

Universidade Estadual de Campinas
Instituto de Química

Estudos Teóricos da Relação entre Estrutura Química e Atividade
Biológica de uma Série de Derivados de Hexaidroindenopiridina
como Contraceptivos Masculinos

Maximiliano Segala

Orientador: Prof. Dr. Yuji Takahata

Dissertação de Mestrado

9 de julho de 2001

Agradecimentos

- Agradeço de maneira especial ao Professor Yuji Takahata pela orientação;
- Aos amigos Cristiano Krug e Paula Kipper pela grande ajuda pessoal e ao trabalho;
- Aos amigos e colegas André Severo Pereira Gomes, Ataulpa Albert Carmo Braga, Cristiano Gomes de Faria, Edilson Grunheidt Borges, Edilton de Souza Barcellos, Hermes Fernandes de Souza, Kesley Moraes Godinho de Oliveira, Lucimara Ramos Martins e Marlon Martins dos Reis;
- Aos amigos e professores Marcia Miguel Castro Ferreira, Pedro Antonio Muniz Vazquez e Rogério Custódio;
- Ao amigo Klaus Steding-Jessen e a todos os desenvolvedores de *softwares* livres;
- A família pelo grande apoio dado;
- Ao assessor do projeto pelas excelentes sugestões apresentadas;
- Ao Centro Nacional de Supercomputação no Rio Grande do Sul e ao Centro Nacional de Processamento de Alto Desempenho em São Paulo pelos recursos computacionais;
- A Fundação de Amparo à Pesquisa do Estado de São Paulo pela concessão da bolsa pelo projeto 99/04558-7.

Resumo

Estudos Teóricos da Relação entre Estrutura Química e Atividade Biológica de uma Série de Derivados de Hexaidroindenopiridina como Contraceptivos Masculinos, por Maximiliano Segala e Prof. Dr. Yuji Takahata.

Palavras-chave: contraceptivos masculinos, cálculos quânticos, SAR, PCA, KNN, NN.

Neste trabalho investigou-se a relação entre estrutura química e atividade biológica para uma série de derivados de 2,3,4,4a,5,9b-hexaidroindeno[1,2-c]piridina, compostos com atividade antiespermatogênica, sendo assim, potenciais contraceptivos masculinos orais. Primeiramente determinaram-se as conformações de mínimo local para um esqueleto base, por análise conformacional sistemática. Os mínimos obtidos deste processo foram utilizados como sementes para análise conformacional estocástica. Em ambas as análises utilizou-se o operador PM3, sendo os resultados idênticos. Os mínimos locais encontrados também foram otimizados com os métodos AM1, HF e DFT. Com isso, determinou-se a conformação mais estável do esqueleto base, sendo esta utilizada como ponto de partida para a construção dos derivados. A otimização geométrica dos derivados foi obtida com o método HF/3-21G e propriedades eletrônicas foram calculadas em HF/3-21G, HF/6-31G e HAM/3. Também foram calculadas propriedades estéricas, como volume molecular e coeficiente de partição octanol/água. Estas variáveis foram submetidas a tratamento com PCA e, ao final do processo, três variáveis (duas cargas atômicas e o volume molecular) mostraram-se úteis para a correlação entre estrutura química e atividade biológica. Com PCA obteve-se boa separação entre os grupos de ativos e inativos. O método SIMCA não foi capaz de gerar um modelo satisfatório para compostos de baixa e média atividade, contudo produziu um bom modelo para os compostos ativos. Os métodos KNN e NN apresentaram muito bons resultados, com aproximadamente 70 % de acerto. Ao final destes processos novos compostos foram propostos com base nos modelos estatísticos.

Abstract

Theoretical Studies of the Relationships between Chemical Structure and Biological Activity of a series of Hexahydroindenopiridine Derivatives as Male Contraceptives, by Maximiliano Segala and Yuji Takahata.

Keywords: male contraceptives, quantum chemical calculations, SAR, PCA, KNN, NN.

In this work we investigated the relationships between chemical structure and biological activity of a series of 2,3,4,4a,5,9b-hexahydroindeno[1,2-c]pyridine derivatives which present antispermato-genic activity. Firstly, we determined the local minima conformations for the basic skeleton, by means of a sistematic conformation analysis. The minima obtained by this process was used as starting points for stocastic conformation analysis. In both case we used the MNDO-PM3 operator. The results were identical. The four local minima that were found were also optimized by the AM1, HF, and DFT methods, leading to the determination of the most stable conformation of the basic skeleton. It was employed as a starting conformation for building the derivatives. The geometric optimization of the derivatives was made by the use of the HF/3-21G, and the electronic properties were calculated by the methods HF/3-21G, HF/6-31G, and HAM/3. Steric properties, molecular volume, and the octanol/water partition coefficient were also calculated. These variables were explored with PCA, which revealed three variables (two atomic charges and the molecular volume) as being useful for the SAR. The PCA gave a good separation between the active and non-active groups. The SIMCA method was not able to generate a satisfactory model for the low- and medium-activity groups; however, it produced a good model for the actives group. The KNN and NN methods presented very good results, with approximately 70 % correct classifications. Finally new compounds were proposed based on the statistical models.

Currículo

Formação

- Universitária: Universidade Federal do Rio Grande do Sul; Bacharel em Química.

Trabalhos em Congressos

- Maximiliano Segala, Edilson Grünheid Borges e Yuji Takahata, Análise de SAR em Hexaidroindenopiridinas Antiespermatogênicas, 24^a Reunião Anual da Sociedade Brasileira de Química, em maio de 2001;
- Maximiliano Segala e Yuji Takahata, Estudo de SAR/QSAR de Hexaidroindenopiridinas Antiespermatogênicas, XXVI Congresso Internacional de Químicos Teóricos de Expressão Latina em setembro de 2000;
- Sandro Hillebrand, Maximiliano Segala e Valter Stefani, Síntese e Caracterização Fotofísica de um Novo Benzoxazolil Desenhado para Óptica Não Linear, 23^a Reunião Anual da Sociedade Brasileira de Química em maio de 2000;
- Maximiliano Segala, Dione S. Correa e Valter Stefani, Estudo da Fotofísica do 4-(2'-benzoxazolil)-2,5-diidroxibenzoato de etila por Métodos Espectroscópicos e Teóricos, 23^a Reunião Anual da Sociedade Brasileira de Química em maio de 2000;
- Maximiliano Segala e Yuji Takahata, Skeleton Conformational Analysis of Antispermatogetic Hexahydroindenopyridines, X Simpósio Brasileiro de Química Teórica em novembro de 1999;
- Sandro Hillebrand, Maximiliano Segala e Valter Stefani, Cálculos de Hiperpolarizabilidade de 1^a Ordem em 2-(2'-hidróxifenil)benzoxazóis, 22^a Reunião Anual da Sociedade Brasileira de Química em maio de 1999;

- Maximiliano Segala, Nei S. Domingues Jr., Paolo R. Livotto e Valter Stefani, Avaliação de Métodos Semi-empíricos para a Determinação da Fotofísica de Benzimidazóis Substituídos, 22^a Reunião Anual da Sociedade Brasileira de Química em maio de 1999;
- Maximiliano Segala, Nei S. Domingues Jr., Paolo R. Livotto e Valter Stefani, Estudo Teórico do Efeito de Substituintes em Benzoxazóis que apresentam Transferência Protônica Intramolecular no Estado Excitado, 21^a Reunião Anual da Sociedade Brasileira de Química em maio de 1998;
- Michele Schuch, Maximiliano Segala e Valter Stefani, Uma Nova Rota Sintética para a Preparação da 2,3-Diamino-1,4-Naftoquinona e seu Emprego na Obtenção de Heterociclos, 20^a Reunião Anual da Sociedade Brasileira de Química em maio de 1997.

Publicações

- Sandro Hillebrand, Maximiliano Segala, Tiago Buckup, Ricardo B. Correia, Flávio Horowitz e Valter Stefani, *First Hyperpolarizability in Proton-transfer Benzoxazoles: Computer-aided Design, Synthesis and Study of a New Model Compound*, *Chemical Physics*, aceito para publicação;
- Rogério Custodio, José Roberto dos Santos Politi, Maximiliano Segala, Roberto Luiz Andrade Haiduke, Quatro Alternativas para Resolver a Equação de Schrödinger para o Átomo de Hidrogênio, *Química Nova*, aceito para publicação;
- Maximiliano Segala, Nei S. Domingues Jr., Paolo R. Livotto e Valter Stefani, *Heterocyclic Dyes Displaying Excited-State Intramolecular Proton-Transfer Reactions (ESIPT): Computational Study of the Substitution Effect on the Electronic Absorption Spectra of 2-(2'-hydroxyphenyl)benzoxazole Derivatives*, *Journal of the Chemical Society Perkin Transactions 2*, 1999, (6), 1123.

Sumário

Lista de Tabelas	xiv
Lista de Figuras	xv
1 Introdução	1
1.1 Motivação do Projeto	1
1.2 Sistema Reprodutivo e Espermatogênese	4
1.3 Métodos Contraceptivos	6
1.4 Os Compostos Estudados	12
1.5 Objetivos	15
2 Materiais e Métodos	17
2.1 Materiais	17
2.1.1 Programas	17
2.1.2 Computadores	19
2.2 Métodos	19
2.2.1 Método Hartree-Fock	20
2.2.2 Métodos Semi-empíricos AM1 e PM3	23
2.2.3 Teoria do Funcional de Densidade (DFT)	25
2.2.4 <i>Hydrogenic Atoms in Molecules</i> versão 3 (HAM/3)	27
2.2.5 Somatório de contribuições atômicas (log P)	30
2.2.6 Método de Higo e Go para o Cálculo de Volume e Superfície Moleculares	30
2.2.7 Análise de Componentes Principais (PCA)	31
2.2.8 <i>Soft Independent Modeling of Class Analogy</i> (SIMCA)	32
2.2.9 K Vizinhos Mais Próximos (KNN)	33
2.2.10 Redes Neurais	34
3 Análises Conformacionais e Propriedades Moleculares	36
3.1 Trans/Cis	36
3.1.1 Análise Conformacional	36
3.1.2 Efeito da Base	41
3.1.3 Efeito de Solvente	41
3.2 Trans/Trans	42

3.3	Cis/Cis	42
3.4	Outros	44
3.4.1	Composto 8a	44
3.4.2	Composto 13	44
3.4.3	Composto 14	45
3.4.4	Composto 15	46
3.4.5	Compostos 16a e 16b	46
3.4.6	Composto 17	47
3.5	Parâmetros Físico-químicos	48
3.5.1	Variação com a Conformação	48
3.5.2	Variação com a Base	50
4	Correlação Estrutura Química-Atividade Biológica	53
4.1	Dados Experimentais	53
4.2	Seleção de Variáveis pelo Método de PCA	58
4.2.1	Variáveis dos Métodos HF/6-31G** e HAM/3	58
4.2.2	Variáveis do Método HF/3-21G	62
4.3	SIMCA	66
4.4	KNN	69
4.5	Redes Neurais	70
5	Novos Compostos	73
5.1	KNN	74
5.2	Redes Neurais	75
6	Conclusões	77
A	Programa de PCA	79
B	Programa em PERL	82
	Referências Bibliográficas	84
	Índice Remissivo	88

Lista de Tabelas

1.1	Relação dos compostos estudados e seus dados experimentais.	13
2.1	Quadro resumo dos materiais e métodos.	17
3.1	Normalização para os calores de formação e energias total.	40
3.2	Variação da geometria da cadeira 1 com a base.	41
3.3	Calores de formação para o composto 13	45
3.4	Calores de formação para os compostos 16a e 16b	47
3.5	Valores de PI e EA, em eV , para as conformações da família trans/cis.	49
3.6	Variação de μ com o método e a conformação.	49
3.7	Variação de volume e superfície molecular com a conformação.	50
4.1	Dados experimentais dos compostos estudados.	53
4.2	Correlação entre cargas de Mulliken do HAM/3 e do HF/3-21G.	62
4.3	Propriedades do HF/3-21G selecionadas pela PCA.	64
4.4	Estrutura e parâmetros das redes neurais.	71
4.5	<i>Leave one out</i> das redes neurais.	71
5.1	Proposta para novos compostos	74
5.2	Previsão das redes neurais para os compostos propostos.	76

Lista de Figuras

1.1	Relação histórica entre teoria e experimento.	2
1.2	Sistema reprodutivo masculino.	4
1.3	Etapas da espermatogênese.	5
1.4	Esqueleto das estruturas em estudo.	12
2.1	Descrição gráfica da PCA.	32
2.2	Descrição gráfica do SIMCA.	33
3.1	Esqueleto da família trans/cis.	37
3.2	Histogramas das conformações únicas.	38
3.3	Conformações da família trans/cis.	39
3.4	Sobreposição de estruturas otimizadas com diferentes bases.	42
3.5	Estrutura do composto 3c	43
3.6	Estrutura do composto 9b	43
3.7	Estrutura do composto 8a	44
3.8	Estrutura do composto 13	45
3.9	Estrutura do composto 14	46
3.10	Estrutura do composto 15	46
3.11	Estrutura dos compostos 16a e 16b	47
3.12	Estrutura do composto 17	48
3.13	Variação das cargas CHELPG com a base.	51
3.14	Variação das cargas Mulliken com a base.	51
4.1	PCA para os dados de 3–10 <i>mg/Kg</i>	56
4.2	Gráfico do log da dose pela resposta para o índice espermatogênico.	57
4.3	PCA com propriedades do HAM/3 e volume de van der Waals.	61
4.4	PCA com propriedades do HF/3-21G e volume de van der Waals.	63
4.5	Ampliação do grupo dos ativos.	65
4.6	Modelos de SIMCA para classe dos inativos.	67
4.7	Modelos de SIMCA para classe dos ativos.	67
4.8	Modelos de SIMCA para classe dos médios.	68
4.9	Modelo de KNN.	70
5.1	Esqueleto base dos novos compostos.	73

5.2	Substituintes em R_3 para os novos compostos.	74
5.3	Modelo de KNN para os compostos propostos.	75

Capítulo 1

Introdução

1.1 Motivação do Projeto

O processo de descoberta e desenvolvimento de uma droga eficiente é longo e árduo. A estatística industrial sugere que cerca de 10000 compostos são sintetizados e testados; 100 são avaliados por segurança; e até 10 compostos são testados clinicamente em humanos para cada droga que é aprovada para uso medicinal. O custo do desenvolvimento de uma nova droga envolve investimentos da ordem de US\$ 230 milhões e requer aproximadamente 12 anos de pesquisa, segundo informação divulgada pelo Centro de Pesquisa de Desenvolvimento de Drogas dos Estados Unidos da América (EUA). Cerca de 16 % do orçamento total do Banco Mundial para saúde, nutrição e planejamento populacional,¹ têm sido utilizados para empréstimos ou crédito para atividades farmacêuticas entre 1989 e 1995 [1]. Os gastos com desenvolvimento de novos produtos químicos têm requerido investimentos crescentes de recursos e muitos anos de pesquisa; desta forma, a necessidade de aumentar a eficiência do processo e reduzir os custos é vital. Uma das opções para tanto é a modelagem molecular com ajuda de computadores “*Computer Assisted Drug Design* (CADD)” [2]. Historicamente, a relação entre a teoria e a experimentação pode ser vista na Figura 1.1 [3].

Pode-se observar que até os anos 70 a teoria fornecia apenas idéias para a realização de experimentos, isso porque não era possível a realização de cálculos devido a pequena capacidade dos computadores da época. A partir dos anos 70, com o aumento da capacidade de cálculo dos computadores e o surgimento de métodos semi-empíricos mais razoáveis, possibilitaram o cálculo de propriedades antes só obtidas experimentalmente. Assim, os químicos teóricos forneceram mais uma ferramenta para a síntese planejada de novos produtos. Numerosos exemplos [4–6] demonstram que o CADD tem uma contribuição fundamental na tarefa de desenvolver novas drogas. Nosso grupo, por trabalhar nesta linha de pesquisa há cerca de sete anos, acredita que ela ainda apre-

¹aproximadamente US\$ 1,3 bilhões

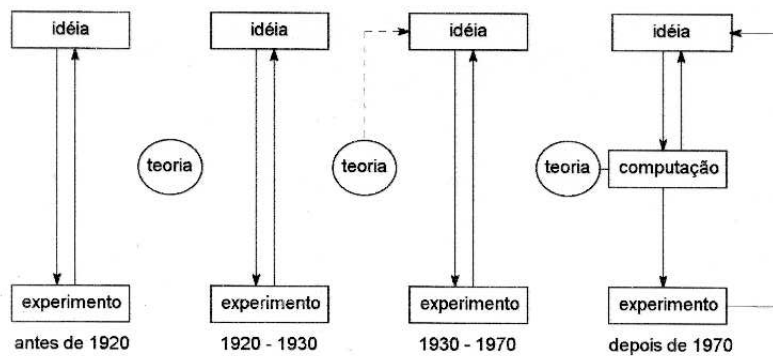


Figura 1.1: Relação histórica entre teoria e experimento [3].

sentam grandes possibilidades de desenvolvimento em virtude das necessidades do homem em obter respostas rápidas a um baixo custo. Existe uma grande variedade de assuntos de pesquisa nesta área interdisciplinar em que este tipo de abordagem pode ser aplicada [4–6].

Há cerca de 1000 anos, quando começou o segundo milênio, a população mundial era menos que 300 milhões de habitantes, com uma expectativa de vida de aproximadamente 30 anos. Na década de 50 esses dados apontavam uma população de 2500 milhões e uma expectativa de vida de 46 anos. No ano 2000, quando da entrada do terceiro milênio, a população mundial é projetada em mais de 6200 milhões de habitantes, com uma expectativa de vida de 68 anos, ou seja, a população do planeta aumentou mais de 20 vezes em um século. Como vai ficar a população no futuro? A projeção da população feita pelas Nações Unidas para daqui a 20 anos é de 7100 milhões. A cada ano que passa, a população mundial é acrescida de 86 milhões de pessoas.

O conteúdo do artigo “Unicef faz sombrio balanço do subdesenvolvimento no mundo”, de dezembro de 93 [7], ainda tem atualidade. Nele, o Unicef relata que o crescimento demográfico agrava a espiral pobreza/pressão demográfica/meio ambiente (PPM). Em realidade, o efeito do crescimento demográfico ameaça não só o meio ambiente, mas também o próprio futuro da humanidade. Dados indicam que a falta de água potável, alimentos, recursos naturais, moradias, empregos, escolas é iminente. Mesmo em países desenvolvidos o problema da fome já é muito grave. Cerca de 31 milhões de norte-americanos não podem comprar a comida que precisam para se alimentar, segundo o relatório do *Bread for the World Institute*, em Washington. Baseado em estatísticas do Departamento de Agricultura, o estudo afirma que 3,7 milhões de famílias norte-americanas (3,6 %) passaram fome em 1998, enquanto 19 milhões de adultos e 12 milhões de crianças nem sempre puderam comprar a comida necessária. O relatório, intitulado “Programa para Terminar com a fome: Fome 2000”, afirma que o número de pessoas que estão padecendo de fome poderia se reduzir a metade, 16 milhões, com aproximadamente U\$ 5 bilhões por ano, ou U\$ 18 para cada cidadão do país [8]. Certamente esse problema não é puramente relacionado com a superpopulação, mas já revela um sério problema

social justamente na nação mais rica do mundo.

Resultados preliminares do censo 2000 indicam que o Brasil tem 167 milhões de habitantes e no ano 2020 deverá ter 208 milhões. A queda da natalidade e o consequente envelhecimento da população brasileira são as primeiras constatações do Censo 2000. A taxa de crescimento populacional, de 3 % nos anos 50 e 60, caiu no ano 2000 para 1,4 %. O número de brasileiros com idade superior a 60 anos deve chegar a 14 milhões, o dobro do registrado há 20 anos. São algarismos preocupantes para o Ministério da Previdência. O Censo indica também predominância das mulheres em 25 dos 27 estados, sendo mais acentuada nos centros urbanos [9]. Esse aumento da proporção de mulheres, principalmente nos centros urbanos, é decorrente do aumento na mortalidade de homens, em especial jovens, decorrentes de traumas.²

No futuro, é muito provável que o nível de vida sócio-econômico, tanto dos brasileiro como de estrangeiros, tenda a baixar gradativamente. O desenvolvimento econômico através da expansão da produção de alimentos e artigos de utilidades talvez eleve o nível de vida a curto prazo, porém o aumento de produtividade e desenvolvimento econômico será consumido imediatamente pela pressão da expansão demográfica. Deixando-se a situação como está, sem qualquer atitude, a espiral PPM vai se intensificar, o que pode levar, daqui a 100 ou 200 anos, à inabitabilidade do planeta.

Uma das maneiras de evitar esta tragédia anunciada é frear o ritmo de crescimento demográfico e estabilizar a população dentro de um limite suportável. A Organização Mundial de Saúde “*World Health Organization (WHO)*” está apoiando o desenvolvimento de métodos contraceptivos que sejam mais eficientes, seguros e de fácil manipulação [10]. Já existem programas de controle de natalidade em diversos países, mesmo assim, a população mundial continua crescendo a um ritmo recorde. Não se têm espaço aqui para entrar em detalhes quanto a discussão de que tipos de programas de controle de natalidade são ideais, contudo sabem-se que métodos de controle de natalidade através de contraceptivos orais (COs) têm apresentado excelentes resultados. Cerca de 60 milhões de mulheres utilizam contraceptivos orais no mundo, no entanto este tipo de modelo para homens ainda não tem produtos comerciais [11]. Apesar de muitos esforços feitos por pesquisadores [12,13] os contraceptivos orais masculinos ainda não estão bem desenvolvidos, isto porque eles envolvem dificuldades maiores que os femininos [14,15] como o fato da espermatogênese ser um processo contínuo, complexo, onde há falta de entendimento de muitas etapas envolvidas. A princípio, os métodos contraceptivos mais utilizados pelos homens são os de barreira, como é o caso do preservativo e da vasectomia, esta muitas vezes irreversível.

Independentemente dos dados apresentados, é clara a necessidade de um efetivo controle populacional, não só pelos nocivos efeitos sociais, mas também ambientais. Para tanto, a utilização de modelagem molecular para o desenvolvimento de contraceptivos orais nos parece ser a melhor proposta.

²mortes provocadas pela violência urbana

1.2 Sistema Reprodutivo e Espermatogênese

Os principais órgãos do sistema reprodutivo masculino são os testículos, o epidídimo, os canais deferentes, as vesículas seminais, a próstata e o pênis. A espermatogênese ocorre nos testículos, onde o espermatozoide se diferencia a partir de espermatogônias, que se multiplicam por etapas sucessivas formando espermatócitos primários, espermatócitos secundários e, finalmente, espermátides, que maturam espermatozoides.

Os testículos são o lugar onde é formado o espermatozoide e é secretada a testosterona. O epidídimo é um sistema de tubos enovelados por onde passam os espermatozoides durante a sua maturação, como será apresentado melhor na descrição da espermatogênese. Os canais deferentes são onde o espermatozoide fica armazenado, sendo também condutores do espermatozoide do epidídimo para a uretra; a parte superior desses canais é dilatada, formando a âmpula do canal deferente, onde uma quantidade adicional de espermatozoide fica armazenada antes da ejaculação. A vesícula seminal é uma glândula tubular que secreta o líquido seminal na parte superior do canal deferente durante o ato sexual. A próstata é uma grande massa glandular que circunda a uretra no ponto em que sai do colo da bexiga e emite o líquido prostático, também durante o ato sexual. O pênis é o órgão sexual externo. Todo esse sistema pode esquematicamente ser visto na Figura 1.2.

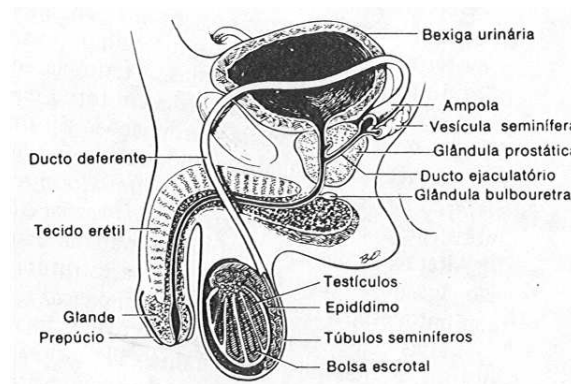


Figura 1.2: Sistema reprodutivo masculino [16].

A espermatogênese é um processo complexo que ocorre em uma série de etapas nos testículos. Para tanto, cada testículo é formado por cerca de 900 tubos seminíferos, onde no interior desses tubos existe um grande número de células chamadas espermatogônias ou células germinais masculinas, por ser delas que se derivam todos os espermatozoides. Como mencionado brevemente acima, a primeira etapa da espermatogênese é a formação dos espermatócitos primários, a partir das espermatogônias, cada um dos espermatócitos primários forma dois espermatócitos secundários que, por sua vez, dão origem, cada um, a duas espermátides. É justamente no intervalo das etapas de espermatócito primário a secundário que os 23 pares de cromossomos da espermatogônia se dividem para formar

23 cromossomos individuais. É a espermátide então que matura a espermatozóide. Situadas entre as células germinativas dos tubos seminíferos, existem as células de Sertoli. A superfície destas células envolve os espermatócitos, as espermátides e os espermatozóides em maturação; desta forma, ela desempenha um papel vital no processo da espermatogênese. Em qualquer localização dos tubos seminíferos uma ou mais gerações de espermatogônias, espermatócitos e espermátides estão presentes. Como resultado disso a espermatogênese é um processo contínuo, o que dificulta sua manipulação para efeitos contraceptivos [17]. Esse processo de formação leva cerca de dez semanas e suas etapas são apresentadas na Figura 1.3.

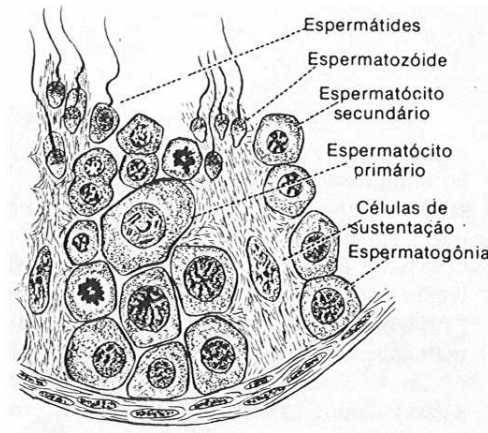


Figura 1.3: Etapas da espermatogênese [16].

Após a formação nos tubos seminíferos, o espermatozoide começa sua viagem pelos cerca de seis metros do epidídimo, o que em geral leva alguns dias. É durante essa viagem que ocorre o processo de maturação do espermatozoide, no qual ele adquire a capacidade de fecundar o óvulo. Parte do espermatozoide permanecerá armazenado no epidídimo até o momento da ejaculação, mas outra parte passa para o canal deferente e para a âmpula do canal deferente onde vai ficar armazenada. Este processo leva cerca de 2 semanas e, nessas condições de armazenamento, o espermatozoide pode ficar fértil por até um mês.

Todos esses processos são controlados por um complexo sistema hormonal, do qual as principais etapas são apresentadas a seguir. O hormônio sexual masculino, a testosterona, é secretado pelas células intersticiais de Leydig, nos testículos. Tanto a formação do espermatozoide como a secreção de testosterona pelos testículos são controladas por hormônios secretados pela hipófise anterior: o hormônio foliculoestimulante (FSH), que é o responsável principal pelo início da espermatogênese, e o hormônio luteinizante (LH), que é o estímulo primordial para a secreção de testosterona [16].

1.3 Métodos Contraceptivos

O planejamento familiar através do uso de métodos contraceptivos já é utilizado por mais da metade das famílias em idade reprodutiva em países do primeiro mundo. Os mais utilizados atualmente são:

- Amenorréia láctea;

Amenorréia láctea é o uso da amamentação como método contraceptivo, baseado no efeito fisiológico da amamentação para suprimir a ovulação. O uso efetivo da amamentação como um método contraceptivo requer que a mãe alimente seu filho com grande frequência; além disso, a criança precisa ter menos de seis meses de idade e a menstruação não pode ter retornado.

- Coito interrompido;

O coito interrompido é um método de planejamento familiar no qual retira-se completamente o pênis do canal vaginal, antes da ejaculação. Este método é extremamente falho, pois antes da ejaculação o homem elimina fluidos (Seção 1.2), os quais podem conter pequena quantidade de espermatozóides.

- Combinados injetáveis femininos;

Os combinados injetáveis são injeções de hormônios estrogênio e progesterona, os quais são administrados geralmente uma vez por mês. Os hormônios suprimem a ovulação, misturam-se ao muco cervical impedindo a penetração do espermatozoide e alteram o endométrio para dificultar a passagem do espermatozoide para as trompas de falópio.

- Combinados orais femininos;

Os contraceptivos orais são pílulas as quais contêm hormônios estrogênicos e progestagênicos. Elas são tomadas diariamente e produzem os mesmos resultados dos injetáveis, como supressão da ovulação, mistura ao muco cervical e alteração no endométrio.

- Diafragma;

Um diafragma é um copo em forma de abóboda feito de látex, o qual é inserido na vagina antes da cópula e serve para tampar o cérvix.³ Ele impede que o espermatozoide tenha acesso ao canal cervical e conseqüentemente ao útero e às trompas. Serve também como suporte para espermicidas.

³região delimitadora entre a vagina e o canal cervical

- Dispositivo intra-uterino;

O dispositivo intra-uterino (DIU) é um pequeno objeto flexível em forma de **T**, o qual é inserido dentro da cavidade uterina. Os DIU podem ser inertes, liberadores de cobre, ou liberadores de progesterona. Os liberadores de cobre interferem na habilidade do esperma passar pela cavidade uterina e com o processo reprodutivo antes do óvulo atingir a cavidade uterina. Os liberadores de progesterona também misturam-se com o muco cervical e mudam o tecido endometrial.

- Espermicidas;

Espermicidas são produtos químicos (usualmente nonoxinol-9) que inativam ou matam o esperma. Eles são disponíveis em aerossóis (espuma), cremes, tabletes vaginais, supositórios e filmes dissolvíveis, sendo frequentemente utilizados em conjunto com métodos de barreira com o preservativo masculino. Os espermicidas causam a ruptura da membrana celular do esperma, o que diminui sua movimentação e habilidade de fertilizar o óvulo.

- Oclusão tubular;

A oclusão tubular é um procedimento cirúrgico para terminar com a fertilidade da mulher. Ela pode ser feita por minilaparatomia ou laparoscopia e bloqueia as trompas de falópio (amarrando e cortando, ou por meio de anéis, cliques ou eletrocauterização) evitando assim que o esperma chegue até o óvulo.

- Planejamento familiar;

O planejamento familiar consiste em evitar a cópula durante a fase fértil do ciclo menstrual. Existem quatro tipos de planejamento familiar: método do calendário, temperatura corporal basal (TCB), muco cervical e associação de TCB e sintomas sugestivos de ovulação.

- Preservativo feminino;

O preservativo feminino é uma fina capa feita de poliuretano com anéis nas bases. Ele é inserido na vagina e, como o masculino, impede o acesso do esperma ao sistema reprodutivo feminino, bem como evita que microorganismos passem de um parceiro para o outro.

- Preservativo masculino;

O preservativo masculino é uma fina capa feita de borracha, vinil ou produtos naturais, a qual é colocada no pênis quando ereto. Este preservativo pode ser tratado com espermicidas para maior proteção. Esta barreira impede que o esperma tenha acesso ao sistema reprodutivo

feminino, assim como impede que microorganismos (como HIV) passem de um parceiro para o outro.⁴

- Vasectomia.

A vasectomia é um procedimento cirúrgico que termina, permanentemente, em muito casos, com a fertilidade masculina. Ela pode ser feita por métodos padrão ou por técnicas sem cortes, as quais são preferidas. A vasectomia bloqueia o canal deferente (Figura 1.2), impedindo que o espermatozoide esteja presente no ejaculado.

Estudo recente, realizado ao longo de cinco anos, comparou quinze métodos contraceptivos (coito interrompido, combinados orais femininos, combinados injetáveis femininos, diafragma, DIU liberador de cobre, DIU liberador de progesterona, espermicidas, implantes, método do calendário, oclusão tubular, preservativo feminino, preservativo masculino, vasectomia, esponja e capa cervical) através de modelos econômicos, e constatou que os contraceptivos orais têm a melhor relação custo/benefício que todos os outros métodos, a exceção do DIU, vasectomia e implantes. Entretanto, estes últimos três métodos apresentam contra-indicações, como o custo financeiro [18], além disso a vasectomia muitas vezes é irreversível e pode estar correlacionada com doenças cardiovasculares [19] e urolitíases [20].

Como visto na Seção 1.1, os contraceptivos femininos são largamente utilizados, sendo desenvolvido desde a década de 40. Em 1957, o *Food and Drug Administration* (FDA) já havia liberado para uso clínico a noretindrona e o noretinodrel, dois dos compostos pioneiros em contracepção feminina, para tratamento de problemas de infertilidade e distúrbios menstruais. Até hoje a noretindrona é utilizada em medicamentos comerciais. Mesmo após várias gerações de desenvolvimento, os contraceptivos femininos ainda apresentam efeitos colaterais indesejáveis [21]. Comercialmente, métodos contraceptivos para mulheres têm sido disponibilizados desde os anos 60, isso se deve em parte ao fato da contracepção feminina ser baseada no mecanismo natural que evita a ovulação durante a gravidez. Já a espermatogênese é um processo contínuo, complexo, onde há falta de entendimento de muitas etapas envolvidas, sendo esse o grande responsável pela falta de contraceptivos masculinos. Além disso, há o fator sócio-econômico, pois ainda existe dúvida sobre se os homens estariam interessados em drogas desta natureza. Com o aumento na ênfase em conceitos como igualdade, compartilhamento de responsabilidade para com a contracepção criou-se um aumento de demanda e uma maior aceitação para contraceptivos masculinos.

Maior progresso tem sido obtido com métodos hormonais, todavia esse procedimento leva à necessidade de complementação de outros hormônios [22], além do processo ser ainda pouco conhecido. Métodos não hormonais podem agir em qualquer ponto ao longo do processo da espermatogênese,

⁴somente para preservativos feitos de latex ou vinil

transporte, maturação e interação com o óvulo. Uma incrível variedade de compostos têm apresentado efeitos contraceptivos, sendo muitos desses produtos naturais ou medicamentos feitos com outros propósitos. Uma coletânea destes compostos [15,23] é apresentada abaixo por classes:

- Análogos do GnRH;

- Cetorelix e Nal-Glu.

Utilizados no combate ao câncer de próstata eles não apresentam atividade quando administrados por via oral. Requerem co-administração de androgênio e os estudos estão sendo realizados em humanos.

- Antibióticos;

- Ornidazole;

Apresenta antifertilização por mecanismo ainda desconhecido, além disso tem atividade antimicrobial. Os testes estão sendo feitos em animais.

- Pirimetamina e seus derivados Etoprino e a Metoprino;

Atuam por inibição enzimática da diidrofoloreductase. Os estudos estão sendo realizados em animais. A Pirimetamina apresenta fraca atividade, contudo seus derivados (por mínimas modificações) apresentam boa atividade.

- RS-90847;

Altera a mobilidade do espermatozoide, sendo os testes realizados *in vitro*.

- Sulfapiridinas e Sulfasalazina.

Sulfonamidas utilizadas no combate de algumas infecções bacterianas. A sulfapiridina precisa de altas doses para apresentar modesta atividade, e estudos estão sendo feitos em animais. A sulfasalazina é utilizada no tratamento de inflamações ulcerativas de grande intensidade e na doença de Crohn. Apresenta baixa atividade e muitos efeitos colaterais, sendo os testes realizados em humanos.

- Canal de cálcio;

- Fenoxibenzamina;

Atua na diminuição da capacidade de ejaculação, por um processo reversível. É utilizado também como anti-hipertensivo, sendo todos os testes realizados em animais.

- Nifedipina e Verapamil;

Além da atividade antiesperma atuam como bloqueadores de canais de cálcio e anti-hipertensivos. Os testes estão sendo realizados em humanos.

- Tamulosin.

Interfere na ejaculação, sendo a administração por via parenteral e os testes realizados em animais.

- Esteróides;

- Acetato de ciproterona;

Administrado oralmente, requer co-administração de androgênio. Seus estudos já estão sendo feitos em humanos.

- Acetato de Medroxiprogesterona;

Hormônio administrado por via parenteral, requer co-administração de androgênio. Utilizado com contraceptivo feminino, e os testes realizados em humanos.

- Ésteres de testosterona;

Hormônio administrado por via parenteral, cujos estudos são feitos em humanos.

- ICI 182,780;

Anti-hormônio utilizando inicialmente no combate ao câncer. Seus efeitos colaterais para uso prolongado ainda são desconhecidos, e os estudos são feitos em animais.

- Levonorgestrel;

Hormônio administrado por via oral ou parenteral que requer co-administração de androgênio. Inicialmente utilizado como contraceptivo feminino, seus estudos são realizados em humanos.

- Mifepristona;

Anti-hormônio, com atividade antiprogesterônica e antiglicocorticóide, utilizado como abortivo, com eficácia e mecanismos incertos.

- 7α -metil-19-nortestosterona.

Hormônio administrado por via parenteral com teste em humanos.

- Miscelâneos;

- 1-formil-4-dicloroacetamidopiperazina;

Atua como antiespermatogênico, sendo que outras dicloroacetilamidas têm efeitos sobre o fígado. Os ensaios são feitos em animais.

- Flunixin;

Utilizado também como anti-inflamatório e analgésico, ele apresenta múltiplos efeitos, atuando de forma geral nos testículos diminuindo o níveis hormonais. Os estudos são feitos em animais.

- 1,6-dicloro-1,6-dideoxi-D-frutose e 6-cloro-6-deoxi-D-glucose;

Pro-droga para o gliceraldeído-3-fosfato inibidor de deidrogenase. Seus produtos de clivagem são neurotóxicos, e os ensaios estão sendo feitos em animais.

- Carbaril;

Provoca a diminuição na contagem e mobilidade do esperma, e é utilizado também como inseticida. Seu principal problema é a toxicidade, e os testes são realizados em humanos.

- DL-111;

Atua com inibidor da biosíntese de androgênio. É improvável que possa ser útil, devido ao seu mecanismo de ação. Testes em animais.

- Tolnidamine;

Composto antiespermatogênico irreversível. Testes em animais.

- Bis-triclorometil sulfona.

Atua na mobilidade do esperma, sendo reversível e com moderada toxicidade para estudos em animais.

- Produtos naturais.

- Ácido oleanólico e Solasodina;

Antiespermatogênicos de eficácia e mecanismos incertos, com testes realizados em animais.

- Aloin, Hidróxido de palmatina e 5,7,3'-Trihidroxi-6,8,4'-trimetoxi flavona;

Antiespermatogênicos com testes realizados em animais. O Aloin afeta as células de Leydig reversivelmente. Já o hidróxido de palmatina, além de afetar as células de Leydig, tem baixa atividade. A flavona afeta as últimas etapas da espermatogênese.

- Gossipol;

Antiespermatogênico e potencialmente anti-câncer, sendo o principal problema sua toxicidade. Os testes são realizados em larga escala em humanos.

- Triptolidio.

Atua na redução da mobilidade do esperma. É utilizado na medicina herbária chinesa para o tratamento de inflamações e problemas dermatológicos.

Além destes, têm-se as Hexaidroindenopiridinas, as quais são o assunto deste trabalho e são melhor tratada na Seção 1.4.

1.4 Os Compostos Estudados

Ebenother et al. sintetizaram o derivado (4aRS,5SR,9bRS)-2-etil-2,3,4,4a,5,9b-hexaidro-7-metil-5-p-toluil-1H-indeno[1,2-c]piridinas (SANDOZ 20-438, **10a**), o qual apresentou antagonismo frente à serotonina e atividade analgésica. O estudo de toxicidade mostrou forte atividade antiespermatogênica em cães, com efeitos similares também para ratos. O composto mostrou-se não tóxicos nem mutagênicos. Uma única dose de 30 *mg/kg* para ratos produz efeito no teste de massa dos testículos dentro de 24 horas. Mudanças degenerativas nos tubos seminíferos são observadas. As espermátides aproximam-se, ocasionalmente formando associações multinucleadas. As células de Sertoli permanecem dentro de um controle normal. O LH e o FSH em ratos tratados cronicamente aumenta sensivelmente, indicando que o composto não atua por supressão de gonadotrofina. As funções das células de Leydig não são afetadas, uma vez que não há redução de massa da vesícula seminal. Administração crônica para cães induziu completa inibição da espermatogênese, sem sinal de toxicidade em geral. As capacidades de ereção e ejaculação não foram afetadas, apesar da contagem de espermatozóides ter caído a zero. A retirada da droga é seguida pelo retorno a níveis normais da contagem de espermatozóides e aparência histológica normal passadas, de oito a onze semanas [24].

Cook et al. [14, 25] apresentam a síntese para uma série de análogos do 2,3,4,4a,5,9b-hexaidro-indeno[1,2-c]piridinas (Figura 1.4 e Tabela 1.1) e seu comportamento frente a testes biológicos em ratos quanto a redução do índice espermatogênico (IE) de massa testicular (MT).

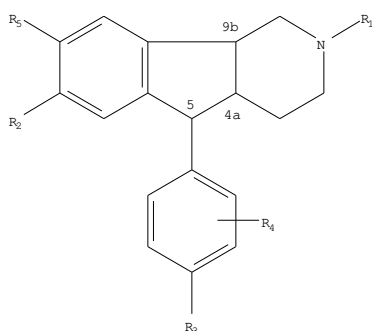


Figura 1.4: Esqueleto das estruturas em estudo.

A série de compostos apresentou efeito em ambos os testes, sendo o índice espermatogênico utilizado para estimar semi-quantitativamente a habilidade na produção de esperma, baseado na observação histológica das células espermatogênicas nos tubos seminíferos através de uma ou mais secções de corte [26]. É observado que a atividade é fortemente influenciada pelo estereoisomerismo e quiralidade dos compostos, o que indica um mecanismo de ação envolvendo interações com macromoléculas específicas. Foi também observado que substituições em posições específicas como

no nitrogênio e, mais fortemente, na posição *para* do anel 5-aril são mais eficientes. A natureza dos substituintes é também de vital importância para a atividade do composto. Os resultados demonstram que os derivados do referido composto são agentes antiespermatogênicos não hormonais muito eficientes. Testes toxicológicos foram reportados recentemente [27] para o RTI 4587-056 (**12b**) o qual mostrou-se um contraceptivo muito eficiente sem efeitos colaterais observáveis. A Tabela 1.1 apresenta as substituições feitas no esqueleto base (Figura 1.4), além dos dados experimentais para os derivados.

Tabela 1.1: Relação dos compostos estudados e seus dados experimentais [14, 25].

nome	R ₁	R ₂	R ₃	R ₄	R ₅	4a,5/4a,9	dose ¹	MT ²	IE ³
2a ⁴	Et	Me	COOH	H	I	trans/cis	2	-	67
							6	-	66
							20	-	76
2b ⁴	Et	Me	COOMe	H	I	trans/cis	2	-	57
							6	-	69
							20	-	74
<i>(l)</i> 2b ⁴	Et	Me	COOMe	H	I	trans/cis	0,6	-	34
							2	-	66
							6	-	71
3a ⁴	Et	Me	COOH	H	Br	trans/cis	19	-	72
							6	-	69
							21	-	71
3c	Et	Me	Me	H	H	trans/trans	62	-	72
4a ⁴	Et	Me	COOH	H	Cl	trans/cis	90	-18	0
							2	-	55
							7	-	66
8a	Et	Me	Me	H	H	dupla 4a-5	25	-	72
9a	Et	Me	Me	H	H	cis/cis	90	5	-9
9b	Me	Me	Me	H	H	cis/cis	90	2	-9
10a	Et	Me	Me	H	H	trans/cis	30	12	7
							3	10	5
							10	10	31
<i>(d)</i> 10a	Et	Me	Me	H	H	trans/cis	30	24	50
							1	11	7
							3	13	31
<i>(l)</i> 10a	Et	Me	Me	H	H	trans/cis	10	25	59
							30	43	62
							30	10	9
10b	Me	Me	Me	H	H	trans/cis	90	4	-9
							3	8	-3

continua na próxima página

Tabela 1.1: Relação dos compostos estudados e seus dados experimentais (continuação).

nome	R ₁	R ₂	R ₃	R ₄	R ₅	4a,5/4a,9	dose ¹	MT ²	IE ³
							10	7	26
							30	18	38
10c	Et	Me	H	o-Me	H	trans/cis	90	3	-9
10d	Et	Me	H	m-Me	H	trans/cis	90	-11	-9
10e	Et	Me	F	H	H	trans/cis	90	5	-5
10f	Et	Me	OMe	H	H	trans/cis	3	-7	2
							10	-17	-6
							90	5	-5
10g	Et	H	H	H	H	trans/cis	90	-2	2
10h	Et	Me	H	H	H	trans/cis	90	-8	-9
10i	Et	H	Me	H	H	trans/cis	10	5	0
							30	3	20
							90	20	35
10j	Et	Me	Me	Me(2'-)	H	trans/cis	10	1	0
							30	6	3
							90	11	30
10k	Et	Me	COOH	H	H	trans/cis	3	13	2
							10	24	36
							30	28	51
<i>(d)</i> 10k	Et	Me	COOH	H	H	trans/cis	3	-5	21
							10	26	40
							30	16	67
							90	27	70
<i>(l)</i> 10k	Et	Me	COOH	H	H	trans/cis	90	-7	0
10l	Et	Et	Me	H	H	trans/cis	3	3	-2
							10	2	7
							30	26	33
10m	Et	Et	COOMe	H	H	trans/cis	3	-18	4
							10	0	-4
							30	8	23
10o	Et	Me	OH	H	H	trans/cis	30	-1	-5
11a	H	Me	Me	H	H	trans/cis	90	-3	-5
11b	n-Pr	Me	Me	H	H	trans/cis	30	11	0
11c	i-Pr	Me	Me	H	H	trans/cis	10	7	10
							30	15	28
							90	23	45
12b	Et	Me	COOMe	H	H	trans/cis	3	5	5
							10	17	27
							30	32	57

continua na próxima página

Tabela 1.1: Relação dos compostos estudados e seus dados experimentais (continuação).

nome	R ₁	R ₂	R ₃	R ₄	R ₅	4a,5/4a,9	dose ¹	MT ²	IE ³
(<i>d</i>) 12b	Et	Me	COOMe	H	H	trans/cis	90	3	-5
(<i>l</i>) 12b	Et	Me	COOMe	H	H	trans/cis	3	11	28
							10	26	47
							30	30	61
							90	34	68
12c	Et	Me	COO-n-Pr	H	H	trans/cis	3	11	2
							10	3	33
							30	31	56
12d	Et	Me	CN	H	H	trans/cis	30	13	-4
12e	Et	Me	CONH ₂	H	H	trans/cis	30	-7	-2
12f	Et	Me	CONMe ₂	H	H	trans/cis	30	-3	3
12g	Et	Me	CH ₂ OH	H	H	trans/cis	3	4	-5
							10	13	20
							30	23	57
12h	Et	Me	CH ₂ CONHMe	H	H	trans/cis	30	7	8
12i	Et	Me	CHO	H	H	trans/cis	3	7	0
							10	6	2
							30	7	31
13	(Figura 3.8)						3	2	2
							30	9	-5
14	(Figura 3.9)						90	-5	0
15	(Figura 3.10)						90	-21	0
16a	(Figura 3.11)						90	-9	0
16b	(Figura 3.11)						90	0	0
17	(Figura 3.12)						30	0	2

¹ mg de composto por Kg de rato (*mg/Kg*)² massa dos testículos (% redução em relação ao controle)³ índice espermato gênico (% redução em relação ao controle)⁴ μ g de composto por Kg de rato (μ g/*Kg*)

1.5 Objetivos

Relações estrutura química-atividade biológica “*Structure-Activity Relationships* (SAR)”, foram estudadas teoricamente com uma série de derivados 2,3,4,4a,5,9b-hexaidroindeno[1,2-c]piridinas a fim de procurar quais propriedades moleculares estão correlacionadas com a atividade biológica, no caso, índice espermato gênico realizado por Cook [14,25] em ratos. Baseados nos resultados obtidos, formularam-se modelos estatísticos para SAR e os aplicaram-se para propor novos compostos ainda não sintetizados potencialmente ativos.

Este trabalho é uma das peças necessárias para atingirmos nosso objetivo final, que é obter contraceptivos masculinos ideais. Contraceptivos ideais devem apresentar uma atividade expressiva com pequenas doses, atuar em um processo de esterilização reversível, sem efeitos colaterais e com longa duração. Adicionalmente, é desejado que as doses sejam de fácil administração, como por via oral [28].

Capítulo 2

Materiais e Métodos

A Tabela 2.1 apresenta um quadro resumo dos recursos utilizados neste trabalho.

Tabela 2.1: Quadro resumo dos materiais e métodos.		
programa	plataformas	métodos utilizados
CADPAC	Cray T-94	DFT
GAMESS	IBM, PC-UNIX e SGI	HF e PCM
GIMP	PC-UNIX	tratamento de imagens
Grace	PC-UNIX	produção de gráficos 2D
HAM/3	PC-UNIX e SGI	CI, EA, ESCA e PES
MATLAB	IBM	KNN e SIMCA
MOLDEN	PC-UNIX	montagem e visualização de estruturas
MOPAC	PC-UNIX e SGI	AM1, PCM e PM3
Scilab	PC-UNIX	seleção de variáveis e PCA
Spectrum	PC-UNIX	espectros de UV
Surf	PC-UNIX	volumes e superfícies moleculares
PSDD	PC-UNIX	redes neurais
PERL	PC-UNIX	programas para a manipulação de dados
TINKER	PC-UNIX	distâncias geométricas e sobreposição
XLOGP	PC-UNIX	coeficiente de partição octanol/água

O objetivo da Tabela 2.1 é apresentar uma visão geral dos recursos, sendo esses devidamente apresentados nas seções 2.1 e 2.2.

2.1 Materiais

2.1.1 Programas

A seguir é apresentada uma pequena descrição de cada programa utilizado e a forma de obtenção.

- CADPAC: *The Cambridge Analytic Derivatives Package* [29] versão 6.3, como parte integrante do UniChem [30] versão 4.1 licenciado para o Centro Nacional de Supercomputação (CESUP) [31];
- GAMESS: *The General Atomic and Molecular Electronic Structure System* [32] versões de 18 março de 1997, 1 dezembro de 1998 e 25 março de 2000, livremente distribuído para o meio acadêmico pelo grupo de pesquisa do professor Gordon da *Iowa State University*, USA;
- GIMP: *The GNU Image Manipulation Program* [33] versão 1.1.10, livremente distribuído pela *Free Software Foundation* [34];
- Grace: Grace [35] versão 5.1.2, livremente distribuído pelo grupo de desenvolvedores;
- HAM/3: *Hydrogenic Atoms in Molecules* versão 3, comprado do *Quantum Chemistry Program Exchange* (QCPE) [36] do Departamento de Química da *Indiana University*, USA;
- MATLAB: *MATrix LABoratory* [37] versão 5.3.1.29215a (R11.1), licenciado para o Centro Nacional de Processamento de Alto Desempenho em São Paulo (CENAPAD) [38];
- MOLDEN: *MOlecular DENsity* [39] versão 3.6, livremente distribuído pelo autor;
- MOPAC: *Molecular Orbital PACkage* [40] versão 93, licenciado para CESUP;
- Scilab: *Scientific software package for numerical computations in a user-friendly environment* [41] versão 2.5, livremente distribuído pelo *Institut National de Recherche en Informatique et en Automatique* (INRIA). Sob esta plataforma desenvolveram-se programas para seleção de variáveis e análise de componentes principais. Para exemplificar, trechos do programa para PCA são apresentados no Apêndice A.
- Spectrum: versão de 1997, livremente distribuído pela Central de Pesquisa e Desenvolvimento da DuPont, USA;
- Surf: *Molecular Surface Area and Volume Calculation Software* [42] versão 1.2, desenvolvido pelo grupo;
- PSDD: *Perceptron-type Neural Network Simulator* versão de novembro de 1991, comprado do QCPE;
- PERL: *Practical Extraction and Report Language* [43] versão 5. Sob esta plataforma desenvolveram-se programas para a produção de arquivos de entrada e a retirada de propriedades calculadas com o GAMESS, HAM/3, MOPAC, Spectrum e Surf. Para exemplificar, trechos do programa desenvolvido para o GAMESS são apresentados no Apêndice B.

- TINKER: *Software Tools for Molecular Design* [44] versão 3.8, livremente distribuído por Jay William Ponder, sendo o desenvolvimento mantido pela *National Science Foundation* e pelo *United States National Institutes of Health*, USA;
- XLOGP: É um programa desenvolvido pelo Professor Renxiao Wang [45], versão 2.0, livremente distribuído pelo ftp da *Pekin University*, China.¹

2.1.2 Computadores

Abaixo é apresentada brevemente cada plataforma utilizada neste trabalho:

- Cray T-94: Supercomputador Cray modelo T-94 equipado com duas CPU de 1,8 *GFlop*, 2 *Gb* de memória RAM, 53 *Gb* de disco SCSI e operando com sistema UNICOS 10.0. A máquina está instalada no CESUP;
- IBM: Indica o ambiente do CENAPAD, o qual é baseado em máquinas RISC e sistema operacional Unix. São 59 máquinas, totalizando uma capacidade de processamento em torno de 99,46 *GFlop*, 137,5 *Gb* de memória RAM e 330 *Gb* de disco externo;
- PC-UNIX: Microcomputador Intel com duas CPU de 700 *MHz*, 512 *Mb* de memória RAM, 27 *Gb* de disco SCSI e operando com FreeBSD 4.2 ou Microcomputador Intel com uma CPU de 500 *MHz*, 128 *Mb* de memória RAM, 12 *Gb* de disco e operando com FreeBSD 3.5;
- SGI: Servidor *Silicon Graphics Inc* modelo Origin 200 com quatro CPU 180 *Mhz* RISC R10000, 256 *Mb* de memória RAM e 12 *Gb* de disco ou estações de trabalho modelo Octane com duas CPU 175 *MHz* MIPS RISC R10000, 128 *Mb* de memória RAM e 4 *Gb* de disco. Todas as máquinas operam com sistema IRIX 6.5 e estão instaladas no CESUP.

2.2 Métodos

Começou-se com os métodos quânticos, os quais foram utilizados para as primeiras etapas do trabalho como análise conformacional e otimização geométrica. O segundo grupo de métodos apresentados é o daqueles utilizados para a obtenção de propriedades físico-químicas e o terceiro grupo é o dos métodos estatísticos. Apesar de bem extensa, esta seção não apresenta todos os métodos utilizados, pois a descrição rigorosa de todos seria inviável em um trabalho como este.

¹<ftp://ftp2.ipc.pku.edu.cn/pub/software/xlogp/>

2.2.1 Método Hartree-Fock

Conforme Erwin Schrödinger postulou em 1926, o movimento dos elétrons é descrito por uma equação diferencial conhecida com Equação de Schrödinger, sendo a Equação 2.1 sua forma unidimensional e independente do tempo.

$$-\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} + V(x)\psi(x) = E\psi(x) \quad (2.1)$$

Como a equação de Schrödinger só tem solução analítica para átomos hidrogenóides,² devido ao problema da inseparabilidade de variáveis (existente na forma multidimensional) causado pelo termo de repulsão eletrônica,³ quando se trabalha com sistemas multieletrônicos algumas aproximação precisam ser feitas, sendo a mais importante delas o campo auto-consistente, proposto por Douglas Hartree em 1928.

O método do campo auto-consistente (SCF)

De acordo com o modelo de Hartree, a movimentação de cada elétron em um campo efetivo dos outros $N - 1$ elétrons é governada por uma equação de Schrödinger de uma partícula sujeita a esse campo, ou seja, a auto-consistência da distribuição de carga no campo produz uma série de equações acopladas (Equações de Hartree) para N funções de onda de uma partícula. Em outras palavras o método auto-consistente vem para substituir a equação de Schrödinger para N partículas, por N equações de uma partícula, as quais apresentam solução. Desta forma, precisa-se resolver as N equações até que a distribuição de carga não varie significativamente; quando isso ocorre, diz-se que o campo atingiu a auto-consistência “*Self-Consistent Field* (SCF)”.

O determinante de Slater

Como a equação de Schrödinger não tem solução para sistemas multieletrônicos, pela Equação 2.1 não se podem determinar as funções de onda para estes mesmos sistemas. Assim, precisa-se supor uma função de onda inicial, sendo a utilização do determinante de Slater o melhor processo para esse fim. Isso porque este determinante já garante a antissimetrização da função de onda, de acordo como o princípio de Pauli, como apontado por J. C. Slater em 1929. Todos os elementos dados nas colunas do determinante envolvem o mesmo *spin*-orbital, enquanto que elementos da mesma linha envolvem o mesmo elétron. Outra diferença entre o método proposto originalmente por Hartree é que V. Fock adicionou em 1930 o termo de troca ao operador Hamiltoniano, termo esse necessário

²átomos que apresentem apenas um elétron

³contido em $V(x)$

também devido ao princípio de Pauli. Por ser um termo relativístico ele não tem correspondente clássico. Neste ponto já se tem estruturado o método dito Hartree-Fock (HF).

As equações de Roothaan-Hall

Com o método Hartree-Fock os cálculos de orbitais moleculares tornaram-se equivalentes ao problema de resolver a equação:

$$\hat{F}(1)\psi_i(1) = \varepsilon_i\psi_i(1) \quad (2.2)$$

a qual é uma equação de autovalor-autovetor, a exemplo da equação de Schrödinger, só que trocando-se o operador Hamiltoniano pelo operador Fock. Ela pode ser interpretada como a equação de Schrödinger de uma partícula submetida a um campo efetivo.

Como a resolução desta equação é muito custosa computacionalmente, a contribuição de C. C. J. Roothaan em 1951 foi mostrar que, pela introdução de um conjunto de funções de base especiais conhecidas, a equação diferencial 2.2 poderia ser convertida a um conjunto de equações algébricas que seriam resolvidas por técnicas de operações matriciais. A expansão dos orbitais em termos das funções de base é apresentada na Equação 2.3.

$$\psi_i = \sum_{\nu}^k c_{\nu i} \phi_{\nu}(1) \quad (2.3)$$

Substituindo-se 2.3 em 2.2 têm-se:

$$\hat{F} \sum_{\nu}^k c_{\nu i} \phi_{\nu} = \varepsilon_i \sum_{\nu}^k c_{\nu i} \phi_{\nu}. \quad (2.4)$$

Multiplicando-se por ϕ_{μ}^* e integrando-se:

$$\sum_{\nu}^k c_{\nu i} \underbrace{\int \phi_{\mu}^* \hat{F} \phi_{\nu} d\tau}_{F_{\mu\nu}} = \varepsilon_i \sum_{\nu}^k c_{\nu i} \underbrace{\int \phi_{\mu}^* \phi_{\nu} d\tau}_{S_{\mu\nu}}, \quad (2.5)$$

ou equivalentemente

$$\sum_{\nu}^k F_{\mu\nu} c_{\nu i} = \varepsilon_i \sum_{\nu}^k S_{\mu\nu} c_{\nu i}. \quad (2.6)$$

A Equação 2.6 é conhecida como equação de Roothaan-Hall, a qual é a forma matricial, com base na expansão em orbitais atômicos, da equação de Hartree-Fock. As grandes vantagens desta formulação são a maior facilidade na resolução por técnicas matriciais e na implementação compu-

tacional.

O limite Hartree-Fock

Sendo a expansão em funções de base dada pela Equação 2.3, quanto mais completo for o ajuste das funções de bases ϕ_ν , maior a flexibilização na expansão dos *spin*-orbitais e menor o valor esperado da energia.

$$E_0 = \langle \psi_0 | \hat{H} | \psi_0 \rangle \quad (2.7)$$

Aumentando-se a expansão das bases, mantém-se a tendência de diminuição da energia E_0 , até se chegar a $k = \infty$. Na prática, qualquer conjunto finito de bases produz uma energia acima do limite Hartree-Fock.

A energia de correlação

A energia de correlação é um erro do limite Hartree-Fock, sendo definida como:

$$E_{\text{correlação}} \equiv E_{\text{exata}} - E_{\text{HF}}. \quad (2.8)$$

Este erro é causado justamente pela troca da equação para N partículas por N equações de uma partícula. Com essa aproximação, cada elétron só sente um campo médio causado pelos outros $N - 1$ elétrons, não sendo possível descrever interações instantâneas. Como resultado têm-se um erro da ordem de 1 % na energia. Aparentemente, esse erro é pequeno; no entanto, a energia total do átomo de carbono é aproximadamente 1000 eV, sendo 0,5 % disso 5 eV. 5 eV/molécula é o valor de uma ligação simples, de modo que, sem a energia de correlação, não se pode calcular energias de ligação corretamente.

Neste ponto encerra-se a breve apresentação dos fundamentos do método Hartree-Fock. Uma leitura mais apurada pode ser obtida nas referências [46–48]. O método acima é conhecido como aproximação *ab initio*,⁴ ou primeiros princípios. Os métodos ditos semi-empíricos são muito utilizados por representarem um custo computacional menor do que o método *ab initio*, isso por que as etapas mais caras do método *ab initio* são substituídas por parâmetros. Assim, toda arte destes métodos está em estimar as integrais de recobrimento ($S_{\mu\nu}$) e em modificar o operador de Fock ($F_{\mu\nu}$) na Equação 2.6, de forma a compensar as perdas decorrentes da estimativa das integrais $S_{\mu\nu}$. Os dois métodos semi-empíricos baseados na teoria Hartree-Fock que se utilizaram neste trabalho são apresentados na seção 2.2.2.

⁴latim: do começo

2.2.2 Métodos Semi-empíricos *Austin Model 1* (AM1) e *Modified Neglect of Diatomic Overlap-Parametric Method 3* (MNDO-PM3)

Como os métodos *Austin Model 1* (AM1) [49] e *Modified Neglect of Diatomic Overlap-Parametric Method 3* (MNDO-PM3) [50] têm rigorosamente a mesma estrutura teórica, a apresentação conjunta é a melhor forma de se perceber as diferenças. Os cálculos de todas as integrais são aproximados, os únicos orbitais considerados são o s , p_x , p_y e p_z e as integrais de recobrimento são ignoradas. Tanto o AM1 como o PM3 utilizam parâmetros derivados exclusivamente de átomos livres, entretanto o PM3 otimiza alguns deles.⁵ As aproximações descritas acima levam o nome de *Neglect of Diatomic Overlap* (NDO). Os métodos AM1 e PM3 utilizam essa aproximação com pequenas modificações, como será apresentado adiante.

Originalmente, têm-se as equações seculares como sendo:

$$|F - \varepsilon_i S|c_i = 0 \quad (2.9)$$

onde c_i são os autovetores, F é a matriz de Fock, ε_i são os autovalores e S a matriz de sobreposição. Os elementos da matriz de Fock são definidos como:

$$F_{\mu\nu} = H_{\mu\nu} + \sum_{\lambda} \sum_{\sigma} [P_{\lambda\sigma} \langle \mu\nu | \lambda\sigma \rangle - \frac{1}{2} P_{\lambda\sigma} \langle \mu\lambda | \nu\sigma \rangle] \quad (2.10)$$

onde H é a matriz de um elétron e P a matriz de densidade.⁶

Recobrimento

Como apresentado anteriormente todas as integrais de recobrimento de dois orbitais atômicos diferentes são ortogonais, com isso a equação secular fica:

$$|F - \varepsilon_i|c_i = 0 \quad (2.11)$$

As integrais de três e quatro centros também são negligenciadas. Os elementos da matriz de Fock, caso os orbitais estejam em centros diferentes, se reduzem a:

$$F_{\mu\nu} = H_{\mu\nu} - \sum_{\lambda}^A \sum_{\sigma}^B P_{\lambda\sigma} \langle \mu\lambda | \nu\sigma \rangle. \quad (2.12)$$

Se os orbitais são diferentes, mas estão no mesmo centro, então as integrais do tipo $\langle \mu\nu | \lambda\sigma \rangle$ são zero por ortogonalidade, exceto se $\mu = \nu$ e $\lambda = \sigma$ ou $\mu = \lambda$ e $\nu = \sigma$. Nesse casos, os elementos de

⁵daí o nome *Parametric Method*

⁶somatório dos produtos de c_i

um centro da matriz de Fock tornam-se:

$$F_{\mu\nu} = H_{\mu\nu} + P_{\mu\nu} \langle \mu\nu | \mu\nu \rangle - P_{\mu\nu} (\langle \mu\nu | \mu\nu \rangle + \langle \mu\mu | \nu\nu \rangle). \quad (2.13)$$

dois elétrons e um centro

Para o AM1 estas expressões são derivadas de dados experimentais de átomos isolados. Já para o PM3 os valores são otimizados para reproduzirem propriedades moleculares experimentais.

dois elétrons e dois centros

Em um arranjo diatômico local há somente 22 integrais únicas de dois elétrons e dois centros para cada par de átomos pesados.⁷ Cada integral representa a interação da densidade eletrônica do primeiro elétron com a densidade eletrônica do segundo elétron. Todas essas integrais são calculadas por diferentes técnicas, sendo as distribuições de densidade eletrônica aproximadas por uma série de cargas pontuais.

um elétron e um centro ($H_{\mu\mu}$)

Esta integral representa a energia de um elétron em um orbital atômico φ_μ se todos os demais elétrons tivessem sido retirados do sistema. A energia de um elétron é obtida parametricamente e é simbolizada por $U_{\mu\mu}$. A integral tem então a seguinte expressão:

$$H_{\mu\mu} = U_{\mu\mu} - \sum_{B \neq A} Z_B, \quad (2.14)$$

onde Z_B é a carga nuclear.

um elétron e dois centros ($H_{\mu\nu}$)

Algumas vezes também chamada de integral de ressonância ($H_{\mu\nu}$), é aproximada utilizando-se as integrais de recobrimento $S_{\mu\nu}$. É interessante notar que isso viola a aproximação NDO, mas, uma vez que as integrais de ressonância têm valores expressivos, estas não podem ser negligenciadas. Esta é a origem do termo *Modified* no MNDO.

Utilizando-se orbitais do tipo Slater

$$\varphi_\mu = N r^{n-1} \exp^{-\xi r}, \quad (2.15)$$

a integral de recobrimento é dada por

⁷átomos diferentes do hidrogênio

$$S_{\mu\nu} = \langle \varphi_\mu | \varphi_\nu \rangle \quad (2.16)$$

e a integral de ressonância tem então a forma

$$H_{\mu\nu} = S_{\mu\nu} \frac{1}{2} (\beta_\mu + \beta_\nu) \quad (2.17)$$

onde β são parâmetros.

repulsão núcleo-núcleo

Este termo consiste basicamente na repulsão eletrostática entre os núcleos; no entanto, alguns parâmetros são adicionados para corrigir distorções causadas pelas aproximações descritas acima. O termo de repulsão para ambos os métodos é

$$E_N(A, B) = E_N^{MNDO}(A, B) + \frac{Z_A Z_B}{R_{AB}} \left(\sum_k a_{kA} \exp^{-b_{kA}(R_{AB} - c_{kA})^2} + \sum_k a_{kB} \exp^{-b_{kB}(R_{AB} - c_{kB})^2} \right), \quad (2.18)$$

onde

$$E_N^{MNDO}(A, B) = Z_A Z_B \langle s_A s_A | s_B s_B \rangle (1 + \exp^{-\alpha_A R_{AB}} + \exp^{-\alpha_B R_{AB}}). \quad (2.19)$$

Uma demonstração mais aprimorada sobre os métodos semi-empíricos AM1 e PM3 pode ser encontrada na referência [51].

2.2.3 Teoria do Funcional de Densidade (DFT)

A Teoria do Funcional de Densidade [52, 53] “*Density Functional Theory* (DFT)” tem grande vantagem sobre os métodos baseados nas equações de Hartree-Fock, devido ao menor custo computacional. Como não utiliza parâmetros, além de constantes universais da física, ela pode ser considerada um método do tipo *ab initio* a exemplo do método Hatree-Fock descrito na Seção 2.2.1. Considerando-se um sistema com n funções de base, o esforço computacional utilizando-se DFT é de n^3 , enquanto que para os métodos HF este valor é de n^4 . A vantagem sobre os métodos semi-empíricos está no fato do Hamiltoniano do DFT ser bem definido e suas características não serem obscurecidas por frequentes e injustificáveis aproximações no procedimento computacional (Seção 2.2.2). Outra grande vantagem é que se pode introduzir a correlação eletrônica sem alterar

drasticamente o aspecto formal do modelo [54], através do funcional de troca e correlação.

Hohenberg e Kohn provaram, em 1964, que o potencial externo é função da densidade eletrônica (Primeiro teorema de Hohenberg-Kohn). Eles mostraram também que a energia, função de onda e outras propriedades moleculares são univocamente determinadas por essa densidade de probabilidade eletrônica $\rho[x, y, z]$ (Segundo Teorema de Hohenberg-Kohn). A energia para o estado fundamental é dada pela expressão:

$$E_0 = -\frac{1}{2} \sum_{i=1}^n \langle \psi_i(1) | \nabla_1^2 | \psi_i(1) \rangle - \sum_{\alpha} \int \frac{Z_{\alpha} \rho(1)}{r_{1\alpha}} d\nu_1 + \frac{1}{2} \int \int \frac{\rho(1)\rho(2)}{r_{12}} d\nu_1 d\nu_2 + E_{xc}[\rho] \quad (2.20)$$

onde $\psi_i(1), i = 1, 2, \dots, n$ são os orbitais Kohn-Sham e $E_{xc}[\rho]$ é a energia de troca-correlação. Kohn e Sham também mostraram que ρ exato para o estado fundamental pode ser determinado pelos orbitais ψ_i :

$$\rho = \sum_{i=1}^n |\psi_i|^2. \quad (2.21)$$

O terceiro teorema de Hohenberg-Kohn diz que $E_0[\rho] \leq E_0[\tilde{\rho}]$, onde ρ é a densidade exata e $\tilde{\rho}$ a densidade aproximada pela expansão 2.21 para n finito. Este teorema é equivalente ao teorema variacional da teoria Hartree-Fock.

Os orbitais Kohn-Sham podem ser determinados pela expressão

$$\hat{F}_{KS}(1)\psi_i(1) = \varepsilon_{i,KS}\psi_i(1), \quad (2.22)$$

onde $\hat{F}_{KS}$ é o operador de Kohn-Sham apresentado pela Equação 2.23.

$$\hat{F}_{KS} \equiv -\frac{1}{2} \nabla_1^2 - \sum_{\alpha} \frac{Z_{\alpha}}{r_{1\alpha}} + \sum_{j=1}^n \hat{J}_j(1) + V_{xc}(1), \quad (2.23)$$

onde o potencial V_{xc} é a principal diferença entre os métodos HF e DFT [54].

Desta forma, pode-se minimizar a energia em relação à densidade eletrônica através das condições de contorno

$$\frac{\partial E_0}{\partial \tilde{\rho}} = 0, \quad (2.24)$$

sendo a $\int \tilde{\rho} dr = N$, onde N é o número de elétrons do sistema.

O principal problema do método é a falta de um processo sistemático para determinar $E_{xc}[\rho]$, assim, vários tipos de funcionais já foram propostos [46]. Este funcional é, na verdade, dividido em duas partes, uma de troca e outra de correlação. Um dos funcionais de troca mais utilizados é o B3

proposto por Becke em 1993 [55], o qual se utilizou neste trabalho.

$$E_x = (1 - a_0)E_x^{LSDA} + a_0E_x^{HF} + a_x\Delta E_x^{B88} \quad (2.25)$$

Para correlação, o funcional proposto por Lee, Yang e Parr [56] foi utilizado.

$$E_c = -a \int \frac{1}{1 + d\rho^{1/3}} \{ \rho + b\rho^{-2/3} \left[C_F \rho^{5/3} - 2t_w + \left(\frac{1}{9}t_w + \frac{1}{18}\nabla^2\rho \right) \right] \exp(-c\rho^{-1/3}) \} dr \quad (2.26)$$

2.2.4 *Hydrogenic Atoms in Molecules* versão 3 (HAM/3)

O método HAM/3 é baseado no modelo de átomos hidrogenóides obtido pelo uso de constantes de blindagem. A idéia de constantes de blindagem, primeiramente proposta por Slater, produz a seguinte expressão para energia total de átomos

$$E = - \sum_{\mu} \left(\frac{P_{\mu\mu}}{n_{\mu}^2} \right) \left[Z_A - (P_{\mu\mu} - 1)\sigma_{\mu\mu} - \sum_{\nu \neq \mu} P_{\nu\nu}\sigma_{\nu\mu} \right]^2, \quad (2.27)$$

onde $P_{\mu\mu}$ é a matriz de densidade, Z_A a carga nuclear, n_{μ} o número quântico principal e $\sigma_{\nu\mu}$ é a constante de blindagem.

Por tentativas obteve-se a seguinte expressão para a constante de blindagem

$$\sigma_{\nu\mu} = a_{\nu\mu} - \frac{b_{\nu\mu} + c_{\nu\mu}Z_A}{\zeta_{\mu}}, \quad (2.28)$$

onde $a_{\nu\mu}$, $b_{\nu\mu}$ e $c_{\nu\mu}$ são constantes empíricas. O termo ζ_{μ} é expresso por:

$$\zeta_{\mu} = \frac{Z_A - s_{\mu}}{n_{\mu}}. \quad (2.29)$$

Este modelo é bem mais elaborado que o proposto originalmente por Slater, onde a constante de blindagem é expressa por um número fixo. No modelo original, não se podia descrever corretamente a complicada interdependência entre os elétrons, mas na equação proposta para o método HAM/3 isso foi aperfeiçoado.

Os excelentes resultados obtidos para os cálculos espectroscópicos atômicos motivaram a extensão do método para moléculas. A boa precisão obtida nos espectros é devida a incorporação de grande parte da correlação eletrônica [57].

Para sistemas moleculares, as equações de Roothaan-Hall têm a forma

$$\sum_{\nu} (F_{\mu\nu} - \varepsilon_i S_{\mu\nu}) c_{\nu i} = 0, \quad (2.30)$$

onde $c_{\nu i}$ são os coeficientes de expansão para a função de onda. Os elementos da matriz de Fock são dados por

$$F_{\mu\nu} = \frac{\partial E}{\partial P_{\mu\nu}}, \quad (2.31)$$

onde a matriz de densidade $P_{\mu\nu}$ tem a forma

$$P_{\mu\nu} = \sum_i q_i c_{\mu i} c_{\nu i}. \quad (2.32)$$

A expressão para a energia total E é dividida em diversos termos

$$E = \sum_{\mu\nu} E_{\mu\nu}, \quad (2.33)$$

onde $E_{\mu\nu}$ são as energias das cargas eletrônicas $P_{\mu\nu}S_{\mu\nu}$ na região $\mu\nu$. Os elementos de energia dentro e fora da diagonal principal são estimados de maneira distinta.

Os termos de energia para os elementos dentro da diagonal principal são dados por

$$E_{\mu\mu} = -P_{\mu\mu}\zeta_\mu^2, \quad (2.34)$$

onde ζ_μ^2 são os mesmos do cálculo atômico.

Para os elementos fora da diagonal principal os termos de energia ficam

$$E_{\mu_A\nu_B} = -P_{\mu\nu}S_{\mu\nu} \left[\frac{1}{2}(\zeta_\mu^2 + \zeta_\nu^2) \right] f_{\mu\nu}, \quad (2.35)$$

onde $f_{\mu\nu}$ precisa ser diferente para os diferentes tipos de ligação química, além de ser dependente da distância internuclear R_{AB} .

Em ambas as equações 2.34 e 2.35 não é considerada a blindagem feita por elétrons que estejam em outros núcleos, com isso, as interações cruzadas são dadas por

$$\sum_{A>B} Q_A Q_B \gamma_{AB}, \quad (2.36)$$

onde γ_{AB} é calculado segundo o método de Ohno e Klopman [58]. A expressão da energia eletrônica tem a forma

$$E = - \sum_{\mu} T_{\mu} \zeta_{\mu}^2 + \sum_{A>B} Q_A Q_B \gamma_{AB}, \quad (2.37)$$

onde

$$T_\mu = P_{\mu\mu} + \sum_{A \neq B} \sum_{\nu_B}^B \frac{1}{2} (P_{\mu\nu} S_{\mu\nu} + P_{\nu\mu} S_{\nu\mu}) f_{\mu\nu}. \quad (2.38)$$

A matriz de Fock dada na Equação 2.31 tem a forma

$$F_{\mu\mu} = -\zeta_\mu^2 + \sum_{\nu}^A \sigma_{\mu\nu} (2/n_\nu) \zeta_\nu T_\nu - \sum_B Q_B \gamma_{AB} \quad (2.39)$$

e

$$F_{\mu_A \nu_B} = \frac{1}{2} S_{\mu_A \nu_B} [F_{\mu_A \mu_A} + F_{\nu_B \nu_B} - (\zeta_{\mu_A}^2 + \zeta_{\nu_B}^2) (f_{\mu_A \nu_B} - 1)]. \quad (2.40)$$

Adicionando-se o termo de repulsão nuclear

$$\sum_{A>B} Z_A Z_B (R_{AB}^{-1} - \gamma_{AB}) \exp^{-\alpha_{AB} R_{AB}}, \quad (2.41)$$

à Equação 2.37 obtém-se a expressão da energia total do sistema.

O método HAM/3 foi desenvolvido para o estudo de potenciais de ionização. O teorema de Koopman não pode ser utilizado devido a grande incorporação da energia de correlação, a exemplo do que ocorre na Teoria do Funcional de Densidade, entretanto a equação para a energia E_μ de um elétron em um orbital ϕ_μ em um átomo A apresenta valores corretos para a energia do orbital

$$E_\mu = -\frac{1}{n_\mu^2} \left[Z_A - (P_{\mu\mu} - 1) \sigma_{\mu\mu} - \sum_{\nu \neq \mu} P_{\nu\nu} \sigma_{\nu\mu} \right]^2, \quad (2.42)$$

a exemplo do que já foi visto na Equação 2.27 [59].

Retirando-se uma pequena fração de carga dq_i de uma molécula obtém-se a seguinte equação diferencial:

$$\frac{\partial E}{\partial q_i} = \sum_{\mu\nu} \frac{\partial E}{\partial P_{\mu\nu}} \frac{\partial P_{\mu\nu}}{\partial q_i} = \sum_{\mu\nu} F_{\mu\nu} \frac{\partial P_{\mu\nu}}{\partial q_i} = \sum_{\mu\nu} F_{\mu\nu} c_{\mu i} c_{\nu i} = \epsilon_i. \quad (2.43)$$

Por integração obtém-se o potencial de ionização (PI)

$$PI_i = \int_{q_i=0}^1 \epsilon_i dq_i \approx -\epsilon_i, \quad (2.44)$$

calculado com $q_i = 3/2$. A carga fracionária é a forma matemática resultante para o conceito de estado de transição primeiramente introduzido por Slater. Adicionalmente, é conveniente remover o meio elétron de todos os orbitais ocupados, por um processo dito ionização difusa. Este processo

acaba por produzir todos os PI da molécula em um único cálculo chamado *Photoelectron Spectroscopy* (PES). O cálculo da afinidade eletrônica (EA) é feito pelo mesmo método de estado de transição, onde agora adiciona-se meio elétron à molécula, também de maneira difusa.

As energias de excitação podem ser obtidas pelo método de *configuration interaction* (CI).

Outro recurso do HAM/3 é o cálculo de *Electron Spectroscopy for Chemical Analysis* (ESCA), onde um elétron é retirado do orbital 1s de átomos pesados. A técnica é conhecida por esse nome; pois, as energias dos elétrons de níveis internos são características de cada átomo. Em verdade, a técnica é sensível à presença de diferentes espécies químicas do mesmo átomo, que aparecem como pequenos deslocamentos nos espectros. Esses deslocamentos podem ser correlacionados com a constante de reatividade de Hammett [60], sendo esse o nosso interesse.

2.2.5 Somatório de contribuições atômicas (log P)

O coeficiente de partição octanol/água (log P) foi calculado pelo programa xlogp desenvolvido pelo Professor Renxiao Wang [45]. Este programa adiciona contribuições atômicas para calcular o coeficiente de partição de um dado composto, sendo essa sistemática apresentada na Equação 2.45.

$$\log P = \sum_i a_i A_i + \sum_j b_j B_j \quad (2.45)$$

A_i e B_j são as ocorrências para os i -ésimos e j -ésimos tipos atômicos, enquanto a_i e b_j são os fatores de correção para estes tipos atômicos. Os coeficientes são derivados por regressão multivariada para um grande número de compostos orgânicos. Um total de 90 tipos atômicos são utilizados para classificar carbono, nitrogênio, oxigênio, enxofre, fósforo e halogênios.

2.2.6 Método de Higo e Go para o Cálculo de Volume e Superfície Moleculares

Os parâmetros estéricos calculados foram volume de van der Waals, superfície de van der Waals, volume acessível ao solvente (água) e superfície acessível ao solvente, utilizou-se o programa Surf [42].

O método de cálculo por quadraturas é baseado no método de Higo e Go [61], onde a molécula é colocada em uma caixa retangular, sendo os lados desta caixa escolhidos como múltiplos de 2 Å. Cada lado é dividido de 2 em 2 Å, formando então cubos, isso é chamado nível 1. Cada cubo é analisado e classificado em três categorias: interno, externo ou superficial à molécula. Se o cubo é interno ou externo ele é descartado, se ele é superficial, então é novamente dividido em oito cubos de 1 Å, o que é chamado de nível 2. No programa esse processo pode ser repetido até o nível 8, com cubos de 0,015625 Å de lado, o qual utilizou-se aqui. Desta forma, pode-se ter uma relação linear entre a área da esfera que está contida no cubo e o lado do cubo.

$$A_{esfera\ contida} = f A_{lado\ do\ cubo} \quad (2.46)$$

onde f é um fator de proporcionalidade. Através de algumas manipulações matemáticas, pode-se determinar o fator de proporcionalidade como sendo:

$$f = \frac{A_{esfera}}{N_c A_{lado\ do\ cubo}} = \frac{8\pi r^2}{N_c l^2} \quad (2.47)$$

onde r é o raio da esfera, N_c é o número de cubos que interceptam a superfície molecular e l é o lado do cubo. Assim, pode-se calcular a superfície de van der Waals. Para os cálculos com solvente adiciona-se 1,5 Å ao raio da esfera. Já o volume de van der Waals é determinado pelo somatório dos cubos contidos na molécula.

2.2.7 Análise de Componentes Principais (PCA)

A análise de componentes principais “*Principal Component Analysis* (PCA)” é a principal forma de análise multivariada de dados, provendo uma aproximação dos dados originais em termos de duas matrizes (*scores* e *loadings*). A idéia básica é agrupar as variáveis que estão altamente correlacionadas combinando-as linearmente de forma a produzir novas variáveis, chamadas de componentes principais, as quais são ortogonais entre si. Isso proporciona a redução da dimensionalidade dos dados, detecção de *outlier*,⁸ seleção de variáveis, classificação dos objetos, entre outras melhorias na visualização dos dados.

Graficamente, a PCA pode ser vista como uma rotação nos eixos de coordenadas originais, de forma a produzir novos eixos, que passem através das direções de maior variância, sempre perpendiculares entre si. A Figura 2.1 apresenta a primeira componente principal na direção da maior variância das três variáveis originais, a segunda componente principal seria obrigatoriamente perpendicular a esta e passando pela segunda direção de maior variância e assim sucessivamente até o número máximo de variáveis originais do sistema. A distância t_{i1} representa o valor (*score*) da amostra x_i no novo eixo de coordenadas P_1 .

Matematicamente, existem várias formas de se obter as matrizes de *scores* e *loadings*. Aqui utilizou-se o método de Decomposição em Valor Singulares [63] “*Singular Value Decomposition* (SVD)”. Este método reúne uma eliminação LU de Gauss e uma ortogonalização de Gram-Schmidt, de forma que a matriz de dados X_{mn} é expressa por

$$X = \underbrace{US}_T V', \quad (2.48)$$

⁸dado anômalo que pode ou não estar errado

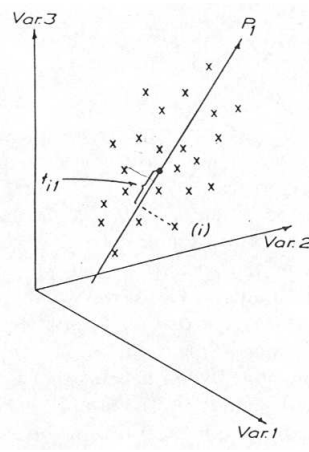


Figura 2.1: Descrição gráfica da PCA [62].

onde U_{mm} apresenta os autovetores da matriz XX' , S_{mn} os valores singulares da matriz X e V_{nn} os autovetores da matriz $X'X$, os quais são chamados de *loadings*, ou seja, o peso que cada variável original tem para a composição das componentes principais. A matriz T_{mn} é chamada matriz de *scores* que são as posições das amostras no novo sistema de coordenadas. Graficando-se as colunas de T umas pelas as outras tem-se então a visualização dos dados em duas ou três dimensões. A manipulação matemática pode ser acompanhada no código apresentado no Apêndice A.

2.2.8 *Soft Independent Modeling of Class Analogy (SIMCA)*

A capacidade de classificar amostras em categorias conhecidas é o principal recurso que o método SIMCA apresenta. Mais do que isso, o SIMCA é capaz de classificar objetos em categorias, utilizando-se, para tanto, um limite de confiança, normalmente de 95 %. No método SIMCA um modelo de PCA é feito para cada categoria e envelopes de confiança (linhas, planos, volumes) são construídos ao redor do modelo para conter os pontos dos dados. Graficamente, isso pode ser observado na Figura 2.2, onde a linha, o plano e o paralelepípedo representam modelos com uma, duas e três componentes principais, respectivamente.

Matematicamente, um dado x_{ik} em uma categoria q é representado como

$$(x_{ik})_q = (\bar{x}_k)_q + \sum_{j=1}^{NC} (v_{kj})_q (y_{ij})_q + (e_{ik})_q, \quad (2.49)$$

onde $\bar{x}_k$ é a média da categoria para a variável k , v_{kj} são os *loadings* das NC componentes principais utilizadas no modelo, y_{ij} são os *scores* e e_{ik} são os resíduos que expressam como o modelo descreve os dados.

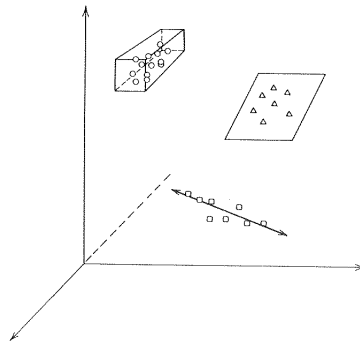


Figura 2.2: Descrição gráfica do SIMCA [64].

A variância residual na classe q , a qual mede as tensões no *cluster* da categoria, é expressa por

$$s_0^2 = \sum_{i=1}^{NP} \sum_{k=1}^{NV} \frac{(e_{ik})^2}{(NP - NC - 1)(NV - NC)}, \quad (2.50)$$

onde NP é o número de amostras e NV o número de variáveis do sistema.

A variância residual do objeto i para o modelo da categoria q é dada por

$$s_i^2 = \sum_{k=1}^{NV} \frac{(e_{ik})^2}{(NV - NC)}. \quad (2.51)$$

Se s_i^2 é do mesmo tamanho que s_0^2 o objeto é tipicamente pertencente à categoria, caso contrário ele é estranho aquela classe. Quantitativamente, um intervalo normal é estabelecido para os *scores* com respeito a cada componente principal, utilizando-se o desvio padrão s_j , onde j é o intervalo de 1 até o número total de componentes principais do modelo. As fronteiras são dadas por

$$y_j(max) + s_j c \quad e \quad y_j(min) - s_j c, \quad (2.52)$$

onde c é um coeficiente ajustado pelo usuário ou dependente do número de objetos em um grupo através de um teste estatístico [64,65].

2.2.9 K Vizinhos Mais Próximos (KNN)

O método dos k vizinhos mais próximos [64] “*K-Nearest Neighbor* (KNN)” é uma técnica muito simples e poderosa. Este método classifica as amostras através de uma matriz de votos dados pelos vizinhos mais próximos. Calculam-se as distâncias entre as amostras, agrupando-as pela proximidade, assim, os vizinhos mais próximos dão seus votos para a sua classe, a classe que tiver mais votos fica então com a amostra.

Neste método todas as amostras serão obrigatoriamente classificadas em uma única classe, diferentemente do que ocorre no método SIMCA. Uma grande vantagem do KNN é que ele não depende da distribuição dos dados, sendo mais indicado para a classificação de dados assimétricos [66].

2.2.10 Redes Neurais

A idéia básica das redes neurais ou “*Neural Network* (NN)” remonta dos anos 40 e, pelo fato de sua operação envolver distribuição paralela de processos, ela tem recebido especial atenção nos últimos anos. Há vários tipos de conexões para as redes, no entanto o programa PSSD faz uso somente da conexão do tipo *perceptron*, que é um modelo de auto-organização que aprende por reconhecimento. Para redes de duas camadas de neurônios este método apresenta algumas falhas; contudo, com o algoritmo de retro-propagação e, utilizando-se três camadas de neurônios, estas falhas foram corrigidas. Devido ao seu caracter não linear, a princípio a capacidade de classificação e ajuste é muito grande. Entretanto, isso impossibilita a rede neural de apresentar coeficientes de correlação simples para QSAR como são obtidos por métodos de regressão linear.

A informação que é acumulada na fase de aprendizagem é mantida pela força entre as ligações dos neurônios, a qual é guardada na matriz de pesos. Na primeira camada, o número de neurônios é igual ao número de variáveis mais uma;⁹ na terceira camada, o número de neurônios é igual ao número de classes e, na segunda camada, o número de neurônios é arbitrário, muitas vezes determinado através de testes de tentativas. O valor dos neurônios da segunda e terceira camadas é expresso pela Equação 2.53.

$$O_j = \frac{1}{1 + \exp(-\alpha y_j)} \equiv f(y_j) \quad (2.53)$$

para

$$y_j = \sum W_{ij}x_i - \theta_j, \quad (2.54)$$

onde x_i é um dos valores de um neurônio da camada $n - 1$, W_{ij} é o elemento da matriz de pesos que expressa a conexão entre os neurônios i e j , α é o parâmetro que expressa a não-linearidade das operações dos neurônios e θ_j é o valor inicial do neurônio j .

O treinamento da rede é baseado nas seguintes equações:

$$\delta W_{ij} = -d_j x_i \varepsilon, \quad (2.55)$$

$$d_j = (O_j - t_j) f'(y_j) \quad (2.56)$$

⁹dita variável bias

e

$$d_j = (\sum W'_{ij} d'_i) f'(y_j), \quad (2.57)$$

onde ε é um parâmetro o qual determina o deslocamento para a correção da retro-propagação. A Equação 2.56 é utilizada para corrigir a conexão entre a segunda e a terceira camadas, enquanto a Equação 2.57 é utilizada para a outra conexão onde W'_{ij} e d'_j da segunda camada são W_{ij} e d_j da terceira camada.

Se $f(y_j)$ é sigmoidal, sua derivada na Equação 2.55 é:

$$f'_{y_j} = f(y_j)[1 - f(y_j)]\alpha. \quad (2.58)$$

Nas Equações 2.55 e 2.58 tanto ε como α podem ser ajustados independentemente da camada. O treinamento ocorre até que a função erro (Equação 2.59) fique tão pequena quanto determinada.

$$E = \sum (O_j - t_j)^2 \quad (2.59)$$

Capítulo 3

Análises Conformacionais e Propriedades Moleculares

Os quarenta e seis compostos apresentados na Tabela 1.1 podem ser divididos em quatro grupos:

- trans/cis (**2a** ao **3a**, **4a** e do **10a** ao **12i**);
- trans/trans (**3c**);
- cis/cis (**9a** e **9b**);
- outros (**8a** e do **13** ao **17**).

Onde as relações de estereoquímica¹ se referem ao par de átomos 4a com o 5 e ao par 4a com o 9 (Figura 1.4), como apresentado na Seção 1.4. É interessante notar que os compostos apresentam isômeros ópticos, entretanto só é utilizado aqui o esqueleto mais ativo (4aS,5R,9bS) [67], apresentado na Figura 3.1 e as atividades dos racematos.

3.1 Trans/Cis

Como vinte e nove estruturas apresentam a estereoquímica trans/cis, começou-se o trabalho pela montagem de um esqueleto base para esse grupo.

3.1.1 Análise Conformacional

A proposta inicial era utilizar a metodologia MM2 para o cálculo dos aspectos conformacionais, entretanto este campo de força não apresenta parâmetros para alguns ângulos diedros. Os diedros

¹trans/cis, trans/trans e cis/cis

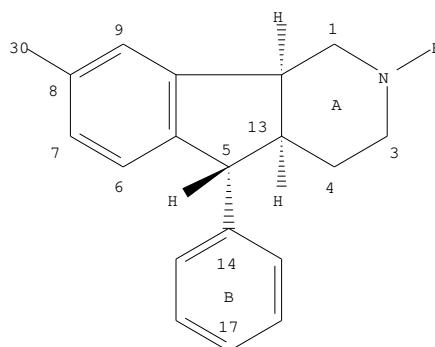


Figura 3.1: Esqueleto da família trans/cis (numeração utilizada nos cálculos).

que apresentaram problema foram principalmente os que envolvem o átomo do tipo N sp^3 . A segunda metodologia empregada foi a PM3, por esta apresentar resultados de calor de formação mais coerentes com os dados experimentais, quando comparada com outros métodos semi-empíricos [68].

Sistemática

Iniciou-se com a análise conformacional sistemática, uma vez que são somente três as variáveis de interesse, no caso ângulos diedro. Realizaram-se a análise conformacional pela variação dos ângulos diedros C-4-C-3-N-2-C-1 e H-33-C-13-C-4-H-25 (H-25 esta ligado ao C-4 na Figura 3.1) obrigando-se o anel A a percorrer todas as suas possíveis conformações, tanto cadeira como barco. Primeiramente, partiu-se de estruturas cadeiras e barcos com os diedros ajustados de forma a produzirem conformações completamente eclipsadas. Com isso, obtiveram-se coordenadas de reação que apresentavam degraus, isso porque, em determinados pontos, ocorria a passagem rápida entre estrutura barco e cadeira. Este processo indicou a existência de quatro mínimos locais: duas cadeiras e dois barcos. O próximo passo foi então o de otimizar completamente os quatro mínimos locais, e repetir novamente a análise conformacional, só que dessa vez, partindo-se destes mínimos locais. Com esse procedimento, obtiveram-se coordenadas de reação suaves; pois, a energia é sempre crescente. Começando-se a partir dos quatro mínimos locais obtidos nas análises anteriores, rodou-se o anel B através do eixo da ligação C-5-C-14. Os mínimos obtidos do giro do anel B foram considerados como os verdadeiros mínimos locais e, conseqüentemente, utilizados para os cálculos na análise estocástica.

Estocástica

Novamente com o PM3, partiu-se para a análise conformacional estocástica com o pacote Tinker. Os quatro mínimos locais obtidos na análise conformacional sistemática foram utilizados como sementes no método de distâncias geométricas [69]. Quinhentas conformações foram então gera-

das para cada uma das quatro sementes. Em separado, as conformações dos quatro grupos foram completamente otimizadas. Muitas das quinhentas estruturas acabou por evoluir, durante as otimizações, para o mesmo mínimo; de forma, que a retirada das estruturas repetidas fez-se necessária. O resultado final é que cada grupo acaba por ter um conjunto de estruturas únicas, sendo estas então agrupadas em intervalos de $3 \text{ kcal} \cdot \text{mol}^{-1}$. Da literatura [70] sabe-se que se múltiplas procuras a partir de pontos não idênticos produzem o mesmo ajuste final de confôrmeros, isso é um indício de que o espaço conformacional foi corretamente explorado. Isto pode ser graficamente observado pela sobreposição dos histogramas destas procuras, apresentado na Figura 3.2.

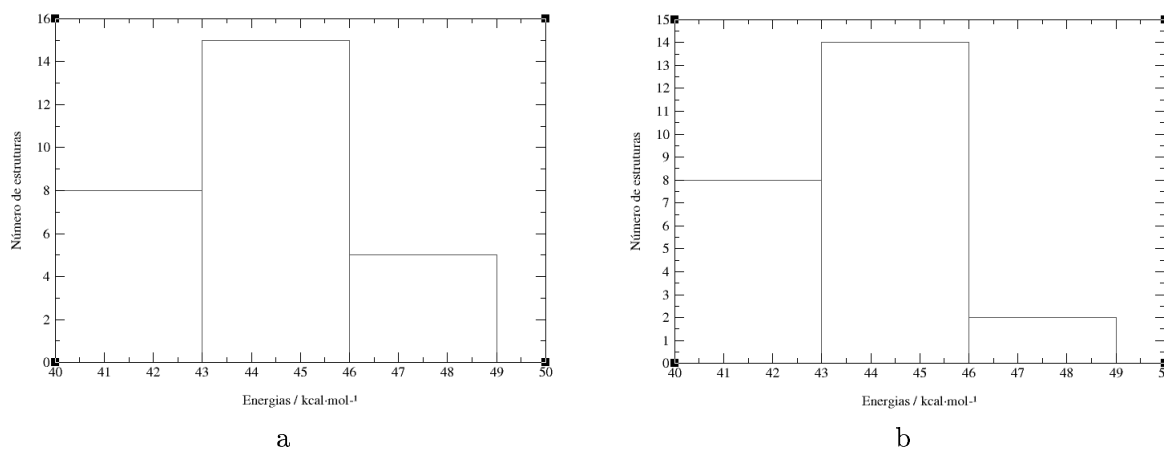


Figura 3.2: Histogramas das conformações únicas, sementes: a) cadeira 2; b) barco 2.

Na Figura 3.2 vê-se que os dois gráficos apresentam a mesma forma. Tem-se assim, indícios para acreditar que o espaço conformacional para o esqueleto base trans/cis foi corretamente explorado.

Mínimos Locais

Como ambas as análises conformacionais apresentaram os mesmos resultados, estes serão descritos conjuntamente. Nos dois casos, o método PM3 apresentou dois pontos inesperados:

- conformações barco mais estáveis que conformações cadeira;
- conformações com o hidrogênio em axial são mais estáveis que equatorial (H ligado ao N na Figura 3.1).

Esta ordem de estabilidade não pareceu confiável, já que estruturas cicloexenóides geralmente apresentam conformações cadeira torcida como sendo as mais estáveis. A discrepância na estabilidade relativa dos confôrmeros poderia ser atribuída à incorreta descrição feita pelo PM3 para a

substituição cíclica apresentada no anel *A* (Figura 3.1) e, para se elucidar isso, uma segunda metodologia deveria ser empregada. Apesar do grande custo computacional, optou-se pelo uso de um cálculo HF/6-31G.

Assim, os oito mínimos locais (cadeira 1, cadeira 2, barco 2 e barco 1), com o hidrogênio em axial e equatorial, foram otimizados com o HF/6-31G. Os cálculos indicaram que, independentemente da posição do hidrogênio amínico, as conformações cadeira são mais estáveis que as conformações barco, com a cadeira 1 equatorial sendo a mais estável delas. Em relação à posição do hidrogênio, os cálculos indicam que a posição axial é mais estável que a equatorial em média $2 \cdot 10^{-4}$ *hartree* para cadeira 1 e barco 1; é menos estável em média $1,3 \cdot 10^{-3}$ *hartree* para cadeira 1 e barco 2. Mesmo com estes resultados, optou-se por utilizar a posição equatorial em todas as conformações, isso porque a conformação mais estável, cadeira 1, é equatorial. Assim, todas as estruturas apresentadas daqui em diante são com o hidrogênio amínico em posição equatorial.

Uma vez que os cálculos HF/6-31G são muito caros computacionalmente, outras teorias mais rápidas deveriam ser utilizadas. Assim, os resultados do HF/6-31G, com posição equatorial, (Figura 3.3) foram utilizados como geometrias de entrada para os métodos AM1, DFT e HF/3-21G.

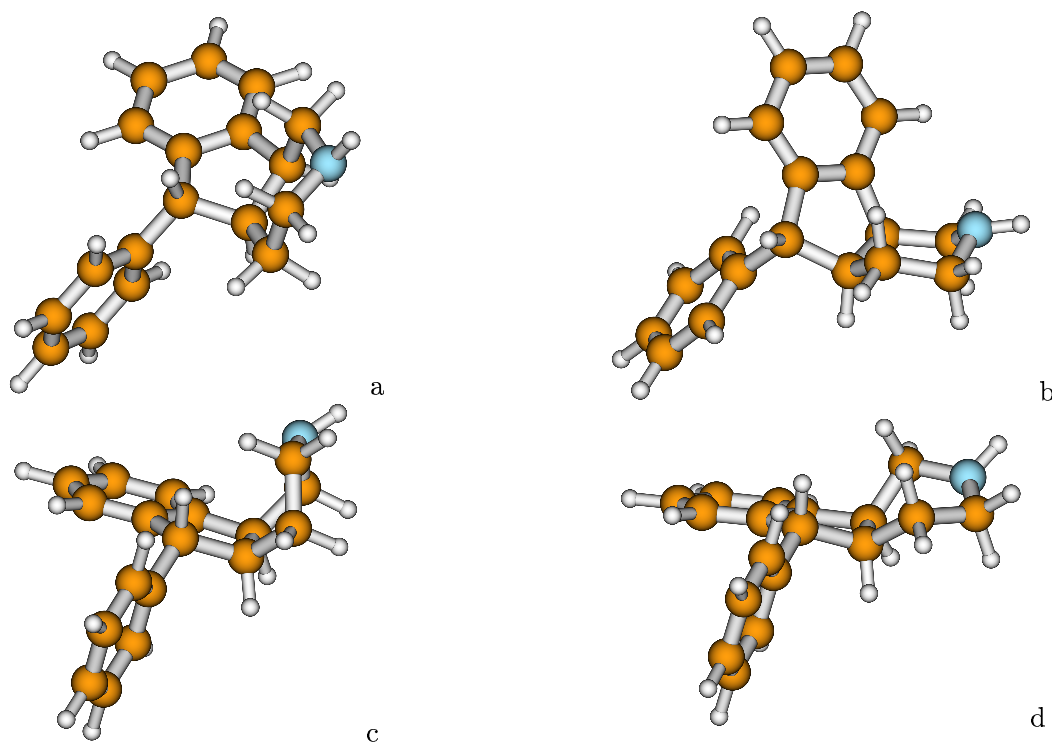


Figura 3.3: Conformações da família trans/cis: a) cadeira 1; b) cadeira 2; c) barco 2 e d) barco 1.

Apesar de ambos os métodos AM1 e PM3 usarem a mesma aproximação semi-empírica, eles diferem quanto ao processo de parametrização como foi visto na Seção 2.2.2. Este fato justifica a aplicação de ambos os métodos no mesmo grupo de moléculas. Na Teoria do Funcional de Densidade utilizou-se o funcional de troca-correlação B3LYP e base 3-21G (DFT/3-21G). Esta metodologia é tida como mais rápida que cálculos Hartree-Fock que incorporam correlação eletrônica, mas com resultados igualmente confiáveis. Além disso, em uma publicação recente, Hoffman e colaboradores correlacionam linearmente orbitais resultantes de cálculos semi-empíricos, HF e DFT, o que poderia justificar a utilização de algumas propriedades dadas pelo DFT em SAR [71].

Utilizando-se o operador AM1, novamente observa-se uma relação de estabilidade que não concorda com nenhum dos métodos anteriores. Isso indica que o AM1 também não é próprio para a descrição da estabilidade relativa entre os confôrmeros. Com os resultados do DFT, pela primeira vez teve-se a concordância entre dois cálculos. A coerência entre o DFT e o HF foi o primeiro indício quanto à correta relação entre as estabilidades teóricas das conformações. Apesar do DFT ter produzido resultados semelhantes quanto a estabilidade, o mesmo não é capaz de gerar posteriormente muitas das propriedades físico-químicas utilizadas em cálculos de SAR. Neste sentido um cálculo HF ainda se faz necessário, só que desta vez com base 3-21G. Com os cálculos HF/3-21G, observou-se novamente concordância entre resultados, o que leva a acreditar que a estrutura mais estável, em vácuo, é efetivamente a cadeira 1.

Como a comparação direta entre os valores de formação e energias total não é possível, uma normalização foi feita segundo a Equação 3.1

$$N = \frac{E_i - E_m}{E_M - E_m}, \quad (3.1)$$

onde E_i é a energia de interesse, E_m e E_M são as energias mínima e máxima, respectivamente. Esse processo elimina as unidades e cria uma faixa entre 0 e 1 para cada método, quanto menor o número mais estável a conformação. Os resultados da normalização são apresentados na Tabela 3.1.

Tabela 3.1: Normalização para os valores de formação e energias total.

método	cadeira 1	cadeira 2	barco 2	barco 1
AM1	0,62	1	0,53	0
PM3	0,47	1	0,87	0
HF/3-21G	0	0,47	0,72	1
HF/6-31G	0	0,64	0,95	1
DFT/3-21G	0	0,48	0,71	1

Como considerações finais sobre esta seção pode-se dizer que, conforme esperado [68], o PM3 apresenta melhores resultados que o AM1 quando se trata de calor de formação de compostos orgâ-

nicos; no entanto, nenhum dos dois métodos apresenta resultados satisfatórios. Devido à completa concordância entre os cálculos HF e DFT (cadeira 1, cadeira 2, barco 2 e barco 1) decidiu-se finalmente por utilizar cálculos HF/3-21G com a conformação cadeira 1 para todos os membros da família trans/cis, em fase gasosa.

3.1.2 Efeito da Base

A fim de se observar o efeito da base na geometria molecular, a Tabela 3.2 apresenta o cálculo das distância RMS, em Å, para a cadeira 1. As distâncias RMS, apresentadas na Tabela 3.2, consistem em determinar a distância entre os átomos correspondentes² nas duas conformações. Eleva-se ao quadrado cada uma das distâncias, faz-se a média do quadrados e tira a raiz quadrada dessa média.

Tabela 3.2: Variação da geometria da cadeira 1 com a base (valores em Å).

	3-21G	6-31G	6-31G*	6-31G**
3-21G	0	0,0618	0,0637	0,0635
6-31G		0	0,0265	0,0260
6-31G*			0	0,0008
6-31G**				0

Apesar de, visualmente (Figura 3.4), o nitrogênio amínico apresentar um aumento em sua piramidalização quando adiciona-se polarização, esta mudança não altera expressivamente as geometrias como um todo. Analisando-se as contribuições de cada átomo para os desvios RMS,³ observa-se que a mudança da base 3-21G para as demais bases alterou principalmente os hidrogênios do anel *B*. É na mudança de 6-31G para a 6-31G* que se observa a variação esperada, onde a maior parte da dissimilaridade entre as estruturas é devida ao hidrogênio ligado o N (Figura 3.1). Já a passagem de 6-31G* para 6-31G** não produziu nenhum efeito relevante nas geometrias.

3.1.3 Efeito de Solvente

Utilizando-se o método *Polarizable Continuum Model* (PCM) [72,73], tentou-se modelar o efeito de solvente, no caso a água. Esta metodologia foi adicionada ao PM3 e HF, através dos programas MOPAC e GAMESS, respectivamente. Pelo método PM3 há praticamente uma completa inversão entre a estabilidade no vácuo e em água. No vácuo, a ordem de estabilidade é barco 1, cadeira 1, barco 2 e cadeira 2. Já em água, a ordem é cadeira 2, barco 2, barco 1 e cadeira 1. No

²distância entre os átomos com a mesma numeração

³dados não apresentados

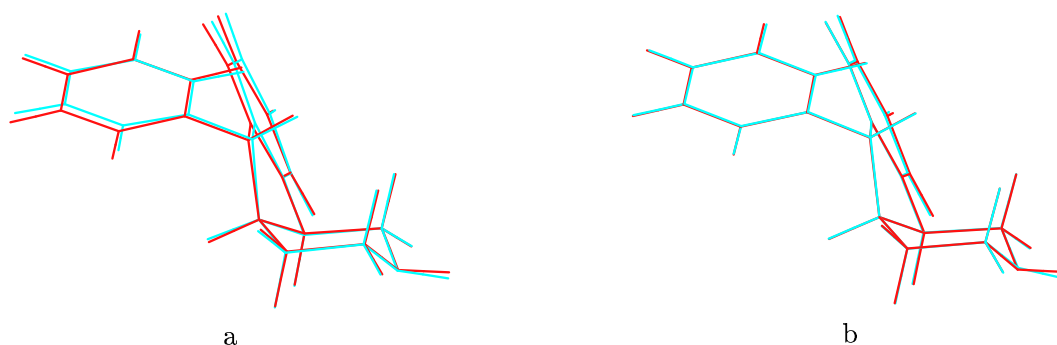


Figura 3.4: Sobreposição de estruturas otimizadas com diferentes bases: a) 3-21G e 6-31G**; b) 6-31G e 6-31G* (alinhamento feito pelos átomos 1, 3 e 13; ver Figura 3.1).

HF/3-21G a cadeira 1 se manteve como a mais estável, todavia as demais conformações trocaram substancialmente de lugar. No HF/6-31G observa-se que há uma inversão entre as duas cadeiras e os dois barcos. A sequência de estabilidade é cadeira 2, cadeira 1, barco 2 e barco 1. As inversões observadas podem ser atribuídas ao efeito do volume molecular no método PCM e a variações no momento de dipolo (μ). Tentou-se também utilizar o PCM com o AM1 e o DFT, no entanto isso não foi possível devido a limitações dos pacotes MOPAC e GAUSSIAN, respectivamente. Devido à não concordância do PCM nos diversos métodos, e à impossibilidade de otimizar as geometrias com efeito de solvente, os dados não são apresentados e nem utilizados neste trabalho.

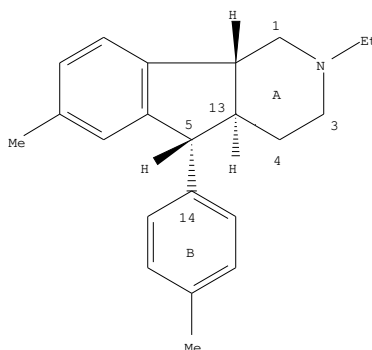
3.2 Trans/Trans

A família trans/trans é composta por um único membro, o composto **3c**. Utilizando-se o isômero óptico que mais se assemelha ao da família trans/cis, e pela análise com modelos moleculares, observa-se que somente a conformação equivalente à cadeira 2 da família trans/cis é viável. Além da forma cadeira, outras duas formas barco também mostram-se viáveis; todavia, optou-se pela forma cadeira por questões de uniformidade com a família trans/cis. Assim, a estrutura utilizada para os cálculos é apresentada na Figura 3.5.

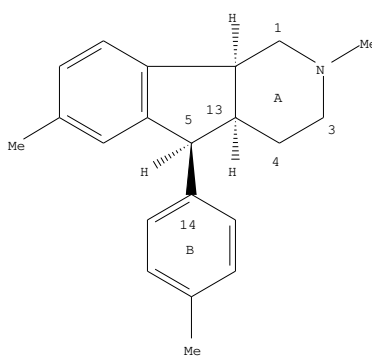
Da mesma forma que para o esqueleto trans/cis, o anel *B* do composto **3c** foi girado. A estrutura de mínimo obtida do giro do anel *B*, foi então otimizada com o HF/3-21G.

3.3 Cis/Cis

A família cis/cis é formada pelos compostos **9a** e **9b**. Utilizando-se o enantiômero semelhante ao obtido nos cálculos do esqueleto trans/cis, e com a ajuda de modelos moleculares, observou-se que as estruturas apresentam a mesma liberdade no anel *A* que o esqueleto trans/cis. Devido a

Figura 3.5: Estrutura do composto **3c**.

essa liberdade, resolveu-se testar as conformações cadeira 1 e 2, utilizando-se o composto **9b** como modelo.

Figura 3.6: Estrutura do composto **9b**.

A partir das conformações do composto **9b**, previamente otimizadas, procedeu-se ao giro do anel *B*, exatamente como descrito anteriormente. O giro do anel *B* para a cadeira 1 procedeu normalmente, contudo para a cadeira 2 observaram-se modificações no anel *A*, durante o giro de *B*, produzindo somente estruturas menos estáveis que a inicial. Os calores de formação obtidos, com o PM3 para a completa otimização dos mínimos, foram 24,5 e 26 $kcal \cdot mol^{-1}$, respectivamente para cadeiras 1 e 2. Devido à pequena diferença energética e aos problemas observados para a cadeira 2, as conformações correspondentes aos mínimos foram então otimizadas com o HF/3-21G. Os valores de energia total foram $-859,448$ e $-859,450$ *hartree* para cadeiras 1 e 2, respectivamente. Observa-se assim mais uma discordância entre os cálculos PM3 e HF/3-21G. As estruturas resultantes tiveram a posição R_1 substituída com Et para produzir a estrutura **9a**, e então, calculadas com o HF/3-21G. As energias total obtidas para **9a** foram $-898,268$ e $-898,270$ *hartree*, respectivamente para cadeiras 1 e 2. Desta forma, optou-se pela cadeira 2 como sendo a conformação mais estável para os compostos **9a** e **9b**.

3.4 Outros

Nesta seção são apresentadas estruturas que não têm um padrão comum quanto a estereoquímica dos átomos 4a com o 5 e do 4a com o 9 (Figura 1.4), como descrito na Seção 3.1.

3.4.1 Composto 8a

Utilizando-se o enantiômero semelhante ao dos cálculos trans/cis e assumindo-se que o hidrogênio da posição 9, (Figura 1.4) conserva a mesma configuração que a apresentada pelos compostos **7** (intermediário não apresentado) e **9** do esquema reacional 1 da referência [14], a estrutura utilizada para os cálculos é apresentada na Figura 3.7.

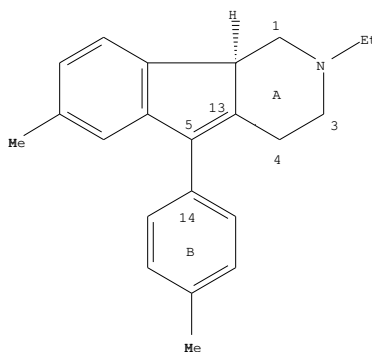


Figura 3.7: Estrutura do composto **8a**.

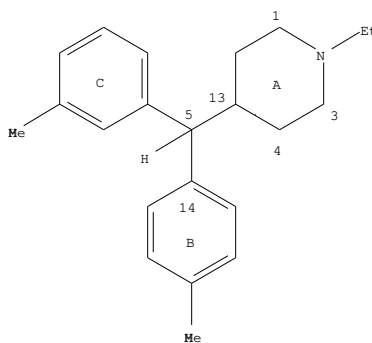
Com a ajuda dos modelos moleculares observa-se que o composto só deve apresentar conformações equivalentes à forma cadeira 1 e a quatro barcos. Para fins de comparação com os demais compostos optou-se por utilizar a forma cadeira 1 no anel *A*. Assim, procedeu-se ao giro do anel *B* como já descrito anteriormente, com a conformação mais estável utilizada para o cálculo HF/3-21G.

3.4.2 Composto 13

O composto **13**, por apresentar um maior número de graus de liberdade, sofreu um procedimento mais amplo que os anteriores.

Foram montadas quatro estruturas, duas barco e duas cadeira, com o carbono 5 posicionado em axial e equatorial (Figura 3.8). Todas as estruturas foram otimizadas com PM3, sendo os resultados de calor de formação, em $kcal \cdot mol^{-1}$, apresentados na Tabela 3.3.

Apesar da pequena diferença energética, as duas conformações barco foram eliminadas. As formas cadeira tiveram seus anéis *B* e *C* rodados de 10 em 10 °, a exemplo do que foi feito anteriormente. Começou-se pelo anel *C* passando a metila pelo lado contrário ao anel cicloexenoide,

Figura 3.8: Estrutura do composto **13**.Tabela 3.3: Calores de formação para o composto **13** (valores em $kcal \cdot mol^{-1}$).

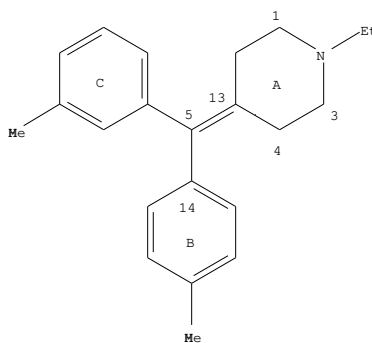
conformero	calor de formação
barco axial	14,2
barco equatorial	13,5
cadeira axial	12,1
cadeira equatorial	12,0

então, o ponto de mínimo teve seu anel *B* rodado. Após, as estruturas foram completamente otimizadas com o PM3, sendo os calores de formação $11,6$ e $11,2 \text{ kcal} \cdot \text{mol}^{-1}$, respectivamente para as conformações axial e equatorial. Devido à pequena diferença energética, ambas conformações foram submetidas ao HF/3-21G, sendo as energias total $-899,424$ e $-899,422 \text{ hartree}$ para axial e equatorial, respectivamente. Desta forma, a conformação axial foi utilizada para o cálculo das propriedades.

3.4.3 Composto 14

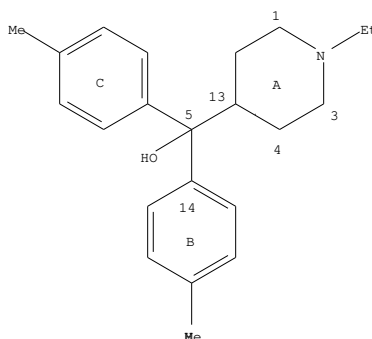
A exemplo do composto **13**, o **14** também apresenta um grande número de graus de liberdade. Da observação dos modelos moleculares e das estruturas otimizadas **8a** e **13**, propôs-se uma conformação inicial para **14**, basicamente posicionando o anel *A* de forma semelhante à estrutura **8a** e os anéis *B* e *C* conforme a estrutura **13**. Mesmo o anel cicloexenoide apresentando todas as estruturas cadeira e barco, optou-se por usar a cadeira 1 para fim de comparação com os demais compostos. A estrutura utilizada é dada na Figura 3.9.

A estrutura proposta foi otimizada com o PM3 e posteriormente teve seu anel *C* girado de 10° em 10° . A posição de mínimo obtida foi utilizada para o giro do anel *B*, também de 10° em 10° . Ambas as análises conformacionais não levaram a estruturas mais estáveis que a inicial, sendo a estrutura inicial utilizada no HF/3-21G.

Figura 3.9: Estrutura do composto **14**.

3.4.4 Composto **15**

O composto **15** foi montado sob grande influência dos resultados obtidos para o **13**. O HF/3-21G do composto **13** foi modificado de forma a produzir **15**.⁴ Feito isso, o grupo OH foi girado de 10 em 10 °, sendo a conformação de mínimo completamente otimizada. O cálculo HF/3-21G procedeu normalmente e a Figura 3.10 apresenta a estrutura.

Figura 3.10: Estrutura do composto **15**.

3.4.5 Compostos **16a** e **16b**

As estruturas **16a** e **16b** foram montadas com base na observação de modelos moleculares e levando-se em conta a descrição sintética apresentada na referência [14]. O carbono 5 é racêmico e os dois ligantes apresentados do anel *A* são que atribuem a denominação de *trans* para **16a** e *cis* para o **16b**. Um esqueleto válido para os dois compostos é apresentado na Figura 3.11.

Para o composto **16a**, os anéis *B* e *C* podem estar ambos em equatorial ou ambos em axial; da mesma forma, para a **16b**, o anel *B* pode estar em axial e o *C* em equatorial ou vice versa.

⁴comparar as figuras 3.8 e 3.10

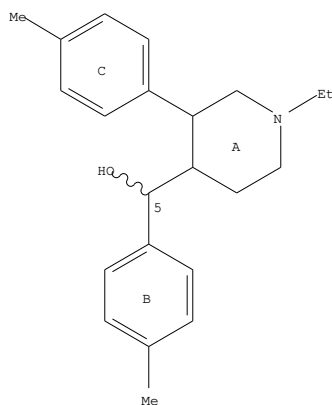


Figura 3.11: Estrutura dos compostos **16a** (trans) e **16b** (cis).

As quatro conformações foram completamente otimizadas com PM3, e a Tabela 3.4 apresenta os resultados dos calores de formação, em $kcal \cdot mol^{-1}$.

Tabela 3.4: Calores de formação para os compostos **16a** e **16b** (valores em $kcal \cdot mol^{-1}$).

conformero	calor de formação
16a (<i>B</i> e <i>C</i> equatoriais)	-30,4
16a (<i>B</i> e <i>C</i> axiais)	-25,5
16b (<i>B</i> axial e <i>C</i> equatorial)	-22,7
16b (<i>B</i> equatorial e <i>C</i> axial)	-24,8

Apesar das pequenas diferenças, adotou-se o primeiro caso da Tabela 3.4 para representar **16a** e o quarto caso para representar **16b**.

A primeira análise conformacional realizou-se no **16a** pelo giro de 10 em 10 ° do grupo OH, com a conformação de mínimo então utilizada para o giro do anel *B*; no entanto, esta última não produziu nenhum conformero mais estável que o primeiro. Finalmente a estrutura foi completamente otimizada com o PM3, sendo o resultado utilizado para o HF/3-21G. Procedeu-se de forma idêntica para o composto **16b**.

3.4.6 Composto 17

O composto **17** foi montado fazendo-se as trocas necessárias no **11a**. A Figura 3.12 apresenta a estrutura.

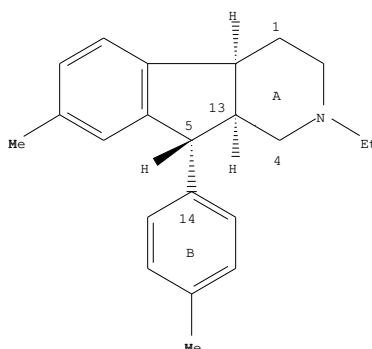


Figura 3.12: Estrutura do composto **17**.

A estrutura foi diretamente submetida ao cálculo HF/3-21G.

Nesse ponto, independentemente da procedência inicial da estrutura, todas apresentam uma conformação, a mais estável, otimizada com HF/3-21G. Excetuando-se a Seção 3.5.1, estas são as únicas geometrias consideradas daqui em diante para propriedades dos derivados.

3.5 Parâmetros Físico-químicos

Como mencionado anteriormente, o objetivo principal com os cálculos HF/3-21G era obter geometrias mais confiáveis que as semi-empíricas, devido aos problemas apresentados na Seção 3.1. Por não significar um aumento expressivo do tempo computacional, os cálculos de algumas propriedades foram realizados conjuntamente. Além das propriedades fornecidas pelos cálculos HF, utilizou-se a metodologia semi-empírica HAM/3. As geometrias obtidas no HF/3-21G foram utilizadas para os cálculos de propriedades com o HAM/3 (HAM/3//HF/3-21G).⁵ Volume molecular e coeficiente de partição octanol/água foram calculados sob as geometrias fixas do HF/3-21G. Uma pequena coletânea dessas propriedades é apresentada abaixo.

3.5.1 Variação com a Conformação

Mesmo tendo-se chegado à conclusão, na Seção 3.1.1, de que a cadeira 1 é a estrutura mais viável para a maioria dos compostos em estudo, nesta seção se quer observar como algumas propriedades comportam-se quando muda a forma de cálculo e a conformação utilizada. Assim, pode-se avaliar qualitativamente o erro que se comete quando utiliza-se informação apenas de uma conformação.

Exceto para o HAM/3, a Tabela 3.5 apresenta os potenciais de ionização (PI), segundo o teorema de Koopmans, para as quatro conformações assumidas pelo esqueleto base da família trans/cis.

⁵propriedade//geometria

Nesta mesma tabela, pode-se observar também eletroafinidade (EA), com todos os dados expressos em eV .

Tabela 3.5: Valores de PI e EA, em eV , para as conformações da família trans/cis.

propriedade	método	cadeira 1	cadeira 2	barco 1	barco 2
PI	AM1	9,207	9,133	9,235	9,235
	PM3	9,221	9,201	9,133	9,016
	HF/3-21G	8,574	8,580	8,623	8,621
	HF/6-31G	8,460	8,446	8,487	8,493
	HAM/3	7,941	7,921	7,601	7,596
EA	HAM/3	0,255	0,096	0,199	0,198

Como era esperado, o PI é fortemente dependente do método utilizado. À medida em que se aumenta o rigor da teoria utilizada, os resultados para o PI diminuem em valor, sendo que o método HAM/3 apresenta os menores valores. Isso certamente é devido a sua especial parametrização para cálculos de PI. Comparando-se exclusivamente os cálculos HF observa-se que a diferença entre eles é de $0,1 eV$, o que fica abaixo do erro experimental de $0,2 eV$ obtido, por exemplo, com Luz Síncrotron no Laboratório Nacional de Luz Síncrotron (LNLS). Isso indica que, pelo menos neste ponto, a utilização de propriedades de HF/3-21G não produziria nenhuma perda significativa. Analisando-se as linhas referentes a cada método, observa-se que, a exceção do HAM/3, o PI varia abaixo do erro experimental com a conformação, o que significa que esta propriedade é insensível à conformação pelos métodos utilizados.

A Tabela 3.6 apresenta resultados de momento de dipolo (μ), em Debye, obtidos pelos diversos métodos.

Tabela 3.6: Variação de μ com o método e a conformação (valores em D).

método	cadeira 1	cadeira 2	barco 1	barco 2
AM1	1,029	1,372	1,258	1,258
PM3	1,074	1,415	1,259	1,550
HAM/3	1,821	2,382	1,867	1,872
HF/3-21G	0,933	1,428	1,067	1,069
HF/6-31G	0,921	1,387	0,944	0,956
DFT/3-21G	0,974	1,409	1,090	1,093

Medidas de μ feitas por microondas podem chegar a uma precisão de $0,01 \%$ ou melhor, como pode ser visto na literatura [74]. Desta forma, observa-se grande influência no momento de dipolo pela metodologia utilizada. É interessante notar que, a exceção do PM3, a cadeira 2 apresenta o maior momento de dipolo para todos os métodos. A discrepância apresentada pelo método PM3

talvez possa ser atribuída ao erro apresentado por este método quando do cálculo de cargas atômicas de nitrogênios, como pode ser visto na referência [51].

Outras propriedades moleculares que apresentam grande interesse em trabalhos de relação estrutura atividade biológica são volume de van der Waals (vol. vdw), superfície de van der Waals (sup. vdw), volume acessível ao solvente (vol. sol.) e superfície acessível ao solvente (sup. sol.), nesse caso água. As superfícies são expressas em $\text{\AA}^2$, e os volumes em $\text{\AA}^3$.

Tabela 3.7: Variação de volume ($\text{\AA}^3$) e superfície ($\text{\AA}^2$) molecular com a conformação.

propriedade	cadeira 1	cadeira 2	barco 2	barco 1
vol. vdw	35,40	36,55	37,56	36,16
sup. vdw	49,45	54,05	58,76	51,65
vol. sol.	117,65	169,13	156,63	126,54
sup. sol.	69,15	127,80	104,68	82,75

Na Tabela 3.7 observa-se que a cadeira 1 apresenta os menores valores para as quatro propriedades. Já os maiores volumes e superfícies de van der Waals são apresentados pelo barco 2. Quanto a acessibilidade a solvente, a cadeira 2 apresenta os maiores valores. O comportamento semelhante das propriedade deve-se, a princípio, à correlação existente entre elas.

3.5.2 Variação com a Base

Como visto na Seção 3.1.1, as geometrias moleculares não variam significativamente com a mudança da base. Nesta seção se pretende observar como algumas propriedades variam quando a base do *ab initio* muda, para tanto assumiu-se que as pequenas mudanças observadas na página 41 são desprezíveis. Utilizou-se exclusivamente a cadeira 1, otimizada com HF/3-21G, 6-31G, 6-31G* e 6-31G**.

Para a energia total observa-se uma expressiva estabilização do sistema quando aumenta-se a base, ou seja, diminuição do valor numérico da energia. Isso é devido ao fato que uma maior base acomoda de melhor forma a densidade eletrônica, o que estabiliza o sistema. O mesmo comportamento foi observado para o momento de dipolo.

Das diversas propriedades eletrônicas calculadas, as que mais interessam são as cargas atômicas, pois com elas fazem-se estudos de reatividade química, efeito de substituintes, constantes de ionização e, frequentemente, apresentam excelentes correlações com a atividade biológica [75]. A Figura 3.13 apresenta as cargas CHELPG para a conformação cadeira 1 da família trans/cis.

Da Figura 3.13 observa-se que as bases 3-21G, 6-31G* e 6-31G** apresentam um comportamento muito semelhante na descrição das cargas atômicas ao longo de toda a molécula; todavia, a base 6-31G apresenta comportamento diverso, principalmente na descrição das cargas dos átomos 5,

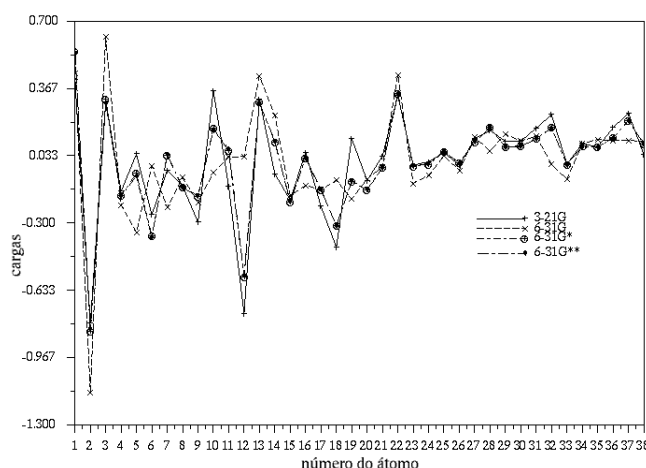


Figura 3.13: Variação das cargas CHELPG com a base.

6, 7, 10, 12, 16, 18 e 19. De forma geral, não se esperava grande variação das cargas, uma vez que é conhecido da literatura que cargas derivadas do potencial eletrostático são independentes da base [76].

Diferentemente das cargas CHELPG, as cargas de Mulliken são muito dependentes da base utilizada [77], sendo isso seguramente observado aqui. A Figura 3.14 apresenta as cargas de Mulliken calculadas da mesma forma que as CHELPG para a cadeia 1.

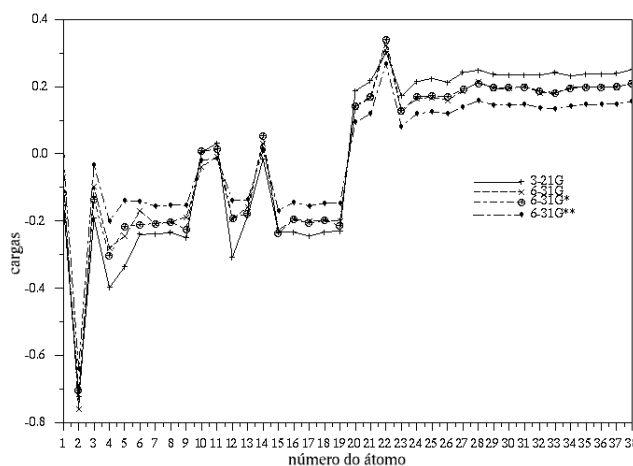


Figura 3.14: Variação das cargas Mulliken com a base.

Do gráfico, pode-se observar que as bases 6-31G e 6-31G* apresentam o mesmo comportamento, indicando que a polarização nos átomos pesados não é necessária para a descrição, dessa propriedade. Quando comparam-se as bases 3-21G e 6-31G**, vê-se que o comportamento delas é oposto em

relação a descrição de átomos leves e pesados. Desta forma, para os átomos pesados (até número 20) a 3-21G apresenta valores mais negativos que a 6-31G**, sendo que para os átomos leves essa situação se inverte. Isso é facilmente explicado pela presença da polarização nos átomos leves, o que abre espaço para que mais densidade eletrônica fique sobre estes átomos.

Capítulo 4

Correlação Estrutura Química-Atividade Biológica

Como apresentado na página 58, o elevado número de variáveis produzidas pelos métodos quânticos não deixa outra escolha, senão o tratamento automático dos dados. Para tanto, programas em linguagem PERL foram desenvolvidos (Apêndice B). Além disso, desenvolveu-se também um programa para análise de variável *versus* variável e outro para análise de componentes principais (Apêndice A). Antes do tratamento estatístico, todas as variáveis foram autoescaladas devido à grande variedade na natureza dos dados.

4.1 Dados Experimentais

O primeiro passo na análise estatística foi a observação dos dados experimentais. O objetivo é observar a relação entre a atividade biológica, medida pelo índice espermatogênico, e o efeito colateral, medido pelo teste de massa. Observa-se que na Tabela 4.1 têm-se os mesmos valores de atividade apresentados na Tabela 1.1, porém reorganizados para facilitar a visualização.

Tabela 4.1: Dados experimentais dos compostos estudados.

	3 mg/Kg		10 mg/Kg		30 mg/Kg		90 mg/Kg		μ g/Kg
	MT ¹	IE ²	MT	IE	MT	IE	MT	IE	IE ³
2a									76
2b									74
(l)2b									72
3a									72

continua na próxima página

Tabela 4.1: Dados experimentais dos compostos estudados (continuação).

	3 mg/Kg		10 mg/Kg		30 mg/Kg		90 mg/Kg		$\mu\text{g/Kg}$
	MT ¹	IE ²	MT	IE	MT	IE	MT	IE	IE ³
3c							-18	0	
4a									72
8a							5	-9	
9a							2	-9	
9b					12	7			
10a	10	5	10	31	24	50			
<i>(d)</i> 10a	13	31	25	59	43	62			
<i>(l)</i> 10a					10	9	4	-9	
10b	8	-3	7	26	18	38			
10c							3	-9	
10d							-11	-9	
10e							5	-5	
10f	-7	2	-17	-6			5	-5	
10g							-2	2	
10h							-8	-9	
10i			5	0	3	20	20	35	
10j			1	0	6	3	11	30	
10k	13	2	24	36	28	51			
<i>(d)</i> 10k	-5	21	26	40	16	67	27	70	
<i>(l)</i> 10k							-7	0	
10l	3	-2	2	7	26	33			
10m	-18	4	0	-4	8	23			
10o					-1	-5			
11a							-3	-5	
11b					11	0			
11c			7	10	15	28	23	45	
12b	5	5	17	27	32	57			
<i>(d)</i> 12b							3	-5	
<i>(l)</i> 12b	11	28	26	47	30	61	34	68	
12c	11	2	3	33	31	56			
12d					13	-4			
12e					-7	-2			
12f					-3	3			
12g	4	-5	13	20	23	57			
12h					7	8			
12i	7	0	6	2	7	31			
13	2	2			9	-5			

continua na próxima página

Tabela 4.1: Dados experimentais dos compostos estudados (continuação).

	3 mg/Kg		10 mg/Kg		30 mg/Kg		90 mg/Kg		$\mu\text{g/Kg}$
	MT ¹	IE ²	MT	IE	MT	IE	MT	IE	IE ³
14							-5	0	
15							-21	0	
16a							-9	0	
16b							0	0	
17					0	2			

¹ massa dos testículos (% redução em relação ao controle)

² índice espermatogênico (% redução em relação ao controle)

³ maior dose reportada para índice espermatogênico em [25]

Separaram-se então os compostos apresentados na Tabela 4.1 em três grupos, pela proximidade da dose utilizada:

- com dados para 3 e 10 mg/Kg: (**10a**, (*d*)**10a**, **10b**, **10f**, **10k**, (*d*)**10k**, **10l**, **10m**, **12b**, (*l*)**12b**, **12c**, **12g** e **12i**);
- com dados para 10 e 30 mg/Kg: (**10a**, (*d*)**10a**, **10b**, **10i**, **10j**, **10k**, (*d*)**10k**, **10l**, **10m**, **11c**, **12b**, (*l*)**12b**, **12c**, **12g** e **12i**);
- com dados para 30 e 90 mg/Kg: ((*l*)**10a**, **10i**, **10j**, (*d*)**10k**, **11c**, (*l*)**12b**).

O agrupamento dos compostos foi necessário para se ter um número maior de variáveis no sistema; desta forma, podem-se utilizar pelo menos duas componentes principais para sua descrição. Os três grupos foram então analisados com base nas duas primeiras componentes principais. Da observação dos *scores* e *loadings* têm-se:

- 3–10 mg/Kg (Figura 4.1);

As variâncias acumuladas para PC1 e PC2 foram 69,8 e 21,1 %, respectivamente. O gráfico dos *scores* (Figura 4.1a) mostra a separação em três grupos distintos: a) formado pelos compostos (*d*)**10a**, (*d*)**10k** e (*l*)**12b**; b) formado pelos compostos **10f** e **10m**; c) demais compostos. No gráfico dos *loadings* (Figura 4.1b) vê-se que a PC1 descreve principalmente os dados de 10 mg/Kg, enquanto a PC2 os dados de 3 mg/Kg. Este gráfico também indica que as PCs agruparam os compostos com maiores valores de índice espermatogênico na parte inferior esquerda, como observado no gráfico dos *scores* pelo posicionamento dos compostos (*d*)**10a**, (*d*)**10k** e (*l*)**12b** (notadamente os mais ativos). Desta forma, puderam-se determinar quais características são mais marcantes em cada grupo formado pelas componentes principais, sendo esta uma das grandes virtudes do método de PCA.

- 10–30 mg/Kg (figura omitida);

A análise não apresentou grupos perfeitamente distintos como no caso anterior, mas a primeira PC evidentemente separa os compostos mais ativos dos menos ativos. Isso por que a PC1 posiciona em um lado compostos como o (*d*)**10a** (ativo), e no outro lado compostos como o **10j** (inativo).

- 30–90 mg/Kg (figura omitida).

A análise produziu três grupos distintos, como no caso para 3–10 mg/Kg ; todavia, o reduzido número de compostos acaba por prejudicar a validade estatística do processo.

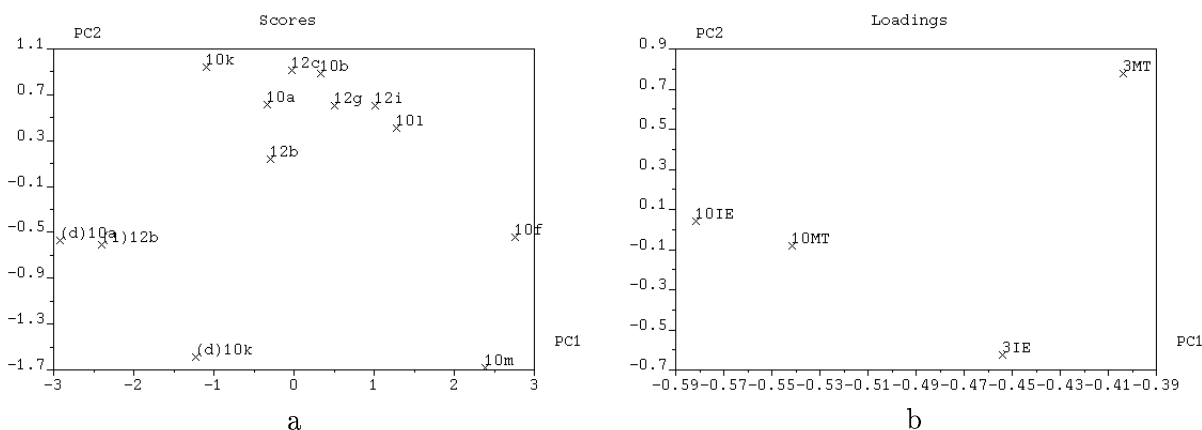


Figura 4.1: PCA para os dados de 3–10 mg/Kg : a) *Scores*; b) *Loadings*.

Como esperado, a PCA dos dados experimentais agrupa os compostos em famílias e apresenta a relação entre atividade e efeito colateral, possibilitando assim a observação dessas duas propriedades conjuntamente. Também pôde-se familiarizar com a análise de componentes principais, etapa importante antes do tratamento estatístico de um número elevado de variáveis.

Na Tabela 4.1, vê-se que o número de dados faltantes é muito grande. Entretanto, dois compostos ((*d*)**10k** e (*l*)**12b**) apresentam valores para todas as doses. O gráfico relacionando log da dose pela resposta obtida para o índice espermatogênico desses dois compostos é apresentado na Figura 4.2.

A partir do gráfico, pode-se observar que, para 30 e 90 mg/Kg ,¹ têm-se o comportamento assintótico da sigmóide, indicando que a eficácia da droga chegou ao limite. Assim sendo, aumentando-se a dose de 30 para 90 mg/Kg não se produz mais um aumento significativo na atividade biológica. Utilizando-se o conceito de dose limite, puderam-se classificar todos os compostos de uma única vez, pelo seguinte critério: alta atividade quando o índice for maior do que 45; média atividade, para índices variando entre 23 e 38; e inativos, para índices menores do que 8.

¹dois últimos pontos do gráfico, que está em escala logarítmica

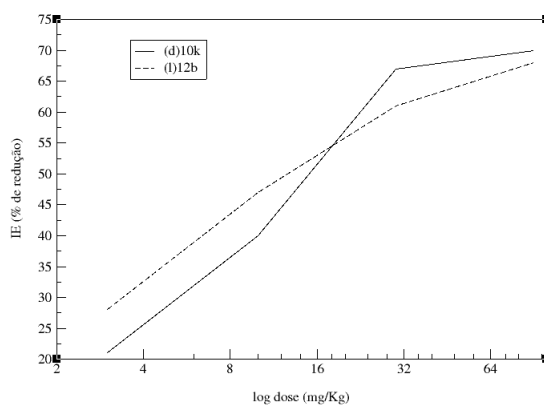


Figura 4.2: Gráfico do log da dose pela resposta para o índice espermatogênico dos compostos $(d)10k$ e $(l)12b$.

- baixa atividade: $(3c, 8a, 9a, 9b, (l)10a, 10c, 10d, 10e, 10f, 10g, 10h, (l)10k, 10o, 11a, 11b, (d)12b, 12d, 12e, 12f, 12h, 13, 14, 15, 16a, 16b$ e 17);
- média atividade: $(10b, 10i, 10j, 10l, 10m$ e $12i)$;
- alta atividade: $(2a, 2b, (l)2b, 3a, 4a, 10a, (d)10a, 10k, (d)10k, 11c, 12b, (l)12b, 12c$ e $12g)$.

Apesar de se dispor de valores numéricos para a atividade biológica, esses não são próprios para análises quantitativas, pois não se encaixam no modelo de Hansch [78]. O modelo de Hansch considera que as principais variáveis para a descrição de um fármaco em meio biológico são a concentração aplicada do medicamento (C), a probabilidade desse medicamento alcançar o sítio de interação (A) e a constante de equilíbrio da etapa limitante do processo (K_x). Desta forma, a variação da resposta da droga com o tempo é dada por:

$$\frac{d}{dt}(\text{Resposta}) = ACK_x. \quad (4.1)$$

Assumindo-se que o número dos sítios de reação é grande e que esses permanecem essencialmente constantes durante o ensaio, a probabilidade A segue uma distribuição normal.

$$A = a \exp^{-(\pi - \pi_0)^2/b}, \quad (4.2)$$

onde a e b são constantes, π_0 é a constante de reatividade de Hammet [60] para o benzeno e π é essa mesma constante, quando substitui-se um dos hidrogênios do benzeno por um grupo X.

Se a variação da resposta com o tempo for constante (*cte*), e substituindo-se A , da Equação 4.2, na Equação 4.1 têm-se:

$$a \exp^{-(\pi-\pi_0)^2/b} C K_x = cte. \quad (4.3)$$

Isolando-se a concentração do fármaco, e aplicando-se logaritmo, a expressão final fica:

$$\log 1/C = -a\pi^2 + b\pi + \log K_x + cte. \quad (4.4)$$

A Equação 4.4 descreve, matematicamente, o modelo de Hansch. A justificativa para a sua utilização vem da comparação com a termodinâmica [79].

$$\log \frac{1}{K_x} = \frac{1}{2,303RT} \Delta G \quad (4.5)$$

Comparando-se as Equações 4.4 e 4.5, percebe-se que a resposta da atividade biológica precisa ser constante para que o modelo de Hansch possa ser comparável à expressão da energia livre de Gibbs. Assim sendo, somente respostas constantes como IC_{50} ou LD_{50} são válidas. Neste trabalho se têm porcentagem de redução, o que não é próprio para análises quantitativas.

Outra questão importante é que, exaustivamente, tentaram-se formular modelos de PCA com todos os quarenta e seis compostos (Tabela 1.1); todavia, isso não foi possível, provavelmente devido às grandes diferenças existentes entre a estereoquímica dos compostos. Além disso, somente estruturas do tipo trans/cis são ativos biologicamente, o que leva à conclusão de que esta estereoquímica é condição necessária para a atividade dos compostos. Desta forma, todas as análises apresentadas daqui em diante são exclusivamente com compostos da família trans/cis. Considerando-se apenas esta família, têm-se treze compostos inativos, seis médios e dez ativos.²

4.2 Seleção de Variáveis pelo Método de PCA

4.2.1 Variáveis dos Métodos HF/6-31G** e HAM/3

Efetuiu-se a primeira PCA com os compostos da família trans/cis, sendo os compostos **2a**, **2b**, **3a** e **4a** (Tabela 1.1) excluídos devido à falta de parâmetros para o HAM/3, o que totaliza vinte e cinco compostos. O HAM/3 só tem parâmetros para o hidrogênio, carbono, nitrogênio, oxigênio e flúor. O objetivo foi utilizar o maior número de variáveis, calculadas pelos métodos HF/6-31G**//HF/3-21G, HAM/3//HF/3-21G e Surf//HF/3-21G, para testar possíveis correlações com a atividade biológica. As variáveis testadas foram:

1. HF/6-31G**//HF/3-21G;

(a) densidade de elétrons em HOMO sobre os trinta e oito átomos do esqueleto base;

²excluindo-se os enantiômeros (*l*)**2b**, (*d*)**10a**, (*d*)**10k** e (*l*)**12b** da página 57

- (b) cargas CHELPG sobre os trinta e oito átomos do esqueleto base;
- (c) momento de dipolo;
- (d) energia total;
- (e) energia dos dez últimos orbitais ocupados.

2. HAM/3//HF/3-21G;

- (a) cargas de Mulliken sobre os trinta e oito átomos do esqueleto base;
- (b) momento de dipolo;
- (c) calor de formação;
- (d) dez menores diferenças de energia entre orbitais;
- (e) dez menores energias de transição singlete;
- (f) dez menores energias de transição triplete;
- (g) dez menores potencial de ionização;
- (h) diferença de energia entre os dez potenciais de ionização;
- (i) dureza molecular;
- (j) dez menores eletroafinidades;
- (k) diferença de energia entre as dez eletroafinidades;
- (l) eletronegatividade de Mulliken;
- (m) ESCA dos átomos 6, 7, 8, 9 e 17 (Figura 3.1).

3. Estéricos (Surf//HF/3-21G).

- (a) volume de van der Waals;
- (b) superfície de van der Waals;
- (c) volume acessível ao solvente (água);
- (d) superfície acessível ao solvente (água);
- (e) coeficiente de partição octanol/água.³

Devido ao elevado número, 208 no total, as variáveis foram agrupadas da seguinte forma: 1(a), 1(b-c), 1(d-e), 2(a-c), 2(d-f), 2(g-i), 2(j-l), 2(m), 3(a-d) e 3(e). Isso é necessário, pois o processo de variável *versus* variável gera um elevado número de gráficos (Equação 4.6).

³calculado com o programa xlogp

$$\text{número de gráficos} = \sum_{i=1}^n (i - 1), \quad (4.6)$$

onde n é o número de variáveis. Neste caso ter-se-á de examinar 21944 gráficos, caso todas as variáveis tivessem sido graficadas umas pelas outras.

Para cada um desses dez grupos, graficaram-se as variáveis umas contra as outras, a exemplo do que é feito com as PCs na Figura 4.1, observando-se então quais variáveis separam os compostos quanto a sua atividade. Esse processo é útil para evitar que variáveis que não tenham correlação com a atividade biológica façam parte dos modelos de PCA, pois uma variável que tenha grande variância e nenhuma correlação com a atividade biológica pode ficar “ocupando espaço” nas primeiras componentes, dificultando assim a visualização dos dados. Deste processo selecionaram-se vinte variáveis entre todos os grupos. Com essas vinte variáveis realizaram-se a análise de componentes principais.

Graficando-se PC1 *versus* PC2, não se observou uma separação efetiva. Sendo assim, retiraram-se uma variável de cada vez, para observar sua influência no sistema. As variáveis que aparentemente não influenciavam o sistema foram então sendo retiradas, novamente observando-se o comportamento do sistema a cada passo. Nem todas variáveis puderam ser retidas; pois, mesmo não alterando o sistema quando retirada sozinha, alteravam-no quando da falta de alguma outra. A próxima etapa foi então tentar recolocar algumas das variáveis retiradas e retirar algumas daquelas que não influenciavam o sistema quando retirada sozinhas, mas que o perturbavam quando da falta de alguma outra. Neste ponto onze variáveis mostraram-se importantes.

O processo de variável *versus* variável não foi geral, devido a divisão nos dez grupos apresentados na página anterior. Desta forma, não fizeram-se muitas das combinações entre variáveis, e alguma importante pode não ter sido selecionada. Para amenizar isso, variáveis que tradicionalmente apresentam correlação com atividades biológicas, e variáveis que descrevem posições importantes das moléculas foram arbitrariamente adicionadas ao grupo das onze, tendo-se agora trinta e quatro variáveis. Com esse grupo, reiniciou-se o processo a partir dos gráficos de variável *versus* variável, seguindo-se até a PCA.

Esse processo, de graficar-se variável *versus* variável e seguido do teste e seleção na PCA, foi repetido por mais duas vezes, até que as variáveis que eram colocadas no sistema não mais permaneciam após a etapa do teste de retirar uma de cada vez na análise de PCA. Ao final, cinco foram as variáveis selecionadas, sendo elas cargas de Mulliken do HAM/3 para os átomos 17 (M17), 18 (M18) e 30 (M30); ESCA do átomo 8 (ESCA8) e volume de van der Waals (VDW). A Figura 4.3 apresenta a análise de componentes principais com estas cinco variáveis.

As variâncias capturadas foram 55,9 e 24 % para PC1 e PC2, respectivamente. Utilizando-se apenas PC1 separam-se todos os seis compostos ativos, quatro médios (**10m**, **10j**, **10l** e **12i**) e quatro

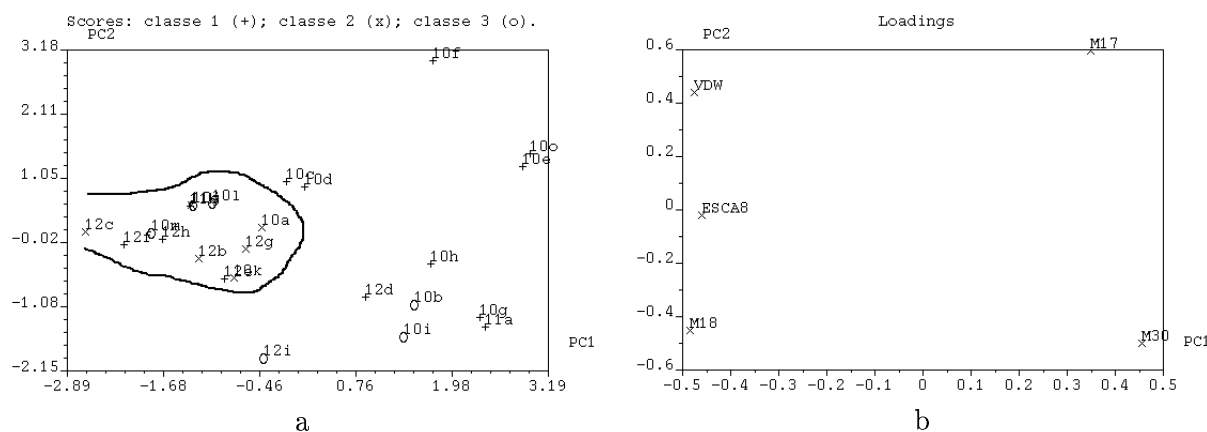


Figura 4.3: PCA com propriedades do HAM/3 e volume de van der Waals: a) *Scores*; b) *Loadings*.

inativos (**12f**, **12h**, **11b** e **12e**) dos demais, como pode ser visto no gráfico dos *scores* (Figura 4.3a). Da mesma forma que foi apresentado na Seção 4.1, podem-se determinar as características de cada grupo analisando-se o gráfico dos *loadings* (Figura 4.3b).

A principal separação realizada pela PC1 é entre compostos com ou sem carbonila em R₃, comportamento descrito basicamente pela variável M17. Entretanto, as demais variáveis apresentam *loadings* da mesma ordem de grandeza que a M17, demonstrando que a separação é mais complexa do que compostos com ou sem carbonila. Isso pode ser observado pela presença dos compostos **12i** no meio da PC1 e dos compostos **11b**, **11c** e **10j** no seu lado direito. Apesar da boa separação obtida, ficam três questões:

- por que as cargas de Mulliken funcionaram e as CHELPG não?
- as cargas de Mulliken calculadas por outros métodos apresentam o mesmo efeito?
- qual a correlação entre as variáveis?

Para responder a estas questões, as cargas de Mulliken para os átomos 17, 18 e 30 foram comparadas com as calculadas por outros métodos. Desta forma, as cargas de Mulliken e de Lowdin obtidas pelos métodos HF/3-21G e HF/6-31G**//HF/3-21G foram adicionadas ao conjunto de variáveis. Calcularam-se os coeficientes de correlação entre essas cargas, sendo a melhor correlação obtida com as cargas de Mulliken do HF/3-21G. A Tabela 4.2 apresenta a correlação, em unidades de elétron ao quadrado, entre as cargas do HAM/3 e do HF/3-21G.

Da Tabela 4.2 percebe-se uma boa correlação entre o método HAM/3 e o método HF/3-21G para os átomos 17 e 18; contudo, não há correlação para o átomo 30. Nenhum dos outros métodos reproduz as cargas descritas pelo HAM/3. Especificamente para a carga do átomo 30, a melhor

Tabela 4.2: Correlação entre cargas de Mulliken do HAM/3 e do HF/3-21G.

		HAM/3		
		17	18	30
HF/3-21G	17	0,812	-0,710	0,082
	18	-0,657	0,928	-0,098
	30	-0,114	0,248	0,187
HAM/3	17	1	-0,678	0,172
	18		1	-0,320
	30			1

correlação foi obtida com a carga CHELPG do HF/6-31G//HF/3-21G, com um coeficiente de 0,517. Mesmo com a boa correlação para as cargas dos átomos 17 e 18, a tentativa de substituí-las pelas do HF/3-21G não produziu boa separação. Na mesma tabela, observa-se que as cargas 17 e 18 do HAM/3 apresentam grande correlação; no entanto, todas as tentativas de construir modelos sem uma dessas variáveis não produziram melhores resultados que o original.

O modelo apresentado na Figura 4.3 é perfeitamente aceitável, uma vez que separa na primeira componente principal ativos de inativos, errando apenas quatro compostos, o que dá uma taxa de 16 % de erro.⁴ Todas as variáveis estão em posições importantes do esqueleto base,⁵ o que valida a utilização. A questão mais importante é que, mesmo que nenhuma das moléculas desse modelo apresente substituição na posição R₅ (Figura 3.1), as variáveis M30 e ESCA8 mostraram-se importantes na descrição da atividade biológica. Os dados apresentados para os compostos halogenados (R₅) corroboram com isso.

Apesar dos bons resultados, as variáveis desse modelo não servem para fazer previsão, pois os compostos mais ativos são aqueles que apresentam halogênios na posição R₅ e o método HAM/3 não é parametrizado para estes átomos, a exceção do flúor. Da mesma forma, propriedades de HF/6-31G** também não podem ser usadas, pois esta base só é definida até o elemento argônio. Assim, procedeu-se a um novo sistema de seleção de variáveis, no entanto utilizaram-se apenas variáveis do método HF/3-21G.

4.2.2 Variáveis do Método HF/3-21G

Como visto na Seção 4.2.1, o método HAM/3 só é parametrizado para o hidrogênio e mais quatro elementos do segundo período; além disso, a base 6-31G só é definida até o elemento argônio. Desta forma, utilizando-se HF/3-21G, os compostos **2a**, **2b**, **3a** e **4a** podem ser adicionados ao grupo da

⁴excetuando-se os compostos de média atividade

⁵posições próximas as regiões de substituição

Seção 4.2.1, o que totaliza vinte e nove compostos. As variáveis utilizadas nesta seção são descritas nos itens 1 e 3 da página 58, além dessas, as cargas de Mulliken e Lowdin foram incluídas ao grupo. Novamente, devido ao elevado número de variáveis — 169 desta vez — o processo de variável *versus* variável foi dividido nos mesmos grupos apresentados na página 59, além de dois novos grupos para as cargas de Mulliken e Lowdin.

O processo para a seleção de variáveis é o mesmo descrito na página 60, ou seja, procede-se ao variável *versus* variável, PCA das variáveis selecionadas, teste uma a uma no modelo de PCA, colocação de variáveis tradicionais e reinicia-se na etapa de variável *versus* variável. Após quatro ciclos, conseguiu-se um bom modelo, apresentado na Figura 4.4. As variáveis utilizadas foram cargas de Mulliken para os átomos 13 (M13) e 17 (M17), além do volume de van der Waals (VDW) apresentadas na Tabela 4.3. Na tabela, os compostos são apresentados em ordem decrescente de atividade.

Da Tabela 4.3 observa-se claramente que todos os compostos ativos têm volumes moleculares acima de $62 \text{ \AA}^3$ e, exceto pelos compostos **12g**, **10a** e **11c**, todos os ativos têm carga no átomo 17 (M17) em torno de $0,26 \text{ u.e.}$. A carga sob o átomo 13 não apresenta um comportamento bem definido por classe, contudo sua retirada do modelo não é possível; pois, o padrão de separação piora muito.

Na Figura 4.4a é apresentado o modelo de PCA, onde pode ser observado que os compostos ativos agrupam-se em uma linha reta ao lado direito do gráfico, já os compostos inativos e médios espalham-se por toda a figura. A coesão apresentada pelo compostos ativos está em concordância com a literatura [64], isso ocorre por que compostos ativos precisam ter características bem definidas, enquanto que os inativos podem apresentar este comportamento por razões diversas.

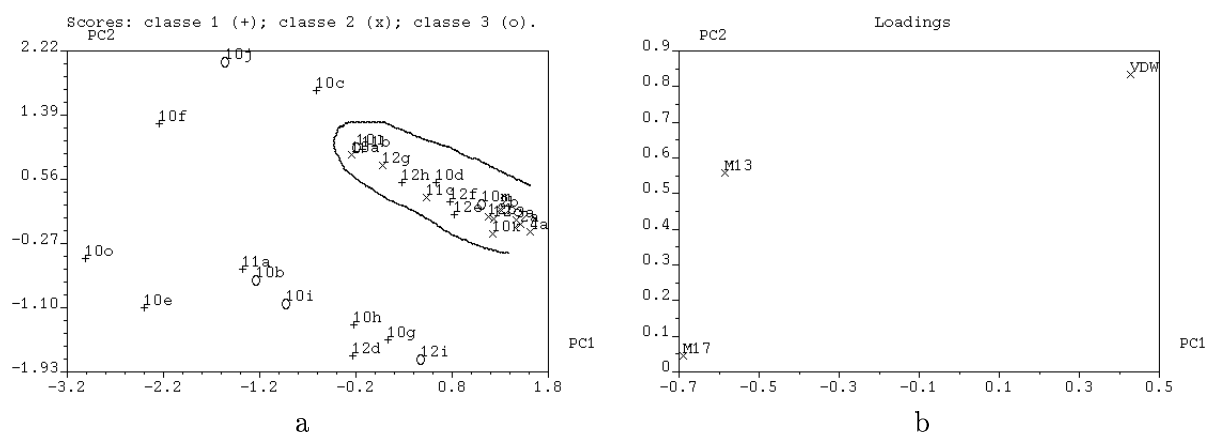


Figura 4.4: PCA com propriedades do HF/3-21G e volume de van der Waals: a) *Scores*; b) *Loadings*.

Adicionando-se os compostos **10c**, **10f** e **10j** ao conjunto assinalado na Figura 4.4a, têm-se todos os compostos com volumes moleculares acima de $62 \text{ \AA}^3$. No entanto, além de um volume molecular

Tabela 4.3: Propriedades do HF/3-21G selecionadas pela PCA.

nome.	M13 (u.e.)	M17 (u.e.)	VDW ($\text{\AA}^3$)
2a	-0,1858	-0,2668	65,31
2b	-0,1854	-0,2632	65,30
3a	-0,1857	-0,2669	65,31
4a	-0,1860	-0,2667	65,25
12b	-0,1853	-0,2638	63,98
12g	-0,1841	-0,1227	64,36
12c	-0,1854	-0,2635	64,13
10k	-0,1855	-0,2674	62,91
10a	-0,1838	-0,0754	64,28
11c	-0,1852	-0,0755	65,35
10b	-0,1836	-0,0754	48,94
10i	-0,1842	-0,0754	48,87
10l	-0,1838	-0,0755	65,13
12i	-0,1856	-0,2690	48,95
10j	-0,1813	-0,0678	64,09
10m	-0,1851	-0,2637	64,62
12h	-0,1845	-0,1221	64,16
12f	-0,1849	-0,2123	63,88
10g	-0,1851	-0,2449	48,88
11b	-0,1839	-0,0755	65,31
12e	-0,1850	-0,2261	62,88
12d	-0,1856	-0,0674	48,94
10e	-0,1845	0,3883	48,85
10f	-0,1832	0,4066	64,32
10o	-0,1832	0,3731	48,86
11a	-0,1833	-0,0755	48,96
10c	-0,1820	-0,2366	64,12
10d	-0,1844	-0,2413	64,00
10h	-0,1845	-0,2451	48,08

acima de $62 \text{ \AA}^3$, para um compostos posicionar-se no conjunto assinalado, ele precisa ter uma carga sob o átomo 17 abaixo de $0,075 \text{ u.e.}$, sendo esse o motivo pelo qual estes três últimos compostos não agrupam-se completamente com os ativos. A Figura 4.5 apresenta a ampliação do grupo ativo.

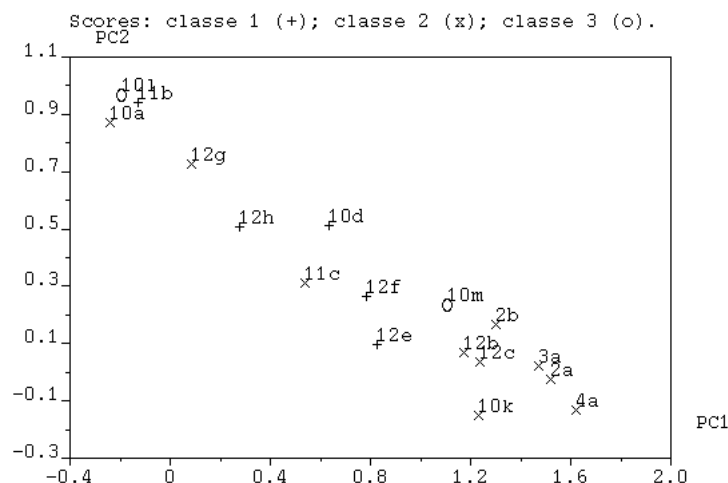


Figura 4.5: Ampliação do grupo dos ativos da Figura 4.4a.

Assim, 5 compostos inativos (**10d**, **11b**, **12e**, **12f** e **12h**) e 2 médios (**10l** e **10m**) foram classificados erradamente. A classificação dos compostos médios é decorrente da semelhança estrutural com os compostos ativos (Tabela 1.1). A classificação dos compostos **12e**, **12f** e **12h** pode ser resultado da influência do grupo amida na carga do átomo 17, entretanto este não é suficientemente reativo para produzir um composto ativo. A classificação do composto **11b** é curiosa. Pela comparação dos compostos **11b** e **11c** a única diferença estrutural observada é a troca, em R_1 , do grupo n -Pr por i -Pr, entretanto as atividades reportadas para 30 mg/Kg são 0 e 28 %, respectivamente. O mesmo vale para o composto **10d**, o qual apresenta propriedades quântica semelhantes as dos compostos ativos, entretanto a atividade reportada é de um composto inativo.

O mesmo comportamento das variáveis, apresentado na Tabela 4.3, pode ser observado graficamente pelos uso dos *Loadings* (Figura 4.4b), onde claramente as variáveis M17 e VDW contribuem de forma oposta na confecção das PC's. Como dito anteriormente, mesmo apresentando baixa variância, a variável M13 é de grande importância, como pode ser observado no mesmo gráfico.

Este modelo apresenta vantagens sobre o apresentado na Seção 4.2.1, pois:

- suporta compostos halogenados, devido à forma de cálculo utilizada;
- as propriedades são obtidas mais rapidamente, devido ao menor tamanho da base;
- utiliza apenas três variáveis para classificar as vinte e nove amostras;

- os erros são mais justificáveis, embora a porcentagem de erro seja de 17,2 %, o que é ligeiramente maior do que o modelo anterior.

4.3 SIMCA

O processo descrito na Seção 4.2.2 produziu, após extenso trabalho, três variáveis para a descrição da atividade biológica. As etapas seguintes são de construção de modelos de classificação para os compostos da família trans/cis. Sendo assim, começou-se pelo modelo de classificação mais parecido com PCA, o SIMCA.

As classes utilizadas são as mesmas apresentadas na página 57. Como se têm três variáveis, pode-se descrever cada grupo com uma ou duas componentes principais. A utilização das três PCs no modelo levaria a não existência da matriz de erro (Seção 2.2.8). Uma vez que os modelos de SIMCA são independentes entre si, cada classe pode ser apresentada separadamente.

Classe dos Inativos

Para a classe dos inativos, as variâncias de PC1 e PC2 são 42,86 e 40,05 %, respectivamente. A utilização de uma ou duas PCs modifica sensivelmente o modelo. Entretanto, não há distinção entre as atividades, sendo todos os compostos admitidos dentro do modelo para a classe dos inativos (Figura 4.6). Lembrando que, para modelos de SIMCA, um composto é considerado pertencente à classe que está sendo descrita, quando ele fica posicionado abaixo dos limites de confiança dos testes T^2 e Q .⁶ A Figura 4.6 apresenta os dois modelos.

Mesmo não apresentando seleção entre as classes, nos dois modelos têm-se o grupo dos ativos extremamente coeso. Ele fica junto à linha de T^2 , no modelo de uma PC (Figura 4.6a), e mais ao centro, para o modelo de duas PCs (Figura 4.6b).

Classe dos Ativos

Para a classe dos ativos, a variância de cada componente principal foi de 60,87 e 34,59 %, respectivamente, para PC1 e PC2. A Figura 4.7 apresenta os modelos com uma e duas PCs.

A principal observação é que o modelo para os compostos ativos não mais aceita todas as estruturas, nitidamente distinguindo-as quanto a sua atividade. Para o modelo com uma PC, todos os dez compostos ativos são classificados corretamente; além destes, o modelo inclui dois médios (**10l** e **10m**) e cinco inativos (**10d**, **11b**, **12e**, **12f** e **12h**) (Figura 4.7a). Já no modelo com duas PCs, o comportamento é bem semelhante, sendo que o composto ativo **11c** e o inativo **10d** ficam fora dos limites de confiança (Figura 4.7b).

⁶quadrante inferior esquerdo das Figuras 4.6, 4.7 e 4.8

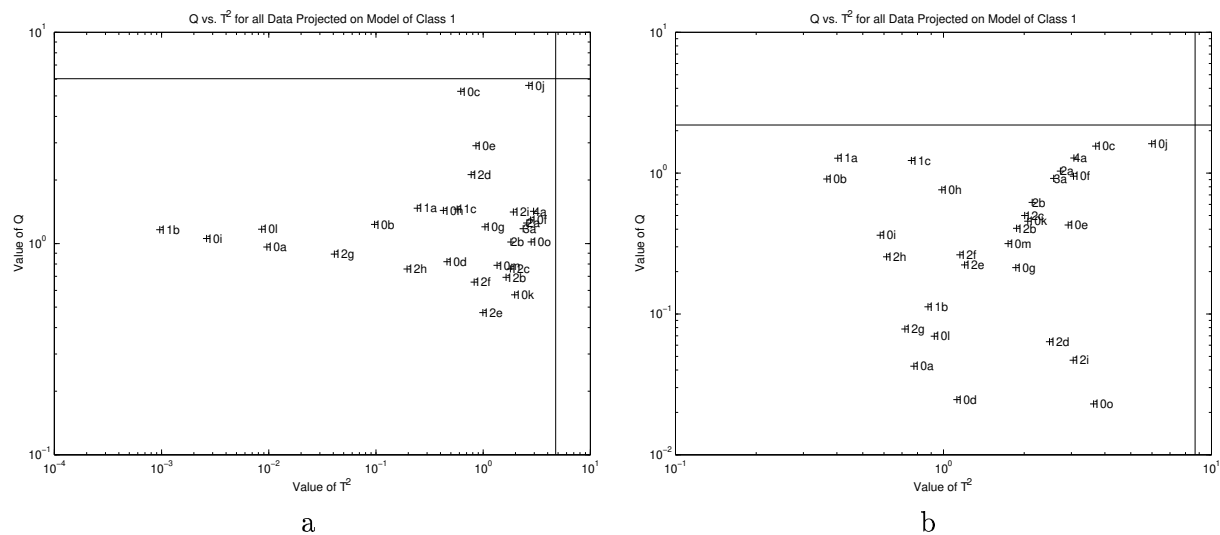


Figura 4.6: Modelos de SIMCA para classe dos inativos: a) uma PC; b) duas PCs.

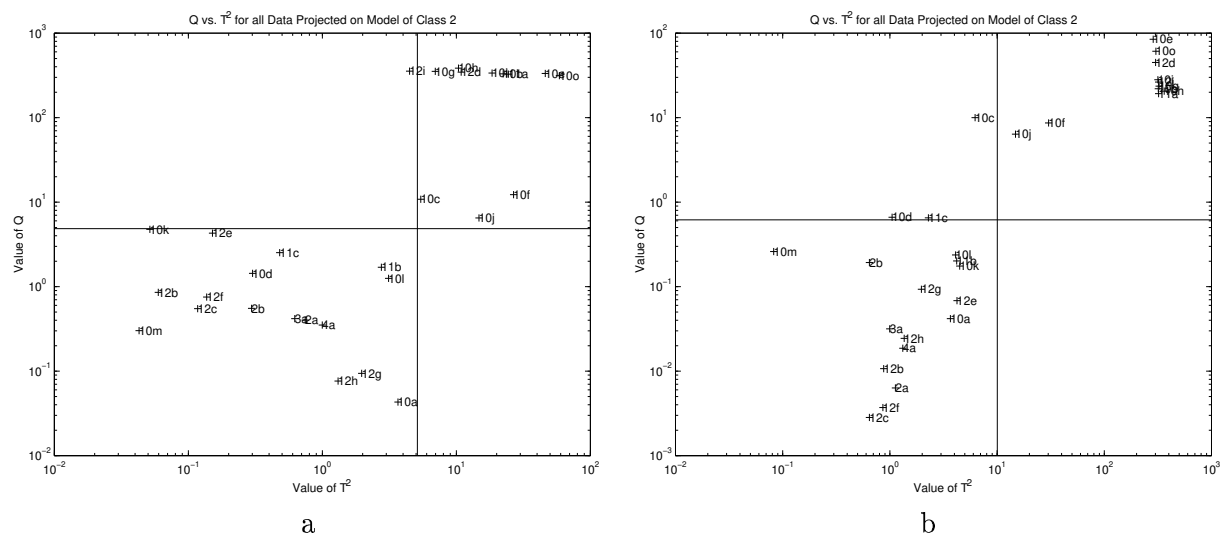


Figura 4.7: Modelos de SIMCA para classe dos ativos: a) uma PC; b) duas PCs.

Esse comportamento mais sensível ao número de componentes principais utilizadas era esperado, uma vez que os compostos ativos se agrupam ao longo de uma linha, como pôde ser visto na Figura 4.5, enquanto as classes inativa e média se espalham ao longo de toda a Figura 4.4. Os erros cometidos aqui são idênticos aos apresentados pela PCA da Seção 4.2.2. O fato do modelo dos ativos apresentar melhor resultado está em conformidade com a literatura [64], onde é relatado que é mais fácil modelar classes de compostos ativos do que inativos.

Classe dos Médios

Para a classe dos compostos de média atividade, a variância de cada componente principal foi de 61,51 e 32,78 %, respectivamente, para PC1 e PC2. A Figura 4.8 apresenta os modelos com uma e duas componentes principais. A exemplo do que foi observado para a classe dos compostos inativos, não há grande distinção quanto à atividade.

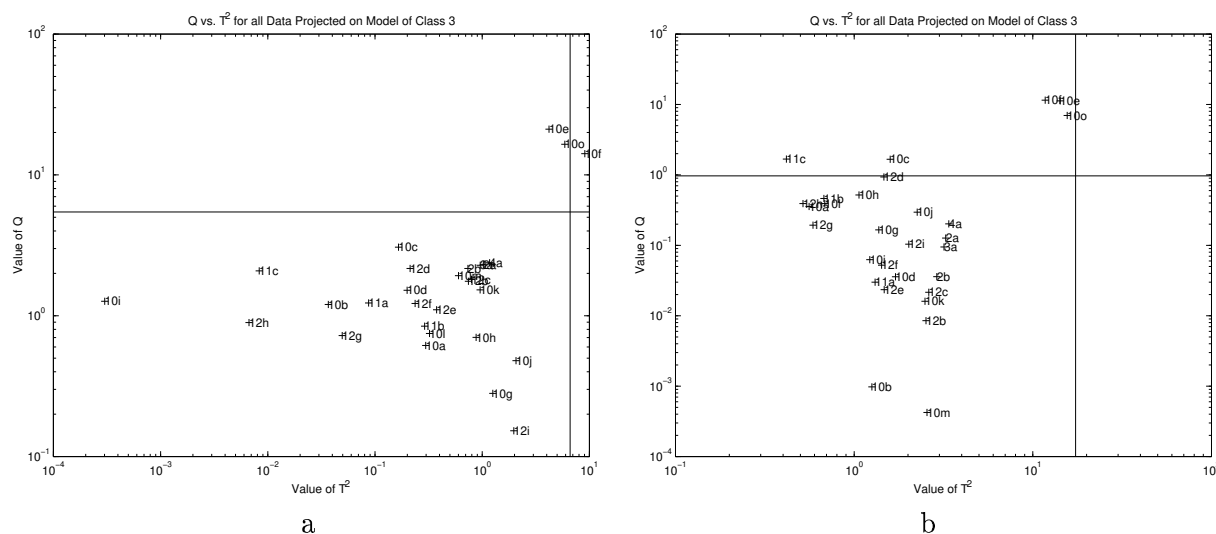


Figura 4.8: Modelos de SIMCA para classe dos médios: a) uma PC; b) duas PCs.

No modelo de uma PC (Figura 4.8a), somente três compostos inativos (**10e**, **10f** e **10o**) são excluídos dos limites de confiança, sendo todos os demais aceitos. No entanto, é interessante notar que, à exceção dos compostos **10a**, **11c** e **12g**, que necessitam ser metabolizados para apresentarem atividade, os compostos ativos ficam extremamente agrupados no canto superior direito da caixa dos limites de confiança. Para o modelo com duas componentes principais (Figura 4.8b) o comportamento é bem semelhante, apesar de, visivelmente, ocorrerem mudanças. Neste modelo, além dos três compostos inativos descritos anteriormente, os compostos **10c** e **11c** também ficam fora do modelo. Apesar desta relativa melhora, nenhum dos dois modelos é aceitável para classificação, a exemplo do que ocorreu para a classe do inativos.

A não adequação dos modelos de inativos e médios pode ser explicada pela dificuldade que o SIMCA tem de descrever dados assimétricos, sendo o KNN um método mais adequado para esse fim [66].

4.4 KNN

Da mesma forma que os modelos de SIMCA, as variáveis utilizadas aqui são as determinadas na Seção 4.2.2, e as classes são as apresentadas na página 57. Realizaram-se alguns testes no conjunto com duas classes,⁷ antes da elaboração do modelo final, sendo eles:

- matriz de votos (Seção 2.2.9) feita com base nas ligações pelos dois vizinhos mais próximos ou pelo centróide;⁸
- tratamento dos dados com PCA e medida das distâncias pelos *scores*.

No primeiro teste as distâncias euclidianas foram calculadas em relação aos dois vizinhos mais próximos ou em relação ao centróide, para a determinação das similaridades. Ambas as formas apresentam os mesmos cinco erros (**10d**, **11b**, **12e**, **12f** e **12h**), contudo a ligação pelos dois vizinhos mais próximos é mais clara, com grupos melhor definidos.

O segundo teste foi para verificar a melhor forma de medir as distâncias, se pelas variáveis originais ou pelos *scores* da PCA. Utilizando-se uma componente principal se têm os melhores resultados, pois somente três dos compostos ativos (**10a**, **11c** e **12g**) ficam de fora, sendo estes erros tradicionais devido a processos de metabolização.

Assim, utilizando-se ligações pelos dois vizinhos mais próximos, com distâncias euclidianas nos *scores* da PCA, obteve-se o melhor resultado com a classificação correta de sete dos dez compostos ativos (70 % de acerto para os ativos), sendo esse modelo apresentado na Figura 4.9.

Analisando-se a figura de cima para baixo, o primeiro *cluster* observado é o que reúne os compostos **10o**, **10e** e **10f**. Os compostos deste *cluster* se caracterizam pela presença de um elemento muito eletronegativo (O e F), ligado a R₃, sendo estes os únicos compostos com valores positivos para M17. O segundo grupo, compreendido entre o composto **2b** e o **2a**, apresenta 7 dos compostos ativos, além do médio **10m**. A não inclusão dos ativos **10a**, **11c** e **12g**, muito provavelmente é resultado do processo de metabolização sofrido por estes compostos no meio biológico. Como os dados experimentais são resultados de testes *in vivo*, a oxidação do grupo R₃ dever ser obtida rapidamente, já as propriedades quânticos não são capazes de modelarem isso. Do composto **11b** ao **12e** formam-se 3 pequenos *clusters* que acabam por agrupar-se para formar um conjunto de inativos. O

⁷exclusão do grupo dos compostos com média atividade

⁸centro de massa dos pontos

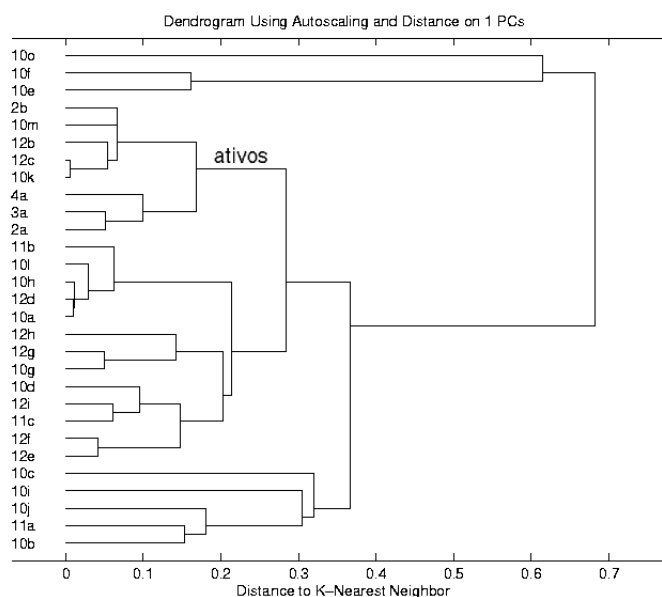


Figura 4.9: Modelo de KNN.

últimos *cluster* compreendido entre o **10c** ao **10b** aparentemente não tem um padrão bem definido, apresentando uma baixa similaridade com relação aos demais compostos.

Como descrito na literatura [66], o KNN apresentou melhores resultados que o SIMCA para a classificação dos compostos, sendo que esse deve ser preferido ao SIMCA.

4.5 Redes Neurais

A utilização da rede neural deu-se primeiramente nos mesmos moldes da PCA da Seção 4.2.2, ou seja, os vinte e nove compostos e três classes da família trans/cis. Vários testes foram realizados, alterando-se o número de neurônios da segunda camada (entre cinco e doze) e o limite de erro entre ciclos (entre 0,01 e 0,001). O melhor resultado foi obtido com cinco neurônios e um limite de erro de 0,001 (Tabela 4.4), sendo esse apresentado na Tabela 4.5.

Neste modelo, ocorrem nove classificações equivocadas (**10a**, **10b**, **10l**, **10m**, **11a**, **11b**, **11c**, **12g** e **12i**), o que dá um 69 % de acerto. Dos oito erros, quatro são de compostos com média atividade, sendo que o **10l** e o **10m** já eram classificados de maneira errada na PCA. À exceção do **11a**, **11c** e do **12g**, que são ativos e foram classificados como inativos, todos demais erros são trocas entre as classes média e inativa ou média e ativa. Isso já era esperado, uma vez que os compostos médios se distribuíram entre as classes inativa e ativa (Figura 4.4) durante a seleção das variáveis. É interessante notar que o termo classificação, utilizado na PCA, não é o mais apropriado, já que

não foi realizado um processo de validação, a exemplo do *leave one out*⁹ realizado para RN.

Apesar de se considerarem aceitáveis os erros cometidos no modelo de três classes, procurou-se melhorá-lo retirando-se os compostos ditos médios, tendo-se agora vinte e três compostos. Com esse processo espaçam-se mais os grupos, o que leva, em geral, a um melhor modelo de previsão. Novamente, vários testes foram realizados, sendo o melhor resultado obtido com oito neurônios na segunda camada e limite de erro de 0,001 (Tabela 4.4). Neste modelo (Tabela 4.5), ocorreram sete erros (**10a**, **10c**, **10f**, **11b**, **11c**, **12g** e **12h**), resultando em uma porcentagem de acerto de 69,6 %. Desta vez, dois erros são coincidentes com os da PCA; entretanto, nenhum outro padrão foi observado. Primeiramente, os três compostos ativos (**10a**, **11c** e **12g**) são erros tradicionais devido aos processos de metabolização. Exceto pelo composto **12h**, os demais não deveriam ter sido classificados equivocadamente.

As estruturas e os parâmetros para os dois melhores modelos são apresentados na Tabela 4.4, cujos parâmetros α , ϵ e θ são descritos na Seção 2.2.10.

Tabela 4.4: Estrutura e parâmetros das redes neurais.

camada	3 classes				2 classes			
	neurônios	α	ϵ	θ	neurônios	α	ϵ	θ
1	4				4			
2	5	1,0	0,001	0,0	8	1,0	0,001	0,0
3	3	1,0	0,001	0,0	2	1,0	0,001	0,0

Tabela 4.5: *Leave one out* das redes neurais.

nome	padrão ¹			previsto			padrão ²		previsto	
	1	2	3	1	2	3	1	2	1	2
2a	0	1	0	0,000	1,000	0,014	0	1	0,002	0,998
2b	0	1	0	0,031	0,954	0,106	0	1	0,004	0,996
3a	0	1	0	0,000	1,000	0,005	0	1	0,002	0,998
4a	0	1	0	0,011	0,999	0,043	0	1	0,002	0,998
10a	0	1	0	0,126	0,000	0,577	0	1	0,619	0,382
10b	0	0	1	0,619	0,007	0,486				
10c	1	0	0	0,946	0,000	0,333	1	0	0,000	1,000
10d	1	0	0	0,949	0,002	0,211	1	0	0,955	0,045
10e	1	0	0	0,999	0,016	0,000	1	0	0,995	0,005
10f	1	0	0	0,614	0,190	0,376	1	0	0,348	0,659
10g	1	0	0	0,898	0,000	0,371	1	0	0,998	0,002

continua na próxima página

⁹constrói-se o modelo desconsiderando-se um dos compostos, depois tenta-se prever o comportamento deste composto

Tabela 4.5: *Leave one out* das redes neurais (continuação).

nome	padrão ¹			previsto			padrão ²		previsto	
	1	2	3	1	2	3	1	2	1	2
10h	1	0	0	0,993	0,000	0,069	1	0	0,987	0,012
10i	0	0	1	0,042	0,003	0,770				
10j	0	0	1	0,007	0,718	0,913				
10k	0	1	0	0,004	0,903	0,087	0	1	0,031	0,968
10l	0	0	1	0,635	0,468	0,145				
10m	0	0	1	0,447	0,697	0,000				
10o	1	0	0	1,000	0,001	0,000	1	0	0,992	0,007
11a	1	0	0	0,001	0,031	0,957	1	0	0,931	0,059
11b	1	0	0	0,124	0,436	0,119	1	0	0,001	0,999
11c	0	1	0	1,000	0,038	0,000	0	1	0,963	0,036
12b	0	1	0	0,045	0,691	0,276	0	1	0,048	0,952
12c	0	1	0	0,041	0,844	0,225	0	1	0,010	0,990
12d	1	0	0	0,700	0,610	0,119	1	0	0,991	0,008
12e	1	0	0	0,674	0,036	0,149	1	0	0,886	0,114
12f	1	0	0	0,862	0,043	0,045	1	0	0,825	0,176
12g	0	1	0	0,508	0,000	0,244	0	1	0,980	0,020
12h	1	0	0	0,957	0,169	0,011	1	0	0,336	0,666
12i	0	0	1	1,000	0,000	0,001				

¹ativos (0 1 0); inativos (1 0 0); médios (0 0 1).²ativos (0 1); inativos (1 0).

Desta seção observa-se que o método de rede neural apresentou um bom desempenho, principalmente por se considerar o processo de validação *leave one out*. Desta forma, esse método é indicado para a classificação de novos compostos, como será apresentado no próximo capítulo.

Capítulo 5

Novos Compostos

Como apresentado no Capítulo anterior, as principais posições de substituição são a R_3 e R_5 . Nessas posições, os melhores resultados foram obtidos com compostos carboxílicos em R_3 e halogênios em R_5 . Baseada nesses dados, a proposta de novos compostos é a substituição, nas posições R_3 e R_5 , da carboxila e dos halogênios por seus bioisósteres.

Bioisósteres são grupos que possuem similaridade química e/ou física e que produzem propriedades biológicas semelhantes. Dentre os bioisósteres, existem os clássicos e os não-clássicos. Os clássicos são aqueles grupos que apresentam estrutura ou valência similares, já para os não-clássicos, o número de ligações e efeitos estereoeletrônicos são diferentes [80].

Os bioisósteres não-clássicos para a carboxila são grupos fosfato, sulfanamida e o sulfônico. Para os halogênios, pode-se utilizar os grupos trifluormetano ou ciano [81]. Além da substituição com base no bioisosterismo, utilizaram-se também os grupos *ter*-butila e nitro. A *ter*-butila se destina para avaliação do volume molecular na estrutura; já o nitro se destina para mimetizar os oxigênios da carboxila. A Figura 5.1 e a Tabela 5.1 apresentam estas propostas.

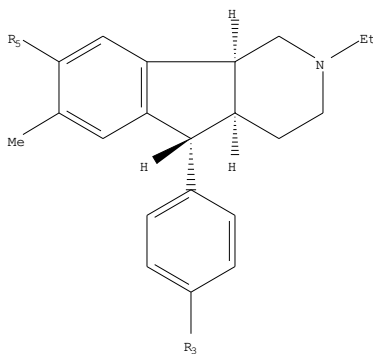
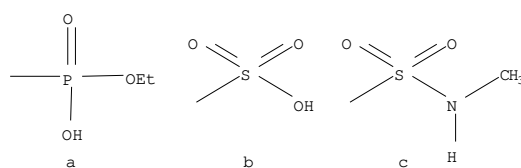


Figura 5.1: Esqueleto base dos novos compostos.

A exemplo da Seção 3.1, a otimização dos novos derivados foi realizada com HF/3-21G. Como

Tabela 5.1: Proposta para novos compostos

composto	R ₃	R ₅
12b-NO₂	NO ₂	H
12b-tBu	COOt-Bu	H
2a-NO₂	NO ₂	I
2a-tBu	COOt-Bu	I
2a-H	H	I
2a-F	COOH	F
2a-CN	COOH	CN
2a-CF₃	COOH	CF ₃
2a-P	Figura 5.2a	I
2a-S	Figura 5.2b	I
2a-SN	Figura 5.2c	I

Figura 5.2: Substituintes em R₃ para os novos compostos.

descrito na Seção 4.2.2, as propriedades quânticas de interesse são as cargas de Mulliken sobre os átomos 13 e 17. Além dessas, o volume de van der Waals também se mostrou importante. Dos diversos métodos estatísticos utilizados, o KNN foi o que apresentou um dos melhores resultados (Seção 4.4), sendo esse utilizado para a previsão da atividade dos compostos propostos. Além do KNN, o método de redes neurais (Seção 4.5) também apresentou bons resultados, qualificando-se, assim, para a classificação dos novos compostos.

5.1 KNN

Como apresentado na Seção 4.4, o KNN apresentou bons resultados para a classificação. A Figura 5.3 apresenta a inclusão dos onze novos compostos (Tabela 5.1) no modelo de classificação do KNN apresentado na Figura 4.9.

Analisando-se a Figura 5.3 de cima para baixo, vê-se que do composto **2a-S** ao composto **12b-tBu** há formação de um *cluster* completamente distinto dos demais, o qual não existia no conjunto de treinamento apresentado na Figura 4.9. Isso é um forte indício de que as alterações feitas no esqueleto da Figura 5.1 foram acima do suportado pelo modelo. Desta forma, não é possível afirmar a correta atividade desses compostos.

Os compostos **2a-CN**, **2a-CF₃** e **2a-F** inserem-se perfeitamente no *cluster* dos compostos ativos,

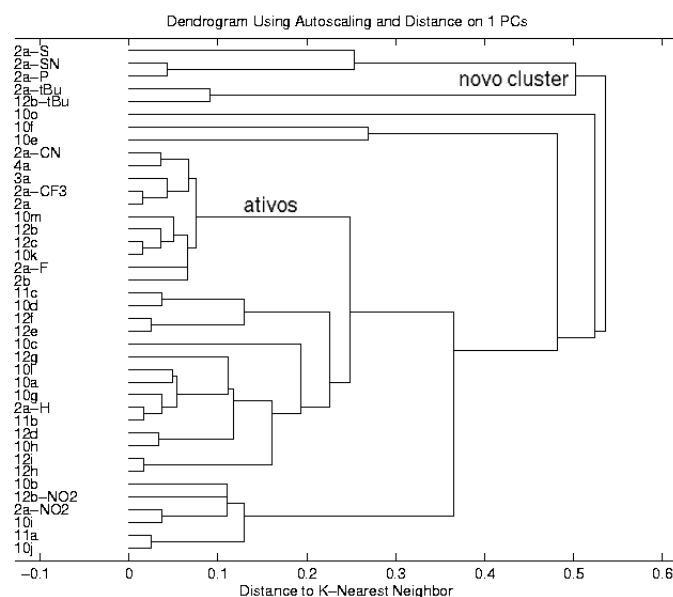


Figura 5.3: Modelo de KNN para os compostos propostos.

previamente formado na fase de construção do modelo na Seção 4.4. Esses compostos são produtos das substituições, em R₅, do iodo por bioisósteres, o que realmente deveria produzir compostos ativos.

O composto **2a-H** é formado pela troca da carboxila por hidrogênio, no R₃ do composto **2a**. Como esperado, a retirada da COOH de R₃ produziu um composto inativo, como pode ser observado pelo posicionamento apresentado na Figura 5.3.

Os compostos **2a-NO₂** e **12b-NO₂** inseriram-se em um *cluster* formado por três compostos médios (**10b**, **10i** e **10j**) e um inativo (**11a**). Apesar do grupo NO₂ ser um retirador, a exemplo da carboxila, este fato não é suficiente para posicioná-lo no grupo dos ativos, mostrando assim que a presença da carboxila é de extrema importância para uma alta atividade.

5.2 Redes Neurais

Como descrito na Seção 4.5, a rede neural apresentou um bom desempenho no processo de validação *leave one out*, qualificando-se assim como um método próprio para a previsão dos novos compostos. Os dois melhores modelos de redes neurais apresentados na Tabela 4.5 foram utilizados para prever os onze novos compostos. A estrutura e os parâmetros da rede são os mesmos descritos na Tabela 4.4.

A exemplo do que ocorreu no modelo de KNN, alguns dos novos compostos têm variáveis que

estão fora do intervalo descrito pelas variáveis do conjunto de treinamento; desta forma, conclusões sobre esses compostos devem ser evitadas. Os compostos que apresentaram esse problema foram **12b-tBu**, **2a-NO₂**, **2a-P**, **2a-SN**, **2a-S** e **2a-tBu**. Excetuando-se o **2a-NO₂**, todos os demais compostos fazem parte do *cluster* isolado apresentado na Figura 5.3. Para esses compostos, nenhuma afirmação é possível.

Na Tabela 5.2 são apresentadas as classificações feitas nos dois modelos de rede neural.

Tabela 5.2: Previsão das redes neurais para os compostos propostos.

nome	atividade ^a	previsto			atividade ^b	previsto	
		1	2	3		1	2
12b-NO₂	ativo	0,002	0,775	0,000	inativo	0,998	0,002
2a-CF₃	ativo	0,000	1,000	0,004	ativo	0,001	0,999
2a-CN	ativo	0,000	1,000	0,006	ativo	0,001	0,999
2a-F	ativo	0,000	0,999	0,011	ativo	0,002	0,998
2a-H	inativo	0,989	0,000	0,087	inativo	0,998	0,002

^aativos (0 1 0); inativos (1 0 0); médios (0 0 1)

^bativos (0 1); inativos (1 0)

Dos cinco compostos restantes (Tabela 5.2), somente o **12b-NO₂** apresenta uma classificação diferente nos dois modelos de rede neural (ativo e inativo), além disso, nenhuma dessas classificações coincide com a descrita pelo KNN (médio).

Para os demais compostos, as classificações da rede neural e do KNN coincidem perfeitamente, sendo considerados ativos os compostos **2a-CF₃**, **2a-CN** e **2a-F** e inativo o composto **2a-H**.

Neste capítulo, vê-se que é possível a previsão de novos compostos potencialmente mais ativos como contraceptivos orais masculinos, desde que esses novos compostos estejam dentro do limite de segurança do modelo. Assim, propões-se os compostos **2a-CF₃**, **2a-CN** e **2a-F** para síntese.

Capítulo 6

Conclusões

- As análises conformacionais (Seção 3.1.1) sistemática e estocástica apresentam os mesmos resultados quanto aos mínimos locais para o esqueleto base para a família trans/cis;
- Os métodos *ab initio* do tipo HF e DFT concordam completamente quanto à estabilidade relativa entre as conformações do esqueleto base para a família trans/cis (Seção 3.1.1).
- Analisando-se a variação da geometria da conformação cadeira 1, quando varia-se a função de base do cálculo HF (apresentado na página 41), conclui-se que essa não é significativa;
- Algumas das propriedades estudadas variam com a conformação (Seção 3.5.1); no entanto, isso não impediu a confecção dos modelos estatísticos;
- As cargas atômicas, pelas metodologias CHELPG e Mulliken, variam com a função de base utilizada (Seção 3.5.2); todavia, mais uma vez, isso não impossibilitou a confecção dos modelos estatísticos;
- A PCA separa eficientemente os compostos ativos dos inativos, como pôde ser atestado na Seção 4.1;
- O conceito de dose limite (Figura 4.2) mostrou-se perfeitamente adequado para a descrição da atividade biológica dos compostos;
- A sistemática de se graficar variável *versus* variável, utilizada conjuntamente com PCA, é uma eficiente forma de seleção de variáveis para modelos de SAR, como pode ser evidenciado nas Seções 4.2.1 e 4.2.2;
- Os processos de metabolização dos compostos **10a**, **11c** e **12g**; devem ser os principais repon-sáveis pelos erros cometidos na PCA;

- Os métodos HF, HAM/3 e Surf são capazes de produzir variáveis que apresentam correlação com a atividade biológica;
- O método SIMCA (Seção 4.3) só apresentou bons resultados na modelagem dos compostos ativos;
- O método KNN (Seção 4.4) apresentou bons resultados na classificação de todos os compostos da família trans/cis;
- O método de Redes Neurais (Seção 4.5) apresentou bons resultados na validação *leave one out*, qualificando-se, assim como o KNN, para a previsão dos novos compostos;
- Apesar de restrições quanto ao grau de modificação das estruturas, os modelos de KNN e Redes Neurais apresentaram a classificação esperada para os compostos propostos;
- A utilização de métodos quânticos em conjunto com métodos estatísticos é capaz de produzir eficientes modelos de SAR.

Apêndice A

Programa de PCA

Da mesma forma que no apêndice B, abaixo são apresentados pequenos trechos do programa para análise de componentes principais desenvolvido para Scilab.

```
function [T,V,S,aval,varp,varpa]=acp(X,rotobj,rotvar,class);
// Algoritmo para análise PCA utilizando decomposição de valor singular (SVD)
// Wold, S.; Esbensen, K.; Geladi, P. Chemometrics and Intelligent Laboratory
// Systems 1987, 2, 37-52
// I/O [scores,loadings,autovalores,variância_percentual,
// variância_perc_acumulada]=acp(X,rótulo_objetos,rótulo_variáveis,classes)
// Para utilizar rótulos do programa coloque 1 no lugar da variável.

// $Id: apend.tex,v 1.6 2001/08/06 14:45:59 msegala Exp $
[...]
// Pré-tratamento
[k] = x_choose(['centrar na média;','autoescalar;','escalar pela
variância;','escalar pela variação.'], ['Tipo de pré-tratamento:'],
'sem pré-tratamento');

if k == 1
    X = cmedia(X);
elseif k == 2
    X = autosc(X);
elseif k == 3
    X = scvar(X);
elseif k == 4
    X = scrange(X);
end

// Decomposição de valor singular (SVD)
// Gilbert Strang, Linear Algebra and its Application, Third Edition;
// Harcourt Brace Jovanovich: San Diego, 1988 (Appendix A).
if ((n-m) > 100)
    // Caso número de colunas X >> números de linhas X (PCA kernel)
    [U S V]=svd(X*X',0);
    V=X'*V;
```

```

for i=1:size(V,2)
    V(:,i)=V(:,i)/norm(V(:,i));
end
T=X*V;
else
    // (PCA Tradicional)
    [U S V]=svd(X,0);
    T=U*S;
end

// Autovalores
aval=diag(S.^2);
[...]
// Controle para as subrotinas
l = 10;
while l ~= 0
    [l] = x_choose(['Scores e Loadings (Draftsman)'; 'Scores e Loadings';
    'Variáveis x Loadings'; 'Biplots'], ['Menu Principal'], 'Fim do programa');

    // Visualização dos dados em forma de escada
    if l == 1
        // dos scores
        xset("window",cont_win);
        if size(T,2) <= 8
            fun_escada(T,'Scores')
        else
            fun_escada(T,'Scores (8 primeiras PCs)')
        end
        cont_win = cont_win + 1;

        // dos loadings
        xset("window",cont_win);
        if size(V,2) <= 8
            fun_escada(V,'Loadings')
        else
            fun_escada(V,'Loadings (8 primeiras PCs)')
        end
        cont_win = cont_win + 1;

        // Scores e Loadings para cada par de PC
    elseif l == 2
        cont_win = scor_load(X,T,V,class,rotobj,rotvar,in,posto,cont_win);

        // Variáveis x Loadings para cada PC de interesse
    elseif l == 3
        cont_win = grafpc(X,V,rotobj,rotvar,posto,cont_win);

        // Biplot (Scores e Loadings)
    elseif l == 4
        cont_win = fun_biplot(X,S,V,U,rotobj,rotvar,posto,cont_win);
    end
end
end

```

```

//////////////////////////////////// SUBROTINAS //////////////////////////////////////
[...]
// Scores e Loadings
function [cont_win]=scor_load(X,T,V,class,rotobj,rotvar,in,posto,cont_win)
[...]
    // Janela de opção para as PC's
    pcpar1 = ('PC'+string((1:posto)'));
    pcx = x_choose(pcpar1,'Componente do eixo X','Menu Principal');
    // Interrupção do loop caso Menu Principal seja selecionado
    if pcx == 0
        break;
    end
    pcy = x_choose(pcpar1,'Componente do eixo Y','Menu Principal');
    if pcy == 0
        break;
    end
[...]
    // Plot dos Scores com classes
    xset("window",cont_win)
    xset('font',0,rep(1,1))
    X2 = [T(:,pcx) T(:,pcy)];
    fun_class(X2,class,pcx,pcy);
    if rep(1,2) == 1
        for i=1:m
            xstring(T(i,pcx),T(i,pcy),rotobj(i,:))
        end
    end
    cont_win = cont_win + 1;

    // Plot dos Loadings
    xset("window",cont_win)
    xset('font',0,rep(1,1))
    xset("mark",-2,3)
    plot2d(V(:,pcx),V(:,pcy),[-2])
    xtitle(['Loadings'],['PC'+string(pcx)],['PC'+string(pcy)])
    if rep(1,3) == 1
        for i=1:n
            xstring(V(i,pcx),V(i,pcy),rotvar(i,:))
        end
    end
    cont_win = cont_win + 1;
end
[...]

```

Na primeira linha vê-se a definição da função e suas variáveis de entrada e saída. Em seguida vem as opções de pré-tratamento dos dados e o SVD, sendo que após as matrizes “V” e “T” são montadas¹. Tem-se então disponíveis as formas que visualização dos dados, as quais serão realizadas por diversas sub-rotinas, sendo que a única sub-rotina apresentada é a que grafita os *scores* e *loadings*.

¹conforme a PCA utilizada

Apêndice B

Programa em PERL

Abaixo são apresentados pequenos trechos do programa para a retirada de propriedades do GAMESS. As propriedades implementadas são apresentadas na página 58 e o programa funciona com o conceito de expressão regular.¹

```
#!/usr/bin/perl
# $Id: apend.tex,v 1.6 2001/08/06 14:45:59 msegala Exp $
# Programa para retirada de propriedades do GAMESS

# Help das opcoes para a linha de comando:
if ($opt_h) {
    print <<EOF;
Uso: prop_gms.pl [-nevh] saida_gameess > arquivo
    -n  retira cargas e densidade do atomos 1 ateh n;
    -e  necessario quando for um calculo de energia;
    -v  modo verboso;
    -h  help.
EOF
    exit 0;
}
[...]
# Procurando linha a linha no arquivo_saida_gameess.out
while ($line=<>) {
    chop($line);
    [...]
    # Determinando a regioao dos orbitais moleculares
        if (($line =~ /MOLECULAR\sORBITALS/) || ($line =~ /EIGENVECTORS/)) {
            $reg_mol_orb = 1;
        }

    # Entrando na regioao dos orbitais moleculares
        if ($reg_mol_orb == 1) {
    [...]
    # Retirando as energia determinadas pelo HOMO
```

¹expressão posicionada junto a uma propriedade desejada

```

        if ($energias_homo == 1) {
            $_ = $line;
            @ENER_HOMO_LINHA = /^s+(.+\.d+)\s+(.+\.d+)\s+(.+\.d+)\s+(.+\.d+)\s+(.+\.d+)/;
            $energias_homo = 0;
#            print "HOMO (Energias): @ENER_HOMO_LINHA \n";

# Retirando as energias de interesse do vetor ENER_HOMO_LINHA
            $i = 0;
            foreach $posicao (@POSICAO_HOMO) {
                $ENERGIAS_HOMO_INTERESSE[$i] = $ENER_HOMO_LINHA[$posicao];
#                print "HOMO (Energias de interesse): @ENERGIAS_HOMO_INTERESSE \n";
                $i++;
            }
        }

# Achou o HOMO
        if ($line =~ /^.+s\s$homo/) {
            $_ = $line;
            @No_HOMO = /^s+(\d+)\s+(\d+)\s+(\d+)\s+(\d+)\s+(\d+)/;
            $energias_homo = 1;
#            print "HOMO (Numeros): @No_HOMO \n";

# Determinando as posicao de interesse no vetor No_HOMO
            $i = 0;
            $chave = 1;
            while ($chave > 0) {
                $POSICAO_HOMO[$i] = (($homo - $i) - $No_HOMO[0]);
                $chave = $POSICAO_HOMO[$i];
#                print "HOMO (Posicoes de interesse): @POSICAO_HOMO \n";
                $i++;
            }
        }

        if ($line =~ /ENERGY\sCOMPONENTS/) {
            $reg_mol_orb = 0;
            $reg_ener_com = 1;
        }
    }

[...]
```

Imprimindo a matriz de resultados

```

if (!$opt_v) {
    print "$tot_ener $elec_den $debye $chelpg $ener_interesse_ev $mulliken $lowdin \n";
}

```

Nas primeiras linhas vê-se a versão do programa e suas opções para a linha de comando. Depois ele abre o arquivo de interesse e procura as expressões regulares, como por exemplo “MOLECULAR ORBITALS”. Daí para frente é retirar a energia do HOMO e alguns outros orbitais próximos e finalmente imprimir o resultado. Este mesmo processo é repetido para todas as propriedades.

Referências Bibliográficas

- [1] Falkenberg, T.; Tomson, G. *Health Policy and Planning* **2000**, *15*, 52-58.
- [2] Cohen, N. C. *Guidebook on Molecular Modeling in Drug Design*; Academic Press: New York, 1996.
- [3] Csizmadia, I. G.; Arnaut, L. G.; Formosinho, S. J. *Theoretical and Computational Models for Organic Chemistry*; Kluwer Academic Publishers: Dordrecht, 1991.
- [4] Moye-Sherman, D.; Jin, S.; Ham, I.; Lim, D. Y.; Scholtz, J. M.; Burgess, K. *J. Am. Chem. Soc.* **1998**, *120*, 9435-9443.
- [5] Liang, G. Y.; Chen, X. N.; Dustman, J. A.; Lewin, A. H.; Bowen, J. P. *J. Comp. Chem.* **1997**, *18*, 1371-1391.
- [6] Costa, M. C. A.; Takahata, Y. *J. Comp. Chem.* **1997**, *18*, 712-721.
- [7] “Unicef faz sombrio balanço do subdesenvolvimento no mundo”, Jornal Correio Popular, Campinas - SP, 21 dezembro de 1997.
- [8] “Fantasma da fome atormenta os EUA”, Correio do Povo, Porto Alegre - RS, 13 de fevereiro de 2000.
- [9] “Instituto Brasileiro de Geografia e Estatística”, <http://www.ibge.gov.br>, 2000.
- [10] “World Health Organization”, <http://www.who.int/rht>, 1999.
- [11] Oga, S.; Zanini, A. C. *Farmacologia Aplicada*; Editora São Paulo: São Paulo, 2^a ed.; 1982.
- [12] Buchter, D.; von Eckardstein, S.; von Eckardstein an A. Kamischke, A.; Simoni, M.; Behre, H. M.; Nieschlag, E. *J. Clin. Endocrinol. Metab.* **1999**, *84*, 1244-1249.
- [13] Wu, F. C. W.; Balasubramanian, R.; Mulders, T. M. T.; Coelingh-Bennink, H. J. T. *J. Clin. Endocrinol. Metab.* **1999**, *84*, 112-122.
- [14] Cook, C. E.; Wani, M. C.; Jump, J. M.; Lee, Y.-W.; Fail, P. A.; an Yu-Q. Gu, S. A. A.; Petrow, V. *J. Med. Chem.* **1995**, *38*, 753-763.
- [15] Cook, C. E. *Emerging Drugs: The Prospect for Improved Medicines* **1999**, *4*, 291-308.

- [16] Guyton, A. C. *Fisiologia Humana*; Guanabara: Rio de Janeiro, 6^a ed.; 1985.
- [17] Zirkin, B. R. *Cell & Developmental Biology* **1998**, 9, 417-421.
- [18] Trussell, J.; Leveque, J. A.; Koenig, J. D.; London, R.; Borden, S.; Henneberry, J.; LaGuardia, K. D.; Stewart, F.; Wilson, G.; Wysocki, S.; Strauss, M. *American Journal of Public Health* **1995**, 85, 494-503.
- [19] Manson, J. E.; Ridker, P. M.; Spelsberg, A.; Ajani, U.; Lotufo, P. A.; Hennekens, C. H. *Contraception* **1999**, 59, 181-186.
- [20] Kronmal, R. A.; Krieger, J. N.; Coxon, V.; Wortley, P.; Thompson, L.; Sherrard, D. J. *American Journal of Kidney Diseases* **1997**, 29, 207-213.
- [21] Vendrame, R. *Estudo das Relações entre Estrutura-Atividade de Progestogênios Contraceptivos, Através da Utilização de Parâmetros Físico-Químicos Calculados e Métodos de Reconhecimento de Padrão*, Thesis, Universidade Estadual de Campinas, 1998.
- [22] Bonn, D. *The Lancet* **1998**, 353, 302-302.
- [23] Combs, D. W. *Annu. Rep. Med. Chem.* **1997**, 32, 191-200.
- [24] Ebenother, A.; Bastian, J.-M.; Gradient, F. USP 3,678,057, 1972.
- [25] Cook, C. E.; Jump, J. M.; Zhang, P.; Stephens, J. R.; Lee, Y.-W.; Fail, P. A.; Anderson, S. A. *J. Med. Chem.* **1997**, 40, 2111-2112.
- [26] Whitsett, J. M.; Noden, P. F.; Cherry, J.; Lawton, A. D. *J. Reprod. Fertil.* **1984**, 72, 277-286.
- [27] Fail, P. A.; Anderson, S. A.; Cook, C. E. *Reprod. Toxicol.* **2000**, 14, 265.
- [28] Alexander, N. J. *Scientific American* **1995**, 273, 136-141.
- [29] "The Cambridge Analytic Derivatives Package", <http://ket.ch.cam.ac.uk/software/cadpac.html>, 2001.
- [30] "UniChem4.1", <http://www.oxmol.co.uk/prods/unichem>, 1999.
- [31] "Centro Nacional de Supercomputação", <http://www.cesup.ufrgs.br>, 1999.
- [32] "The General Atomic and Molecular Electronic Structure System", <http://www.msg.ameslab.gov/GAMESS/GAMESS.html>, 1999.
- [33] "The GNU Image Manipulation Program", <http://www.gimp.org/>, 2001.
- [34] "Free Software Foundation", <http://www.gnu.org>, 2001.
- [35] "Grace Home", <http://plasma-gate.weizmann.ac.il/Grace/>, 2001.
- [36] "Quantum Chemistry Program Exchange", <http://www.chem.indiana.edu/facilities/qcpe/%5Ffront.htm>, 2001.

- [37] “The MathWorks Home”, <http://www.mathworks.com>, 1999.
- [38] “Centro Nacional de Processamento de Alto Desempenho em São Paulo”, <http://www.cenapad.unicamp.br>, 1999.
- [39] “MOlecular DENsity”, <http://www.caos.kun.nl/~schaft/molden/molden.html>, 2001.
- [40] “Molecular Orbital PACkage”, <http://www.schrodinger.com/mopac.html>, 1999.
- [41] “Scilab Home Page”, <http://www-rocq.inria.fr/scilab/>, 2000.
- [42] Gaudio, A. C.; Takahata, Y. *Computers Chem.* **1992**, *16*, 277-284.
- [43] O'REILLY, <http://www.perl.com>, 2001.
- [44] “TINKER - Software Tools for Molecular Design”, <http://dasher.wustl.edu/tinker/>, 2001.
- [45] Wang, R. *J. Inf. Comput. Sci.* **1997**, *37*, 615-621.
- [46] Levine, I. N. *Quantum Chemistry*; Prentice Hall: New Jersey, 4^a ed.; 1991.
- [47] Ostlund, N. S.; Szabo, A. *Modern Quantum Chemistry, Introduction to Advanced Electronic Structure Theory*; Dover Publications: New York, 1996.
- [48] Blinder, S. M. *Am. J. Phys.* **1965**, *33*, 431-443.
- [49] Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902-3909.
- [50] Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209-220.
- [51] Stewart, J. J. P. “MOPAC 93 MANUAL Revision Number 2”, Fujitsu Limited, 1^a ed.; 1994.
- [52] Hohenberg, P.; Kohn, W. *Phys. Rev. B* **1964**, *136*, 864.
- [53] Kohn, W.; Sham, L. J. *Phys. Rev. A* **1965**, *140*, 1133-1138.
- [54] Morgon, N. H.; Custódio, R. *Quim. Nova* **1995**, *18*, 44-54.
- [55] Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648-5652.
- [56] Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785-789.
- [57] Takahata, Y. *Energias de Ionização, Excitação e Afinidade Eletrônica de Bases de DNA Estudadas pelo Método Semiempírico HAM/3*, Thesis, Universidade Estadual de Campinas, 1995.
- [58] Murrell, J. N.; Harget, A. J. *Semi-empirical Self-consistent-field Molecular Orbital Theory of Molecules*; Wiley-Interscience: New York, 1972.
- [59] Livotto, P. R. “Avaliação do Método Semiempírico HAM/3 no Cálculo de Propriedades Moleculares”, Master's thesis, Universidade Estadual de Campinas, 1986.

- [60] Hammett, L. P. *J. Am. Chem. Soc.* **1937**, *59*, 96-103.
- [61] Higo, J.; Go, N. *J. Comp. Chem.* **1989**, *10*, 376-379.
- [62] Wold, S.; Esbensen, K.; Geladi, P. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37-52.
- [63] Strang, G. *Linear Algebra and its Applications*; Harcourt Brace Jovanovich: San Diego, 3^a ed.; 1988.
- [64] Sharaf, M. A.; Illman, D. L. *Chemometrics*; volume 82 of *Chemical Analysis* John Willey & Sons: New York, 1986.
- [65] van de Waterbeemd, H., Ed.; *Chemometric Methods in Molecular Design*; VCH: Weinheim, 1996.
- [66] Dunn, W. J.; Wold, S. *J. Med. Chem.* **1980**, *23*, 595-599.
- [67] Jump, J. M.; McPhail, A. T.; Cook, C. E. *Tetrahedron Lett.* **1997**, *38*, 3691-3694.
- [68] Stewart, J. J. P. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 1-105.
- [69] Spellmeyer, D. C.; Wong, A. K.; Bower, M. J.; Blaney, J. M. *J. Mol. Graphics Modell.* **1997**, *15*, 18-36.
- [70] Chang, G.; Guida, W. C.; Still, C. *J. Am. Chem. Soc.* **1989**, *111*, 4379-4386.
- [71] Stowasser, R.; Hoffmann, R. *J. Am. Chem. Soc.* **1999**, *121*, 3414-3420.
- [72] Miertus, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117.
- [73] Miertus, S.; Tomasi, J. *Chem. Phys.* **1982**, *65*, 239.
- [74] Townes, C. H.; Schawlow, A. L. *Microwave Spectroscopy*; Dover Publication Inc.: New York-USA, 1975.
- [75] Guadagnini, P. H. *Cargas Atômicas e Energias de Elétrons Internos*, Thesis, Universidade Estadual de Campinas, 2000.
- [76] Chirlian, L. E.; Franci, M. M. *J. Comp. Chem.* **1987**, *8*, 894-905.
- [77] Wiberg, K. B.; Rablen, P. R. *J. Comp. Chem.* **1993**, *14*, 1504-1518.
- [78] Hansch, C.; Fujita, T. *J. Am. Chem. Soc.* **1964**, *86*, 1616-1626.
- [79] Gaudio, A. C. "Relações entre Estrutura e Atividade Biológica", Minicurso realizado de 25 a 27 de outubro na UNICAMP, 2000.
- [80] Silverman, R. B. *The Organic Chemistry of Drug Design and Drug Action*; Academic Press: Washington, 1992.
- [81] Sá, M. M. "Descoberta de Novos Fármacos Através de Planejamento Racional Baseado em Estruturas Tridimensionais de Proteínas", Minicurso Realizado durante 12^a ENEQ em Florianópolis - SC, 1999.

Índice Remissivo

Abreviações

AM1, 23
CADD, 1
CADPAC, 18
CENAPAD, 18
CESUP, 18
CI, 30
CO, 3
DFT, 25
DIU, 7
EA, 30
ESCA, 30
FDA, 8
FSH, 5
GAMESS, 18
GIMP, 18
HAM/3, 18
HF, 21
IBM, 19
IE, 12
INRIA, 18
KNN, 33
LH, 5
LNLS, 49
log P, 30
MATLAB, 18
MNDO, 24
MOLDEN, 18
MOPAC, 18
MT, 12
NDO, 23
NN, 34
PC-UNIX, 19
PCA, 31
PCM, 41

PERL, 18
PES, 30
PI, 29
PM3, 23
PPM, 2
PSDD, 18
QCPE, 18
SAR, 15
SCF, 20
Scilab, 18
SGI, 19
SIMCA, 32
Surf, 18
SVD, 31
TCB, 7
TINKER, 19
WHO, 3

AM1, 23

CADD, 1
CADPAC, 18
CENAPAD, 18
CESUP, 18
CI, 30
CO, 3
Contraceptivos Masculinos, 8–12
Contraceptivos Oraís, 3, 6, 12

DFT, 25
DIU, 7

EA, 30
ESCA, 30

FDA, 8
FSH, 5

- GAMESS, 18
GIMP, 18

HAM/3, 18
HF, 21

IBM, 19
IE, 12
INRIA, 18

KNN, 33

LH, 5
LNLS, 49
log P, 30

Métodos
 AM1, 23
 Clássicos, 30–31
 Contraceptivos, 6–11
 DFT, 25
 Distâncias Geométricas, 37
 Estatísticos, 31–35
 HAM/3, 27
 Hansch, 57–58
 HF, 20
 KNN, 33
 NN, 34
 PCA, 31
 PM3, 23
 Quânticos, 19–30
 SCF, 20
 SIMCA, 32
 Xlogp, 30
MATLAB, 18
MNDO, 24
MOLDEN, 18
MOPAC, 18
MT, 12

NDO, 23
NN, 34

PC-UNIX, 19
PCA, 31
PCM, 41
PERL, 18
PES, 30
PI, 29
PM3, 23
PPM, 2
Programas, 17–19
Propriedades, 58–59
PSDD, 18

QCPE, 18

SAR, 15
SCF, 20
Scilab, 18
SGI, 19
SIMCA, 32
Surf, 18
SVD, 31

TCB, 7
TINKER, 19

WHO, 3