_H + E_L)/2]$.

E_{NUCL} : energia referente aos íons (núcleo mais elétrons de caroço).

ES: Equação de Schrödinger.

ESIT: Equação de Schrödinger independente do tempo.

E_{TOTAL} : energia total.

HF: Hartree-Fock.

LDA: Análise Discriminante Linear (*Linear Discriminant Analysis*).

LDOS: Densidade Local de Estados (*Local Density of States*).

m : parâmetro de corte.

MM: Mecânica Molecular.

NC: número de combinações.

OL: ordem de ligação.

p : grau de significância estatística.

PCA: Análise de Componentes Principais (*Principal Components Analysis*).

PEP_H : população eletrônica parcial do HOMO.

PEP_L : população eletrônica parcial do LUMO.

PET : população eletrônica total.

r : parâmetro de correlação.

r^2 : Variabilidade total das informações.

RHF: *Restricted Hartree-Fock*.

s : desvio padrão.

SCF: self-consistent field.

SPSS: *Statistical Package for the Social Sciences.*

V_i : variável independente.

$\Delta E_{(H,H1)}$: diferença da energia entre HOMO e HOMO-1.

$\Delta E_{(L,H)}$: diferença da energia entre LUMO e HOMO.

$\Delta E_{(L1,H1)}$: diferença da energia entre LUMO+1 e HOMO-1.

$\Delta E_{(L1,L)}$: diferença da energia entre LUMO+1 e LUMO.

ΔH_F : Calor de Formação.

ΔMST : *difference of mean survival time.*

μ_{TOTAL} : momento de dipolo elétrico total.

SUMÁRIO

1	Introdução	16
2	Metodologias	19
2.1	Mecânica Molecular	24
2.2	Métodos Semi-empírico	25
2.2.1	O Método semi-empírico Austin Model 1	26
2.3	Correlação Linear	27
2.4	Métodos Estatísticos	30
2.4.1	Regressão Linear Simples	30
2.4.2	Regressão Linear Múltipla	33
2.4.3	Análise Discriminante Linear	35
2.5	Método Quimiométrico	42
2.5.1	Análise dos Componentes Principais	43
3	Moléculas com Atividade Antimalarial	48
3.1	Sulfonamidas	51
3.1.1	Resultados	57
3.1.1.1	Regressões Lineares Simples e Múltiplas	57
3.1.1.2	Análise dos Componentes Principais	59
3.1.1.3	Análise Discriminante Linear	62
3.1.2	Geometrias de equilíbrio das sulfonamidas	68
3.2	Anilinoquinolinas	70
3.2.1	Resultados	76
3.2.1.1	Regressões Lineares Simples e Múltiplas	76
3.2.1.2	Análise dos Componentes Principais	78
3.2.1.3	Análise Discriminante Linear	84
3.2.2	Geometrias de equilíbrio das anilinoquinolinas	94

4	Discussão e conclusões	101
----------	-------------------------------------	-----

5	Referências	111
----------	--------------------------	-----

Apêndices

Apêndice A	115
------------	-------	-----

Apêndice B	120
------------	-------	-----

Apêndice C	129
------------	-------	-----

CAPÍTULO 1

Introdução

Nas últimas décadas, as conquistas obtidas nas áreas computacionais auxiliaram o desenvolvimento de novos materiais através de um processo que interliga diversos campos do conhecimento humano. Esses avanços levaram essas pesquisas, que antes eram quase em sua totalidade empíricas a seguirem novos rumos, tornando-se um processo mais racional a partir do uso da modelagem de materiais, que consiste na previsão das propriedades através de um conjunto de métodos que utilizam conceitos de física quântica ou clássica [1].

Com esses métodos é possível construir um modelo inicial do material e investigar suas possíveis conformações. Na maioria das vezes o que se pretende é encontrar a estrutura de mínima energia para então obter previsões de suas propriedades. Para tanto, empregam-se cálculos de estrutura eletrônica: uma área de investigação teórica onde se aplica uma grande variedade de métodos para descrever a distribuição espacial e os níveis de energia dos elétrons de um sistema, que pode ser, desde a mais simples molécula diatômica até sólidos extensos com mais de 10^{23} átomos.

Os cálculos fornecem informações sobre o sistema e esses resultados podem ser aplicados para a previsão e explanação dos mais diversos tipos de dados experimentais de natureza quântica. Como por exemplo, a absorção e emissão de luz de um material podem ser previstas por essa teoria. Em sistemas extensos, o estudo da estrutura eletrônica pode prever as propriedades de condução elétrica do material: se este é um isolante, semicondutor ou condutor de eletricidade.

Enfim, com os cálculos de estrutura eletrônica pode-se investigar e caracterizar um material, com vista a compreender suas propriedades, avaliando e prevendo certas características como a conformação de mínima energia e as propriedades físico-químicas, que são de fundamental importância na compreensão das correlações estrutura-atividade, permitindo o desenvolvimento racional de novos materiais que possuam um perfil mais adequado para as necessidades atuais.

Dada a sua natureza complexa, o estudo da estrutura eletrônica é em sua grande parte desenvolvida com o uso de computadores e isto de certa forma delimita os sistemas que podem ser estudados. Em geral, quanto maior o sistema em estudo, mais simples ou com mais aproximações é a teoria empregada. Simplicidade, no entanto, não é sinônimo de menor eficácia: alguns métodos mais simplificados são bem sucedidos na previsão de algumas propriedades. Existe uma grande variedade de métodos de cálculo e de diferentes graus de aproximações. Dentre esses, os três principais utilizados para se calcular as propriedades dos materiais classificam-se em empíricos, semi-empíricos e *ab initio* [2].

Nesse trabalho de modelagem molecular, foram feitos cálculos da estrutura eletrônica com métodos empíricos e semi-empíricos para moléculas de sulfonamidas e anilinoquinolinas [3, 4], com propriedades antimalariais, *in vacuo*. O objetivo foi de

fornecer indicações sobre a correlação estrutura-atividade, tais informações podem servir para o desenvolvimento de novas drogas.

No estudo das moléculas antimalariais há dois tipos de análises que podem ser feitas: a primeira, chamada direta, onde são conhecidos o receptor (ou enzima) e suas características tridimensionais; a segunda, conhecida como indireta, onde a geometria do receptor não é conhecida e as novas moléculas são elaboradas através de um modelo obtido pela análise de um conjunto conhecido de moléculas ativas/inativas. Na abordagem indireta, o modelo inicial da molécula, muitas vezes, é obtido de estudos experimentais como cristalografia por difração de raio X ou podem ser construídos, quando os dados experimentais não estão disponíveis, através de metodologias de modelagem molecular. Nesse estudo sobre antimalariais, utiliza-se a análise indireta, pois o que se quer verificar é se a informação da atividade biológica das moléculas está em suas próprias estruturas eletrônicas [5].

CAPÍTULO 2

Metodologias

Neste trabalho a Física Clássica e Quântica são utilizadas como ferramentas de apoio na análise e interpretação de diversos dados experimentais. A teoria aplicada é voltada para modelar estruturas, buscando informações que muitas vezes não são possíveis de serem obtidas diretamente dos experimentos e que são de fundamental importância para prever diversas propriedades dos materiais estudados. Utilizando-se dessa teoria, tem-se como objetivo investigar a estrutura eletrônica de dois grupos de moléculas: as sulfonamidas e as anilinoquinolinas, os quais tiveram os mais variados descritores eletrônicos calculados através de métodos empíricos e semi-empíricos.

Existem vários métodos de cálculos (empíricos, semi-empíricos e *ab initio*) aplicados aos materiais, que com suas características próprias se adaptam a um determinado tipo de aplicação. Dentre esses foram escolhidos a Mecânica Molecular (MM), a qual foi utilizada para uma “pré-otimização” das moléculas e o *Austin Model 1* (AM1), ambos por já terem sido amplamente avaliados [6, 7, 8, 9], requerer baixo custo computacional, visto o grande número de moléculas analisadas neste trabalho (mais de 80) e pelos resultados obtidos em um estudo prévio

apresentado no Apêndice C. O que se procura é a conformação de menor energia para cada molécula e uma possível correlação da estrutura eletrônica com a atividade de interesse.

Para os grupos estudados realizou-se uma busca conformacional detalhada, devido ao fato de não se encontrarem relatos sobre as estruturas dessas moléculas nos bancos de dados. O processo de busca conformacional consiste na otimização da geometria e sua verificação através do cálculo de frequências vibracionais. A estrutura de menor energia ausente de frequências vibracionais negativas é o confômero mais estável (geometria de equilíbrio), sendo este o mais provável de ser encontrado na natureza [10].

O processo de otimização da geometria desconhecida foi dividido em duas partes: na primeira se realiza uma “pré-otimização” das geometrias de todas as moléculas e seus confômeros através de mecânica molecular (MM), um método empírico que utiliza a mecânica clássica. Na segunda, as geometrias geradas pela MM eram então “reotimizadas” através de cálculos quânticos através do método semi-empírico *Austin Model 1* (AM1) com o auxílio do pacote computacional MOPAC versão 6 (MOPAC 6) [11]. Este método utiliza a aproximação Hartree-Fock (HF). Neste trabalho, adotamos a *Restricted Hartree-Fock* (RHF) *in vacuo* (Apêndice B). O valor máximo da norma do gradiente (das coordenadas geométricas em função da energia) para a geometria ser considerada estável foi tomado como 0,01 kcal/(Å.mol), valor considerado usual.

A fim de verificar se a geometria obtida tratava-se realmente da geometria mais estável, efetuou-se, com o auxílio do MOPAC 6, o cálculo das frequências de vibração de cada molécula. Neste tipo de cálculo, a Hessiana, matriz de segundas derivadas da energia com respeito ao deslocamento de todos os pares de átomos

nas direções x , y e z , é obtida. O sistema pode ser caracterizado como o estado fundamental ou um estado de transição pela observação das constantes de força resultantes da diagonalização da Hessiana. Se dentre as constantes de força não for observado nenhum valor negativo, o sistema é dito estar no estado fundamental, ou seja, a disposição dos átomos é tal que um mínimo de energia foi alcançado e a geometria provavelmente é a mais estável.

Após a otimização das moléculas antimalariais utilizou-se um programa (*biomat_02*) desenvolvido no grupo de pesquisa BioMat, criado em linguagem FORTAN 90, que lê os arquivos de saída do MOPAC 6 e de posse dos autovalores e autovetores, o programa calcula diversos descritores eletrônicos provenientes dos cálculos da estrutura eletrônica. A maioria desses índices está relacionada aos orbitais moleculares de fronteira, ou seja:

- O último orbital molecular ocupado (*highest occupied molecular orbital*, HOMO);
- O penúltimo orbital molecular ocupado, HOMO-1;
- O primeiro orbital molecular desocupado (*lowest unoccupied molecular orbital*, LUMO);
- E o segundo orbital molecular desocupado, LUMO+1.

Estes índices eletrônicos foram escolhidos por terem sido empregados até agora com sucesso em diversos outros grupos de moléculas biológicas ativas. Os descritores de maior relevância utilizados na verificação de algum tipo de correlação com os dados experimentais estão citados na Tabela 2.1.

Tabela 2.1: Descritores eletrônicos calculados para cada molécula.

Descritores	Símbolos
Energia total	E_{TOTAL}
Energia total referente aos elétrons de valência da molécula	E_{ELET}
Energia total referente aos íons (núcleo e elétrons de caroço)	E_{NUCL}
Calor da Formação das moléculas	ΔH_F
Momento de dipolo elétrico total	μ_{TOTAL}
Energia do último orbital molecular ocupado	E_H
Energia do penúltimo orbital molecular ocupado	E_{HI}
Energia do primeiro orbital molecular desocupado	E_L
Energia do segundo orbital molecular desocupado	E_{LI}
Diferença da energia entre HOMO e HOMO-1	$\Delta E_{(H,HI)}$
Diferença da energia entre LUMO e HOMO	$\Delta E_{(L,H)}$
Diferença da energia entre LUMO+1 e LUMO	$\Delta E_{(LI,L)}$
Diferença da energia entre LUMO+1 e HOMO-1	$\Delta E_{(LI,HI)}$
Eletronegatividade de Mulliken $[(E_H + E_L)/2]$	EM
População Eletrônica Total	PET
População Eletrônica Parcial do HOMO	PEP_H
População Eletrônica Parcial do LUMO	PEP_L
Ordem de Ligação	OL
Carga Líquida	CAR
Densidade de Estados	$LDOS$

Os descritores PET , PEP_H , PEP_L , OL , CAR e $LDOS$ foram calculados para cada átomo componente da estrutura básica das moléculas antimalariais. Essas estruturas serão apresentadas posteriormente.

Em posse desses parâmetros foi realizada uma busca por um modelo de alta correlação linear entre os índices acima citados e a atividade antimalarial, através de

regressão linear simples e múltipla. Caso não se encontrasse uma correlação nessa primeira etapa, realizar-se-ia um tratamento estatístico de análise exploratória utilizando métodos de reconhecimento de padrões.

Pretende-se propor nesse trabalho um modelo matemático que consiga não apenas descrever, mas também prever relações estrutura-atividade, explanando a atividade biológica dessas moléculas. Seguem as cinco regras estabelecidas por Unger e Hansch para estudos de regressão linear na área de *Quantitative Structure-Activity Relationship* (QSAR) [12, 13, 14]:

1. Seleção de variáveis independentes: deve-se testar grande número de variáveis e as selecionadas na melhor equação devem ser essencialmente independentes;
2. Validação estatística das variáveis selecionadas: cada variável incluída na melhor equação precisa ser validada por testes estatísticos apropriados, tais como o Teste-F;
3. Princípio da parcimônia (Navalha de Occam): quando houver dúvida na escolha de um entre muitos modelos (aproximadamente) equivalentes, deve-se escolher o mais simples;
4. Número de variáveis em cada modelo de regressão: para minimizar a ocorrência de correlação por coincidência deve haver, no mínimo, cerca de cinco ou seis moléculas para cada variável incluída no modelo. Um número maior poderia culminar num ajuste forçado das regressões (*overfitting*).
5. Modelo qualitativo para o mecanismo de ação das moléculas: é essencial que o modelo quantitativo de relação entre estrutura e atividade seja consistente com o mecanismo de ação em nível molecular.

Por fim, tomou-se também o cuidado de só se apresentar regressões múltiplas compostas de variáveis independentes pouco correlacionadas, com parâmetro de correlação entre elas inferior a 0,6 [15]. Apesar da idéia de que essas regras fossem utilizadas para estabelecer uma metodologia na elaboração de modelos clássicos de QSAR, o que não é o caso aqui, a utilização delas neste trabalho tem como finalidade melhorar a consistência do modelo matemático proposto, validando os resultados desta pesquisa.

2.1 Mecânica Molecular

A MM [16] é um método utilizado para calcular as geometrias e energias conformacionais de um material, baseado em uma combinação de campos de força empíricos (funções e parametrizações que definem as energias de comprimento e do ângulo das ligações de uma molécula), cuja forma é proveniente da mecânica clássica.

O método pressupõe que as ligações covalentes podem ser descritas por funções de potencial tipo “mola” e que as interações entre átomos não ligados podem ser descritos pelas forças de van der Waals e eletrostáticas repulsivas e atrativas. As funções que descrevem a conformação geométrica são parametrizadas em comparação com dados experimentais. A energia então é minimizada e o sistema tende para a geometria mais estável [16]. O princípio do método é que toda molécula possui comprimentos e ângulos “naturais” de ligações e que o sistema tende a esses valores.

Os métodos de MM são geralmente empregados em estudos de sistemas moleculares relativamente grandes (centenas de átomos), já que apresentam um “baixo custo computacional” quando comparados aos métodos semi-empíricos e *ab initio*. Os métodos de MM são hoje bastante utilizados na área de materiais, uma vez que sistemas com centenas ou até mesmo, milhares de átomos podem ser modelados. Nos estudos feitos, adotou-se, entre os vários métodos de mecânica molecular, o MM+ por possuir melhor parametrização para moléculas orgânicas.

Como a MM não trata de forma explícita os elétrons de um sistema, esses métodos não são apropriados para descrever fenômenos eletrônicos, como transições ópticas, por exemplo. Assim utilizam-se as geometrias moleculares obtidas com a MM somente como geometrias iniciais para os cálculos semi-empíricos, onde os elétrons são tratados explicitamente.

2.2 Métodos Semi-empíricos

Um método é dito semi-empírico quando utiliza a teoria da Mecânica Quântica [17, 18] e resolve a Equação de Schrödinger de forma aproximada, levando em conta somente os elétrons de valência. São utilizados alguns parâmetros provenientes de dados empíricos para simplificar os cálculos computacionais, em vez de resolver todas as integrais.

É um método mais sofisticado que a Mecânica Molecular, podendo-se obter dos cálculos propriedades associadas com a estrutura eletrônica das moléculas. Por esse motivo, adota-se esta teoria nos cálculos da conformação final. Dentre os diferentes métodos, será descrito o AM1 [19], pois foi este o método semi-empírico adotado. Sua escolha é devido aos resultados apresentados no apêndice C.

2.2.1 O Método Semi-empírico Austin Model 1

O AM1 é um método Hartree-Fock ou de campo autoconsistente (Apêndice B) baseado na aproximação *Neglect of Diatomic Differential Overlap* na qual, além de desprezar as integrais de sobreposição entre orbitais atômicos, todas as demais integrais são calculadas por meios aproximados. Ele tem se mostrado altamente eficiente na previsão de calores de formação e para otimizar a geometria de equilíbrio de moléculas orgânicas [11]. Embora nesta linha nenhum método substitua completamente o anterior, pode-se dizer que ele é o sucessor direto do MNDO (*Modified Neglect of Diatomic Overlap*) [20] e precursor do PM3 (*Parametric Method 3*) [11].

Por se tratar de um método semi-empírico, o AM1 possui várias simplificações e aproximações [19] que serão descritas a seguir:

1. A base empregada constitui-se de um orbital s e três orbitais p (p_x , p_y e p_z) por átomo (exceção feita ao hidrogênio, que emprega somente um orbital).
2. A matriz sobreposição é considerada uma matriz identidade, ou seja, $S_{\mu\nu} = \delta_{\mu\nu}$ (*neglect of diatomic overlap integral*).
3. Todas as integrais originárias da sobreposição de dois orbitais atômicos da base centrados em átomos diferentes são ignoradas.
4. Todas as integrais de dois elétrons envolvendo a sobreposição de dois orbitais atômicos centrados em átomos diferentes são ignoradas.

O cálculo das integrais remanescentes se dá de modo aproximado por diversas formas. Para se ter uma idéia, citam-se alguns aspectos:

1. As integrais de um centro e dois elétrons são substituídas por dados experimentais obtidos de átomos isolados.
2. A repulsão entre os íons deriva de um cálculo simples de eletrostática (repulsão entre duas cargas pontuais) adicionada de alguns termos de correção.
3. Os parâmetros empíricos são dependentes dos átomos (e não de pares de átomos).
4. Alguns parâmetros são otimizados e outros tomados de dados experimentais.

2.3 Correlação Linear

Após a aplicação dos métodos citados acima, os dados colhidos foram então analisados, procurando-se uma correlação entre a propriedade de interesse apresentada pelas moléculas estudadas e os dados obtidos das respectivas estruturas eletrônicas.

Pode-se dizer que existe uma correlação entre duas (ou mais) variáveis quando uma delas está, de alguma forma, relacionada à outra (ou outras) de forma sistematizada. Encontrar tal correlação é muitas vezes de grande relevância, pois freqüentemente permite a determinação de uma das variáveis em função do conhecimento da outra.

Existem diversos tipos de correlação entre variáveis. Na Figura 2.1 pode-se observar alguns exemplos através de diagramas de dispersão. Em (a) e (b) nota-se

a presença de uma correlação linear positiva entre as variáveis x e y , sendo mais evidente em (b); em (c) e (d) onde há uma correlação linear negativa, sendo esta mais evidente em (d); no entanto em (e) nota-se a ausência de qualquer correlação entre x e y enquanto que em (f) pode-se observar uma certa correlação não linear entre as variáveis.

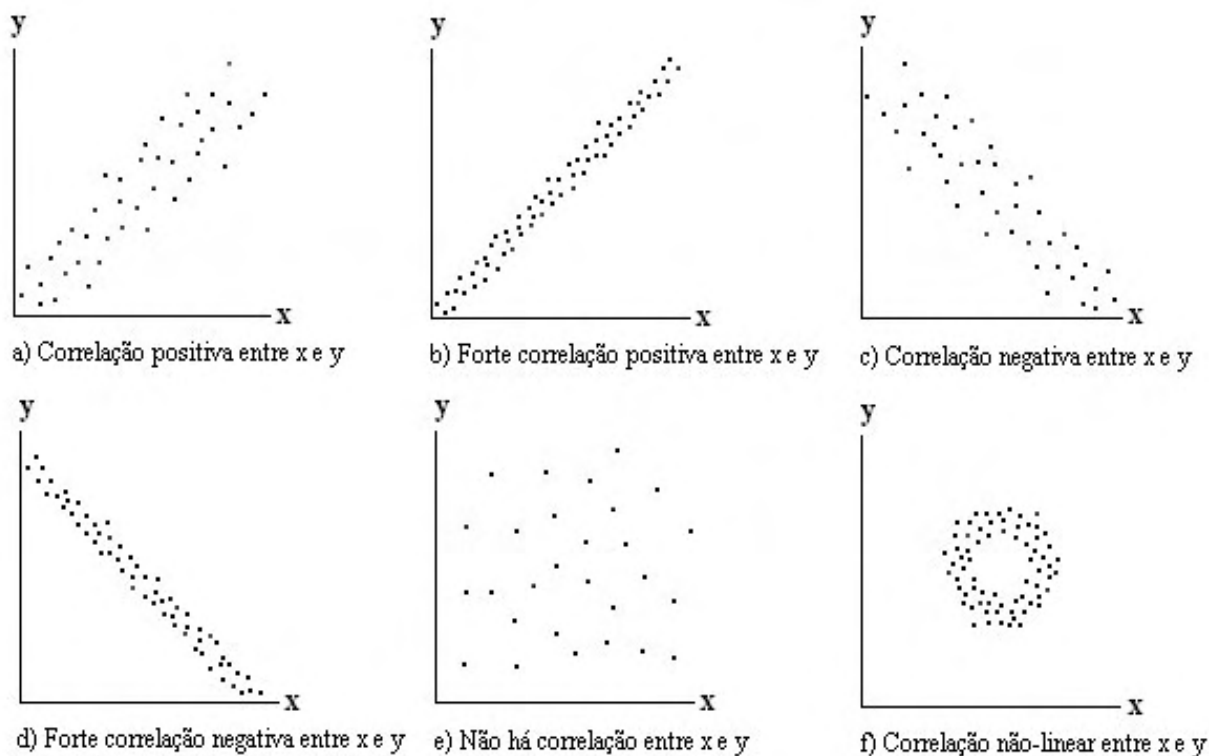


Figura 2.1: Representação qualitativa de correlações observadas.

As conclusões tiradas de diagramas de dispersão geralmente são subjetivas. Devido a isto, como método mais preciso e objetivo de análise, utiliza-se o valor r . Tal coeficiente mede o grau de relacionamento linear entre os valores do par x e y em uma amostra e, se for admitida uma correlação linear entre duas variáveis, r pode ser escrita como [21]:

$$r = \frac{n \sum XY - (\sum X)(\sum Y)}{\sqrt{n \sum X^2 - (\sum X)^2} \sqrt{n \sum Y^2 - (\sum Y)^2}} \quad (2.1)$$

sendo:

- n o tamanho da amostra em estudo;
- X uma das variáveis em questão, considerada como variável independente;
- Y a outra variável em questão considerada como dependente da variável X .

O uso da equação 2.1 é muito interessante, pois acaba por facilitar muito o cálculo computacional de r . Seu valor varia entre -1 e 1 inclusive, de modo que, se o valor obtido se aproximar de zero, diz-se que não há correlação entre as variáveis analisadas. Caso contrário, um valor próximo de -1 ou 1 , diz-se que há uma correlação linear negativa ou positiva, respectivamente. Além disto, r^2 corresponde à fração da variabilidade total dos dados explicada pelo modelo.

Ao se encontrar uma correlação significativa, utiliza-se métodos estatísticos para o ajuste da curva e sua validação (Regressões Lineares Simples e Múltiplas). Caso não fosse encontrada tal correlação, outros métodos estatísticos (Análise Discriminante Linear e Análise de Componentes Principais) seriam empregados na análise dos conjuntos de dados, reconhecendo padrões e buscando uma correspondência entre eles [22, 23].

2.4 Métodos Estatísticos

Conjunto de procedimentos ou técnicas que utilizam teorias probabilísticas para explicação dos mais variados eventos. Tem a finalidade de obter, analisar e agrupar dados, buscando uma correlação entre esses, produzindo a melhor informação possível a partir das disponíveis. São freqüentemente utilizados na análise da correlação dos resultados obtidos através do cálculo da estrutura eletrônica de substâncias e as propriedades demonstradas por elas.

Entre os métodos estatísticos, são descritos as Regressões Lineares Simples e Múltiplas, a Análise Discriminante Linear (LDA) e o método estatístico aplicado principalmente em química (conhecido como método quimiométrico) Análise de Componentes Principais (PCA), por estes terem sido adotados neste estudo [21, 24].

2.4.1 Regressão Linear Simples

Verificada a existência de uma correlação significativa, busca-se expressar uma das variáveis em função da outra. Para tanto, recorre-se a métodos de ajuste de curvas, procurando uma forma de relacionar as variáveis em questão de maneira mais simples. Para isto utiliza-se uma linha reta como curva de ajuste. Assim sendo, a equação de regressão tentada possui a forma:

$$Y_{est} = a_0 + a_1 X \quad (2.2)$$

sendo que:

- Y_{est} e X representam respectivamente as variáveis dependente e independente da regressão linear;
- a_0 e a_1 são constantes a serem determinadas de forma a se obter uma reta que mais se ajuste à disposição dos pontos no diagrama de dispersão.

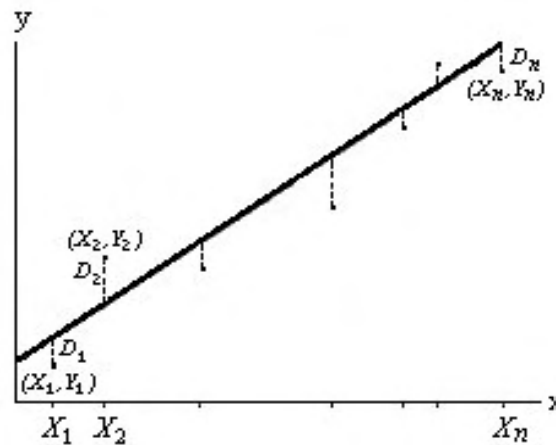


Figura 2.2: Regressão linear simples para pontos não igualmente espaçados.

Considerando a Figura 2.2 onde os dados de uma amostra de n elementos são representados por (X_1, Y_1) , (X_2, Y_2) , ..., (X_n, Y_n) , pode-se notar que, para um dado valor de X , por exemplo, X_i , haverá uma certa diferença D_i entre o valor de Y_i e o valor de Y_{est} dado pela reta t . Sendo:

$$D_i = (a_0 + a_1 X_i - Y_i) \quad (2.3)$$

que define:

$$s = \sum_{i=1}^n D_i^2 = \sum_{i=1}^n (a_0 + a_1 X_i - Y_i)^2 \quad (2.4)$$

Tem-se que uma medida da qualidade do ajustamento da reta t aos pontos apresentados está de certa forma relacionada ao desvio padrão s . De modo que, para um valor mínimo desta variável, tem-se a reta que melhor se ajusta aos pontos do diagrama de dispersão. Tal procedimento é conhecido como “mínimos quadrados”. Visando tal condição de minimização de s , pelo cálculo diferencial, sendo s uma função de a_0 e a_1 , obtém-se:

$$\frac{\partial s}{\partial a_0} = 2 \sum_{i=1}^n (a_0 + a_1 X_i - Y_i) = 0 \quad (2.5)$$

$$\frac{\partial s}{\partial a_1} = 2 \sum_{i=1}^n [(a_0 + a_1 X_i - Y_i) X_i] = 0 \quad (2.6)$$

ou seja:

$$\begin{cases} \sum_{i=1}^n Y_i = n a_0 + a_1 \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i Y_i = a_0 \sum_{i=1}^n X_i + a_1 \sum_{i=1}^n X_i^2 \end{cases} \quad (2.7)$$

Tais equações são denominadas equações normais da reta obtida por métodos de mínimos quadrados e a solução para o sistema 2.7 fornece os valores de a_0 (interceptação no eixo dos Y) e a_1 (coeficiente angular ou declividade da reta) da reta que melhor se ajusta aos pontos no gráfico. A nova equação é devidamente

validada através do cálculo do coeficiente linear (r), do desvio-padrão (s) e do grau de significância estatística (p), medida através do Teste-F [21, 25].

O Teste-F se constitui no cálculo de um valor (F), que é usado para determinar o valor de p a partir de variáveis dependentes e independentes, o qual é a probabilidade de rejeitar a hipótese de que não exista uma relação linear entre a variável dependente e a independente. Usualmente, valores de p abaixo de 0,05 são considerados muito bons, sendo que valores abaixo de 0,01 são considerados ótimos [21, 25].

2.4.2 Regressão Linear Múltipla

Freqüentemente, o uso de uma só variável independente (X) não é suficiente para se descrever um conjunto de dados (Y) através de uma regressão linear simples. A variável dependente, em muitos destes casos, é uma função de mais de um valor de X (X_1, X_2, \dots).

A fim de se obter uma aproximação satisfatória não muito complexa, admite-se que Y varia linearmente com as variáveis X_i , optando desta forma por uma regressão linear múltipla entre as variáveis independentes.

Tomando-se o caso mais simples de duas variáveis independentes X_1 e X_2 , supondo uma relação linear, a equação que relaciona Y_{est} com estas variáveis seria:

$$Y_{est} = a_0 + a_1 X_1 + a_2 X_2 \quad (2.8)$$

sendo a_0 , a_1 e a_2 constantes a serem determinadas. De maneira similar à regressão linear simples:

$$S = \sum_{i=1}^n (a_0 + a_1 X_{1i} + a_2 X_{2i} - Y_i)^2 \quad (2.9)$$

sendo S uma medida da qualidade do ajuste entre os valores de Y e Y_{est} , para cada X_1 e X_2 .

Da mesma forma, pela imposição do critério de mínimos quadrados, as derivadas referentes à a_0 , a_1 e a_2 devem ser nulas. Tal condição leva ao sistema de equações:

$$\begin{cases} \sum_{i=1}^n Y_i = na_0 + a_1 \sum_{i=1}^n X_{1i} + a_2 \sum_{i=1}^n X_{2i} \\ \sum_{i=1}^n X_{1i} Y_i = a_0 \sum_{i=1}^n X_{1i} + a_1 \sum_{i=1}^n X_{1i}^2 + a_2 \sum_{i=1}^n X_{1i} X_{2i} \\ \sum_{i=1}^n X_{2i} Y_i = a_0 \sum_{i=1}^n X_{2i} + a_1 \sum_{i=1}^n X_{1i} X_{2i} + a_2 \sum_{i=1}^n X_{2i}^2 \end{cases} \quad (2.10)$$

De maneira análoga, no caso geral, com um número k de variáveis independentes analisadas, a equação procurada toma a forma:

$$Y_{est} = a_0 + a_1 X_1 + a_2 X_2 + \dots + a_k X_k \quad (2.11)$$

onde, aplicando o método de mínimos quadrados tem-se:

$$\left\{ \begin{array}{l} \sum_{i=1}^n Y_i = na_0 + a_1 \sum_{i=1}^n X_{1i} + a_2 \sum_{i=1}^n X_{2i} + \dots + a_k \sum_{i=1}^n X_{ki} \\ \sum_{i=1}^n X_{1i} Y_i = a_0 \sum_{i=1}^n X_{1i} + a_1 \sum_{i=1}^n X_{1i}^2 + a_2 \sum_{i=1}^n X_{1i} X_{2i} + \dots + a_k \sum_{i=1}^n X_{1i} X_{ki} \\ \sum_{i=1}^n X_{2i} Y_i = a_0 \sum_{i=1}^n X_{2i} + a_1 \sum_{i=1}^n X_{1i} X_{2i} + a_2 \sum_{i=1}^n X_{2i}^2 + \dots + a_k \sum_{i=1}^n X_{2i} X_{ki} \\ \vdots \\ \sum_{i=1}^n X_{ki} Y_i = a_0 \sum_{i=1}^n X_{ki} + a_1 \sum_{i=1}^n X_{1i} X_{ki} + a_2 \sum_{i=1}^n X_{2i} X_{ki} + \dots + a_k \sum_{i=1}^n X_{ki}^2 \end{array} \right. \quad (2.12)$$

Existem muitas equações que permitem identificar qual o grau de correlação linear entre o plano obtido pela regressão linear múltipla e os gráficos de dispersão da amostra.

2.4.3 Análise Discriminante Linear

A Análise Discriminante Linear, conhecida como LDA (do inglês *Linear Discriminant Analysis*), é uma técnica estatística multivariada de reconhecimento de padrões. Frequentemente é utilizada em problemas que exigem a separação de um ou mais subconjuntos amostrais que apresentam determinadas características distintas. Ao se fazer uma LDA tem-se como principais objetivos:

1. Observar as variáveis que melhor discriminam (separam) os diferentes grupos (variáveis discriminantes);
2. Através destas variáveis discriminantes, desenvolver uma função que possibilite representar as diferenças entre os subconjuntos considerados (função discriminante);

3. E por fim, através da função obtida, ser capaz de classificar os demais casos, não presentes necessariamente no conjunto de dados iniciais, como pertencentes a um ou a outro grupo.

Seja um conjunto de variáveis independentes (descritores) $\mathbf{X}_k = \{X_{k1}, X_{k2}, \dots, X_{kj}\}$, que representam a k -ésima componente de um grupo de n amostras. Suponha que, segundo um determinado critério, este grupo de amostras esteja dividido em dois subgrupos distintos A e B , com respectivamente, n_1 e n_2 componentes. Através da LDA, procura-se por uma função (função discriminante) que possibilite diferenciar ao máximo os subgrupos A e B utilizando a informação contida em \mathbf{X}_k .

Dois métodos podem ser utilizados para a obtenção da função discriminante: o método simultâneo e o método *stepwise*. No método simultâneo, todas as variáveis são necessariamente utilizadas na obtenção da função, sendo que todo o conjunto de dados é empregado a fim de se observar uma distinção entre os subgrupos. O método *stepwise* é um método alternativo de confecção da função discriminante, onde as variáveis independentes são inseridas uma a uma conforme seu poder de separação dos grupos (tal separação pode ser calculada por diversos métodos, sendo o mais utilizado o Mahalanobis D^2 [26]). Supondo, por exemplo, que num primeiro passo a variável X_1 oferecesse a separação mais significativa entre os grupos, no segundo passo seria avaliada qual a variável que melhor discrimina os grupos em um conjunto sem a variável X_1 , e assim por diante, até não se observar mais nenhuma separação significativa (segundo um determinado critério de teste estatístico). Deste modo a LDA permite a redução do número de variáveis utilizadas para um número q de variáveis mais significativamente relacionadas com a separação entre os grupos. No presente caso, tal método é bastante interessante,

pois se deseja descobrir quais índices melhor se relacionam com a atividade antimalarial além do fato de que o número de descritores eletrônicos calculados é muito grande.

A função discriminante, deste modo, é construída a partir da combinação linear dos q descritores X_{ki} mais significativos obtidos através do método *stepwise* de maneira semelhante a uma regressão linear múltipla.

$$Z_k = a_0 + a_1 X_{k1} + a_2 X_{k2} + \dots + a_p X_{kp} = a_0 + \sum_{i=1}^p a_i X_{ki} \quad , \quad k=1, 2, \dots, n \quad (2.13)$$

sendo que:

- Z_k são denominados *scores* da função discriminante para a k-ésima amostra;
- a_i são os coeficientes da i-ésima variável independente utilizada na construção da função discriminante;
- X_{ki} são os i-ésimos descritores da k-ésima amostra.

Na Regressão Linear Múltipla, procura-se por uma combinação entre as variáveis independentes que promova o melhor ajuste dos valores estimados com os valores da variável dependente, ou seja, procura-se por coeficientes que minimizem os desvios entre os valores previstos e os valores reais da variável dependente. Na LDA, a combinação linear descrita, visa obter os coeficientes que retratem a máxima diferença entre os subgrupos *A* e *B*.

Define-se:

$$Z_{kA} = a_0 + a_1 X_{k1A} + a_2 X_{k2A} + \dots + a_p X_{kpA} = a_0 + \sum_{i=1}^p a_i X_{kiA}, \quad k=1, 2, \dots, n_1 \quad (2.14)$$

$$Z_{kB} = a_0 + a_1 X_{k1B} + a_2 X_{k2B} + \dots + a_p X_{kpB} = a_0 + \sum_{i=1}^p a_i X_{kiB}, \quad k=1, 2, \dots, n_2 \quad (2.15)$$

sendo que:

- Z_{kA} e Z_{kB} são denominados *scores* da função discriminante para a k-ésima amostra do subgrupo *A* e do subgrupo *B*, respectivamente;
- X_{kiA} e X_{kiB} são os i-ésimos descritores da k-ésima amostra do subgrupo *A* e do subgrupo *B*, respectivamente.

As médias dos *scores*, de acordo com 2.14 e 2.15, são dadas por:

$$\bar{Z}_A = \frac{1}{n_1} \sum_{k=1}^{n_1} Z_{kA} = \frac{1}{n_1} \sum_{k=1}^{n_1} \left(a_0 + \sum_{i=1}^p a_i X_{kiA} \right) = \frac{a_0}{n_1} + \sum_{i=1}^p a_i \left(\frac{1}{n_1} \sum_{k=1}^{n_1} X_{kiA} \right) = \frac{a_0}{n_1} + \sum_{i=1}^p a_i \bar{X}_{iA} \quad (2.16)$$

$$\bar{Z}_B = \frac{1}{n_2} \sum_{k=1}^{n_2} Z_{kB} = \frac{1}{n_2} \sum_{k=1}^{n_2} \left(a_0 + \sum_{i=1}^p a_i X_{kiB} \right) = \frac{a_0}{n_2} + \sum_{i=1}^p a_i \left(\frac{1}{n_2} \sum_{k=1}^{n_2} X_{kiB} \right) = \frac{a_0}{n_2} + \sum_{i=1}^p a_i \bar{X}_{iB} \quad (2.17)$$

É importante frisar que, numa distribuição normal, a média dos *scores* (centróide) das amostras de um subgrupo indica a localização mais provável dos componentes pertencentes a este subgrupo. A comparação entre os centróides, desta forma, permite dizer o quão afastado se encontra um subgrupo do outro.

Assim sendo, na LDA, a combinação linear para a construção da função discriminante tem como objetivo principal obter o conjunto $\mathbf{a} = \{a_1, a_2, \dots, a_p\}$ de

coeficientes que maximize a diferença entre os valores de \bar{Z}_A e \bar{Z}_B , dividida pelo desvio padrão total (s_z), ou seja, procura-se por uma combinação linear que maximize a distância entre os centróides dos subgrupos com relação ao espaço todo observado:

$$MAX \left[\frac{(\bar{Z}_A - \bar{Z}_B)}{s_z} \right] \quad (2.18)$$

Considerando que tal valor pode ser negativo, usualmente considera-se o seu quadrado, obtendo-se desta forma a variância dos dados (s_z^2) no denominador:

$$MAX \left[\frac{(\bar{Z}_A - \bar{Z}_B)^2}{s_z^2} \right] \quad (2.19)$$

A expressão 2.18 pode ser escrita de maneira mais prática da seguinte forma:

$$MAX \left[\frac{\tilde{a}^T (\tilde{X}_A - \tilde{X}_B) (\tilde{X}_A^T - \tilde{X}_B^T) \tilde{a}}{\tilde{a}^T \tilde{S}_c \tilde{a}} \right] \quad (2.20)$$

sendo que \tilde{a} é o vetor dos coeficientes a_i ; \tilde{X}_A e \tilde{X}_B são os vetores das médias dos descritores X_i do subgrupo A e B , respectivamente; \tilde{S}_c é a matriz covariância comum a ambos os subgrupos sendo dada por:

$$\tilde{S}_c = \frac{1}{n_1 + n_2 - 2} [(n_1 - 1) \cdot \tilde{S}_1 + (n_2 - 1) \cdot \tilde{S}_2] \quad (2.21)$$

sendo que \tilde{S}_1 e \tilde{S}_2 são respectivamente as matrizes de covariância amostrais referentes a cada subgrupo. Os índices T indicam que se trata da matriz transposta.

O ponto máximo da equação 2.20 ocorre para os valores de a_i dados por:

$$\tilde{a} = (\tilde{X}_A - \tilde{X}_B)^T \tilde{S}_c^{-1} \quad (2.22)$$

e é dada por:

$$\tilde{Z} = (\tilde{X}_A - \tilde{X}_B)^T \tilde{S}_c^{-1} \tilde{X} \quad (2.23)$$

sendo que:

- \tilde{Z} é a matriz que contém os valores dos *scores* discriminantes Z_k ; \tilde{X} é a matriz que contém os descritores X_j para cada amostra k .

Os coeficientes a_i , também denominados pesos discriminantes, indicam o quanto a variável independente contribui para a construção da função discriminante, sendo uma medida do poder de classificação que cada variável possui.

Após a obtenção da função discriminante, visando obter um valor crítico de classificação que permita diferenciar amostras pertencentes ao grupo A das pertencentes ao grupo B , define-se o parâmetro de corte m . Tal parâmetro serve como um critério de classificação, sendo definido por:

$$m = \frac{1}{2}(\bar{Z}_A + \bar{Z}_B) \quad (2.24)$$

Amostras que apresentarem um valor de *scores* $Z \geq m$ serão classificadas num determinado subgrupo, enquanto que variáveis que apresentem *scores* $Z < m$ serão classificadas no outro subgrupo.

Na realidade, através da LDA transforma-se a informação multivariada contida em X , de dimensão q , numa informação univariada Z , que acaba por facilitar imensamente a análise dos dados. Um exemplo se encontra na Figura 2.3. Considerando um conjunto de apenas duas variáveis independentes (X_1 e X_2), encontra-se uma função que permite, através de um parâmetro de corte, diferenciar e classificar dois subgrupos descritos por variáveis independentes que obedecem a uma distribuição normal.

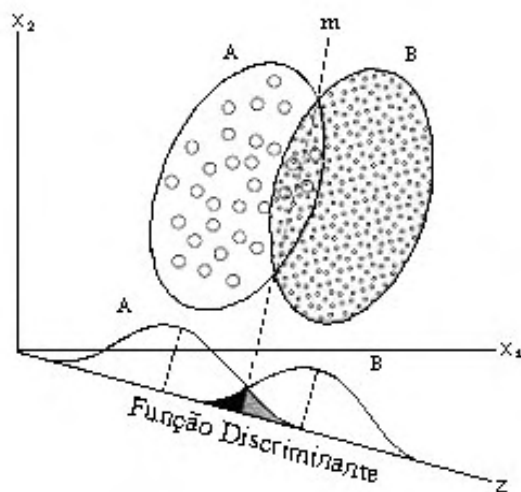


Figura 2.3: Representação gráfica do resultado obtido através da LDA [26], linha tracejada representa o parâmetro de corte m .

O parâmetro de corte é calculado de forma a tornar a região marcada de intersecção entre os dois subgrupos, a menor possível.

Na explicação aqui descrita considerou-se apenas o problema de classificação de dois subgrupos dentro do conjunto de amostras. Existe ainda a possibilidade de se considerar três ou mais subgrupos. Neste caso, utiliza-se a denominada Análise Discriminante Linear Múltipla - MLDA (*Multiple Linear Discriminant Analysis*) como método de classificação, onde são obtidas mais de uma função discriminante. Como no trabalho aqui descrito, utilizou-se apenas da LDA simples; optou-se por não se descrever em mais detalhes a MLDA.

É importante frisar que, tanto na LDA como na MLDA, o conjunto de dados deve satisfazer alguns critérios aos quais a análise é muito sensível [26]:

1. Distribuição normal das variáveis independentes;
2. Igualdade das matrizes variância-covariância dos grupos.

2.5 Método Quimiométrico

Quimiometria é definida pela Sociedade Internacional de Quimiometria (*International Chemometrics Society - ICS*) como sendo:

“Uma ciência relacionada a medidas realizadas em um sistema ou processo químico, na qual é possível obter informações sobre o estado do sistema através da aplicação de métodos matemáticos ou estatísticos”.

ou seja, métodos quimiométricos são métodos estatísticos aplicados especificamente em processos químicos. Os diferentes métodos quimiométricos surgiram da necessidade de melhorar as análises quantitativas que eram realizadas na maioria das vezes por “via úmida”, como por exemplo: titulação, precipitação e reações específicas. Essas eram muitas vezes demoradas, pouco precisas e não

permitted a direct information about the problem, but produced a large quantity of data about the system, which had to be treated for elaboration and interpretation of mathematical models.

Chemometric methods serve for interpretation of these data and for the extraction of the maximum of information about the systems analyzed. Among the most used methods we can cite: *Principal Component Analysis (PCA)*; *Hierarchical Cluster Analysis (HCA)*; *Principal Component Regression (PCR)*; *Partial Least Squares (PLS)*; *Kth Nearest Neighbor (KNN)*; *Independent Modeling Of Class Analogy (SIMCA)*.

For the analysis of the results obtained, the chemometric method of Principal Component Analysis (PCA) [24].

2.5.1 Análise de Componentes Principais

The Principal Component Analysis, known as PCA (from the English *Principal Component Analysis*), is a statistical technique of great relevance in the analysis of data sets, being widely used as a method of pattern recognition in chemometrics [22].

In PCA, one seeks to rewrite the coordinates (independent variables) of the samples in a system of axes more convenient for the analysis of the data. The n original variables generate, through their linear combinations, n principal components (PC's), whose main characteristic, besides orthogonality, is the fact of being obtained in order of decreasing maximum variance. So that the PC1 (principal component 1) brings more statistical information than the PC2, the latter in its turn has more statistical information than the PC3 and so on.

Assim sendo, são suprimidas todas as correlações possíveis entre as variáveis originais, uma vez que o conjunto das PC's é ortogonal, ou seja, cada nova variável traz uma informação estatística diferente das demais. Devido ao decréscimo de importância das PC's, freqüentemente pode-se diminuir a dimensionalidade do problema inicial desprezando as componentes que possuam menor informação estatística.

Aplicando-se as condições de máxima variância residual e ortogonalidade no cálculo das componentes principais, chega-se a um sistema de equações matriciais que pode ser descrito de um modo elegante como exposto a seguir.

Define-se uma matriz de dados $m \times n$ \tilde{D} , onde n representa o número de medidas de diversas propriedades efetuadas sobre m amostras quaisquer:

$$\tilde{D} = \begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{m1} & d_{m2} & \cdots & d_{mn} \end{pmatrix} \quad (2.25)$$

Pode-se afirmar que, freqüentemente, as propriedades que irão constituir \tilde{D} não possuem as mesmas grandezas, de forma que seus valores numéricos podem ser acentuadamente diferentes. A fim de se facilitar a comparação direta entre as variáveis, sem perder a informação estatística, realiza-se sobre a matriz \tilde{D} uma transformação, de modo que cada variável apresente média zero e variância igual a 1. Tal processo é dito auto-escalonamento, sendo uma das várias opções de transformação de \tilde{D} .

A fim de se obter uma medida da similaridade, entre as variáveis que compõem \tilde{D} , calcula-se a Matriz Covariância \tilde{C} , cujos elementos são dados por:

$$c_{kl} = \frac{1}{m-1} \sum_{i=1}^m (d_{ik} - \bar{d}_k)(d_{il} - \bar{d}_l) \quad (2.26)$$

sendo que:

- \bar{d}_j é o valor da média da variável j de \tilde{D} .

Assim definida, a matriz \tilde{C} carrega informações a respeito do tipo de associação entre as variáveis originais. Se os elementos de \tilde{C} apresentarem um grande valor positivo, diz-se que as variáveis apresentam uma correlação positiva, de modo que, ao se aumentar o valor de uma delas, em média é notado um respectivo acréscimo na outra. O caso de uma covariância negativa, da mesma forma indica uma correlação negativa entre as variáveis e a presença de componentes iguais (ou próximos) a zero indica a ausência de correlação. Muitas vezes também se utiliza a Matriz Correlação, cujos elementos são uma covariância padronizada entre -1 e 1 , para se obter tal medida de similaridade.

Como a matriz \tilde{C} é uma matriz quadrada de ordem n , sendo por definição simétrica, ela pode ser diagonalizada de modo a se obter seus autovalores (λ) e autovetores (v):

$$\tilde{C} \tilde{v}_i = \lambda_i \tilde{v}_i, \quad i = 1, \dots, n \quad (2.27)$$

sendo que:

$$\tilde{v}_i = \begin{pmatrix} v_{1i} \\ v_{2i} \\ \vdots \\ v_{ni} \end{pmatrix}; \quad (2.28)$$

e λ é um escalar.

A partir dos n autovetores de 2.27, gera-se uma matriz transformação \tilde{V} , que aplicada sobre a matriz de dados escalonada resultará nas novas variáveis (PC's). Os autovetores constituem as colunas desta matriz, sendo ordenados de acordo com os seus respectivos autovalores, ou seja, na primeira coluna se encontra o autovetor referente ao maior autovalor obtido, na segunda o autovetor associado ao segundo maior autovalor e assim por diante. Tal ordenamento se deve ao fato de que os autovalores λ são as variâncias associadas a cada novo eixo, uma vez que, ordenando-se a matriz \tilde{V} como descrito, as primeiras PC's terão a máxima variância do conjunto de dados (maior informação estatística) como se pretendia.

$$\tilde{V} = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n1} & v_{n2} & \cdots & v_{nn} \end{pmatrix} \quad (2.29)$$

Tendo-se a matriz \tilde{V} , basta aplicá-la sobre os valores escalonados de \tilde{D} para se obter as coordenadas dos dados no novo sistema.

$$\tilde{U} = \tilde{D}\tilde{V} \quad (2.30)$$

Sendo que \tilde{U} é uma matriz $m \times n$ com elementos u_{ij} . As coordenadas u_{ij} do novo sistema são denominadas *scores*. As componentes dos autovetores de \tilde{V} (denominados *loadings*) explicitam a importância da contribuição das variáveis originais na composição em que são obtidas n PC's. No entanto, devido ao decréscimo de informação estatística observada conforme o índice destas componentes vão aumentando (PC1, PC2,...), freqüentemente considera-se somente as primeiras PC's componentes que juntas totalizem 80 ou 90% da variância total. É comum se obter mais de 90% de informação somente tomando-se as PC's 1 e 2, de modo que a plotagem dos *scores* destas duas componentes fornece uma visão estatisticamente privilegiada do conjunto de dados.

CAPÍTULO 3

Moléculas com Atividade Antimalarial

A malária está entre as mais importantes doenças provocadas por parasitas em muitos países tropicais. Segundo a Organização Mundial de Saúde (OMS), atualmente a malária é a maior doença tropical parasitária no mundo, provocando inúmeros problemas sociais e econômicos, sendo superada em número de mortes apenas pela AIDS.

Estima-se que em torno de 40% da população mundial vive em locais onde podem ser infectadas pela malária. Isso abrange cerca de 100 países, principalmente os da América Latina e dos continentes Africano e Asiático. No Brasil, principalmente na região Amazônica, a malária registra por volta de 500 mil novos casos por ano, cifra que pode chegar a 1,5 milhão contando com os casos não-registrados. Felizmente, aqui a letalidade da moléstia é baixa e não chega a 0,1% do número total de enfermos [27].

A malária é causada por quatro espécies de protozoários parasitas do gênero *Plasmodium*: *P. falciparum*, *P. vivax*, *P. ovale* e *P. malariae*, responsáveis por diferentes formas sintomáticas da doença. Dentre estes, o *P. falciparum* é o mais difundido e perigoso, sendo a espécie plasmodial capaz de produzir quadros de

extrema gravidade, podendo culminar em êxito letal se não tratada rapidamente. Infelizmente essa é também a espécie mais resistente às drogas usadas atualmente [28].

Os parasitas são transmitidos de uma pessoa infectada para outra através da fêmea do mosquito do gênero *Anopheles*, possuindo dois ciclos de desenvolvimento distintos: um no hospedeiro (humano) e outro no vetor (mosquito). A contaminação dos humanos se dá quando o mosquito ao procurar nestes o sangue para se alimentar, transporta o parasita de um humano já contaminado para outro. Não existem vacinas disponíveis contra esta doença e o tratamento da malária visa abolir o ciclo de reprodução sanguínea do parasita. Entretanto, as diferentes espécies de plasmódio e os vários estágios evolutivos de cada uma destas espécies apresentam respostas de diversos graus aos vários tratamentos empregados [27].

No combate dessa moléstia diferentes drogas foram utilizadas, como por exemplo: o quinino, a cloroquina e a mefloquina. Todas ao longo do tempo apresentaram redução da eficácia aos diferentes tipos de *Plasmodium*, principalmente em linhagens de *P. falciparum* [29]. O surgimento desses focos resistentes aos medicamentos e sua rápida dispersão nas áreas malarígenas da Ásia, Américas e África, foram uma das razões para o “fortalecimento” da doença. Em alguns pontos dessas regiões endêmicas, os antimalaríais são ineficazes em cerca de dois terços dos casos.

Uma das opções para contornar o problema é a combinação de drogas. Nesse caso, sua adoção está praticamente restrita a países com alto poder aquisitivo devido ao elevado custo destes novos medicamentos. Para tentar amenizar os custos, diversos países são obrigados a adaptar seus sistemas de saúde, forçando os pacientes a arcarem com parte de seu tratamento. O problema desta política está

no fato que a maioria das vítimas dessa doença são crianças da zona rural de países economicamente menos favorecidos. Sem uma vacina efetiva, a quimioprofilaxia e a quimioterapia são atualmente as únicas armas capazes de controlar e combater os parasitas da malária. Estes fatos levam a uma urgência no desenvolvimento de novas substâncias com propriedades antimalariais.

Não só o combate à malária, mas outras doenças graves como a tuberculose e a leishmaniose exigem, por assim dizer, o desenvolvimento de fármacos de interesse nacional. Essas pesquisas são feitas na maioria das vezes por instituições ligadas ao governo, pois não há interesse dos grandes grupos farmacêuticos internacionais em investir em tratamentos para doenças de populações pobres. Assim, por este e outros motivos, é compreensível que o desenvolvimento de novos fármacos é atualmente considerado uma das áreas estratégicas dentro da nova Política Industrial Tecnológica e de Comércio Exterior do Governo Federal.

Atualmente é impensável o estudo de novas drogas sem o auxílio da modelagem molecular, em um esforço que envolve física, química, biologia, matemática, estatística, computação, farmacologia e medicina. Dentre as contribuições que a Física pode dar a este campo de pesquisa, a área de estrutura eletrônica tem papel de destaque, pois agrega dados intrínsecos das substâncias estudadas, ou seja, os dados dizem respeito quase que exclusivamente às propriedades das moléculas em si, uma vez que é bastante difícil incluir as interações com o meio em que estas moléculas agem. É bastante comum, inclusive, se observar uma forte correlação da atividade biológica de interesse com algum resultado da estrutura eletrônica da molécula supondo que esta se encontra isolada no vácuo. Em existindo tal correlação, torna-se claro que isto é de grande valia no projeto de novas substâncias com melhorias na atividade biológica de interesse.

Considerando que muitos fenômenos biológicos ocorrem a partir da interação entre moléculas e os organismos vivos, e que muitas vezes é possível correlacionar a atividade demonstrada por uma dada molécula com a sua estrutura eletrônica (por exemplo, a atividade anticancerígena das elipticinas é modulada pelo momento de dipolo elétrico das moléculas [6]), buscam-se através de uma análise computacional da estrutura eletrônica, informações importantes que possam ser utilizadas no planejamento racional de fármacos.

Neste trabalho são analisados dois grupos de moléculas que já demonstraram potencial antimalarial e que ainda são reportados na literatura: as sulfonamidas e as anilinoquinolinas. Acredita-se que a estrutura eletrônica poderá trazer uma boa contribuição para o desenvolvimento de novos fármacos ligados a essas substâncias, pois uma revisão da literatura mostra que não há ainda estudos para correlacionar a estrutura eletrônica e a atividade destas.

3.1 Sulfonamidas

As sulfonamidas são grupos de antibióticos sintéticos, muito utilizados no combate à malária e a outras doenças infecciosas. Possuem a capacidade de impedir a multiplicação de microorganismos, inibindo a enzima dihidropteroato sintetase (DHPS), responsável por catalisar uma reação necessária à síntese de ácido fólico imprescindível na composição de precursores de RNA e DNA do microorganismo. Como as células humanas obtêm o ácido da dieta, não possuem a enzima, não sendo afetadas.

Um estudo realizado recentemente [4] apontou uma família de moléculas derivadas das sulfonamidas como uma possível alternativa contra a malária. Ainda

que tenham sido registrados vários efeitos colaterais, a continuidade da pesquisa se justifica pelo fato destas serem um caminho possível dentro das terapias disponíveis e pelo baixo custo quando comparadas aos atuais medicamentos.

Neste trabalho será analisado um grupo de 16 moléculas derivadas da estrutura básica 2,4-diamino-6-quinazolina-sulfonamida, mostrada abaixo (Figura 3.1). O grande interesse nesse grupo é devido ao potencial de inibição da multiplicação do agente parasita.

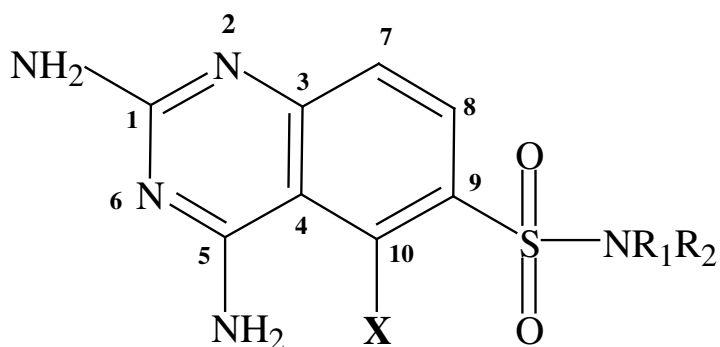


Figura 3.1: Representação esquemática da estrutura básica 2,4-diamino-6-quinazolina-sulfonamida. A simbologia R_1R_2 e X representa os diferentes substituintes que se ligam à molécula.

A propriedade biológica de interesse que se tenta correlacionar com a estrutura eletrônica dessas moléculas é a atividade antimalarial observada contra protozoários parasitas do gênero *Plasmodium*. Essa atividade é fornecida pelo índice experimental Δ MST (*difference of mean survival time*) [4]. Este é a diferença do tempo de vida fornecida em dias de um rato contaminado pelo *P. falciparum*, tratado por uma determinada molécula (Tabela 3.1) e um rato controle, que somente será infectado pelo protozoário e não receberá nenhum tipo de medicamento.

A Tabela 3.1 mostra os substituintes da 2,4-diamino-6-quinazolina-sulfonamida, que geram 16 moléculas, todas abordadas neste estudo e na Tabela 3.2 são representadas as diferentes estruturas dessas moléculas. Adotou-se a mesma nomenclatura empregada no trabalho experimental [4], visando sempre facilitar a comparação dos resultados.

Tabela 3.1: Substituintes da 2,4-diamino-6-quinazolina-sulfonamida e as respectivas propriedades.

Moléculas	Substituintes		Propriedade
	NR ₁ R ₂	X	ΔMST (em dias)
1	N(C ₂ H ₅) ₂	H	3,30
2	N(CH ₂) ₅	Cl	2,30
3	N(CH ₂ CH ₂ CH ₃) ₂	H	0,30
4	N(CH ₂ CH ₂ OH) ₂	H	0,30
5	N(CH ₃)CH(CH ₃) ₂	H	0,70
6	N(CH ₃)CH ₂ CH ₂ N(C ₂ H ₅) ₂	H	0,10
7	N(CH ₂) ₅	H	4,40
8	N(CH ₂) ₄	H	5,00
9	N[(CH ₂) ₂] ₂ O	H	4,70
10	N[(CH ₂) ₂] ₂ S	H	2,50
11	N[(CH ₂) ₂] ₂ NCH ₃	H	1,00
12	N[(CH ₂) ₂] ₂ NC(=O)OC ₂ H ₅	H	0,20
13	4Cl-C ₆ H ₄ NH	H	0,70
14	3Br-C ₆ H ₄ NH	H	0,30
15	4Cl-C ₆ H ₄ NCH ₃	H	0,30
16	C ₆ H ₅ NCH ₃	H	0,50

Tabela 3.2: Representação esquemática dos 16 moléculas derivadas da 2,4-diamino-6-quinazolina-sulfonamida (Figura 3.1).

Moléculas	Representação da estrutura	Nome IUPAC
1		<p>((2,4-diaminoquinazolin-6-yl) sulfonyl) diethylamine</p>
2		<p>((2,4-diaminoquinazolin-5-chloroquinazolin-6-yl) sulfonyl) diethylamine</p>
3		<p>((2,4-diaminoquinazolin-6-yl) sulfonyl) dipropylamine</p>
4		<p>((2,4-diaminoquinazolin-6-yl) sulfonyl) bis(2-hydroxyethyl)amine</p>
5		<p>((2,4-diaminoquinazolin-6-yl) sulfonyl) methyl(methylethyl)amine</p>

Moléculas	Representação da estrutura	Nome IUPAC
6		<i>((2,4-diaminoquinazolin-6-yl) sulfonyl) (2-(diethylamine) ethyl) methylamine</i>
7		<i>6-(piperidysulfonyl) quinazoline-2,4-diamine</i>
8		<i>6-(pyrrolidinylsulfonyl) quinazoline-2,4-diamine</i>
9		<i>6-(morpholin-4-yl-sulfonyl) quinazoline-2,4-diamine</i>
10		<i>6-(4,1-thiazaperhydroinylsulfonyl) quinazoline-2,4-diamine</i>

Moléculas	Representação da estrutura	Nome IUPAC
11		6-((4-methylpiperazinyl) sulfonyl) quinazoline-2,4-diamine
12		ethyl 4-((2,4-diaminoquinazolin-6-yl) piperazinecarboxylate)
13		((2,4-diaminoquinazolin-6-yl) sulfonyl) (4-chlorophenyl) amine
14		((2,4-diaminoquinazolin-6-yl) sulfonyl) (3-bromophenyl) amine
15		((2,4-diaminoquinazolin-6-yl) sulfonyl) (3-chlorophenyl) amine
16		((2,4-diaminoquinazolin-6-yl) sulfonyl) methylphenylamine

3.1.1 Resultados

A seguir, serão analisados os resultados apresentados nas regressões lineares simples e múltiplas, na análise dos componentes principais e na análise discriminante linear para as moléculas com atividade antimalarial derivados da 2,4-diamino-6-quinazolina-sulfonamida.

3.1.1.1 Regressões Lineares Simples e Múltiplas

Inicialmente com o auxílio de um pacote de programas de análise estatística desenvolvidos pelo Grupo de Física Teórica de Biomoléculas e Matérias (BioMat) da UNESP/Bauru e servindo-se dos resultados provenientes dos cálculos de estrutura eletrônica, obteve-se um conjunto de 113 descritores eletrônicos para cada uma das 16 sulfonamidas do estudo. Ainda com ajuda dos programas citados acima, foram efetuadas regressões lineares simples e múltiplas entre os valores de ΔMST (variável dependente) e os diferentes descritores eletrônicos apresentados na Tabela 2.1 (variáveis independentes).

Na Tabela 3.3 são apresentados os parâmetros de correlação linear (r) mais expressivos, obtidos dos 228,133 estudos de regressão linear simples e múltipla, tomando todas as combinações possíveis de duas e três variáveis independentes (V_i). Nesta tabela também está expressa o número de combinações que apresentaram o mesmo parâmetro r (NC), assim como os descritores envolvidos.

Tabela 3.3: Resultado das regressões lineares simples e múltiplas considerando todas as moléculas.

Regressão Linear:	r	NC	Descritores Eletrônicos		
			Vi 1	Vi 2	Vi 3
Simple	0,64	1	E_{ELET}	-	-
Múltipla com 2 Vi	0,72	2	E_{ELET}	$PEP_L(6)$	-
			E_{ELET}	$PEP_L(8)$	-
Múltipla com 3 Vi	0,81	2	E_{ELET}	$PEP_L(6)$	$LDOS(8)$
			E_{ELET}	$PEP_L(8)$	$LDOS(8)$

Os parâmetros r mais expressivos obtidos possuem valores relativamente baixos (valores considerados significativos são superiores a 0,9) não sendo interessante representá-los graficamente. Tal fato retrata a ausência de uma correlação linear forte entre as variáveis consideradas, de modo que a proposição de uma regra de classificação para todas as moléculas não pôde ser definida nesta etapa da análise. No entanto, nota-se que em todas as combinações mais expressivas os descritores eletrônicos se repetem e possuem alguma ligação com os átomos 6 e 8 da estrutura básica (Figura 3.1), indicando uma possível região de interesse a ser estudada.

É curioso observar nas regressões apresentadas na Tabela 3.3 a presença do descritor referente à energia dos elétrons da camada de valência das moléculas (E_{ELET}), ou seja, a energia dos elétrons mais “fracamente” ligados (portanto facilmente retirados), responsáveis pelas ligações químicas. Isto indica que a atividade antimalarial pode estar de alguma forma relacionada com esse descritor.

3.1.1.2 Análise dos Componentes Principais

O PCA é empregado no reconhecimento de padrões, permitindo em muitos casos agrupar determinados objetos de acordo com os aspectos comuns entre eles, reescrevendo as coordenadas (variáveis independentes) das amostras em um sistema de eixos mais convenientes para a análise dos dados.

Com esse método, pretende-se encontrar um possível agrupamento das moléculas ativas. Para isso são analisadas todas as combinações possíveis de 3 variáveis independentes na matriz de dados composta por 16 linhas (moléculas) e 113 colunas (variáveis independentes). Este estudo foi realizado de maneira automática, utilizando o programa *biomat_01*, desenvolvido pelo grupo de pesquisa BioMat. Tal análise é relevante, pois freqüentemente com um número pequeno de variáveis se pode ter uma separação expressiva entre as moléculas.

O programa classifica as combinações mais expressivas, plotando PC2 contra PC1, separando as moléculas de maior atividade daqueles que foram considerados com menor atividade, em dois polígonos distintos.

Deste modo procura-se por combinações nas quais os polígonos não se interceptem e que a distância entre seus centros geométricos (CG) seja a máxima possível. O resultado ideal seria o caso em que todas as moléculas se apresentassem em grupos (“clusters”) distintos. Mas um resultado também pode ser considerado satisfatório desde que tenha uma separação visual entre as classes estudadas. Para separação dos grupos consideramos o índice Δ MST como critério. A seguir as moléculas separadas em cada grupo:

Grupo 1: apresenta o valor Δ MST inferior ou igual a 2,0. São as moléculas: 3, 4, 5, 6, 11, 12, 13, 14, 15 e 16 consideradas de menor atividade antimalarial;

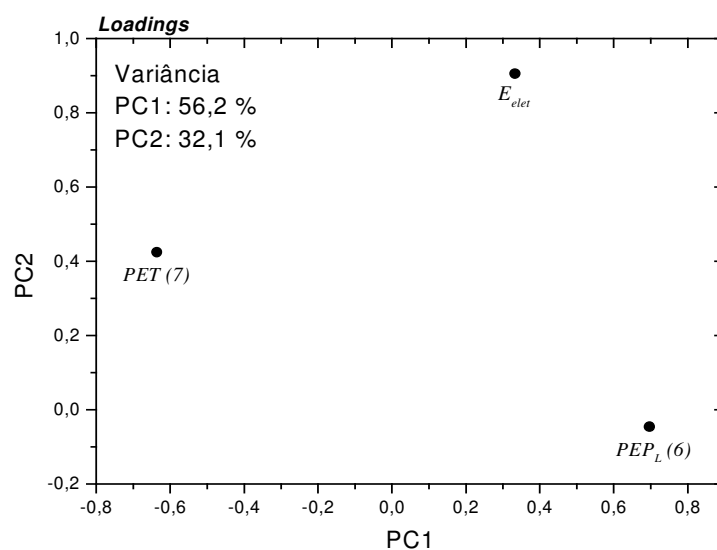
Grupo 2: apresenta o valor Δ MST superior a 2,0. São as moléculas: 1, 2, 7, 8, 9 e 10 consideradas de maior atividade antimalarial.

Através da automação obteve-se mais de 33000 combinações possíveis, considerando combinações de 3 descritores. Foi escolhida a combinação de melhor separação dos grupos. Neste estudo a PC1 e PC2 obtidas detêm juntas 88,3% da informação estatística, sendo 56,2% devido à PC1 e 32,1% devido à PC2. Os resultados podem ser vistos abaixo:

Tabela 3.4: Dados da variância dos componentes principais PC1 e PC2.

	Variância	Porcentagem	Acumulativo
PC1	1,69	56,2 %	56,2 %
PC2	0,963	32,1 %	88,3 %

a)



b)

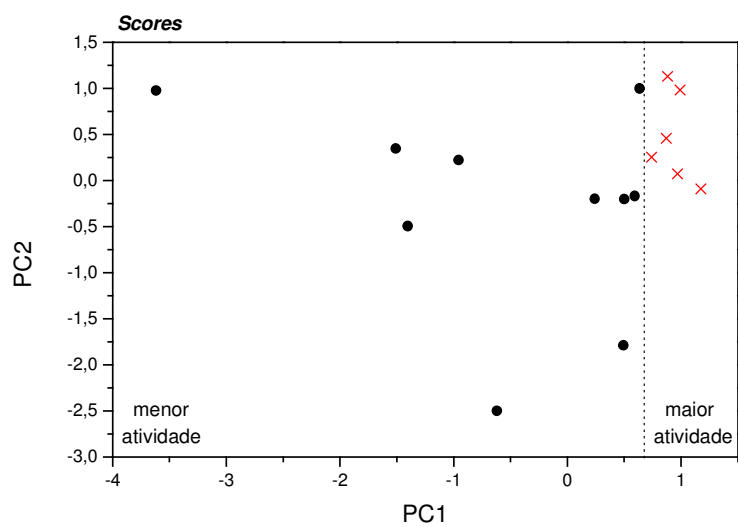


Figura 3.2: a) Contribuição das variáveis independente em PC1 e PC2.

b) Separação obtida para o conjunto das 2,4-diamino-6-quinazolina-sulfonamida.

As equações para as componentes do PCA são dadas por:

$$PC1 = 0,333(E_{ELET}) + 0,697(PEP_L(6)) - 0,636(PET(7)) \quad (3.10)$$

$$PC2 = 0,905(E_{ELET}) - 0,046(PEP_L(6)) + 0,424(PET(7)) \quad (3.11)$$

A Figura 3.2 - a (*Loadings*) representa o quanto cada variável independente (V_i) contribui em PC1 e PC2. Na Figura 3.2 - b (*Scores*) estão plotados os valores de PC1 contra PC2 para cada molécula, valores calculados pelas equações 3.10 e 3.11. Nos resultados para os *Scores*, é possível notar a separação clara de dois grupos de moléculas: um contendo as ativas (1, 2, 7, 8, 9 e 10) e outro contendo as inativas (3, 4, 5, 6, 11, 12, 13, 14, 15 e 16). As variáveis E_{ELET} , $PEP_L(6)$ e $PET(7)$ foram suficientes para a separação das classes de atividades, como pode ser visto pelas equações. Também foram suficientes somente as componentes principais PC1 e PC2 para a separação total das moléculas dos grupos 1 e 2.

Dois descritores de maior relevância nas regressões lineares simples e múltiplas (E_{ELET} , $PEP_L(6)$) aparecem também na PCA. Esse fato indica a importância destes na correlação entre estrutura e atividade.

3.1.1.3 Análise Discriminante Linear

Com a LDA, é possível encontrar outras possíveis separações quanto à atividade biológica, reconhecendo padrões entre os descritores. Os cálculos da LDA foram todos realizados com o auxílio do pacote estatístico SPSS® (*Statistical Package for the Social Sciences*) para Windows, versão 11.0.0 [30]. Como critério de

classificação consideram-se dois grupos distintos usando o Δ MST como referência e adotando os mesmos grupos utilizados no PCA, sendo grupo 1 (menor atividade) e grupo 2 (maior atividade).

A LDA foi realizada com o método *stepwise* para se obter um conjunto mínimo de índices eletrônicos capaz de classificar de maneira satisfatória as moléculas dentro destes dois grupos. Entre os descritores selecionados, os que proporcionaram a melhor separação entre as médias dos grupos foram: E_{ELET} , $PEP_L(9)$.

A Tabela 3.5 mostra os coeficientes de cada variável da função discriminante (Z) o teste lambda de Wilk que serve como uma medida da igualdade entre as médias dos grupos nas referidas variáveis, o valor do Teste-F e o valor da significância p para o teste lambda de Wilk.

O lambda de Wilk é um valor situado entre 0 e 1. Quanto menor o seu valor, menor será a similaridade entre as médias dos dois grupos na referida variável. Tal fato possibilita inferir que o índice discrimina eficientemente moléculas que pertencem a um ou outro grupo.

O Teste-F possuirá um valor F grande para valores de lambda de Wilk pequenos. Quanto mais próximo de 1 for o valor de F , menor será o valor da significância p e quanto menor for o valor de p , melhor será o resultado.

Tabela 3.5: Teste de igualdade das médias entre os grupos 1 e 2.

Variáveis	Coeficientes	Lambda de Wilk	$F_{1, 14}$	p
E_{ELET}	$3,07 \cdot 10^{-4}$	0,659	7,254	0,017
$PEP_L(9)$	143,814	0,743	4,838	0,045
constante	-11,984	-	-	-

Função discriminante obtida:

$$Z = 3,07 \cdot 10^{-4}(E_{ELET}) + 143,814(PEP_L(9)) - 11,984 \quad (3.12),$$

Observa-se que através dos valores de Lambda de Wilk, $F_{1,14}$ e p que o descritor E_{ELET} possui um papel fundamental na separação dos grupos de maior e menor atividade.

A Tabela 3.6 apresenta os centróides de cada grupo considerado, bem como o parâmetro de corte (m) obtido pela média das posições do centróide. Com o referido parâmetro de corte, pode-se classificar as amostras como pertencentes a um ou a outro grupo.

Tabela 3.6: Valor da função discriminante no centróide dos grupos e parâmetro de corte (m).

Grupo	Centróide (\bar{Z})	m
1	-0,729	0,243
2	1,215	

A seguir, elabora-se a seguinte regra de classificação na análise dos *scores* da função discriminante:

- Se $Z \geq m$, então a molécula é classificada como pertencente ao grupo de maior atividade;
- Se $Z < m$, então a molécula é classificada como pertencente ao grupo de menor atividade.

Para avaliar criticamente o modelo obtido, empregou-se o método de validação cruzada das amostras (*leave-one-out classification*), onde cada caso (molécula) é classificado através de uma função discriminante construída a partir de todos os outros casos, exceto o que está em teste. Tal procedimento visa verificar a “robustez” do modelo.

A Tabela 3.7 mostra os resultados da classificação obtida com as duas variáveis independentes (E_{ELET} , PEP_L (9)), considerando todas as moléculas e através da validação cruzada.

Tabela 3.7: Resultados da classificação para as 2,4-diamino-6-quinazolina-sulfonamida, utilizando LDA com o modelo original e de validação cruzada.

		Grupo original	Grupo predito		Total	Porcentagem total de acerto
			1	2		
Modelo Original	Casos	1	9	1	10	93,8 %
		2	0	6	6	
	Porcentagem	1	90,0	10,0	100,0	
		2	0,0	100,0	100,0	
Modelo com Validação Cruzada	Casos	1	9	1	10	93,8 %
		2	0	6	6	
	Porcentagem	1	90,0	10,0	100,0	
		2	0,0	100,0	100,0	

Como se pode observar, a porcentagem de acerto na validação cruzada é bastante grande, o que permite considerar o modelo realmente significativo. Os dois modelos indicaram somente uma única molécula classificada de maneira errônea, a molécula 5.

A Figura 3.3 mostra o gráfico dos *scores* Z do grupo dos inibidores de maior atividade, do grupo de menor atividade e finalmente todos no mesmo gráfico.

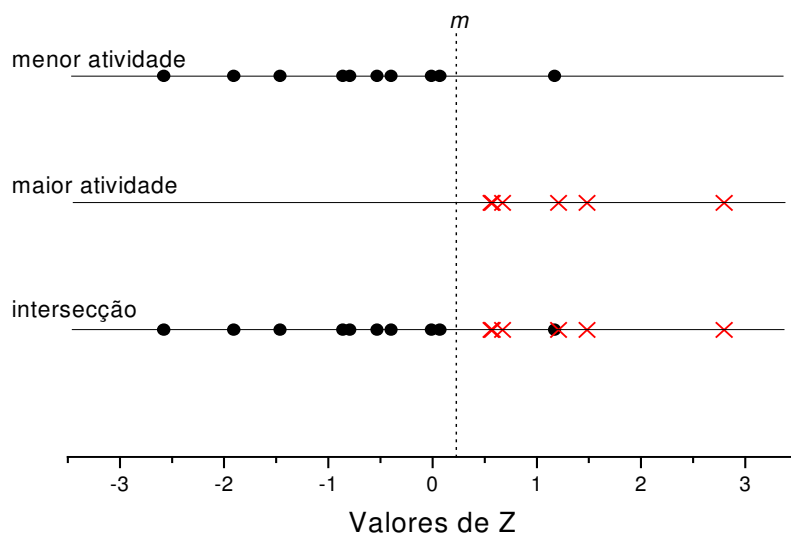


Figura 3.3: Gráfico dos *scores* Z e o parâmetro de corte m .

Em todos os modelos propostos na pesquisa, é possível notar a grande significância do descritor E_{ELET} . Devido a esse fato, foi feito um estudo de LDA somente com essa variável independente, servindo-se dos mesmos grupos definidos anteriormente e os mesmos procedimentos.

A Tabela 3.8 mostra os resultados da classificação obtida com o descritor E_{ELET} , considerando todas as moléculas e através da validação cruzada.

Tabela 3.8: Resultados da classificação, utilizando LDA com o modelo original e de validação cruzada, somente com o descritor E_{ELET} .

		Grupo original	Grupo predito (número de acertos)		Total	Porcentagem total de acerto
			1	2		
Modelo Original	Casos	1	8	2	10	87,5 %
		2	0	6	6	
	Porcentagem	1	80,0	20,0	100,0	
		2	0,0	100,0	100,0	
Modelo com Validação Cruzada	Casos	1	8	2	10	87,5 %
		2	0	6	6	
	Porcentagem	1	80,0	20,0	100,0	
		m	2	0,0	100,0	

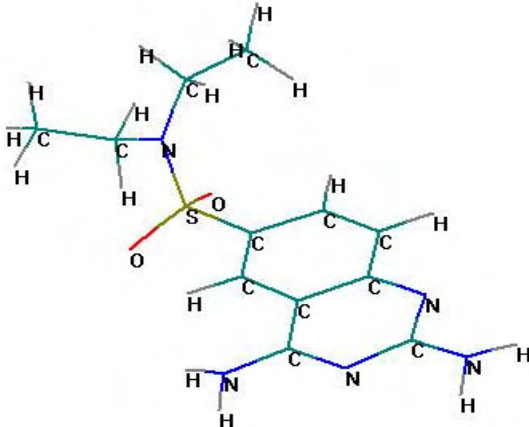
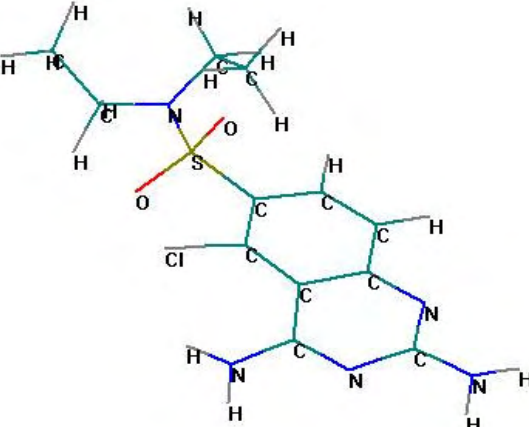
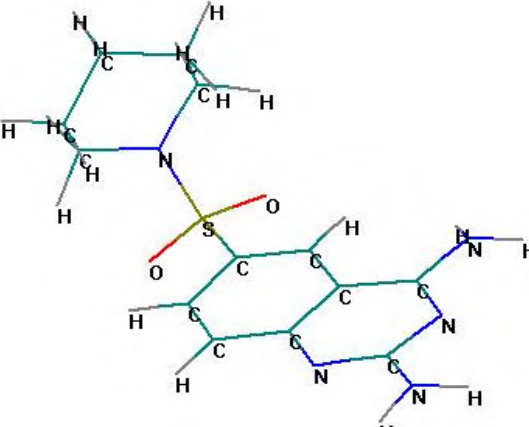
O alto valor de acerto na validação cruzada demonstra a significância do modelo indicando a importância desse descritor na separação dos conjuntos.

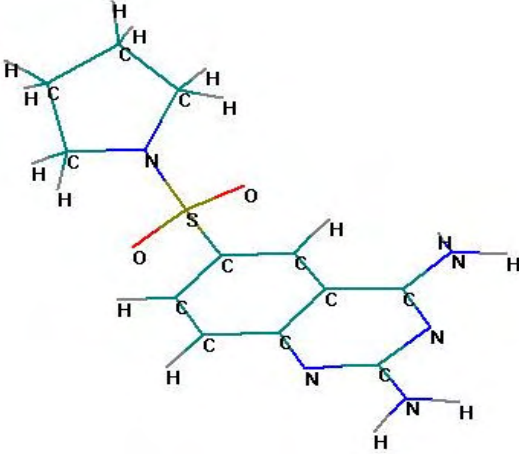
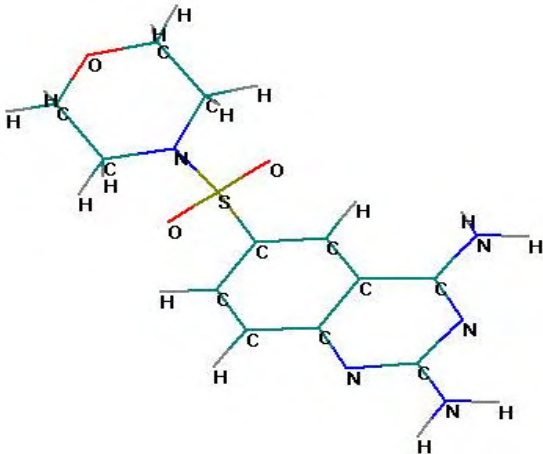
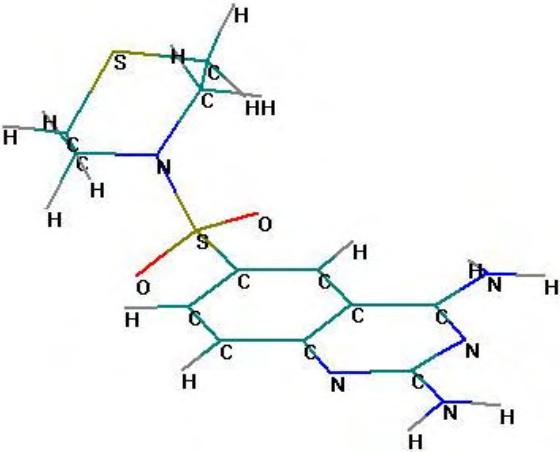
Notou-se que tanto a LDA como a PCA, verificaram o padrão estabelecido pelos descritores eletrônicos em relação à atividade biológica de interesse (atividade antimalarial). Sendo apontado por ambas a importância do descritor E_{ELET} na separação entre moléculas de maior e menor atividade antimalarial.

Na Tabela 3.8 verifica-se que somente este descritor é suficiente para uma boa predição da atividade biológica, acertando 87,5 % da classificação, em um modelo com validação cruzada. A aplicação desse descritor apresentou apenas duas moléculas que estão em desacordo com esta metodologia, as moléculas 5 e 14.

A seguir serão apresentadas as geometrias de equilíbrio para as sulfonamidas de maior atividade antimalarial, otimizadas pelo método semi-empírico AM1.

3.1.2 Geometria de equilíbrio das sulfonamidas de maior atividade antimalarial

Moléculas	Representação em 3D da estrutura em equilíbrio	Nome IUPAC
1		<i>((2,4-diaminoquinazolin-6-yl) sulfonyl) diethylamine</i>
2		<i>((2,4-diaminoquinazolin-5-chloroquinazolin-6-yl) sulfonyl) diethylamine</i>
7		6-(piperidysulfonyl) quinazoline-2,4-diamine

Moléculas	Representação em 3D da estrutura em equilíbrio	Nome IUPAC
8		6-(pyrrolidinylsulfonyl)quinazoline-2,4-diamine
9		6-(morpholin-4-ylsulfonyl)quinazoline-2,4-diamine
10		6-(4,1-thiazaperhydroinylsulfonyl)quinazoline-2,4-diamine

3.2 Anilinoquinolinas

Como é possível notar pelas discussões apresentadas neste trabalho, inúmeras são as drogas aplicadas na tentativa de eliminar ou reduzir a proliferação do *P. falciparum*. Mas por muito tempo o único medicamento utilizado como antimalarial foi o quinino, uma substância natural proveniente de uma planta conhecida popularmente como quina. Infelizmente o longo uso do quinino levou à seleção de linhagens de *Plasmodium* resistentes a essa droga. A preocupação causada pelo avanço dessas linhagens incentivou a procura de substâncias semelhantes ao quinino que também possuísem efeito antimalarial. Surge então a cloroquina, uma substância sintética, utilizada principalmente na década de 80 do século passado.

A falta de investimentos públicos e privados em pesquisas de novos fármacos contra a malária levou ao uso prolongado da cloroquina, o que novamente resultou na seleção de linhagens de *Plasmodium* mais resistentes, sendo uma das razões para o “ressurgimento” da doença. Atualmente, as pesquisas são realizadas buscando substâncias que apresentem propriedades antimalariais, oferecendo baixa toxicidade e menor resistência do *Plasmodium*. Também são levados em consideração os custos desses novos medicamentos, pois como citado anteriormente, a maioria da população afetada por esta moléstia não possui recursos financeiros para o tratamento.

Uma das esperanças no combate à malária é o uso das anilinoquinolinas, uma família de moléculas similar à cloroquina, de fácil síntese e baixo custo, ideal para países com populações pobres [3]. Esta droga apresenta efetiva atividade antimalarial contra diversas espécies de *Plasmodium* resistentes à cloroquina. Entretanto seu uso é muito restrito devido à agranulocitose (diminuição no número

de granulócitos, antes do citoplasma) e a sua hepatotoxicidade (lesões hepáticas que prejudicam o funcionamento do fígado).

A amodiaquina, uma substância derivada das anilinoquinolinas, teve sua síntese efetuada de modo a remover a estrutura causadora desta toxicidade, mantendo a atividade antimalarial e principalmente a fácil síntese, o que promete torná-la, juntamente com outros derivados dessa família, uma nova geração de antimalarials muito mais efetivos e sobretudo mais acessíveis à população. São estes novos análogos que serão objetos de estudo nessa parte do trabalho, novamente buscando uma correlação entre a estrutura eletrônica e a atividade antimalarial.

Nesta parte do trabalho, se fará a análise de um conjunto de 66 moléculas derivadas de uma estrutura básica (Figura 3.4), estas separadas em 3 grupos (Tabela 3.9) de acordo com o critério adotado no trabalho experimental [3], que considerou os substituintes ligados à estrutura básica para elaborar tal divisão. O grande interesse nessas moléculas é novamente o potencial de inibição da multiplicação do agente causador da malária.

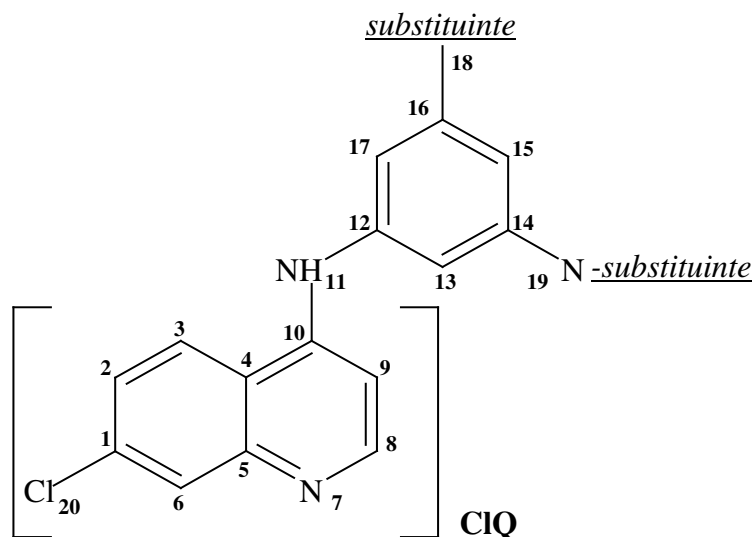
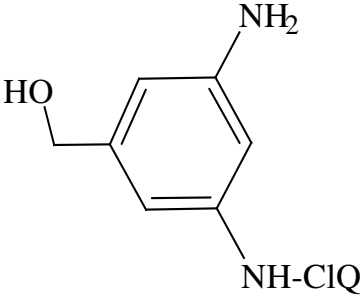
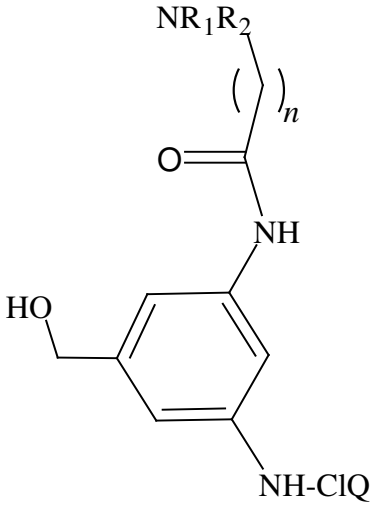


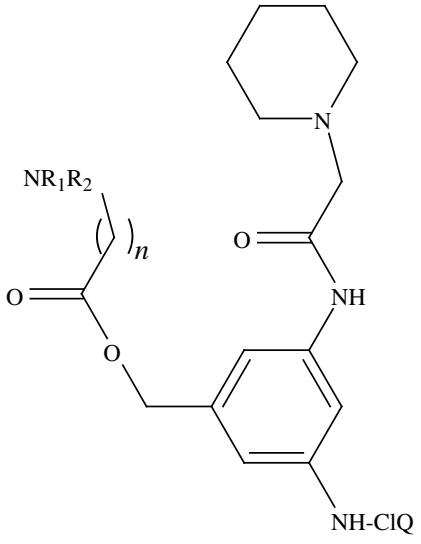
Figura 3.4: Representação esquemática da estrutura básica comum aos 3 grupos das anilinoquinolinas. A parte entre colchetes representa a molécula de cloroquinolina e será representada pela sigla ClQ.

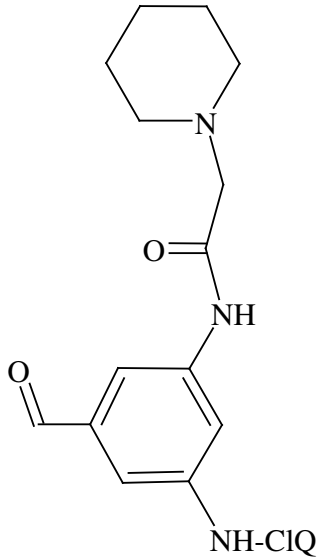
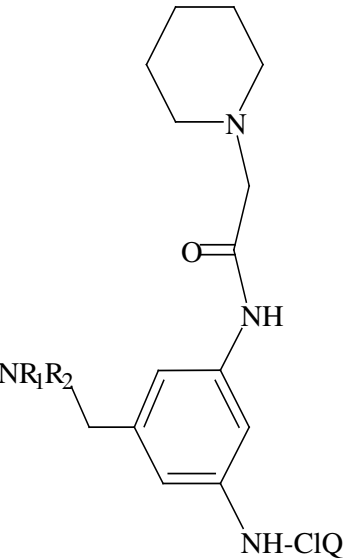
A propriedade de interesse que se tenta correlacionar com os descritores eletrônicos é definida como IC_{50} [12]. Este representa a concentração molar do fármaco capaz de inibir 50% da replicação viral *in vitro*, considerando a relação entre a quantidade administrada do fármaco por quilograma de massa corpórea da cobaia teste [31]. Quanto menor o valor de IC_{50} , maior o potencial da droga.

A seguir a Tabela 3.9 traz a representação esquemática dos 3 grupos, o número da molécula, os substituintes e o valor da propriedade antimalarial (índice antimalarial). Adotou-se a mesma nomenclatura empregada no trabalho experimental [3], visando facilitar a comparação dos resultados.

Tabela 3.9: Representação esquemática dos 3 grupos derivados das anilinoquinolinas, número de carbono (n), substituintes (NR_1R_2) e índice antimalarial (IC_{50}).

Estruturas – grupo 1	Moléculas	n	NR_1R_2	IC_{50} (nM)
<p>Estrutura 1</p> 	1	-	-	278,9
	2	1	4 methylpiperidine	77,7
	3	2	4 methylpiperidine	135,5
	4	4	4 methylpiperidine	465,0
	5	5	4 methylpiperidine	304,5
	6	7	4 methylpiperidine	308,5
	7	11	4 methylpiperidine	86,0
	8	1	piperidine	44,5
	9	1	N-methylpiperazine	344,0
	10	1	morpholine	1000
	11	1	4 hydroxypiperidine	1050,0
	12	1	pyrrolidine	44,0
	<p>Estrutura 2 – 24</p> 	13	1	thiazolidine
14		1	NHC(CH ₃) ₃	50,9
15		1	NEt ₂	41,6
16		1	NHCH ₂ C ₆ H ₅	192,7
17		1	NHCH ₂ C ₆ H ₄ Cl (para)	855,0
18		1	tetrahydroisoquinoline	1450,0
19		1	3-aminopyrazole	806,5
20		1	Cl	1000
21		1	NC ₅ H ₅ ⁺ , Cl ⁻	927,0
22		1	OH	1000
23		0	Phenyl	1000
24		0	H	359,8

Estruturas – grupo 2	Moléculas	<i>n</i>	NR ₁ R ₂	IC ₅₀ (nM)
	25	1	<i>piperidine</i>	54,0
	26	2	<i>piperidine</i>	78,3
	27	4	<i>piperidine</i>	15,5
	28	5	<i>piperidine</i>	27,4
	29	7	<i>piperidine</i>	165,0
	30	4	<i>pyrrolidine</i>	23,0
	31	4	<i>morpholine</i>	14,1
	32	4	<i>N-methylpiperazine</i>	65,1
	33	4	NEt ₂	141,1
	34	4	Br	151,1
	35	0	<i>Phenyl</i>	75,5
	36	0	<i>Quinol-4-yl</i>	25,6

Estructuras – grupo 3	Moléculas	NR ₁ R ₂	IC ₅₀ (nM)	
<p>Estructura 37</p> 	37	-	281,7	
	38	NHCH ₂ CH ₂ NMe ₂	15,5	
	39	NHCH ₂ CH ₂ CH ₂ NMe ₂	134,0	
	40	NMeCH ₂ CH ₂ NMe ₂	5,0	
	41	NHCH ₂ CH ₂ -pyrrolidine	10,2	
	42	NHCH ₂ CH ₂ -pyrrolidine	6,6	
	43	NHCH ₂ CH ₂ -pyrrolidine	7,9	
	44	NEt ₂	8,2	
	45	NH <i>t</i> Bu	3,7	
	46	<i>piperidine</i>	10,2	
	47	<i>pyrrolidine</i>	6,9	
	48	<i>N-methylpiperazine</i>	7,5	
	49	<i>morpholine</i>	9,5	
	50	<i>4 hydroxypiperidine</i>	6,5	
	51	<i>N-(2 hydroxyethyl)piperazine</i>	11,5	
	<p>Estructura 38 – 66</p> 	52	<i>N-phenylpiperazine</i>	5,5
		53	<i>N-benzylpiperazine</i>	13,0
		54	<i>N-(diphenylmethyl)piperazine</i>	12,5
		55	<i>N-(4 chlorobenzyl)piperazine</i>	11,5
		56	<i>N-(4 methoxybenzyl)piperazine</i>	11,0
57		<i>N-(4 nitrobenzyl)piperazine</i>	13,2	
58		<i>N(4diethylaminobenzyl)piperazine</i>	11,9	
59		<i>N-(4 cyanobenzyl)piperazine</i>	13,3	
60		<i>N-piperonylpiperazine</i>	6,9	
61		NHC ₆ H ₅	21,2	
62		NHCH ₂ C ₆ H ₅	8,5	
63		NHCH(C ₆ H ₅) ₂	9,1	
64		NHCH ₂ C ₆ H ₄ Cl (para)	4,6	
65		NHCH ₂ C ₆ H ₄ OMe (para)	4,4	
66		NHCH ₂ C ₆ H ₄ CF ₃ (para)	5,5	

3.2.1 Resultados

As análises dos dados obtidos para estas moléculas seguem o mesmo procedimento feito na seção 3.1.1.

3.2.1.1 Regressões lineares simples e múltiplas

Programas desenvolvidos pelo grupo de pesquisa deste trabalho (*Biomat_01*) foram utilizados para obter um conjunto de 129 descritores eletrônicos para cada uma das 66 moléculas estudadas. Com estes programas, efetuaram-se regressões lineares simples e múltiplas entre os valores de IC_{50} e os descritores eletrônicos.

A seguir os parâmetros de correlação linear r mais expressivos, obtidos das regressões lineares simples e múltiplas, sobre todas as combinações possíveis de duas, três, quatro e cinco variáveis independentes e os descritores envolvidos.

Tabela 3.10: Resultado das regressões lineares simples e múltiplas considerando todas as moléculas.

Regressão Linear:	r	Descritores Eletrônicos				
		Vi 1	Vi 2	Vi 3	Vi 4	Vi 5
Simple	0,55	E_{ELET}	-	-	-	-
Múltipla com 2 Vi	0,58	E_{ELET}	EM	-	-	-
Múltipla com 3 Vi	0,61	E_{ELET}	EM	$PEP_L(16)$	-	-
Múltipla com 4 Vi	0,69	E_{ELET}	EM	$PEP_L(16)$	$CAR(19)$	-
Múltipla com 5 Vi	0,72	E_{ELET}	EM	$PEP_L(16)$	$CAR(19)$	$PEP_H(19)$

Novamente, os parâmetros r possuem valores relativamente baixos, não sendo interessante representá-los graficamente. Como discutido anteriormente, tal fato retrata a ausência de uma correlação linear, de modo que a proposição de uma regra de classificação para todas as moléculas não pôde ser obtida.

Tabela 3.11: Resultados referentes aos grupos dos derivados de anilinoquinolinas.

GRUPO 1				
Regressão Linear:	r	Descritores Eletrônicos		
		Vi 1	Vi 2	Vi 3
Simple	0,52	$PEP_L(4)$	-	-
Múltipla com 2 Vi	0,74	$PET(19)$	$PEP_L(3)$	-
Múltipla com 3 Vi	0,81	$PET(19)$	$PEP_L(2)$	$PEP_L(17)$

GRUPO 2				
Regressão Linear:	r	Descritores Eletrônicos		
		Vi 1	Vi 2	Vi 3
Simple	0,52	$OL(2)$	-	-
Múltipla com 2 Vi	0,65	$OL(3)$	$PEP_H(4)$	-
Múltipla com 3 Vi	0,87	$PET(14)$	$PEP_H(16)$	$PEP_H(18)$

GRUPO 3				
Regressão Linear:	r	Descritores Eletrônicos		
		Vi 1	Vi 2	Vi 3
Simple	0,43	ΔH_F	-	-
Múltipla com 2 Vi	0,52	ΔH_F	$LDOS(2)$	-
Múltipla com 3 Vi	0,60	ΔH_F	$LDOS(2)$	E_{NUCL}

Os resultados para o parâmetro r , considerando as moléculas divididas em 3 grupos apresentaram valores relativamente baixos não sendo novamente interessante representá-los graficamente, contudo mostram em média uma melhora quando comparado com a Tabela 3.10, sendo o grupo 3 o único que apresentou valores menores para r . Assim optou-se por manter a separação para as próximas análises.

3.2.1.2 Análise dos Componentes Principais

Com a PCA tentou-se descobrir um possível agrupamento das moléculas de maior atividade, analisando todas as combinações possíveis de 3 variáveis independentes (mais de 33000 combinações) nos 3 grupos de derivados de anilinoquinolinas, sendo que:

O grupo 1 é representado por uma matriz de dados formada por 24 linhas (moléculas) e 129 colunas (variáveis independentes), sendo classificado segundo seu índice antimalarial como:

- Maior Atividade Antimalarial (A): apresentam um valor de IC_{50} inferior ou igual a 86,0. Assim as moléculas mais ativas são: 2, 7, 8, 12, 14 e 15.
- Menor Atividade Antimalarial (B): apresentam um valor de IC_{50} superior a 86,0. Assim as moléculas menos ativas são: 1, 3, 4, 5, 6, 9, 10, 11, 13, 16, 17, 18, 19, 20, 22, 23 e 24.

O grupo 2 é representado por uma matriz de dados formada por 12 linhas (moléculas) e 129 colunas (variáveis independentes), sendo classificado segundo seu índice antimalarial como:

- Maior Atividade Antimalarial (A): apresentam um valor de IC_{50} inferior ou igual a 26,0. Assim as moléculas mais ativas são: 27, 30, 31 e 36.
- Menor Atividade Antimalarial (B): apresentam um valor de IC_{50} superior a 26,0. Assim as moléculas menos ativas são: 25, 26, 28, 29, 32, 33, 34 e 35.

O grupo 3 representado por uma matriz de dados formada por 30 linhas (moléculas) e 129 colunas (variáveis independentes), sendo classificado segundo seu índice antimalarial como:

- Maior Atividade Antimalarial (A): apresentam um valor de IC_{50} inferior ou igual a 6,5. Assim as moléculas mais ativas são: 40, 45, 50, 52, 64, 65 e 66.
- Menor Atividade Antimalarial (B): apresentam um valor de IC_{50} superior a 6,5. Assim as moléculas menos ativas são: 37, 38, 39, 41, 42, 43, 44, 46, 47, 48, 49, 51, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62 e 63.

Para os grupos 1 e 2, foi criada uma regra para separação das moléculas com maior e menor atividade antimalarial. O grupo 3, a exemplo das regressões lineares simples e múltiplas, não apresentou uma separação significativa entre as moléculas de maior e menor atividade. Os resultados para o grupo 1 e 2 são apresentados a seguir.

Grupo 1

Tabela 3.12: Dados da variância dos componentes principais PC1 e PC2.

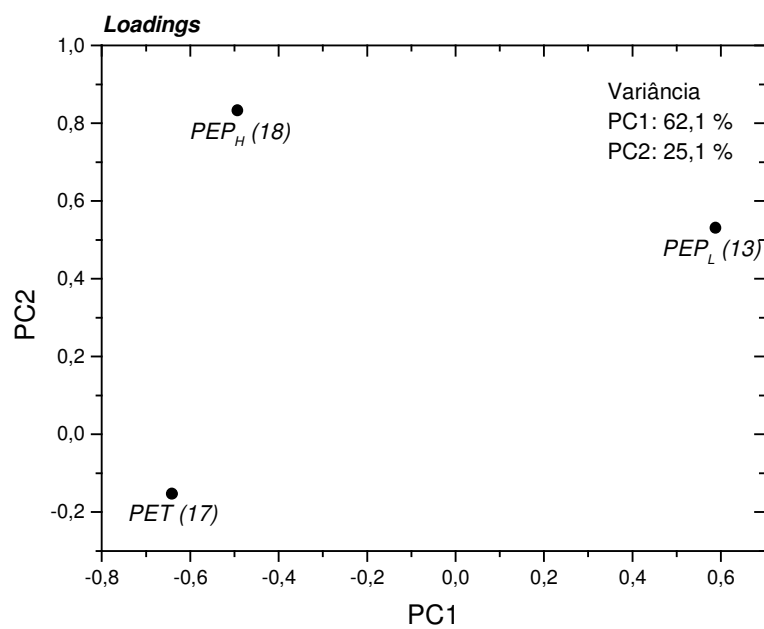
	Variância	Porcentagem	Acumulativo
PC1	1,86	62,1 %	62,1 %
PC2	0,753	25,1 %	87,2 %

As equações para as variáveis (*loadings*) do PCA são dadas por:

$$PC1 = -0,641(PET(17)) - 0,493(PEP_H(18)) + 0,588(PEP_L(13)) \quad (3.13)$$

$$PC2 = -0,153(PET(17)) + 0,833(PEP_H(18)) + 0,5331(PEP_L(13)) \quad (3.14)$$

a)



b)

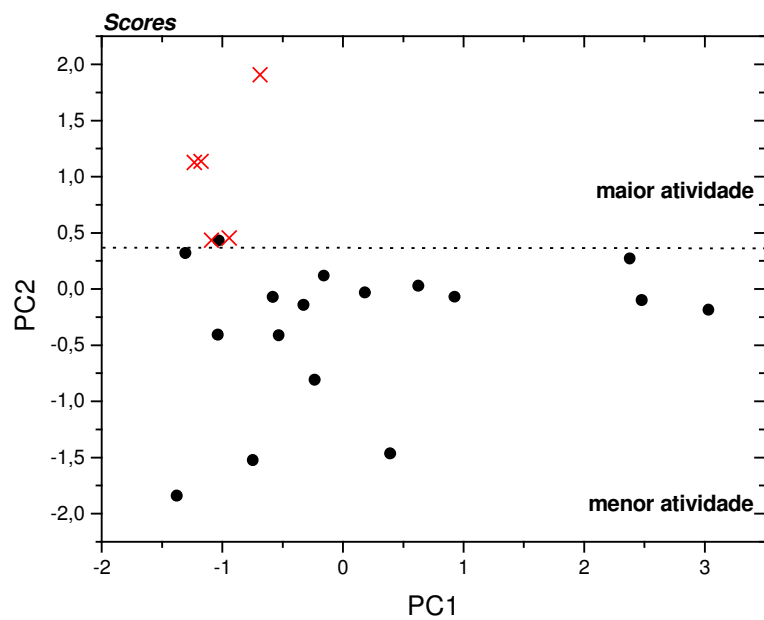


Figura 3.5: a) Contribuição das variáveis independente em PC1 e PC2.

b) Separação obtida para o grupo 1 das anilinoquinolinas.

Grupo 2

Tabela 3.13: Dados da variância dos componentes principais PC1 e PC2.

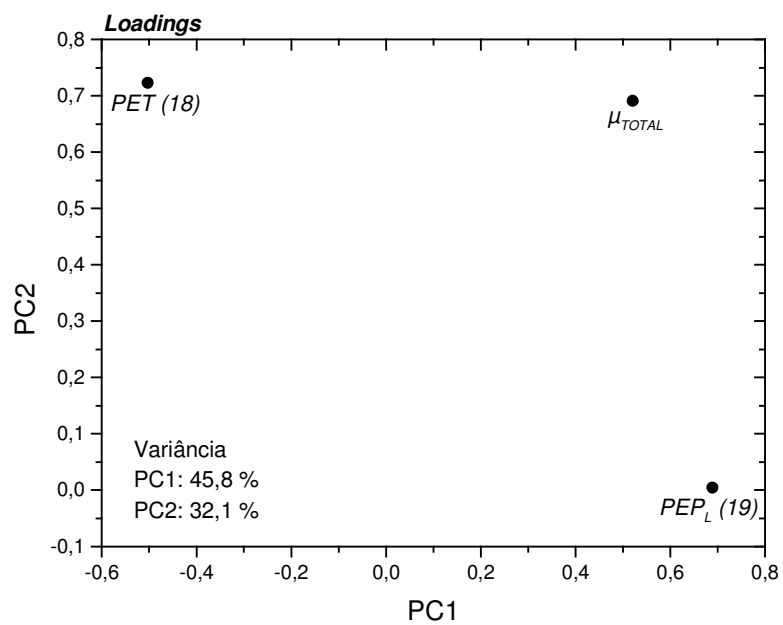
	Variância	Porcentagem	Acumulativo
PC1	1,37	45,8 %	45,8 %
PC2	0,964	32,1 %	78,0 %

As equações para as variáveis de (*loadings*) do PCA são dadas por:

$$PC1 = 0,521(\mu_{TOTAL}) - 0,503(PET (18)) + 0,689(PEP_L (19)) \quad (3.15)$$

$$PC2 = 0,691(\mu_{TOTAL}) + 0,723(PET (18)) + 0,0044(PEP_L (19)) \quad (3.16)$$

a)



b)

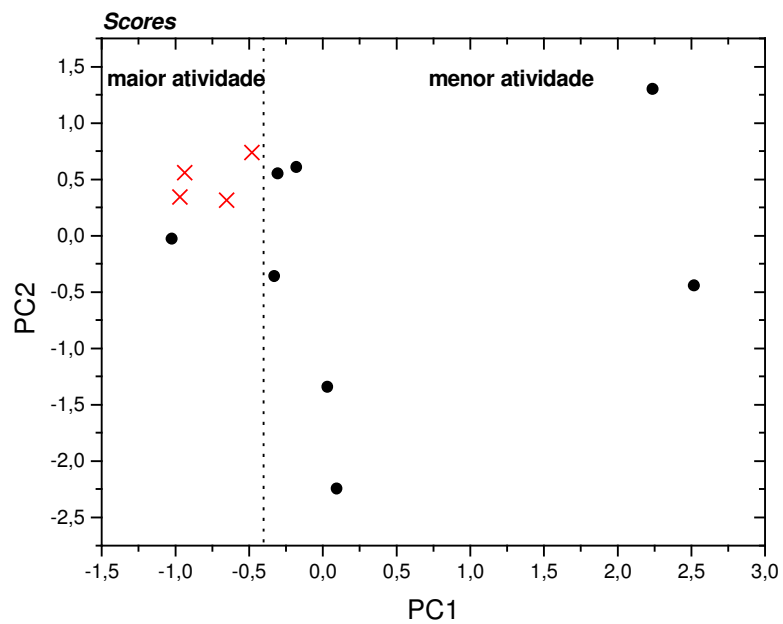


Figura 3.6: a) Contribuição das variáveis independente em PC1 e PC2.

b) Separação obtida para o grupo 2 das anilinoquinolinas.

Nos resultados para o grupo 1 (Figura 3.5), é possível notar pelos padrões fornecidos pela PCA, a separação de dois grupos: um contendo as moléculas de maior atividade e outro contendo as de menor atividade. As variáveis PET (17), PEP_H (18) e PEP_L (13), foram suficientes para a separação das classes de atividade, como pode ser visto pelas equações 3.13 e 3.14, que modelam este grupo. Também foram suficientes somente as componentes principais PC1 e PC2 para a separação das classes, prevendo com 95,7 % de acerto os polígonos das moléculas de maior e menor atividade.

Para os resultados do grupo 2 (Figura 3.6), também foi possível obter uma ótima separação, prevendo com 91,7 % de acerto os polígonos das moléculas de maior e menor atividade, sendo suficientes para conseguir essa separação os descritores μ_{TOTAL} , PET (18) e PEP_L (19). As equações 3.15 e 3.16 modelam o grupo citado.

3.2.1.3 Análise Discriminante Linear

Para cada grupo utilizou-se os mesmo métodos descritos na seção 3.1.1, adotando-se o seguinte critério de classificação: moléculas com maior atividade antimalarial e de menor atividade antimalarial, utilizando a mesma separação apresentada no estudo do PCA, agrupando pelo índice antimalarial (IC_{50}). A seguir, os resultados apresentados pelos 3 grupos utilizando a LDA:

GRUPO 1

Tabela 3.14: Teste de igualdade das médias entre os grupos.

Variáveis	Coefficientes	Lambda de Wilk	$F_{1, 21}$	p
$PET (2)$	489,166	0,560	16,530	0,001
$PET (17)$	-87,965	0,454	12,019	0,0001
$PEP_H (18)$	9861,907	0,346	11,951	0,0001
<i>constante</i>	-1655,206	-	-	-

Na Tabela acima os baixos valores do lambda denotam baixa similaridade entre as moléculas do grupo na referida variável, conforme visto em discussão anterior (seção 3.1.1.3 Análise Discriminante Linear). Esse fato permite concluir que essas variáveis discriminam com eficiência moléculas tidas com maior e menor atividades, para este grupo.

Função discriminante obtida:

$$Z = 489,166(PET (2)) - 87,965(PET (17)) + 9861,907(PEP_H (18)) - 1655,206 \quad (3.17)$$

Tabela 3.15: Valor da função discriminante no centróide dos agrupamentos das moléculas mais ativas (A), menos ativas (B) e parâmetro de corte (m).

Grupo	Centróide (\bar{Z})	m
A	2,209	0,7145
B	-0,780	

Elabora-se a seguinte regra de classificação na análise dos *scores* da função discriminante:

- Se $Z \geq m$, então a molécula é classificada como sendo de maior atividade;
- Se $Z < m$, então a molécula é classificada como sendo de menor atividade.

Tabela 3.16: Resultados da classificação, utilizando LDA com o modelo original e de validação cruzada.

		Grupo original	Grupo predito		Total	Porcentagem total de acerto
			A	B		
Modelo Original	Casos	A	6	0	6	95,7 %
		B	1	16	17	
	Porcentagem	A	100	0	100	
		B	5,9	94,1	100	
Modelo com Validação Cruzada	Casos	A	6	0	6	91,3 %
		B	2	15	17	
	Porcentagem	A	100	0	100	
		B	11,8	88,2	100	

Para este grupo a porcentagem de acerto na validação cruzada é bastante grande, assim é possível considerar o modelo significativo. A Figura 3.7 mostra o gráfico dos *scores* Z das moléculas de maior atividade, menor atividade e todas no mesmo gráfico.

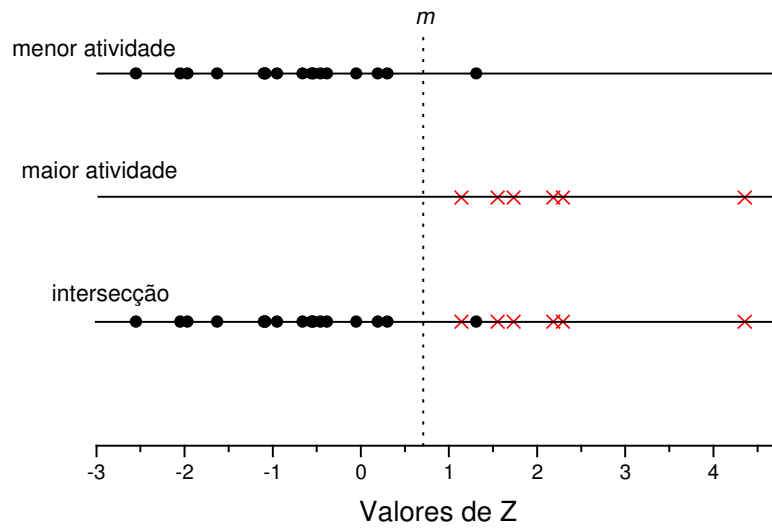


Figura 3.7: Gráfico dos *scores* de Z e o parâmetro de corte m , para o grupo 1.

GRUPO 2

Tabela 3.17: Teste de igualdade das médias entre os grupos.

Variáveis	Coefficientes	Lambda de Wilk	$F_{L,10}$	p
$PEP_L(19)$	16382,33	0,637	5,690	0,003
<i>constante</i>	-2,194	-	-	-

Na Tabela 3.17 o baixo valor de lambda indica que essa variável discrimina com eficiência as moléculas de maior e menor atividade. O baixo valor de p indica a significância estatística do lambda de Wilk.

Função discriminante obtida:

$$Z = 16382,33(PEP_L(19)) - 2,194 \quad (3.18)$$

Tabela 3.18: Valor da função discriminante no centróide dos agrupamentos das moléculas mais ativas (A), menos ativas (B) e parâmetro de corte (m).

Grupo	Centróide (\bar{Z})	m
A	-0,974	-0,244
B	0,487	

Cria-se a seguinte regra de classificação na análise dos *scores* da função discriminante:

- Se $Z \geq m$, então a molécula é classificada como sendo de menor atividade;
- Se $Z < m$, então a molécula é classificada como sendo de maior atividade.

Tabela 3.19: Resultados da classificação utilizando LDA com o modelo original e de validação cruzada.

		Grupo original	Grupo predito		Total	Porcentagem total de acerto
			A	B		
Modelo original	Casos	A	4	0	4	91,7 %
		B	1	7	8	
	Porcentagem	A	100	0	100	
		B	12,5	87,5	100	
Modelo com Validação Cruzada	Casos	A	3	1	4	75,0 %
		B	2	6	8	
	Porcentagem m	A	75,0	25,0	100	
		B	25,0	75,0	100	

No grupo 2, conseguiu-se um elevado percentual de acerto no modelo original. Para a validação cruzada o valor foi menor, mas ainda considerado muito relevante para o modelo. A Figura 3.8 mostra o gráfico dos *scores* Z do grupo 2 para as moléculas de maior atividade, menor atividade e todas no mesmo gráfico.

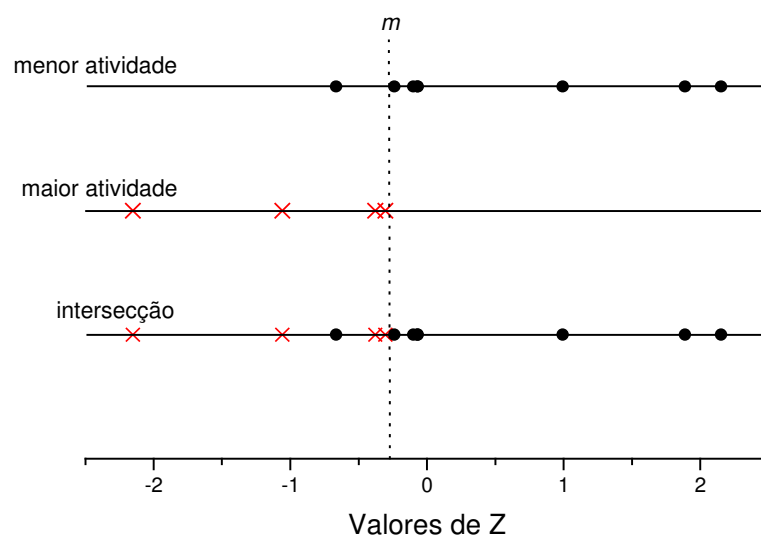


Figura 3.8: Gráfico dos *scores* de Z e o parâmetro de corte m , para o grupo 2.

GRUPO 3

Tabela 3.20: Teste de igualdade das médias entre os grupos.

Variáveis	Coefficientes	Lambda de Wilk	$F_{1, 28}$	p
ΔH_F	0,19	0,876	3,549	0,07
$PEP_H (18)$	3182,995	0,751	3,978	0,03
constante	-2,322	-	-	-

Função discriminante obtida:

$$Z = 0,19(\Delta H_F) + 3182,995(PEP_H (18)) - 2,322 \quad (3.19)$$

Tabela 3.21: Valor da função discriminante no centróide dos agrupamentos das moléculas mais ativas (A), menos ativas (B) e parâmetro de corte (m).

Grupo	Centróide (\bar{Z})	m
A	-0,937	-0,305
B	0,328	

Novamente elaborase uma regra de classificação pela análise dos *scores* da função discriminante:

- Se $Z \geq m$, então a molécula é classificada como sendo de menor atividade;
- Se $Z < m$, então a molécula é classificada como sendo de maior atividade.

Tabela 3.22: Resultados da classificação, utilizando LDA com o modelo original e de validação cruzada.

		Grupo original	Grupo predito		Total	Porcentagem total de acerto
			A	B		
Modelo original	Casos	A	6	1	7	74,1 %
		B	6	14	20	
	Porcentagem	A	85,7	14,3	100	
		B	30	70	100	
Modelo com Validação Cruzada	Casos	A	6	1	7	74,1 %
		B	6	14	20	
	Porcentagem	A	85,7	14,3	100	
		B	30	70	100	

Os resultados da Tabela 3.22 apesar de serem de percentual mais baixo em comparação aos obtidos pelos grupos 1 e 2, ainda são relevantes, podendo fornecer pistas para eventuais confôrmeros de maior atividade biológica. A Figura 3.9 mostra o gráfico dos *scores Z* para o grupo 3.

Na seqüência são apresentadas as geometrias de equilíbrio das anilinoquinolinas de maior atividade antimalarial utilizadas nesse trabalho. Essa geometria foi obtida após busca conformacional do mínimo de energia com o emprego do método semi-empírico AM1.

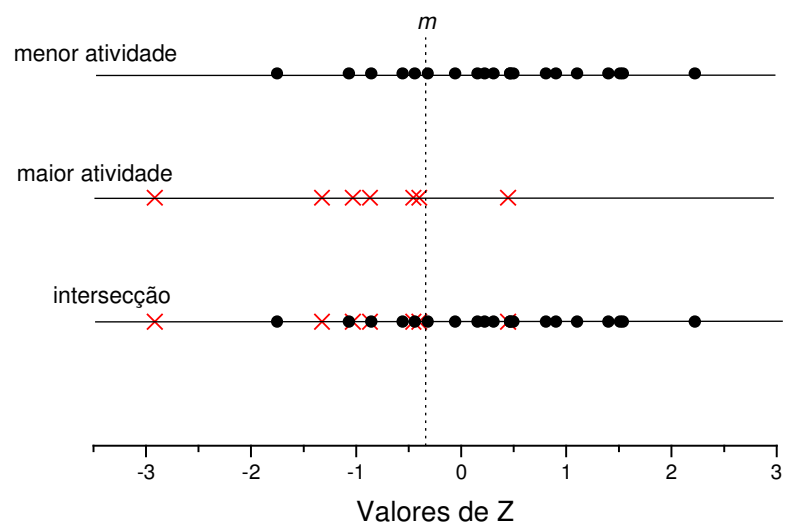


Figura 3.9: Gráfico dos *scores* de Z e o parâmetro de corte m , para o grupo 3.

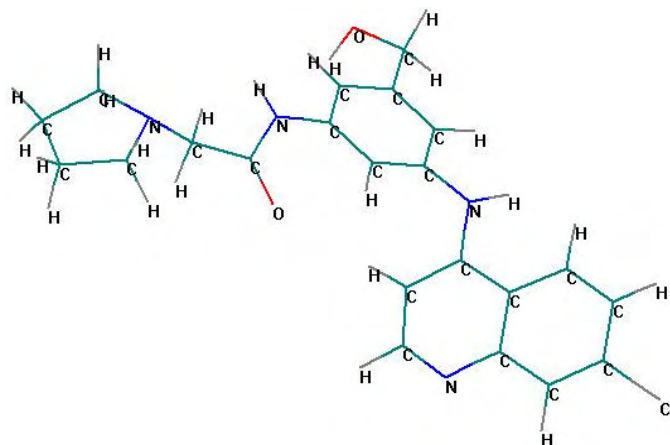
3.2.2 Geometria de equilíbrio das anilinoquinolinas de maior atividade

GRUPO 1

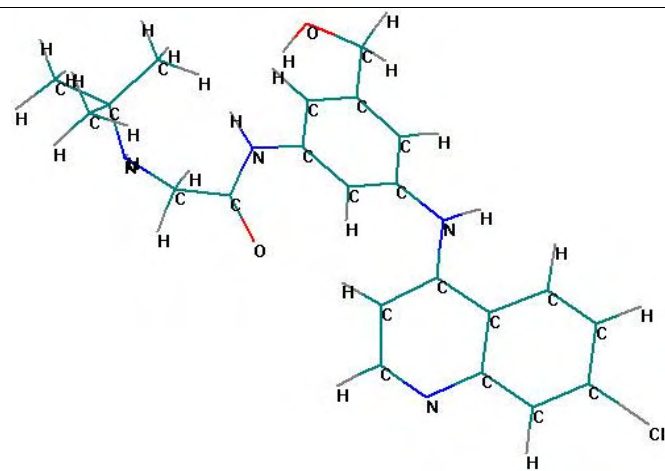
Moléculas	Representação em 3D da geometria de equilíbrio	Substituinte
2		<i>4 methylpiperidine</i>
7		<i>4 methylpiperidine</i>
8		<i>piperidine</i>

Moléculas	Representação em 3D da geometria de equilíbrio	Substituinte
-----------	--	--------------

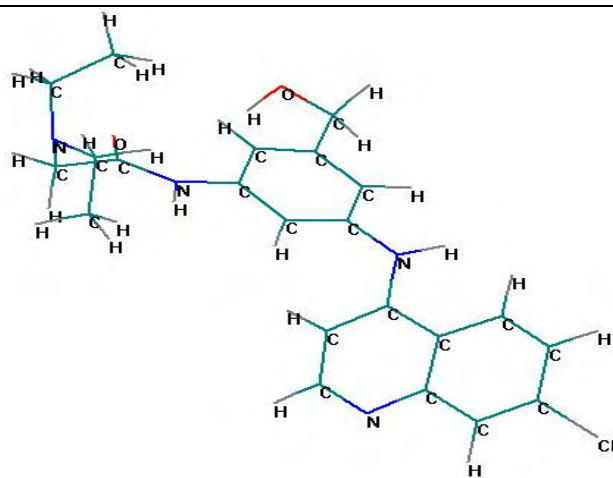
12

*pyrrolidine*

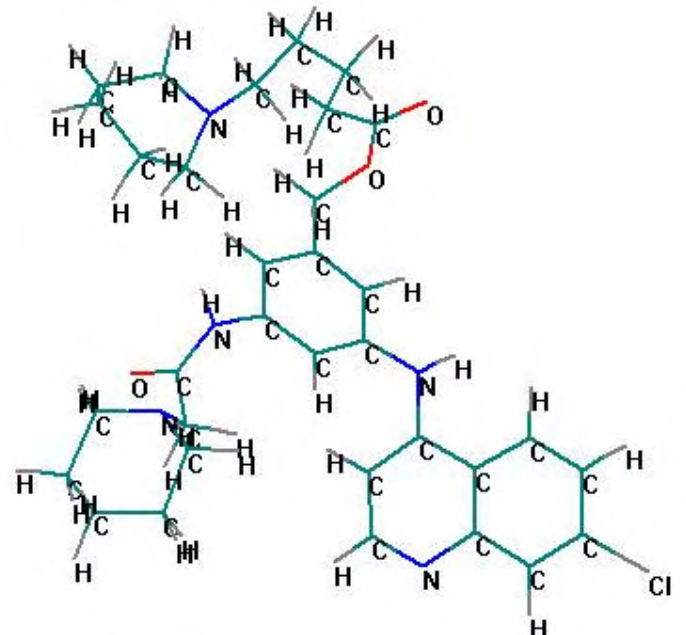
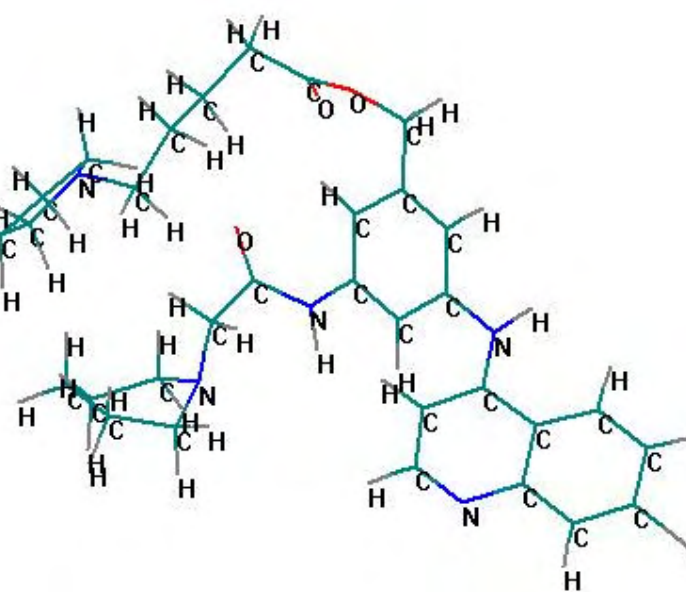
14

 $\text{NHC}(\text{CH}_3)_3$

15

 NEt_2

GRUPO 2

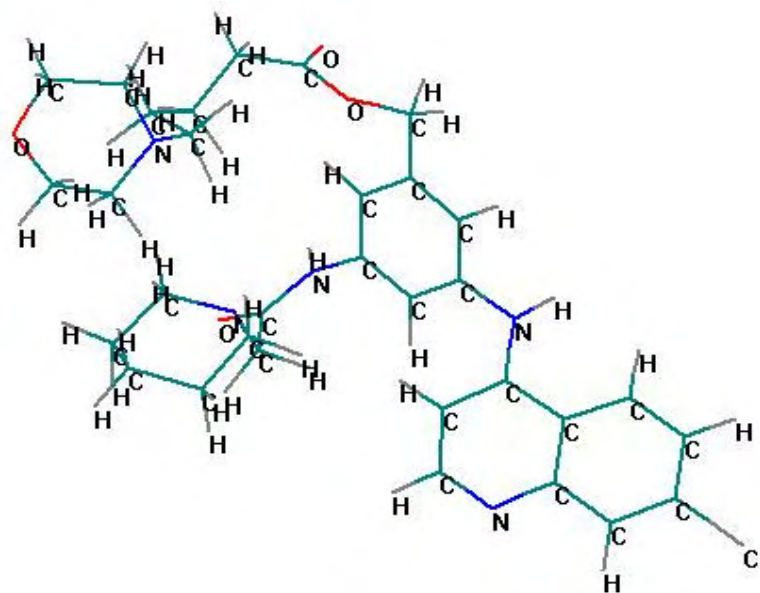
Moléculas	Representação em 3D da geometria de equilíbrio	Substituinte
27		<i>piperidine</i>
30		<i>pyrrolidin</i>

Moléculas

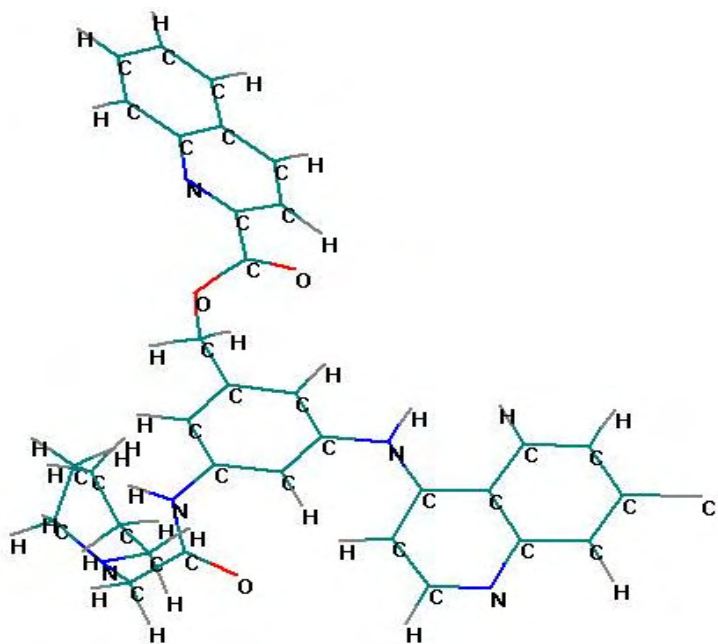
Representação em 3D da geometria de equilíbrio

Substituinte

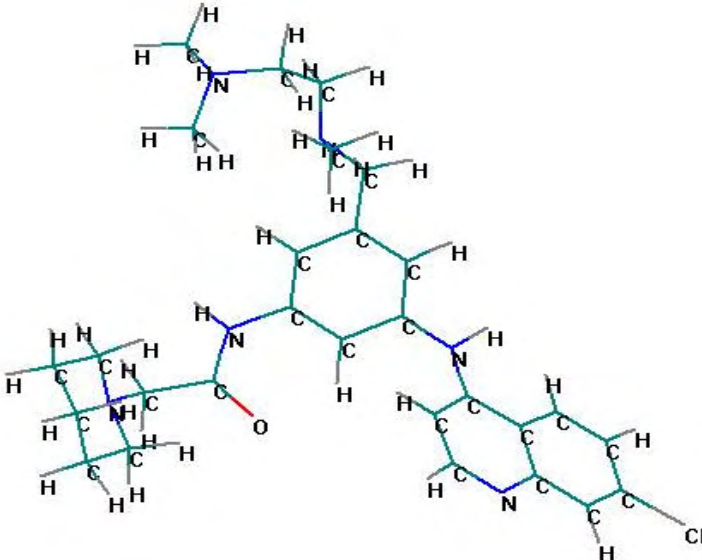
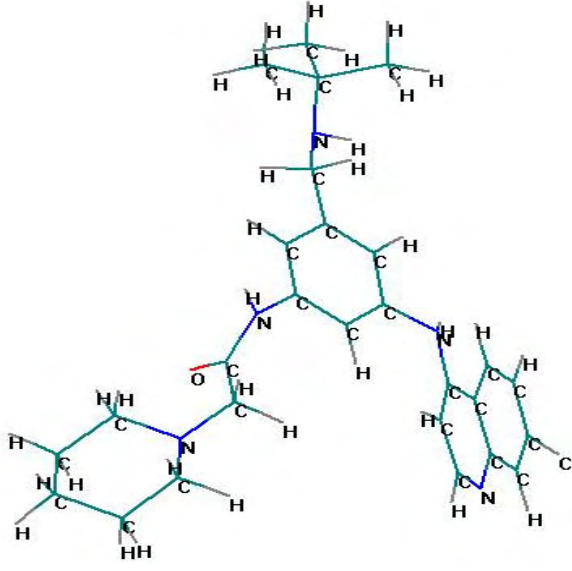
31

*morpholine*

36

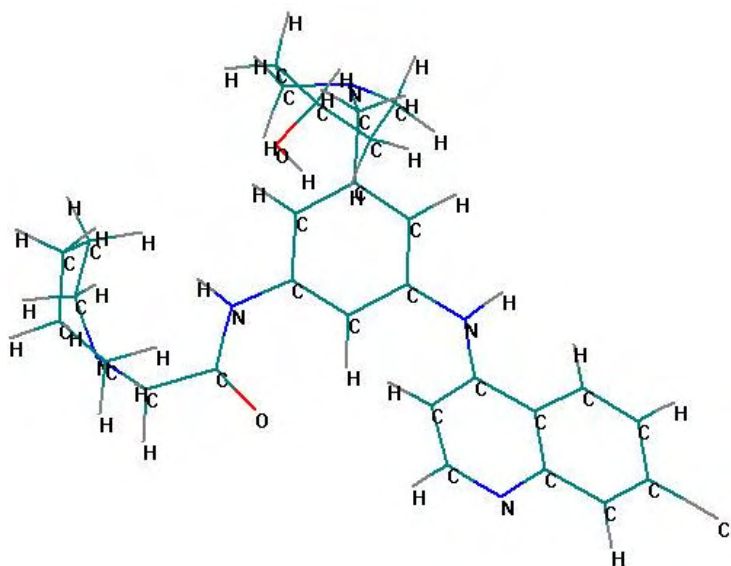
*Quinol-4-yl*

GRUPO 3

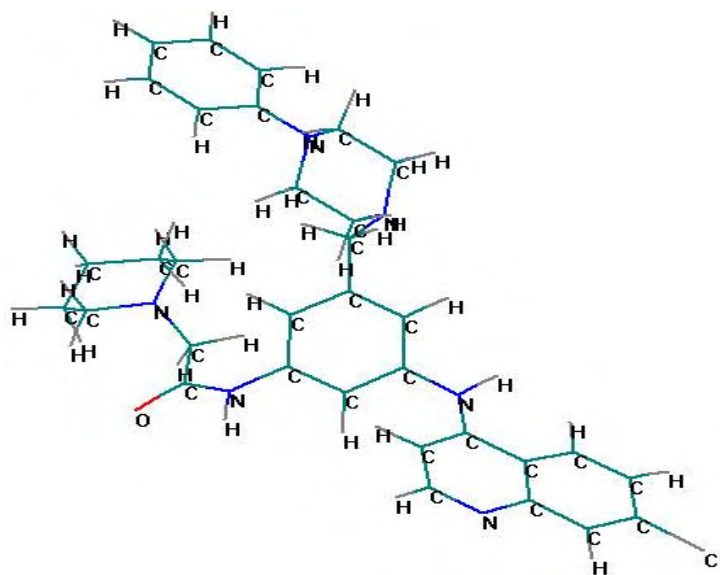
Moléculas	Representação em 3D da geometria de equilíbrio	Substituinte
40		$NMeCH_2CH_2NMe_2$
45		$NHtBu$

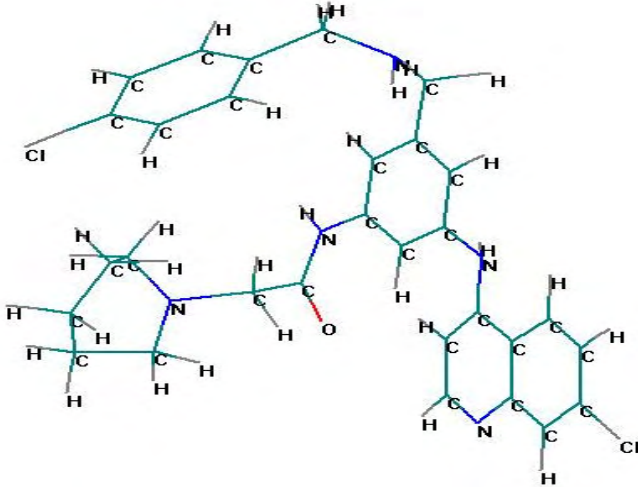
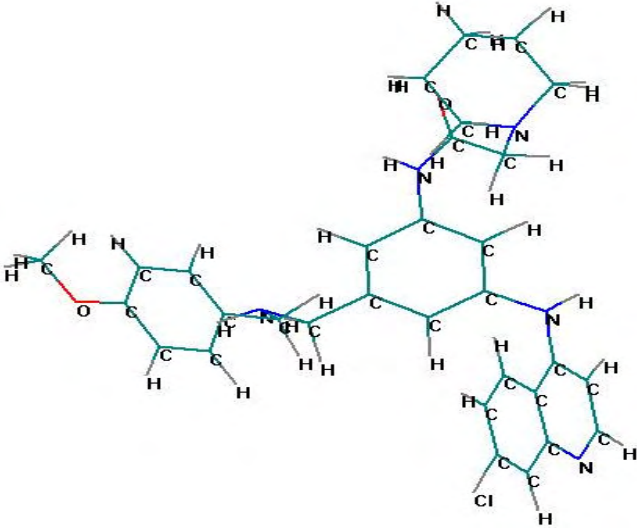
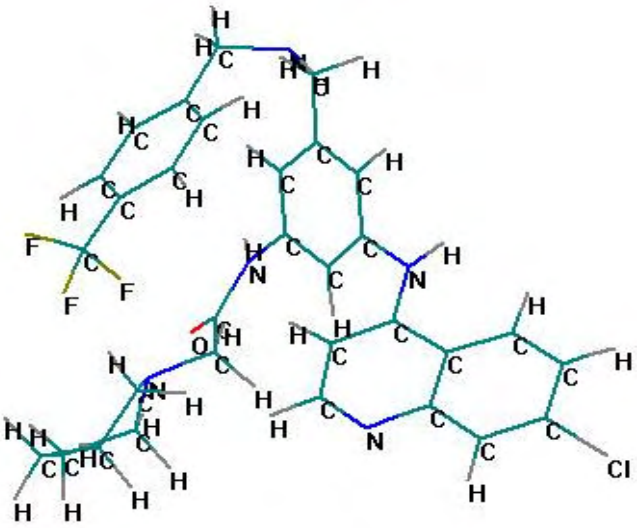
Moléculas**Representação em 3D da geometria de equilíbrio****Substituinte**

50

*4 hydroxypiperidine*

52

*N-phenylpiperazine*

Moléculas	Representação em 3D da geometria de equilíbrio	Substituinte
64		$\text{NHCH}_2\text{C}_6\text{H}_4\text{Cl}$
65		$\text{NHCH}_2\text{C}_6\text{H}_4\text{OMe}$
66		$\text{NHCH}_2\text{C}_6\text{H}_4\text{CF}_3$

CAPÍTULO 4

Discussão e conclusões

Através do estudo da estrutura eletrônica das moléculas apresentadas neste trabalho foi possível obter um grande número de dados, os quais após uma análise estatística resultaram em modelos matemáticos que visam a classificação das moléculas quanto à sua atividade antimalarial. Segue neste capítulo uma apreciação e as conclusões obtidas através dos modelos desenvolvidos, enfatizando os resultados e a aplicação destes na modelagem de novos candidatos com maior atividade antimalarial.

4.1 Sulfonamidas

Nas sulfonamidas, as regressões lineares simples e múltiplas, não resultaram em relações quantitativas com valores do parâmetro de correlação considerados significativos, fornecendo apenas indicações dos possíveis descritores eletrônicos que poderiam ser utilizados para classificar a molécula quanto à sua atividade antimalarial. Nas equações de maior relevância (obtidas na regressão linear), o descritor energia total referente aos elétrons da camada de valência (E_{ELET}) sempre

está presente, indicando sua importância. O PCA, método de análise classificatória exploratória, deixa evidente a relevância desse descritor para a classificação dos grupos de moléculas de maior e menor atividade. Tendência essa mantida na LDA, método de análise classificatório quantitativo, no qual a E_{ELET} desempenha papel importante na equação de discriminação (classificação), sendo que, a inclusão do descritor $PEP_L(9)$ eleva ainda mais o índice de previsões corretas quanto à atividade antimalarial.

A equação de discriminação (ED):

$$Z = 3,07 \cdot 10^{-4}(E_{ELET}) + 143,814(PEP_L(9)) - 11,984 \quad (3.12),$$

apresenta um índice de classificação correta do grau de atividade de 93,8 %, com significância estatística maior que 99,98 %. A ED 3.12 estabelece uma regra matemática com alta previsibilidade para a separação das moléculas de maior e menor atividade: se para uma dada molécula candidata, Z for maior ou igual a 0,243 então a molécula apresenta probabilidade de ser ativa.

No entanto, optou-se por trabalhar com a ED que apresentou somente o descritor E_{ELET} no modelo, como uma alternativa para a seleção do conjunto de moléculas. Outro motivo é devido à simplicidade da equação gerada, o que permite elaborar regras mais fáceis de serem aplicadas, ou seja, modelos mais simples de serem seguidos na procura de novos compostos ativos. A ED em questão:

$$Z = 3,747 \cdot 10^{-4}(E_{ELET}) + 10,4701 \quad (4.1),$$

apresenta um alto índice de acerto na classificação do grau de atividade, com valor de 87,5 %, significância estatística maior que 99,98 % (Lambda de Wilk = 0,659, $F_{1,14} = 7,258$, $p = 0,018$), com parâmetro de corte $m = 0,174$. Com os resultados apresentados por esta equação elaborou-se uma nova regra, a qual também possui uma alta previsibilidade na classificação das moléculas de maior e menor atividade, com a vantagem de ser mais simples a sua aplicação: para uma molécula apresentar uma alta probabilidade de ser ativa é preciso que a ED resulte em um valor maior ou igual a 0,174. Neste modelo a classificação das moléculas está diretamente relacionada com o descritor E_{ELET} , sendo que, quanto menos negativa (ou mais positiva) essa energia, mais ativa se torna a molécula.

A E_{ELET} é constituída pela soma da energia dos elétrons de valência da molécula, sendo esta composta pela energia cinética dos elétrons, atração coulombiana entre elétron/núcleo e a repulsão coulombiana entre os elétrons. Sendo calculada como o valor esperado do operador (em unidades atômicas):

$$E_{ELET} = -\sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} \quad (4.2),$$

sendo que N e M são os números de elétrons e íons, respectivamente.

Esta equação permite entender, por exemplo, porque a molécula **1** é mais ativa que a **2** (Tabela 3.1). Observando-se a fórmula estrutural das duas moléculas na Tabela 3.2, nota-se que a única diferença entre ambas é quanto ao substituinte **X** (Figura 4.1): para a molécula **1** temos um átomo de hidrogênio e para **2** um átomo de cloro.

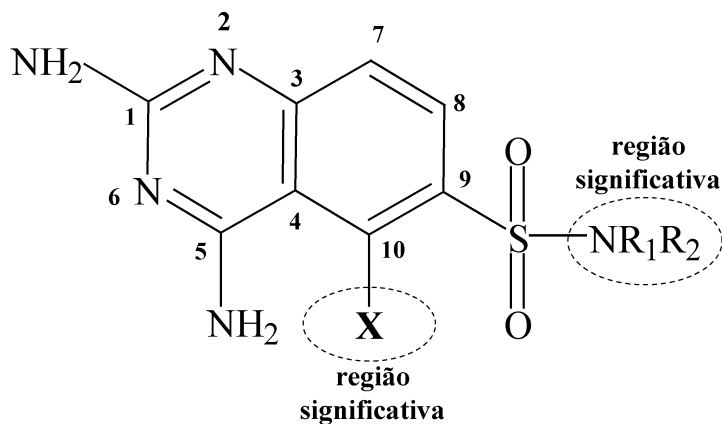


Figura 4.1: Representação da estrutura básica 2,4-diamino-6-quinazolina-sulfonamida.

Como o átomo de cloro possui maior quantidade de elétrons na camada de valência em comparação com o hidrogênio é possível concluir que: o aumento de elétrons faz com que o primeiro termo na equação 4.2 (energia cinética) fique mais negativo, favorecendo para que a molécula apresente menor atividade antimalarial. Outra conclusão possível de obter pela equação 4.2 é referente ao íon cloro: este possui um número atômico (Z_A) maior que o hidrogênio; isto faz com que o segundo termo (atração elétron/núcleo) seja mais forte, portanto mais negativo, também favorecendo para uma menor atividade antimalarial. Analisando novamente o cloro conclui-se que, o seu menor raio atômico em relação ao hidrogênio tende a manter os seus elétrons mais presos ao núcleo, aumentando a distância entre estes e os elétrons dos átomos vizinhos e conseqüentemente diminuindo a energia de repulsão presente no terceiro termo na equação 4.2, ocorrendo um favorecimento para que a molécula tenha menor atividade antimalarial. Assim, analisando os três termos da equação nota-se que a troca do hidrogênio pelo cloro leva a uma E_{ELET} menor (ou mais negativa), que por sua vez, leva a uma menor atividade antimalarial.

Portanto, como a equação 4.2 indica uma relação inversa entre a energia eletrônica total dos elétrons de valência e a quantidade de elétrons que ocupam esta camada, uma diminuição no número de elétrons geralmente implicará em um aumento da E_{ELET} (ou menos negativa), que pelo modelo proposto (equação 4.1) levará a uma maior atividade antimalarial. Com a elaboração dessa regra e a análise feita para a equação 4.2, podemos especular que novos compostos com alta atividade antimalarial deverão preferencialmente:

1. Apresentar no substituinte NR_1R_2 (Figura 4.1) cadeias com o menor número possível de átomos. Em geral, isso auxilia a reduzir o número de elétrons aumentando a E_{ELET} e melhorando a atividade antimalarial da molécula.
2. Possuir substituintes que apresentem átomos com poucos elétrons na camada de valência, o que também favorece um aumento da E_{ELET} e da atividade antimalarial da molécula.
3. Caso os átomos a serem empregados em um determinado substituinte possuam o mesmo número de elétrons na camada de valência optar pelo de menor número atômico. Segundo a equação 4.2, isso favorece uma maior atividade antimalarial.

4.2 Anilinoquinolinas

Nas anilinoquinolinas, foi adotada a mesma divisão do conjunto de moléculas presente no trabalho experimental [3], tendo como critério de agrupamento o tipo de radical ligado à estrutura básica (Figura 4.2).

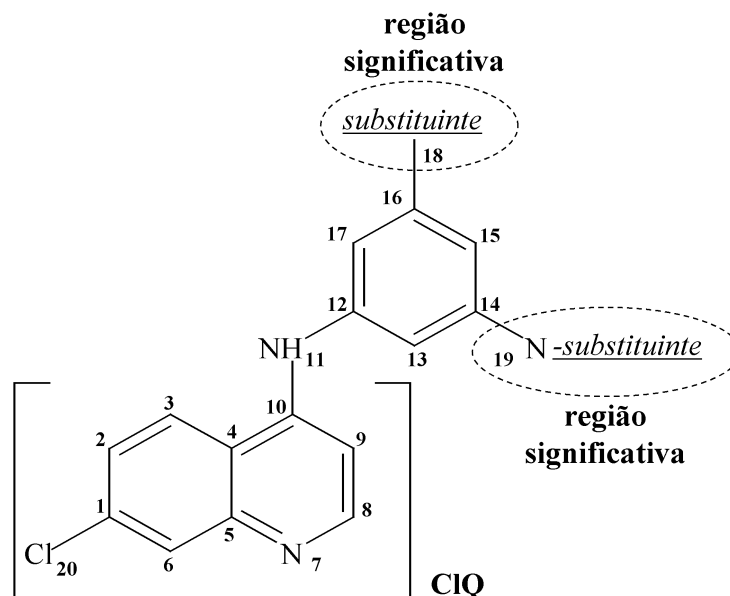


Figura 4.2: Representação da estrutura básica, comum aos 3 grupos das anilinoquinolinas.

Esta separação originou 3 grupos de moléculas, os quais, após serem submetidos a um estudo de regressão linear simples e múltipla, não resultaram em valores do parâmetro de correlação considerados significativos. Porém, as análises realizadas com os métodos estatísticos PCA e LDA apresentaram resultados expressivos.

Com o PCA foi possível obter indicações dos possíveis descritores mais relevantes para a classificação das moléculas em cada um dos grupos, sendo que, alguns desses descritores também foram selecionados pela LDA, confirmando sua importância na elaboração de um modelo. Aplicando a análise discriminante linear para o grupo 1, obteve-se a ED que leva a uma classificação correta do grau de atividade de 95,7 %:

$$Z = 489,166(PET(2)) - 87,965(PET(17)) + 9861,907(PEP_H(18)) - 1655,206 \quad (3.17),$$

sendo que as prováveis moléculas ativas possuem Z maior ou igual a 0,714.

Com o grupo 2 foi possível obter uma equação que leva a uma classificação correta de 91,75 %:

$$Z = 16382,33(PEP_L(19)) - 2,194 \quad (3.18),$$

sendo que as prováveis moléculas ativas possuem Z menor que -0,244.

No grupo 3, a ED obteve uma classificação correta do grau de atividade de 74,1 %:

$$Z = 0,19(\Delta H_F) + 3182,995(PEP_H(18)) - 2,322 \quad (3.19),$$

sendo que as prováveis moléculas ativas possuem Z menor que -0,305.

As ED's 3.17, 3.18 e 3.19, estabeleceram os modelos matemáticos que prevêm com alto índice de acerto as moléculas com maior e menor atividade antimalarial. Essas equações são instrumentos valiosos para o desenvolvimento de novos fármacos, tendo em vista que fornecem indicações quantitativas para a classificação das moléculas.

Contudo, a modelagem molecular é notoriamente uma tarefa muito complexa, envolvendo considerações provenientes das mais diversas áreas de pesquisa. Assim sendo, apesar dos modelos desenvolvidos neste trabalho fornecerem uma regra para sintetizar novos compostos antimalariais, optamos por novamente procurar ED's que permitissem conclusões mais gerais, utilizando modelos matemáticos com menores números de descritores e portanto mais simples de serem aplicados.

No estudo realizado com LDA para as anilinoquinolinas, os átomos 18 e 19 (Figura 4.2), desempenham importante papel na classificação das moléculas, um indicativo de que essas regiões contribuem significativamente para a atividade antimalarial. Não por acaso, estes átomos pertencem à posição de conexão da molécula básica com os diversos substituintes. Devido a esta relevância, foi efetuada uma série de estudos para estes átomos a fim de selecionar os descritores adequados.

Os resultados desses estudos mostraram que os descritores de maior peso na classificação são: a população eletrônica total do átomo 18 ($PET(18)$) e a população eletrônica parcial desse mesmo átomo referente ao orbital molecular mais alto ocupado ($PEP_H(18)$), ou seja, a propriedade física referente à densidade de elétrons presente nesse sítio é um fator importante na previsão da atividade antimalarial dessas moléculas. Nos estudos de LDA, onde se empregam somente estes dois descritores ($PET(18)$, $PEP_H(18)$) para os 3 grupos, nota-se um razoável sucesso de predições. Esses resultados são descritos a seguir.

Para o grupo 1 obteve-se um índice de acerto no grau de atividade de 82,6 % (significância estatística maior que 99,99 %, Lambda de Wilk = 0,559, $F_{1,14} = 16,555$ e $p = 0,003$):

$$Z = -8,328(PET(18)) + 7948,578(PEP_H(18)) - 2,324 \quad (4.3),$$

sendo que as prováveis moléculas ativas possuem Z maior que 0,462.

Para o grupo 2 o índice de acerto foi de 66,7 % (significância estatística = 99,7 %, Lambda de Wilk = 0,894, $F_{1,14} = 12,321$, $p = 0,30$):

$$Z = 5349,617(PEP_H(18)) - 2,153 \quad (4.4),$$

sendo que as prováveis moléculas ativas possuem Z maior que 0,440.

E para o grupo 3 o índice de acerto foi de 63,0 % (significância estatística maior que 99,92 %, Lambda de Wilk = 0,876, $F_{1,14} = 11,445$, $p = 0,07$):

$$Z = 4164,624(PEP_H(18)) - 1,837 \quad (4.5),$$

sendo que as prováveis moléculas ativas possuem Z menor que -0,198.

Com estes modelos foi possível elaborar uma regra mais geral para a classificação das moléculas, relacionando a atividade antimalarial com a população eletrônica presente no átomo 18 (Figura 4.2), indicando uma relação inversa: quanto menor a densidade de elétrons nessa posição, maior a probabilidade da molécula apresentar um caráter ativo. Assim, pelo modelo proposto, uma diminuição na $PEP(18)$, geralmente implicará em um aumento da atividade antimalarial. A partir dessas regras e a análises efetuadas nas ED's 4.3, 4.4 e 4.5, podemos especular que novos compostos com elevada probabilidade de apresentarem propriedades antimalariais deverão preferencialmente apresentar uma estrutura com substituintes nas posições NR_1R_2 (Tabela 3.9) com maior afinidade eletrônica, de maneira a colaborar com a diminuição da PEP (densidade eletrônica) no átomo da posição 18.

4.3 Considerações Finais

Este trabalho teve como finalidade propor um modelo que descreva as relações estrutura/atividade das moléculas sulfonamidas e anilinoquinolinas. Elaborando-se regras para a separação e constituição de novos derivados, que preferencialmente apresentem uma atividade antimalarial elevada.

Para as sulfonamidas, os resultados obtidos pelos métodos PCA e LDA foram qualitativamente iguais, ambos indicam a importância da energia eletrônica (E_{ELET}). Utilizando-se desse descritor foi possível elaborar um modelo com alta previsibilidade para a separação do grupo em maior e menor atividade, este mesmo modelo também pode ser aplicado como regra na síntese de novos derivados dessas moléculas.

Nas anilinoquinolinas, foi possível obter modelos de previsão via LDA utilizando a população eletrônica total (PET (18)) e a população eletrônica parcial do último orbital molecular ocupado do átomo 18 (PEP_H (18)). Como no caso anterior os modelos propostos possibilitam a classificação das moléculas e forneceram uma regra para a síntese de novos derivados.

Os modelos propostos neste estudo são distintos, mas de fácil aplicação, com base nessas regras é possível à separação de sulfonamidas e anilinoquinolinas considerando a sua atividade. Mas principalmente, estas mesmas regras, possibilitam sintetizar derivados dessas moléculas, de maneira a apresentarem uma fórmula estrutural que propicie maior probabilidade de possuírem elevada atividade antimalarial. Isto abre perspectivas para a pesquisa e desenvolvimento de novas moléculas proveniente desses grupos, com a possibilidade de aplicação como fármaco antimalarial.

REFERÊNCIAS

- [1] Cohen, N. C. *Guidebook on Molecular Modeling in Drug Design*. Academic Diego, California, 1996.
- [2] Springborg, M. *Methods of Electronic-Structure Calculations: from molecules to solids*, John Wiley & Sons, 2000.
- [3] Delarue, S.; Girault, S.; Maes, L.; Debreu-Fontaine, M. A.; Labaeïd, M.; Grellier, Sergheraert, P. C. *Synthesis and in vitro and in vivo antimalarial activity of new 4-anilinoquinolines*, J. Med. Chem. 44, 2827, 2001.
- [4] Elslager, E. F.; Colbry, N. L.; Davoll, J.; Hutt, M. P.; Johnson, J. L. Werbel, L. M. *Antimalarial and antibacterial effects of 2,4-diamino-6-quinazolinesulfonamides*, J. Med. Chem. 27, 1740, 1984.
- [5] Santo, L. L. E. *Aplicações de métodos semi-empíricos ao estudo de estrutura eletrônica de compostos bioativos*. Tese de Doutorado; Campinas: UNICAMP, IFGW, 2001.
- [6] Dantas, S. O.; Lavarda, F. C.; Galvão D. S.; Laks B. *An Investigation on the Electronic Structure of Antitumor Drug Ellipticine and their derivatives. Part I: Geometrical AM1 Study*, J. Mol. Struct. (THEOCHEM) 253, 319, 1992.
- [7] Lavarda, F.C.; Galvão, D.S.; Laks, B.; *Electronic Structure of Polyacetylene with Bonded Oxygen*, Synth. Met. 51, 175, 1992.
- [8] Lavarda, F. C.; *Relation between antioxidant activity and electronic structure of phenols*, Int. J. Quant. Chem. 95, 219, 2003.
- [9] Barone, P. M. V.; Camilo, A. Jr.; Galvão, D. S. *Theoretical approach to identify carcinogenic activity of polycyclic aromatic hydrocarbons*, Phys. Rev. Lett. 77, 1186, 1996.
- [10] Santos, H. F.; *Análise conformacional de modelos de lignina*, Quim. Nova, 4, 480, 2001.

- [11] Stewart, J. J. P. *MOPAC program, version 6.0, Quantum Chemistry Exchange Program* No. 455 Comp. J. Chem., 209, 10, 221, 1991.
- [12] Tavares, L. C. *QSAR: a abordagem de Hansch*, Quim. Nova, 27, 631, 2004.
- [13] Montanari, M. L. C.; Montanari, C. A.; Gáudio, A. C. *Validação lateral em relações quantitativas entre estrutura e atividade farmacológica*, QSAR, Quim. Nova, 25, 231, 2002.
- [14] A. C. Gáudio, *Modelos Clássicos de Estudo Quantitativo das Relações entre Estrutura Química e Atividade Biológica*, Quim. Nova, 19, 278, 1996.
- [15] Ferreira, M. M. C.; Montanari, C. A.; Gaudio A. C. *Seleção de variáveis em QSAR*. Quim. Nova. 25, 439, 2002.
- [16] Clark, T. *A Handbook of Computational Chemistry- A Practical Guide to Chemical Structure and Energy Calculations*. John Wiley & Sons, EUA, 1985.
- [17] Lowe, J. P. "Quantum Chemistry". Academic Press, San Diego, California, 1978.
- [18] Pople, J. A.; Beveridge, D. L.; *Approximate Molecular Orbital Theory*. McGraw-Hill, 1970.
- [19] Dewar, M. J. S.; Zoebish, E. G.; Healy, E. F.; Stewart, J. J. P.; *Development and use of quantum mechanical molecular models AM1: a new general purpose quantum mechanical molecular model*, J. Am. Chem. Soc. 107, 3902, 1985.
- [20] Dewar, M. J. S.; McKee, M. L.; *Ground states of molecules MNDO results for molecules containing boron*, J. Am. Chem. Soc. 99, 5231, 1977.
- [21] Spiegel, M. R. *Estatística*. Tradução de Pedro Cosentino. 7 ed. São Paulo. McGraw-Hill, 1974. Título original: *Schaum's outline of theory and problems of statistics*.
- [22] Ferreira, M. M. C. *Multivariate QSAR*. J. Braz. Chem. Soc. 13, 742, 2002.
- [23] Moita Neto, J. M.; Moita, G. C. *Uma introdução à análise exploratória de dados multivariados*. Quim. Nova. 21, 467, 1997.

- [24] Wold, S.; Esbensen, K.; Gelati, P. *Principal component analysis*. Chem. Intell. Lab. Syst. 2, 37, 1987.
- [25] Levin, J.; *Estatística aplicada a ciências humanas*, 2^a ed., Editora Harbra, São Paulo, 1987.
- [26] Hair, J. F. et al. *Multivariate Data Analysis*. 5. ed. Nova Jersey: Prentice Hall, 1998.
- [27] *Malaria*. Disponível em: <<http://www-micro.msb.le.ac.uk/224/Malaria.html>>. Acesso em: 22 jun. 2006.
- [28] Zucker, J. R.; Campbell, C. C.; *Smear negative cerebral malaria due to mefloquine resistant P. falciparum acquired in the Amazon*. J. Inf. Dis. 166, 1458, 1992.
- [29] *Malaria: Drug resistance*. <<http://www-micro.msb.le.ac.uk/224/Bradley/Resistance.html>>. Acesso em: 22 jul. 2006.
- [30] *SPSS for Windows*, Rel. 11.0.1. 2001. Chicago: SPSS Inc.
- [31] Gáudio, A. C.; Zandonate, E.; "Proposição, validação e análise dos modelos que correlacionam estrutura química e atividade biológica"; Quim. Nova, 24, 658, 2001.
- [32] Krungskai, J.; Scozzafava, A.; Reungprapavut, S.; Krungkrai, S. R.; Rattanajak, R.; Kamchonwongpaisan S.; Supuran, C. T.; *Carbonic anhydrase inhibitors. Inhibition of Plasmodium falciparum carbonic anhydrase with aromatic sulfonamides: towards antimalarials with a novel mechanism of action?*, Bio. Med. Chem. 13, 483, 2005.
- [33] Ojala, C. R.; Ostman, J. M.; Ojala, W. H.; Hanson, S. E. *Molecular and crystal structures of N-arylglycopyranosylamines formed by reaction between sulfanilamide and d-ribose, d-arabinose and d-mannose*, Carbohydr. Res. 331, 319, 2001.
- [34] Malesic, M.; Krbavcic, A.; Golobic, A.; Golic, L.; Stanovnik, B. *The Synthesis and Transformations of some 3-Thiocarbamoylthiazolidines*, J. Heterocycl. Chem. 34, 43 1997.

[35] CAChe 5.0, Fujitsu, Japan. <http://www.CACheSoftware.com>.

[36] Schmidt, M. W.; Baldrige, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *General atomic and molecular electronic structure system*, J. Comput. Chem. 14, 1347, 1993.

APÊNDICE A

A Equação de Schrödinger

Sabe-se que as partículas de sistemas microscópicos apresentam um comportamento que não pode ser descrito pela mecânica newtoniana. Nestes casos, todo o comportamento do sistema é descrito através da Equação de Schrödinger (ES):

$$H\Psi(\vec{r}, t) = i\hbar \frac{\partial}{\partial t} \Psi(\vec{r}, t) \quad (\text{A.1})$$

sendo que:

- i é a unidade imaginária e vale $\sqrt{-1}$;
- \hbar é a constante de Planck, h , (constante física associada somente à fenômenos quânticos) dividida por 2π ($h/2\pi$);
- Ψ é a “função de onda” do sistema sendo estudado e é uma função do tempo e das posições dos objetos (partículas) que compõem o sistema;

- H , conhecido como “hamiltoniano” do sistema, é o operador “energia total”, sendo a soma dos operadores para as energias cinéticas (T) e potencial (V);
- t é o tempo;

$$H = T + V \quad (\text{A.2})$$

Supondo que o sistema em estudo seja composto de N partículas:

$$H = \sum_{i=1}^N [T_i + V(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N, t)] \quad (\text{A.3})$$

sendo que:

- V é escrito do mesmo modo como na mecânica clássica.
- T_i é escrito em coordenadas cartesianas como:

$$T_i = -\frac{\hbar^2}{2m_i} \nabla_i^2 \quad (\text{A.4})$$

sendo que m_i é a massa da partícula i .

Substituindo-se (A.4) em (A.3) temos:

$$H = \sum_{i=1}^N \left[-\frac{\hbar^2}{2m_i} \nabla_i^2 + V(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N, t) \right] \quad (\text{A.5})$$

Por fim a ES para um sistema de N partículas pode ser escrita como:

$$\sum_{i=1}^N \left[-\frac{\hbar}{2m_i} \nabla_i^2 \Psi(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N, t) + V(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N, t) \cdot \Psi(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N, t) \right] = i\hbar \frac{\partial}{\partial t} \Psi(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N, t) \quad (\text{A.6})$$

Se a função (ou operador) energia potencial, que descreve as interações entre as partículas do sistema não depender do tempo, é possível separar a ES em duas outras equações, mais simples de serem resolvidas.

$$H \psi(\vec{r}) = E \psi(\vec{r}) \quad (\text{A.7})$$

e

$$\phi(t) = e^{-iEt/\hbar} \quad (\text{A.8})$$

sendo:

- $\Psi(\vec{r}, t) = \psi(\vec{r}) \cdot \phi(t)$
- E a energia do sistema em estudo.

A equação (A.7) é chamada Equação de Schrödinger Independente do Tempo (ESIT) e pode ser expressa de maneira análoga à (A.6), para um sistema de N partículas, por:

$$\sum_{i=1}^N \left[-\frac{\hbar}{2m_i} \nabla_i^2 \psi(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N) + V(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N) \cdot \psi(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N) \right] = E \psi(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N) \quad (\text{A.9})$$

Em equações de Física Atômica e Molecular é relevante a introdução do sistema de unidades atômicas, pois deste modo evita-se a necessidade de se expressar diversas constantes físicas, simplificando em muito as equações a serem tratadas.

Tabela A.1: Sistema de Unidades Atômicas.

Quantidade Física	Unidade	Valor no Sistema Internacional (SI)
Comprimento	a_0	$5,2918^{-11}$ m
Massa	m_e	$9,1095^{-31}$ kg
Carga	E	$1,6022^{-19}$ C
Energia	<i>Hartree</i>	$4,3598^{-18}$ J

Sendo a_0 o raio de Bohr, m_e massa do elétron e E a carga do elétron.

Ao se deparar com um sistema de átomos é conveniente decompor tal sistema em partes, uma primeira referente somente aos núcleos e outra referente aos elétrons. Assim, considerando a energia cinética de núcleos e elétrons e a interação elétrica entre cada par de constituintes, obter-se-á termos um hamiltoniano (em unidades atômicas) dado por:

$$H = -\sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{A=1}^M \frac{1}{2M_A} \nabla_A^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} + \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{R_{AB}} \quad (\text{A.10})$$

sendo que:

- N é o número de elétrons;
- M é o número de núcleos;
- M_A é a razão entre a massa do núcleo A e a massa do elétron;

- Z_A é o número atômico do núcleo A ;
- ∇_i^2 é o Laplaciano que atua sobre as coordenadas do elétron i ;
- ∇_A^2 é o Laplaciano que atua sobre as coordenadas do núcleo A .

O primeiro termo representa a energia cinética dos elétrons; o segundo a energia cinética dos núcleos; o terceiro a atração coulombiana entre os elétrons e o núcleo; o quarto a repulsão coulombiana entre os elétrons e o quinto a repulsão coulombiana entre os núcleos.

Tal separação é interessante, pois, devido ao fato dos núcleos serem mais pesados que os elétrons, eles tenderão a se mover muito mais lentamente. Portanto, é uma boa aproximação considerar os elétrons de uma molécula como se movendo em meio a um campo de núcleos fixos. Deste modo o segundo termo de (A.10) se anularia, e o quinto termo, referente à repulsão entre os núcleos, seria considerado como uma constante, o que levaria ao dito hamiltoniano eletrônico:

$$H_e = -\sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} \quad (\text{A.11})$$

e do mesmo modo a uma ESIT Eletrônica:

$$H_e \psi_e = E_e \psi_e \quad (\text{A.12})$$

APÊNDICE B

O Método Hartree-Fock

As Equações de Hartree-Fock

Encontrar soluções para a Equação de Schrödinger Independente do Tempo (ESIT) eletrônica configura um problema de grande relevância, pois este tipo de equação possui solução analítica somente para casos bem simples como de uma partícula interagindo com um campo, ou duas (e, em casos bem específicos, três) partículas interagentes. Deste modo, num problema de muitos corpos, métodos iterativos são empregados a fim de se obter uma solução aproximada da ESIT.

No trabalho aqui descrito empregou-se a aproximação Hartree-Fock (HF), método muito utilizado na obtenção de soluções aproximadas para a ESIT e que serve também de ponto de partida para outras aproximações mais acuradas.

No método HF, procura-se transformar um problema de N partículas interagentes em N problemas de uma partícula interagindo com o campo produzido pelas outras $N-1$ restantes. Como se pode perceber, tal aproximação acaba por eliminar a correlação entre as partículas. No entanto, apresenta resultados satisfatórios, pois se utiliza um recurso autoconsistente de cálculo (SCF, *self consistent field*).

Devido ao fato de se dividir o problema inicial em N problemas de elétrons não correlacionados, a função de onda ψ pode ser escrita como um produto de funções individuais para cada elétron. Visando preservar a anti-simetria necessária à função de onda dos elétrons, toma-se ψ como um determinante de Slater destas funções.

$$\psi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(1) & \chi_2(1) & \dots & \chi_N(1) \\ \chi_1(2) & \chi_2(2) & \dots & \chi_N(2) \\ \vdots & \vdots & & \vdots \\ \chi_1(N) & \chi_2(N) & \dots & \chi_N(N) \end{vmatrix} \quad (\text{B.1})$$

sendo que $\chi_j(i)$ é a “função de onda individual das variáveis espaciais e de spin do elétron i ” conhecido como “orbital referente à partícula i ”. O conjunto destas funções é imposto como ortonormal e o fator $(N!)^{-1/2}$ visa à normalização de ψ .

Tendo-se H bem determinado e se considerando ψ como o determinante de Slater de um conjunto de funções $\{\chi_i\}$, pode-se obter, através de A.12 (Apêndice A), uma expressão da energia E das partículas em função do conjunto $\{\chi_i\}$, de modo que, o conjunto de funções que representam a solução aproximada de A.12, será aquele para o qual a energia do sistema é minimizada.

Utilizando-se destes argumentos de cálculo variacional e de propriedades de operadores hermitianos, partindo-se da ESIT, chega-se às denominadas Equações de Hartree-Fock (EHF).

$$f(1)\chi_i(1) = \varepsilon_i\chi_i(1), \quad i = 1, 2, \dots, N \quad (\text{B.2})$$

pois:

- $\chi_i(1)$ é a i -ésima função de onda (orbital molecular) que representa a parte espacial e de spin do elétron 1;
- $f(1)$ é o operador de fock operando sobre o elétron 1:

$$f(1) = h(1) + \sum_j^N \{2J_j(1) - K_j(1)\} \quad (\text{B.3})$$

sendo que:

- $h(1)$ é o hamiltoniano de um elétron

$$h(1) = -\frac{1}{2}\nabla_1^2(1) + \sum_{A=1}^M \frac{Z_A}{r_{1A}} \quad (\text{B.4})$$

O primeiro termo representa o operador energia cinética atuando sobre as coordenadas do elétron 1, o segundo a somatória da energia potencial devido aos núcleos.

- J_j e K_j são operadores de Coulomb e de Troca respectivamente. Quando aplicados sobre o orbital molecular χ_i resultam em:

$$J_j \cdot \chi_i(1) = \left[\int \chi_j^*(2) \frac{1}{r_{12}} \chi_j(2) d\tau_2 \right] \chi_i(1) \quad (\text{B.5})$$

$$K_j \cdot \chi_i(1) = \left[\int \chi_j^*(2) \frac{1}{r_{12}} \chi_i(2) d\tau_2 \right] \chi_j(1) \quad (\text{B.6});$$

Em mecânica quântica o estado de energia dos elétrons de um sistema de partículas é definido de acordo com quatro números quânticos, três relacionados à distribuição espacial (número quântico principal, de momento angular e de momento magnético) e um número de spin. Pelo princípio de exclusão de Pauli, dois elétrons não podem possuir o mesmo conjunto de números quânticos. De forma que só podem ocupar uma mesma distribuição espacial (orbital χ_i) se possuírem números quânticos de spin diferentes. Para a resolução das EHF's esta é uma característica muito relevante, pois considerando moléculas de camada fechada (número par de elétrons de valência), pode-se supor que os elétrons, dois a dois, ocupam uma mesma distribuição espacial (possuindo spins diferentes), reduzindo o problema inicial de se calcular N orbitais χ pela metade. Tal método é conhecido como "*Restricted Hartree-Fock*" (RHF) e pôde ser empregado no trabalho aqui descrito devido às moléculas serem todas de camada fechada.

Em caráter prático tal aproximação é interessante uma vez que a comparação entre os resultados obtidos e os dados experimentais mostra-se a melhor possível, além do que os custos computacionais serem bem mais reduzidos.

Deste modo, para os orbitais χ_i^{RHF} , tem-se:

$$\begin{cases} \chi_{2i-1}(\vec{r}, \omega) = \psi_i(\vec{r}) \cdot \phi_\alpha(\omega) \\ \chi_{2i}(\vec{r}, \omega) = \psi_i(\vec{r}) \cdot \phi_\beta(\omega) \end{cases} \quad i = 1, 2, \dots, N/2 \quad (\text{B.7})$$

sendo que ω representa a variável de spin (α ou β) e \vec{r} as variáveis espaciais.

As EHF's pelo método RHF possuem a mesma forma da equação (B.2), diferindo pelo operador de Fock:

$$f(1) = h(1) + \sum_a^{N/2} \{2J_a(1) - K_a(1)\} \quad (\text{B.8})$$

sendo que os operadores de Coulomb e de Troca para camada fechada, quando aplicados sobre o orbital molecular ψ_i , são dados por:

$$J_a(1).\psi_i(1) = \left[\int \psi_a^*(2) \frac{1}{r_{12}} \psi_a(2) d\tau_2 \right] \psi_i(1) \quad (\text{B.9})$$

$$K_a(1).\psi_i(1) = \left[\int \psi_a^*(2) \frac{1}{r_{12}} \psi_i(2) d\tau_2 \right] \psi_a(1) \quad (\text{B.10})$$

É interessante notar que a variável de spin não é levada em conta, pois assume-se que tanto para ϕ_α como para ϕ_β a parte espacial de χ_i^{RHF} será a mesma.

Deste modo as EHF-RHF serão dadas por:

$$f(1)\psi_i(1) = \varepsilon_i\psi_i(1), \quad i = 1, 2, \dots, N \quad (\text{B.11})$$

As Equações de Rothaan-Hall

Como se pode perceber, as EHF-RHF, (B.11) formam um conjunto de equações integro-diferenciais de solução complexa. Devido a isso, muitas vezes procura-se transformar tais equações em um sistema de equações algébricas. Para tanto costuma-se considerar as funções desconhecidas ψ_i como uma combinação linear

desconhecida de um conjunto de funções conhecidas $\{\theta_\nu\}$. Tais funções são denominadas funções de base.

$$\psi_i = \sum_{\nu=1}^k C_{\nu i} \theta_\nu \quad (\text{B.12})$$

sendo que:

- geralmente $k \geq N/2$, de modo que, como qualquer expansão de funções, quanto maior o seu valor, melhor será a aproximação obtida;
- $C_{\nu i}$ é o coeficiente que indica quão grande é a contribuição de θ_ν para a descrição de ψ_i .

Substituindo B.12 na EHF-RHF tem-se:

$$f(1) \sum_{\nu} C_{\nu i} \theta_\nu(1) = \varepsilon_i \sum_{\nu} C_{\nu i} \theta_\nu(1) \quad (\text{B.13})$$

Multiplicando ambos os termos por θ_μ^* e integrando sobre as coordenadas espaciais do elétron 1, tem-se:

$$\sum_{\nu} C_{\nu i} \int \theta_\mu^*(1) f(1) \theta_\nu(1) d\vec{r}_1 = \varepsilon_i \sum_{\nu} C_{\nu i} \int \theta_\mu^*(1) \theta_\nu(1) d\vec{r}_1 \quad (\text{B.14})$$

Definem-se assim duas matrizes hermitianas de ordem k :

1. Matriz Sobreposição \tilde{S} :

$$S_{\mu\nu} = \int \theta_{\mu}^*(1) \theta_{\nu}(1) d\vec{r} \quad (\text{B.15})$$

Por vezes também chamada de Matriz Métrica, \tilde{S} oferece informações a respeito da dependência linear entre os elementos do conjunto de funções de base θ_{μ} e θ_{ν} .

2. Matriz de Fock \tilde{F} :

$$F_{\mu\nu} = \int \theta_{\mu}^*(1) f(1) \theta_{\nu}(1) d\vec{r}_1 \quad (\text{B.16})$$

\tilde{F} é a representação matricial do operador de Fock na base de funções $\{\theta_{\nu}\}$.

Deste modo, pode-se escrever as EHF-RHF como:

$$\sum_{\nu} F_{\mu\nu} C_{\nu i} = \varepsilon_i \sum_{\nu} S_{\mu\nu} C_{\nu i} \quad (\text{B.17})$$

Estas são as Equações de Roothaan-Hall e podem ser escritas de maneira mais compacta, na forma matricial:

$$\tilde{F} \tilde{C} = \tilde{S} \tilde{C} \tilde{\varepsilon} \quad (\text{B.18})$$

pois:

$$\tilde{C} = \begin{pmatrix} C_{11} & C_{12} & \cdots & C_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ C_{k1} & C_{k2} & \cdots & C_{kk} \end{pmatrix} \quad (\text{B.19})$$

$$\tilde{\epsilon} = \begin{pmatrix} \epsilon_{11} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \epsilon_{kk} \end{pmatrix} \quad (\text{B.20})$$

sendo que:

- \tilde{C} é uma matriz quadrada de dimensão k que descreve qual a contribuição de cada função de base θ_ν na construção dos orbitais moleculares ψ ;
- $\tilde{\epsilon}$ é uma matriz diagonal das energias dos orbitais moleculares.

Substituindo a definição de $f(1)$ dada em (B.8), na Equação (B.16) obtém-se:

$$F_{\mu\nu} = H_{\mu\nu}^{car} + G_{\mu\nu} \quad (\text{B.21})$$

uma vez que:

$$H_{\mu\nu}^{car} = \int \theta_\mu^*(1) h(1) \theta_\nu(1) d\vec{r}_1 \quad (\text{B.22})$$

Definindo uma matriz de um “hamiltoniano de caroço” \tilde{H}^{car} , cujos elementos carregam informação sobre a atração dos núcleos (considerados estáticos) sobre cada elétron. Em virtude disso é considerada como parcela referente “a um elétron”

da matriz de Fock. Pelo fato dos elementos de \tilde{H}^{car} não dependerem dos valores de C_{vi} , que são agora as incógnitas do problema, seu cálculo é executado apenas uma vez.

$$G_{\mu\nu} = \sum_{\sigma\lambda} P_{\sigma\lambda} \left[(\mu\nu | \sigma\lambda) - \frac{1}{2} (\mu\sigma | \lambda\nu) \right] \quad (\text{B.23})$$

É a parcela da matriz de Fock referente a dois elétrons. Sendo que:

$$P_{\sigma\lambda} = 2 \sum_{a=1}^{N/2} C_{\sigma a} C_{\lambda a}^* \quad (\text{B.24})$$

define os elementos da Matriz Densidade \tilde{P} , cujos valores fornecem informação a respeito da probabilidade de se localizar os elétrons nos orbitais moleculares. E:

$$(ij | kl) = \iint \theta_i^*(1) \theta_j(1) \frac{1}{r_{12}} \theta_k^*(2) \theta_l(2) d\vec{r}_1 d\vec{r}_2 \quad (\text{B.25})$$

Como se pode perceber, os elementos $G_{\mu\nu}$ dependem dos valores de C_{vi} , de modo que, para se calcular a matriz de Fock é necessário ter em mãos justamente o que se quer calcular, o conjunto $\{C_{vi}\}$. Daí vem o fato do método HF ser dito SFC, “self consistent field” (campo auto consistente): é necessária uma primeira hipótese para o cálculo da matriz \tilde{P} , a fim de se calcular uma nova matriz densidade; o processo se repete até que os elementos de \tilde{P} não se alterem mais.

APÊNDICE C

Estudo do Método Semi-empírico

Conformação Geométrica do Grupo Funcional -SO₂NH₂ em Sulfonamidas

Este estudo tem por objetivo comparar dados experimentais com resultados teóricos provindos de métodos de cálculos de orbitais moleculares nos níveis semi-empírico e *ab initio* de teoria para a conformação molecular do grupo funcional -SO₂NH₂ das sulfonamidas. Surgiu da necessidade de se determinar o método mais adequado para o cálculo da estrutura eletrônica de dois grupos de sulfonamidas que possuem atividade antimalarial [4, 32]. Estes cálculos fornecem os descritores de origem eletrônica empregados em estudos de correlação da estrutura com a atividade antimalarial. Como se partiu do pressuposto que a obtenção de uma boa geometria por parte do método teórico indica a qualidade dos resultados para a estrutura eletrônica, é essencial que o método escolhido reproduza adequadamente as geometrias já determinadas via métodos experimentais, isto é, a conformação molecular obtida seja um critério de avaliação da qualidade do método.

O problema exigiu este estudo em vista do fato da valência do átomo de enxofre neste grupo, perfazendo um total de quatro ligações, em que supostamente os orbitais d desempenham um papel importante nas ligações químicas. Imagina-se que os métodos semi-empíricos mais comumente adotados não seriam capazes de reproduzir adequadamente a geometria de equilíbrio, uma vez que estes utilizam uma base mínima de orbitais atômicos do tipo s e p. Então, foi realizado este estudo para comparar dados experimentais com resultados dos métodos semi-empíricos mais empregados atualmente, com um método semi-empírico que inclui os orbitais d nos cálculos e um método *ab initio*.

A molécula cuja conformação molecular é investigada pelos métodos teóricos é a *cyclohexanesulfonamide*, Figura C.1. Preferiu-se manter a nomenclatura IUPAC das moléculas na sua grafia em inglês para permitir uma mais rápida comparação com outras publicações.

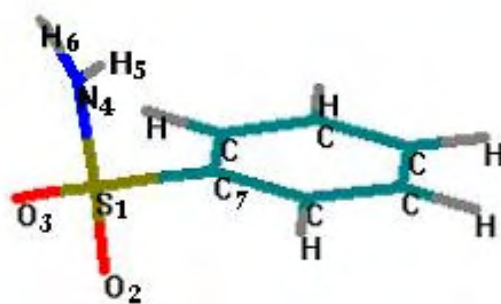


Figura C.1: Sulfonamida empregada nos cálculos semi-empíricos (*cyclohexanesulfonamide*).

Os dados experimentais da geometria de diversas moléculas que contém como parte de sua estrutura a *cyclohexanesulfonamide*, foram encontrados no banco de

dados Cambridge Structural Database (CSD) e originalmente publicados nas referências [33] e [34]. São elas:

1. MESXIR, $C_{11}H_{16}N_2O_6S_1$, *N-(p-Sulfamoylphenyl)- α -D-ribofuranosylamine* [33].
2. MESXOX, $C_{11}H_{16}N_2O_6S_1$, *N-(p-Sulfamoylphenyl) arabinofuranosylamine* [33].
3. MESXUD, $C_{12}H_{16}N_2O_7S_1, H_2O_1$, *N-(p-Sulfamoylphenyl)- β -D-mannofuranosylamine monohydrate* [33].
4. NITRAJ, $C_{17}H_{16}N_4O_3S_3$, *2-(2-Methoxyphenyl) imino-3-(N-(3-sulfamoylphenyl)) thiocarbamoylthiazolidine* [34].

Para o cálculo da estrutura eletrônica no nível de teoria semi-empírico, usa-se o pacote de domínio público Mopac versão 6 [11]. Este oferece quatro tipos de hamiltonianos parametrizados: MNDO (*Modified Neglect of Differential Overlap*), AM1 (*Austin Model 1*), PM3 (*Parametric Method 3*) e MINDO/3 (*Modified Intermediate Neglect of Differential Overlap/3*). Todos eles são métodos Hartree-Fock, de campo autoconsistente, que empregam uma base de orbitais atômicos s e p em que várias simplificações são aplicadas no intuito de se reduzir os recursos computacionais necessários. O MINDO/3 serviu de base para o desenvolvimento dos outros métodos e não foi usado. Além da parametrização, o AM1 e o PM3 diferem do MNDO principalmente pela função que calcula a repulsão nuclear e diferem entre si pela parametrização adotada: o AM1, como o MNDO, possui alguns parâmetros colhidos de experimentos e outros otimizados, enquanto que no PM3 todos os parâmetros são otimizados.

Ainda no nível semi-empírico, o método MNDO/d é uma modificação do método MNDO em que se passa a empregar também os orbitais d na base de orbitais

atômicos para o cálculo da estrutura eletrônica. Os cálculos foram feitos com o programa Cache 5.0 [35].

Para o cálculo da estrutura eletrônica no nível de teoria *ab initio*, usa-se o pacote de domínio público Gamess [36]. Este oferece a possibilidade de cálculos via o método Hartree-Fock, vários métodos pós-Hartree-Fock e o método DFT (*Density Functional Theory*), considerado hoje o método mais adequado para o cálculo de primeiros princípios da estrutura eletrônica de materiais. Este método incorpora de modo mais eficaz os efeitos da correlação eletrônica. Emprega-se o método DFT com uma base extensa e polarizada, 6-31G*, para garantir uma ótima qualidade de resultado.

Apresenta-se na Tabela C.1 os dados da conformação molecular do grupo $-\text{SO}_2\text{NH}_2$ para as diversas moléculas em que se há dados experimentais bem como para aquelas que se calcula a geometria de equilíbrio na fase gasosa.

Tabela C.1: Comparação de dados experimentais e teóricos da conformação geométrica. Os dados experimentais estão identificados de acordo com o nome que recebem no banco de dados Cambridge Structural Database (CSD).

Ligações	Resultados Experimentais					Resultados Teóricos (Média Teórica/Erro Percentual)				
	MESXIR [1]	MESXOX [1]	MESXUD [1]	NITRAJ [2]	Média Exp.	MNDO	MNDO/d	AM1	PM3	DFT, 6-31G*
S1-C7 (Å)	1,759	1,755	1,760	1,771	1,761	1,770 (+1%)	1,777 (+1%)	1,690 (-4%)	1,785 (+1%)	1,80 (+2,3%)
S1-O2 (Å)	1,430	1,434	1,433	1,439	1,434	1,524 (+6%)	1,496 (+4%)	1,411 (-2%)	1,458 (+2%)	1,46 (+2,1%)
S1-O3 (Å)	1,436	1,438	1,430	1,414	1,430	1,524 (+7%)	1,496 (+5%)	1,411 (-1%)	1,457 (+2%)	1,46 (+2,1%)
S1-N4 (Å)	1,598	1,611	1,610	1,597	1,604	1,692 (+5%)	1,689 (+5%)	1,602 (~0%)	1,747 (+9%)	1,68 (+5,2%)
N4-H5 (Å)	0,872	0,852	0,872	-	0,865	1,003 (+16%)	1,001 (+16%)	0,988 (+14%)	0,99 (+15%)	1,1 (+13,1%)
N4-H6 (Å)	1,031	0,898	0,868	-	0,932	1,003 (+8%)	1,001 (+7%)	0,988 (+6%)	0,998 (+7%)	1,1 (+13,1%)
02-S1-C7 (°)	107,6	107,9	107,3	107,4	107,6	110,9 (+3%)	110,1 (+2%)	108,4 (+1%)	109,9 (+2%)	107,4 (-0,3%)
03-S1-C7 (°)	107,5	108,9	107,8	107,3	107,9	110,9 (+3%)	110,1 (+2%)	108,4 (-0%)	109,8 (+2%)	107,4 (-0,3%)
N4-S1-C7 (°)	109,2	108,9	109,4	107,7	108,8	105,4 (-3%)	101,1 (-7%)	103,2 (-5%)	100,9 (-7%)	107,0 (-1,7%)
H5-N4-S1 (°)	107,4	113,2	108,0	-	109,5	115,0 (+5%)	124,7 (+14%)	123,6 (+13%)	118,5 (+8%)	112,4 (+0,2%)
H6-N4-S1 (°)	117,2	113,0	114,2	-	114,8	115,1 (-0%)	124,7 (+9%)	123,6 (+8%)	118,6 (+3%)	112,4 (+0,2%)
Erro Médio Percentual	-	-	-	-	-	5,182 %	6,545 %	4,91 %	5,273 %	3,69 %

Primeiramente, nota-se que (sempre fazendo a análise em relação aos dados experimentais) nenhum método teórico semi-empírico mostrou resultados excepcionalmente melhores que outros. E a comparação destes com os resultados *ab initio* mostra que estes métodos proporcionam bons resultados. Esta última observação permite concluir que, em vista do custo computacional muito maior para a obtenção dos resultados *ab initio*, pode-se empregar métodos semi-empíricos para o cálculo da estrutura eletrônica de sulfonamidas.

Em termos de comprimentos de ligação (CL), pode-se dizer que os mais importantes a serem avaliados são as ligações do átomo de enxofre S-C, S-N e S-O, não necessariamente nesta ordem. Para estes CLs, o método que apresenta, em média, o menor erro é o AM1, que só é pior que os outros métodos na previsão do CL S-C.

Acredita-se que efeitos da cristalização ou dificuldades dos métodos experimentais para a correta determinação devem ser as causas que justificam o fato de que o CL N-H apresenta diferenças altas para todos os métodos teóricos.

Finalmente, se forem considerados todos os CLs presentes no grupo $-\text{SO}_2\text{NH}_2$, mantém-se esta tendência do AM1 apresentar o melhor resultado médio.

Os ângulos O-S-C apresentam bons valores para todos os métodos, sendo que a menor diferença é apresentada pelo método AM1, e o ângulo N-S-C é subestimado em todos os métodos. Os ângulos diedrais não serão comparados com os dados experimentais, uma vez que a cristalização influencia bastante e os métodos teóricos realizam os cálculos considerando a molécula no vácuo.

A análise dos resultados não mostrou entre os métodos semi-empíricos a esperada superioridade do MNDO/d. É conhecido o fato de que muitas vezes uma base mínima de orbitais atômicos de valência não descreve bem as propriedades

moleculares de compostos de enxofre. Chama a atenção que o método MNDO/d, supostamente mais adequado para tratar o enxofre, pois contém orbital atômico d em sua base de orbitais atômico de valência, não produziu resultados melhores. Isto parece demonstrar que os orbitais d não desempenham um papel central nesta configuração do átomo de enxofre.

Entre os métodos semi-empíricos, aquele que apresenta os melhores resultados é o AM1, comparando-se todos os dados. A respeito deste método, pode-se observar que o resultado que não apresenta o menor desvio em relação aos dados experimentais, diz respeito ao CL S-C, sendo este 4,0% menor enquanto o resultado *ab initio* é 2,3% maior, não comprometendo os resultados para a estrutura eletrônica como um todo.

Conclui-se que os métodos teóricos apresentam bons resultados para a conformação geométrica do grupo $-\text{SO}_2\text{NH}_2$, o que indica que os resultados para a estrutura eletrônica devem ser confiáveis. Não se observou, entre os métodos semi-empíricos, a necessidade da utilização de uma base de orbitais atômicos que incluam os orbitais d. O método *ab initio* DFT(6-31G*) indica que a margem de diferença apresentada pelos métodos semi-empíricos é aceitável.

Agradecimentos ao Laboratório de Cristalografia, Estereodinâmica e Modelagem Molecular (UFSCar) pela consulta ao banco de dados Cambridge Structural Database (CSD) que permitiu localizar e obter os estudos de cristalografia contendo as sulfonamidas para a montagem do trabalho. Agradecimentos também ao Grupo de Sólidos Orgânicos e Novos Materiais (UNICAMP) que permitiu calcular a geometria de equilíbrio da *cyclohexanesulfonamide* usando o hamiltoniano MNDO/d no programa Cache 5.0.