

Tese de Doutorado

***“ESTRUTURA ELETRÔNICA DE DROGAS
ANTITUMORAIS”***

Louraine Cláudia de Melo

**Centro Brasileiro de Pesquisas Físicas
2008**

Agradecimentos

Meus agradecimentos:

- ao professor Paulo Barone pela oportunidade, compreensão e orientação, pela confiança, liberdade e apoio,
- à Scheila Furtado Braga Llanes, companheira de trabalho,
- aos amigos que acompanham e incentivam minha caminhada desde longa data,
- aos meus pais, Samuel e Dejanira, e aos meus irmãos Alex e Elisa, pelo apoio e admiração, ainda que silencioso,
- ao José Maria Honorato, pela “quase” interminável paciência e compreensão nos dias de medo, ansiedade e insegurança, pelo amor, atenção e carinho sempre,
- à minha filha muito amada, Lívia Melo Honorato Campos, pelo grande carinho e compreensão, por cuidar de mim sempre.
- aos funcionários da Coordenação de Formação Científica/CBPF pela atenção, presteza e cordialidade,
- aos professores do CBPF com os quais tive contato, pela atenção, apoio e respeito,
- ao CNPq pelo apoio financeiro,
- à UFJF pelo ambiente computacional,
- Em especial a DEUS e seus cooperadores no trabalho de auxílio ao bem.

Resumo

Este trabalho envolveu a investigação sistemática da estrutura eletrônica, através dos métodos semi-empíricos PM3 (Parametric Method 3) e ZINDO/CI (Zerner's Intermediate Neglect of Differential Overlap/Configuration Interaction), e sua correlação com a atividade biológica, por meio de descritores quânticos teóricos, de uma série de compostos pertencentes às famílias das elipticinas e furanocumarinas.

Em uma primeira parte, apresentamos o estudo realizado para um conjunto de 40 elipticinas, compostos orgânicos caracterizados por uma estrutura molecular planar extraídos, principalmente, da planta *Ochrosia elliptica*, com um grau elevado de atividade antitumoral e citotóxica. Investigamos a estrutura destes compostos através de métodos semi-empíricos partindo de sua estrutura básica 5,11-dimethyl-6H-pyrido[4,3] carbazole (elipticina). Investigamos a relação entre a atividade experimental conhecida e descritores teóricos através de três metodologias de reconhecimento de padrões: a metodologia de índices eletrônicos (MIE), a análise de componentes principais (PCA) e análise hierárquica de clusters (HCA). Para estas três investigações distintas pudemos correlacionar a atividade das elipticinas estudadas com parâmetros teóricos, em sua maioria eletrônicos.

Seguindo um tipo de análise similar à aplicada para as elipticinas, estudamos um conjunto de 52 derivados do psoraleno, furanocumarinas isoladas, principalmente, das plantas *Psoralea corylifolia* L. (Leguminosae). Derivados das furanocumarinas têm sido utilizados na fotoquimioterapia e terapia fotodinâmica. Aplicando os três métodos acima citados selecionamos descritores relacionados à resposta biológica destes compostos e construímos regras e padrões que os diferenciasses quanto à atividade biológica.

Numa investigação especulativa, aplicando as regras e padrões construídos, sugerimos a atividade biológica para alguns compostos estudados e que ainda não foram avaliados experimentalmente.

Abstract

This work deals with the investigation electronic structure, using the semiempirical methods PM3 (Parametric Method 3) and ZINDO (Zerner's Intermediate Neglect of Differential Overlap) and the structure-activity relationship, using theoretical quantum descriptors, of ellipticine and furanocoumarin derivatives.

In the first part, we present a study of 40 ellipticines, a planar organic compound isolated from the plant extract of *Ochrosia elliptica*, with a high degree of antitumor and cytotoxic activity. We investigated the compounds structure with semiempirical methods initiating with their basic structure 5,11-dimethyl-6H-pyrido[4,3] carbazole (Ellipticine). Ellipticines experimental activity relationship was investigated with theoretical descriptors applying three pattern recognition methodologies: the electronic indices methodology, principal component analysis and hierarchical cluster analysis. With these three distinct methods we were able to correlate the biological activity of the studied ellipticines with theoretical parameters.

In a similar approach we studied a set of 52 derivatives psoralen, furocoumarins isolated from the plant *Psoralea corylifolia* L. (Leguminosae). Furocoumarins have presented good results for the photochemotherapy treatment and photodynamic therapy applications. Making use of the three methods mentioned above we selected the descriptors related to the compounds biological indices and constructed the rules and patterns able to separate them according to their activity.

Considering a speculative approach, we proposed the biological activity to untested compounds, via the obtained rules.

ÍNDICE

AGRADECIMENTO	2
RESUMO	3
ABSTRACT	4
ÍNDICE DE FIGURAS	7
ÍNDICE DE TABELAS	9
LISTA DE ABREVIACÕES	11
CAPÍTULO 1 – INTRODUÇÃO	13
1.1 – Introdução	14
1.2 – Referências	23
CAPÍTULO 2 – MÉTODOS	27
2.1 – Mecânica Quântica Molecular	28
2.2 – Métodos e Aproximações Semi empírica	38
2.2.1 – Aproximações NDO	39
2.2.2 – Aproximações MNDO, AM1 e PM3	41
2.3 – Metodologia de Índices Eletrônicos – MIE	47
2.4 – Análise de Componentes Principais - PCA	51
2.4.1 – Obtendo as Componentes Principais	54
2.5 – Análise Hierárquica de Clusters – HCA	62
2.6 – Referências	66
CAPÍTULO 3 – FAMÍLIAS DE MOLÉCULAS ESTUDADAS	68
3.1 – Elipticinas	69
3.2 – Furanocumarinas	72
3.3 – Referências	76
CAPÍTULO 4 - ELIPTICINAS	78
4.1 – Introdução	79
4.2 – Resultados	81
4.2.1 - Análise Conformacional.....	81
4.2.2 – Análise de Propriedades Eletrônicas e Óticas	87

4.2.3 - Investigação da Atividade Biológica Através da Metodologia de Índices Eletrônicos – MIE	97
4.2.4 - Investigação da Atividade Biológica Através da Análise de Componentes Principais (PCA) e da Análise Hierárquica de Clusters (HCA).	101
4.3 – Referências.....	114
CAPÍTULO 5 – FURANOCUMARINAS	118
5.1 – Introdução	119
5.2. – Resultados.....	121
5.2.1 - Análise Conformacional.....	121
5.2.2 - Investigação da Atividade Biológica Através da Metodologia de Índices Eletrônicos – MIE	124
5.2.3 - Investigação da Atividade Biológica Através da Análise de Componentes Principais (PCA) e da Análise Hierárquica de Clusters (HCA).	129
5.3 – Referências.....	138
CAPÍTULO 6 – CONCLUSÕES	139
6.1 – Conclusões Finais	140

Índice de Figuras

Figura 3.1 – Estrutura básica da Elipticina

Figura 3.2 - Estrutura Molecular: a) psoraleno ($R_i = H$); b) psoraleno1(P_1); c) psoraleno 2 (P_2) Pseudoisopsoraleno; d) psoraleno 3 (P_3) Allopsoraleno; e) psoraleno 4 (P_4) Isopseudopsoraleno; f) benzopsoraleno (P_{24}); g) allobergapten.

Figura 4.1 – Busca sistemática da conformação mais estável da molécula elipticina E23 no v acuo. (à direita) Superfície do Calor de Formação (Kcal/mol) como uma função da variação dos ângulos diedrais (em destaque na figura) *em* passos de 20 graus. (à esquerda).

Figura 4.2 – Busca sistemática da conformação mais estável da molécula elipticina E37 no v acuo. (à direita) Superfície do Calor de Formação (Kcal/mol) como uma função da variação do ângulo diedral (em destaque na figura) *em* passos de 20 graus. (à esquerda).

Figura 4.3 – Regiões sobre o esqueleto molecular selecionadas para os cálculos de LDOS.

Figura 4.4 – Distribuição Espacial para o HOMO e LUMO (Isosuperfície da função de onda molecular).

Figura 4.5 – Simulação dos espectros de absorção para derivados da elipticina ativos (E11 e E12) e inativos (E32 e E33).

Figura 4.6 - Gráfico SCORE do Grupo G1– distribuição dos compostos no plano das componentes principais PC1 x PC2. Triângulo e círculo representam moléculas ativas e inativas, respectivamente.

Figura 4.7 - Gráfico LOADINGS para grupo G1 – distribuição dos descritores responsáveis pela separação dos compostos em grupos distintos de atividade.

Figura 4.8 - Dendograma da distribuição hierárquica do grupo G1. Os compostos ativos e inativos estão separados em grupos distintos e apresentam uma similaridade nula entre si.

Figura 4.9 - Resultados da análise de PCA. Gráfico de separação dos compostos do grupo G1 e proposição de atividade para os compostos do grupo G2. Os símbolos triângulo, círculo e cruz representam ativo, inativo e não testado, respectivamente.

Figura 4.10 - Dendograma da distribuição hierárquica do conjunto de moléculas. (G1 + G2).

Figura 5.1 – Busca sistemática da conformação mais estável da molécula psoraleno P₁₇ no vácuo. (à direita) Superfície do Calor de Formação (Kcal/mol) como uma função da variação do ângulo diedral (em destaque na figura) em passos de 20 graus. (à esquerda).

Figura 5.2: Divisão das regiões dos psoralenos para as quais os parâmetros η foram calculados.

Figura 5.3 – Resultado da análise PCA. Gráfico de distribuição dos compostos no plano PC1 x PC2, segundo as variáveis do conjunto A.

Figura 5.4 – Resultado da análise HCA. Gráfico de separação dos compostos em agrupamentos, segundo as variáveis do conjunto A.

Figura 5.5 - Resultados da análise de PCA. Gráfico incluindo todos os compostos e proposição de atividade. Conjunto A.

Índice de Tabelas

Tabela 3.1: Elipticina e derivados estudados. Os radicais numerados referem-se a estrutura mostrada na Figura 3.1. *moléculas de camada eletrônica aberta; ** o número que precede o símbolo N indica a posição do átomo de nitrogênio no anel piridínico. A e I refere-se a ativo e inativo, E e EM refere-se a elipticina e elipticinium, e O e OM a olivacina e olivacinium.

Tabela 3.2 – Psoraleno e derivados estudados. Os radicais numerados referem-se a estrutura mostrada na Figura 3.2a. A e I refere-se a ativo e inativo.

Tabela 4.1 – Comparação entre os valores experimentais e teóricos calculados com método semi-empírico PM3 para a molécula E12 (6-metilelipticina).

Tabela 4.2 – Energia, força de oscilador e principais contribuições para o limiar de absorção (primeira transição) e pico máximo (máxima absorção) para espectros simulados pelo método ZINDO. A notação $|A \Rightarrow B\rangle$ indica uma configuração gerada pela remoção de um elétron do orbital molecular A para o orbital molecular B.

Tabela 4.3 – Resultados para o momento de dipolo (DM) e sítios ativos das moléculas do grupo G1. Sítios, radicais e regiões mais ativas. C_n : carbono n, R_n : grupo lateral no carbono n e r_n : vizinhança do carbono n (1Å raio). A e I, refere-se, respectivamente, a composto ativo e inativo, dentro da aproximação proposta por Villemin et al.

Tabela 4.4 – Resultado para o valor do momento de dipolo e sítios ativos para as moléculas do grupo G2. Na última coluna a atividade proposta para o grupo. C_n : carbono n, R_n : grupo lateral no carbono n e r_n : vizinhança do carbono n (1Å raio).

Tabela 4.5 - Descritores ΔH e $CL(rD)$ calculados com o PM3. A. Q. indica a atividade biológica experimental apresentada de forma qualitativa.

Tabela 4.6 - Descritores ΔH e $CL(rD)$ calculados com o PM3. A. P. indica a atividade biológica qualitativa proposta através da metodologia MIE.

Tabela 4.7: Sumário da atividade proposta para os compostos do grupo **G2** das elipticinas segundo diferentes metodologias.

Tabela 5.1: Descritores η_L e $CH(rB)$ calculados com o PM3. A.Q. indica a atividade biológica experimental apresentada de forma qualitativa.

Tabela 5.2: Descritores η_L e $CH(rB)$ calculados com o PM3. A.P. indica a atividade biológica qualitativa proposta através da metodologia MIE.

Lista de Abreviações

MNDO – *Modified Neglect of Differential Overlap*

AM1 – *Austin Method 1*

PM3 – *Parametric Method 3*

ZDO – *Zero Differential Overlap*

ZINDO/S – *Zerner's Intermediate Neglect of Differential Overlap/Spectroscopic*

RHF – Hartree-Fock Restrito (Restricted Hartree-Fock)

UHF – Hartree-Fock não Restrito (Unrestricted Hartree-Fock)

MIE – Metodologia de Índices Eletrônicos

PCA – *Principal Component Analysis* (Análise de Componentes Principais)

PC – *Principal Component* (Componentes Principais)

HCA – *Hierarchical Clusters Analysis* (Análise Hierárquica de Clusters)

LCAO – *Linear Combination of Atomic Orbitals* (Combinação Linear de Orbitais Atômicos)

R-X – Raios-X

DOS – *Density of States* (Densidade de Estados)

LDOS – *Local Density of States* (Densidade Local de Estados)

HOMO – *Highest Occupied Molecular Orbital* (Último Orbital Molecular Ocupado)

LUMO – *Lowest Unoccupied Molecular Orbital* (Primeiro Orbital Molecular Desocupado)

C.F. – Calor de Formação

D.M – Momento de Dipolo

CI – Interação de Configurações

UV- Ultra Violeta

DNA – ácido desoxirribonucleico

A – Ativo

I – Inativo

G1 – Grupo de Controle (compostos com atividade experimental conhecida)

G2 – Grupo de Teste (compostos sem informação experimental)

MEP - potencial eletrostático molecular

ΔH – diferença de energia entre HOMO e HOMO-1

ΔL - diferença de energia entre LUMO e LUMO+1

$CH(r_i)$ - contribuição para a densidade local de estados (LDOS) para a região r_i , na formação do orbital HOMO

$CH-1(r_i)$ - contribuição para a densidade local de estados (LDOS) para a região r_i , na formação do orbital HOMO – 1

$\eta H(r_i)$ – diferença entre $CH(r_i)$ e $CH-1(r_i)$

$CL(r_i)$ - contribuição para a densidade local de estados (LDOS) para a região r_i , na formação do orbital LUMO

$CL+1(r_i)$ - contribuição para a densidade local de estados (LDOS) para a região r_i , na formação do orbital LUMO+1

$\eta H(r_i)$ – diferença entre $CH(r_i)$ e $CH-1(r_i)$

$\eta L(r_i)$ – diferença entre $CL(r_i)$ e $CL+1(r_i)$

CAPÍTULO 1

INTRODUÇÃO

1.1. Introdução

Entre os desafios da área médica está o desenvolvimento de agentes terapêuticos com maior e mais precisa especificidade, isto é, agentes que afetem o tecido alvo e não outros sadios. Muitos antibióticos se aproximam deste ideal. Devido à diferença entre as células bacterianas e as humanas, os antibióticos podem destruir o invasor e deixar os tecidos do corpo ilesos. No entanto, em câncer e em desordens auto-imunes (em que o próprio sistema imunológico ataca os tecidos normais) tal discriminação é muito difícil de ser obtida. Nestes dois tipos de desordens, a causa do dano no tecido não é uma bactéria ou um vírus, mas são as próprias células do corpo que se “degeneram”, e a distinção entre células doentes e outras saudáveis não é tarefa fácil. Em decorrência disto, existem poucas terapias efetivas para doenças auto-imunes, e aquelas que existem para câncer, frequentemente, acarretam severos efeitos colaterais [1].

O câncer se origina de uma alteração genética das células que pode ter sido determinada por um gene herdado ou induzida por cancerígenos (iniciadores) físicos, químicos ou biológicos que alteram a resposta celular ao meio biológico, tornando-as capazes de se multiplicar de modo autônomo. A ação do iniciador é irreversível e admite-se que atue provocando alterações permanentes (mutações) no DNA (ácido desoxirribonucleico). Alguns exemplos de iniciadores são: hidrocarbonetos aromáticos, alguns derivados da anilina, asbesto, raios ultravioleta, radiações ionizantes, como os raios X, dentre outros.

Atualmente, o tratamento pode ser feito através de cirurgia, radioterapia, quimioterapia, e mais recentemente, terapia fotodinâmica. A escolha da técnica ou a

combinação de algumas delas depende do tipo, localização e estágio da doença, além do estado geral do paciente.

A quimioterapia consiste na administração de drogas, que supostamente impedirão a reprodução celular, levando as células malignas à morte. Em particular, os alcalóides, dos quais fazem parte as elipticinas e derivados, alvo desse estudo que serão apresentados nos capítulos 3 e 4, atuam bloqueando a divisão celular durante a mitose. Um alcalóide é uma base obtida da natureza, que contém um nitrogênio em um dos anéis heterocíclicos que compõem sua estrutura [4].

O alcalóide elipticina é um composto orgânico que possui uma estrutura planar simples, cujo núcleo é composto da metade de um carbazol ligado a um anel de piridina (vide capítulo 3). Sua atividade no tratamento quimioterápico contra vários tipos de câncer como a leucemia mieloblástica, adenocarcinoma, câncer de mama e alguns tumores sólidos, associada aos efeitos colaterais limitados e à ausência de toxicidade hematológica [5], reforçam o interesse por suas propriedades biológicas.

De acordo com estudos físico-químicos e biológicos, o alvo principal para as elipticinas é o DNA, que pode ser atacado direta ou indiretamente [6,7]. A forma desta atuação é multimodal, podendo haver ligação ao DNA através de intercalação ou ligação covalente, com a produção de derivados metabólicos oxidados ou ainda interferência na atividade catalítica da topoisomerase II e de outras enzimas.

A intercalação das drogas pode ocorrer entre os pares de bases do DNA através de interações hidrofóbicas, pontes de hidrogênio e forças de Van de Waals [8]. A ligação covalente deve ocorrer por interações eletrostáticas externamente à hélice. Nos dois casos, a associação droga-DNA dificulta a multiplicação celular, uma vez que cria impedimentos para a abertura da dupla hélice e consequente

duplicação do DNA. A hipótese da intercalação da elipticina entre os pares de bases do DNA ser responsável por sua atividade antitumoral é sustentada pelo fato de que as drogas inábeis no processo de intercalação exibem baixa citotoxicidade em culturas de células e atividade antitumoral insignificante [7]. O processo de intercalação elipticina-DNA tem sido observado através de técnicas bioquímicas [9], termodinâmicas [10,11] e espectroscópicas [9,12,13]. Recentemente Stiborová M. et al. mostrou a atividade enzimática das elipticinas por ligação covalente com DNA [14,15].

A enzima topoisomerase II atua separando a dupla-hélice do DNA e formando o complexo DNA-enzima através da ligação de pares de base aos grupos fosfatos das hélices simples para a formação da nova dupla hélice. A elipticina atua ligando-se ao complexo DNA-enzima e inibindo a continuação do processo de ação da topoisomerase II, o que impede que se forme a nova dupla hélice [16].

Muitos estudos de relações estrutura-atividade de drogas anticâncer e também alguns sobre carcinógenos químicos têm sido realizados [2,3]. No entanto, câncer é uma doença dependente de vários fatores, e tais correlações podem ser difíceis de se encontrar.

Na busca por parâmetros críticos que estejam envolvidos num dado efeito biológico são utilizadas aproximações dos estudos experimentais de relação estrutura-atividade. Estas aproximações envolvem: **1.** a síntese de muitos derivados nos quais, idealmente, varia-se somente um parâmetro, e, **2.** o uso adequado de modelos biológicos. A cada nova droga em experiência, testes *in vitro* e *in vivo* devem ser realizados para verificar se esta possui atividade biológica significativa, procedimento que representa grande consumo de recursos e tempo. A correlação da resposta biológica de compostos químicos com parâmetros eletrônicos

moleculares foi sugerida há mais de 50 anos [17], um entendimento da relação entre a atividade biológica e as propriedades eletrônicas das drogas poderia, associado ao processo experimental, auxiliar o processo de síntese de novos compostos promissores para o tratamento quimioterápico [7,18-22]. O grande número de derivados da elipticina sintetizados, estudados e mapeados do ponto de vista da atividade citotóxica e antitumoral, representa uma amostra útil para estes propósitos [23-29].

Outro recurso que se desenvolveu, primeiro como um tratamento improvável para o câncer, para se tornar, em seguida, um recurso sofisticado e efetivo contra uma série de disfunções e, ainda mais, recentemente, contra a degeneração macular e a miopia patológica, é a terapia fotodinâmica, ou TFD. A terapia fotodinâmica consiste na combinação de um agente fotossensível que quando ativado pela luz, reage com o DNA das células em proliferação, levando à morte celular [30,31]. Como exemplo de tratamento, citamos a CTCL (linfoma cutâneo de células T), que não respondia a nenhum tratamento conhecido. A doença em fase inicial é caracterizada por lesões na pele contendo células T malignas, que em fase final são disseminadas pelo corpo. O prognóstico para a sobrevivência média do paciente era de cerca de cinco anos após o diagnóstico confirmado por biópsia. O uso da terapia fotodinâmica, associando um derivado do psoraleno (8 metoxipsoraleno) e radiação ultravioleta A (UVA, de 320 a 400 nm), é capaz de eliminar as lesões causando dano direto sobre células T doentes [32]. O método que tornou-se conhecido como PUVA (abreviação para psoraleno com UVA) é também um dos mais eficazes tratamentos para psoríase e outras condições da pele [33].

O agente químico fotossensível pode apresentar duas utilidades básicas: **(i)** diagnóstico de tecidos neoplásicos através da espectroscopia de fluorescência, que é uma técnica não invasiva de análise de tecidos biológicos, e **(ii)** geração de agentes citotóxicos, capazes de matar a célula doente [34]. Além disso, tem a vantagem de possuir baixa citotoxicidade na ausência de luz, afinidade por localização nas áreas tumorais e interação com luz em comprimentos de onda que podem ter alta penetrabilidade nos tecidos biológicos. Uma estratégia que tem sido explorada, então, é a investigação de drogas que são ativadas por luz, mas são inertes quando não estão expostas ao comprimento de onda correto de radiação e que possam ser ativadas por uma luz de alta penetração, sendo assim mais adequados para o tratamento de tumores internos [35,36].

Produtos extraídos de plantas com atividade dependente da luz têm se mostrado importantes sensibilizadores na terapia fotodinâmica. Em geral, estes compostos tem sistemas estendidos de elétrons π que, sofrendo fotoexcitação com luz visível, são responsáveis pela produção de agentes citotóxicos (por exemplo, o oxigênio singlete). Entre estes constituintes de plantas que mostram atividade fotossensibilizadora induzida por luz UVA estão as furanocumarinas, como o psoraleno (citado acima), a angelicina e seus derivados naturais e sintéticos [37], também alvo deste trabalho que serão apresentados nos capítulos 3 e 5.

Também tem sido testado o potencial antimutagênico de furanocumarinas e os resultados mostram que estes compostos são antimutagênicos muito potentes [38]. Algumas furanocumarinas também têm sido estudadas quanto ao potencial inibitório do citocromo P450 [39] e quanto à atividade antiretroviral [40]. Algumas espécies de vírus supermutantes têm sido fotomodificados por luz UVA na presença de diferentes concentrações de dois psoralenos (8-metoxipsoraleno e 4,5',8-

trimetilpsoraleno) e uma angelicina (4,6,5'-trimetilangelicina). O composto que se mostrou mais ativo foi o trimetilpsoraleno, que testado sobre o vírus HIV-1 provocou a inativação do vírus na presença de UVA [33, 35, 36, 40, 41, 42].

As várias características das furanocumarinas fazem delas também um alvo interessante para aplicações clínicas. Na ausência de luz são inertes, mas uma vez ativadas (mediante absorção de luz) quando associadas ao DNA, interrompem sua funcionalidade, causando danos nas células que se dividem rapidamente.

A interação das furanocumarinas com o DNA, em geral, ocorre em dois passos: **a)** formação de um complexo molecular inicial, no qual as furanocumarinas são intercaladas entre dois pares de bases da macromolécula, e **b)** fotoconjugação covalente com bases pirimídicas do DNA (timina e citosina) [35]. Entre as técnicas utilizadas para avaliar a fotoligação com o DNA estão o dicróismo linear, técnicas químicas e bioquímicas e técnicas calorimétricas, sendo a capacidade de fotorreação com o DNA estudada seguindo a formação do foto-aduto através de medidas de espectrofotometria [43]. O estudo de alguns benzopsoralenos aponta estes compostos como sendo capazes de induzir fortes efeitos antiproliferativos, agindo mediante absorção de luz na faixa do UVA (produzindo oxigênio singlete) ou sem a intermediação de radiação [44]. Estudos com metilpsoralenos mostram que estes compostos se ligam de forma eficiente ao DNA, dependendo da natureza dos substituintes em sítios específicos da molécula, além da influência do próprio arranjo estrutural do DNA [45].

A investigação sobre o mecanismo de ação de furanocumarinas tem mostrado que a ação fotossensibilizadora destes compostos em sistemas biológicos ocorre por dois caminhos de reação: **a)** independente do oxigênio, que envolve a foto-adição do sensibilizador a ácidos nucleicos, proteínas e lípidios, e **b)**

dependente do oxigênio, que inclui as furanocumarinas na categoria de sensibilizadores fotodinâmicos. A ação fotodinâmica de furanocumarinas, estudada usando-se biomoléculas isoladas, eritrócitos e pele humana, aparece envolvendo espécies ativas de oxigênio (oxigênio singleto, ânion superóxido e radicais hidroxil) e espécies de radicais formados por transferência de elétron de ou para furanocumarinas fotoexcitadas. Um outro processo dependente do oxigênio envolve a formação de derivados fotoexcitados, os quais reagem com diversos substratos (em particular componentes da membrana) causando danos irreversíveis às células [46,47].

A estrutura tricíclica de drogas fotoquimioterapêuticas naturais conhecidas é frequentemente tomada como modelo na pesquisa de novos agentes fotossensibilizadores com baixos efeitos fototóxicos e mutagênicos. Estudos mostram que a fotorreatividade e o comportamento fotobiológico depende da localização dos substituintes na molécula [48]. Além disso, as condições externas como pH, temperatura e interação com outros compostos podem modificar as características dos estados eletrônicos desses compostos, portanto influenciando na sua atividade biológica e sua aplicação médica [49].

Um dos objetivos da pesquisa nessa área é a investigação de meios terapêuticos eficientes e pouco tóxicos aos tecidos normais, via técnicas teóricas que viabilizem a síntese de fármacos eficientes com previsão de seus mecanismos de ação. A combinação de estudos experimentais e teóricos pode se tornar uma fonte de informações sobre relações estrutura-atividade, bem como ser um método efetivo e racional para a modelagem de novos compostos eficientes no tratamento. A comparação de resultados teóricos com dados experimentais pode permitir a compreensão de quais propriedades moleculares são mais importantes para a

atividade biológica, além do desenvolvimento de técnicas que direcionem a síntese de derivados mais seletivos e eficientes.

Barone e colaboradores estudaram a relação entre topologia molecular/estrutura eletrônica e atividade carcinogênica de uma família de Hidrocarbonetos Aromáticos Policíclicos (HAP's), por meio do método de Huckel simplificado e dos métodos semi-empíricos PM3 (Parametric Method 3) e ZINDO/CI (Zerner's Intermediate Neglect of Differential Overlap/Configuration Interaction) [50,51]. Nestes trabalhos, foram desenvolvidos critérios, baseados em indicadores eletrônicos das moléculas, capazes de identificar entre os HAP quais têm atividade carcinogênica. O mesmo método tem sido utilizado para estudar relações estrutura-atividade para diferentes classes de compostos (carcinogênicos, antitumorais, antiinflamatórios, inibidores da integrase do HIV-1, e outros), através da investigação sistemática das propriedades moleculares.

Nesse trabalho, investigamos sistematicamente propriedades teóricas eletrônicas de dois conjuntos de moléculas, elipticinas e furanocumarinas, com o objetivo de determinar regras e padrões indicadores de uma correlação entre a atividade biológica destes compostos e características de sua estrutura molecular. Embora os processos envolvidos na interação dos compostos com o meio biológico e em seus mecanismos de ação possam ser muito complexos, as propriedades moleculares são fatores importantes para a definição de sua atividade biológica, já que variações na estrutura química de compostos de uma família podem ser responsáveis pelas mudanças drásticas em sua atividade biológica. Assim, investigações mesmo que qualitativas de atividade biológica, baseadas em indicadores que dependam da estrutura e dos átomos constituintes, representam um passo na formulação/seleção de novos compostos-teste.

No capítulo 2 faremos uma breve descrição dos métodos envolvidos na realização do trabalho, sem a pretensão de que esta seja uma referência completa para iniciantes no assunto: aproximações semi-empíricas utilizadas em cálculos de estrutura eletrônica molecular; Metodologia de Índices Eletrônicos (MIE), Análise de Componentes Principais (PCA) e Análise Hierárquica de Clusters (HCA).

No capítulo 3 apresentamos a estrutura molecular das famílias de compostos estudados neste trabalho - elipticinas e furanocumarinas -, enquanto que nos capítulos 4 e 5 são relatados os resultados para a análise conformacional, análise de propriedades eletrônicas e óticas e a investigação teórica de atividade com métodos de reconhecimento de padrões. Nos capítulos 4 e 5, comparamos os resultados das diferentes metodologias utilizadas para a investigação da correlação entre a atividade biológica e grandezas eletrônicas.

Finalizando, o capítulo 6 é dedicado às conclusões finais.

1.2. Referências

- [1] Richard L. Edelson, Light Activated Drugs, Scientific American, 1988.
- [2] J.Ladik and W.Forner, The Beginnings of cancer in the cell (Springer-Verlag, Berlin, Heidelberg, 1994).
- [3] H.C.Pitot, D.E.Stevenson, R.M.Mcclain, J.A.Poop, J.J.Slaga, J.M.Ward, Mouse Liver Carcinogenesis: Mechanism & Species comparison (Wiley, New York, (1990).
- [4] T.W.G. Solomons, Organic Chemistry, Ed. John Wiley & Sons, 1988.
- [5] Diop, B., Toure, P., Sow, M. T., Toure, M., Halliez, M. L., Castaigne, J. P., Mondesir, J. M., and De Jaeger, R. *Med. Afr. Noire* 1984, 31, 107-110.
- [6] Sizun, P., Auclair, C., Lescot, E., Paoletti, C., Perly, B. and Fermandjian, S.. *Biopolymers* 1988, 27, 1085-1096.
- [7] Auclair, C., *Arch. Biochem. Biophys.* 1987, 259, 1.
- [8] Reha, D., Kabelác, M., Ryjáček ,F., Sponer, J., Sponer, J. E., Elstner, M., Suhai, S. and Hobza, P., *J. Am. Chem. Soc.* 2002, 124, 3366-3376.
- [9] Mauffret, O., Rene, B., Convert, O., Monnot, M., Lescot, E., and Fernandjian, S., *Biopolymers* 1991, 31, 1325.
- [10] Schwaller, M. A., Dodin, G., and Aubard, J., *Biopolymers* 1991, 31, 519-527.
- [11] Adenier, A. , Aubard, J. and Schwaller, M. A., *J. Phys. Chem.* 1992, 96, 8785-8791.
- [12] Aubard, J., Schwaller, M. A, Patigny, J., Marsault, J. P., and Lévi, G.. *J. Raman Spectroscopy* 1992, 23, 373-377.
- [13] Kohn, K. W., Waring, M. J., Glaubiger, D., and Friedman, C., *Cancer Research* 1975, 35, 71.

- [14] Stiborová, M., Stiborová-Rupertová, M., Borjek-Dohalská, L., Wiessler, M. and Frei, E., *Chem. Res. Toxicol.* 2003, 16, 38-47.
- [15] Stiborová, M., Bieler C. A., Wiessler, M. and Frei, E.. *Biochemical Pharmacology* 2001, 62, 1675–1684.
- [16] Gribble, G. W., in *The Alkaloids*, (Academic Press, 1990) Vol. 39, p.239.
- [17] L. P. Hammett, *Physical Organic Chemistry* (McGraw-Hill, New York, 1940).
- [18] Dalton, L. K., Demerac, S. Elmes, B. C., Loder, J. W., Swan, J. M., and Teitei, J.. *Aust. J. Chem.* 1967, 20, 2715-2727.
- [19] JaszoldHoworko, R. *et al.* *J. Med Chem.* 1994, 37, 2445-2452.
- [20] Anderson, W. K., Gopalsamy, A. and Reddy, P. S.. *J. Med. Chem.* 1994, 37, 1955-1963. Jurayj, J., Haugwitz, R. D., Varma, R. K., Paull, K. D., Barret, J. F and Cushman, M.. *J. Med. Chem.* 1994, 37, 2190-2197.
- [21] Mathe, G., Pontiggia, P., Bourut, C., Chenu, E. and Orbach-Arbouys, S.. *Biom. and Pharmacother.* 1994, 48, 51-53.
- [22] Kohn, K. W., Ross, W. E. and Glaubinger, D., in F. E. Hahn (Ed.) *Antibiotics*, vol.2, Springer-Verlag, Berlin, 1979, p. 195
- [23] Le Pecq, J. B., Dat-Xuong, N., Gosse, C., and Paoletti, C. A.. *Proc. Natl. Acad. Sci. USA* 1974, 71, 5078-5082.
- [24] Paoletti, C., Cros, S., Dat-Xuong, N., Lecointe, P., and Moisand, A.. *Chem. Biol. Interact.* 1979, 25, 45-58.
- [25] Chabane, H., Lamazzi, C., Thiéry, V., Guillaumetb, G. and Bessona, T. *Tetrahedron Letters* 2002, 43, 2483–2486.
- [26] Tran-Thi, H. A., Nguyen-Thi, T., Michel, S., Tillequin, F., KOCH, M., Pfeiffer, B., Pierré, A., Trinh-Van-Dufat, H. *Chem. Pharm. Bull.* 2004, 52, 540-545.
- [27] Ergun, Y., Patir, S., Okay, G.. *Synthetic Communications.* 2004, 34, 435-442.

- [28] Ishikura, M., Hino, A., Yaginuma, T., Ágata, I. and Katagiri, N.. *Tetrahedron* 2000, 56, 193–207.
- [29] Diaz, M., Cobas, A., Guitian, E., Castelo, L.. *European Journal of Organic Chemistry* 2001, 23, 4543-4549.
- [30] Ishikura, M., Hino, A., Katagiri, N. *Heterocycles* 2000, 53, 11-14.
- [31] Boehncke WH, König K., Ruck A., Kaufmann R., Sterry W. *Acta Derm Venereol*, 74 (3):201-5, 1994.
- [32] Rook A.H., Suchin K.R., KAO D.M.F., et al. *J. Invest. Derm. Symp P*, 4: (1) 85-90, 1999.
- [33] Gillenvater, A., et al. *Arch. Otolaryngol Head Neck Surg.*, 124: 1251-1258, 1998.
- [34] Zhao J.F., Zhang Y.J., Jin X.H., Athar M., Santella R.M., Bickers D.R., Wang Z.Y. *Journal of Investigative Dermatology*, 113: (6) 1070-1075, 1999.
- [35] Bordin, F., Dall 'Acqua, F. and Guiotto, A.. *Pharmac. Ther.*, Vol. 52, pp 331-363, 1991.
- [36] Towers G.H.N., Page J.E., Hudson J.B.. *Current Organic Chemistry*; 1: (4) 395-414, 1997.
- [37] Ebermann, R., Alth G., Kreitner M. and Kubin A.. *Journal of Photochemistry and Photobiology B-Biology*, 36: (2) 95-97, 1996.
- [38] Edenharder R., Speth C., Decker M., Kolodziej H., Kayser O., Platt K.L.. *Mutation Research-Genetic Toxicology*; 345: (1-2) 57-71, 1995.
- [39] Neal J.J., WU D. *Pesticide Biochemistry and Physiology*, 50:(1)43-50,1994.
- [40] Miolo G., Tomanin R., Derossi A., Dall 'Acqua F., Zacchello F., Scarpa M.. *Journal of Photochemistry and Photobiology B-Biology*, 26: (3) 241-247, 1994.
- [41] Shephard S.E., Panizzon R.G.. *Dermatology*, 199: (2) 106-112, 1999.

- [42] Hannuksela-Svahn A., Pukkala E., Koulu L., Jansen C.T., Karvonen J.. J. of the American Academy of Dermatology; 40: (5) 694-696, part 1, 1999.
- [43] Gia O., Conconi M.T., Uriarte E.. Farmaco, 52: (5) 289-293, 1997.
- [44] Bordin F., Carlassare F., Conconi M.T., et al.. Photochem Photobiol 55: (2) 221-229, 1992.
- [45] Gia O., Uriarte E., Zagotto G., Baccichetti F., Antonello C., Marcianimagno S.. Journal of Photoch. and Photobiology B-Biology, 14: (1-2) 95-104, 1992.
- [46] Potapenko A.Y.. Journal Photoch. Photobiol B, 9: (1) 1-33, 1991.
- [47] Marko J., Vermeersch G., Febvaygarot N., et al.. Photochem. Photobiol. 47: (1) 1-13, 1988.
- [48] Gia O., Anselmo A., Pozzan A., Antonello C., Magno S.M., Uriarte E.. Farmaco, 52: (6-7) 389-397, 1997.
- [49] Otsuki T., J. of Synthetic Organic Chemistry Japan, 49: (9) 809-820, 1991.
- [50] P.M.V.B.Barone, A.Camilo Jr. e D.S.Galvão. Phys. Rev. Lett. 77, 1186, 1996
- [51] P.M.V.B.Barone, R.S.Braga, A.Camilo Jr. e D.S.Galvão. J. Mol. Struct. (Theochem) 505, 55-66, 2000.

CAPÍTULO 2

MÉTODOS DE CÁLCULO

2.1. Mecânica Quântica Molecular [1, 2, 3,12]

A descrição das propriedades estruturais e eletrônicas de moléculas por meio da Mecânica Quântica é baseada numa série de aproximações, que são apresentadas nessa seção. Aproximações adicionais, denominadas semi-empíricas, utilizadas nesse trabalho são apresentadas nas seções seguintes. A motivação para a introdução de aproximações na Mecânica Quântica Molecular é a dificuldade de tratar analiticamente sistemas de partículas correlacionadas. O custo computacional para cálculos envolvendo grande número de átomos conduz ao uso de métodos que simplificam ainda mais os sistemas considerados, como se verá abaixo.

A Teoria Quântica, na **representação de Schrödinger**, descreve os estados estacionários dos sistemas por meio das autofunções, $\Psi(\mathbf{r}, \mathbf{R})$, que são soluções da equação de Schrödinger independente do tempo, $\mathbf{H}\Psi(\mathbf{r}, \mathbf{R}) = E\Psi(\mathbf{r}, \mathbf{R})$, em termos das coordenadas dos elétrons (\mathbf{r}) e núcleos (\mathbf{R}). H é o hamiltoniano do sistema atômico ou molecular, que contém todos os termos da interação para N elétrons e M núcleos e E é a energia total do sistema. Desprezando efeitos relativísticos, em unidades atômicas, H pode ser escrito como:

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

onde a distância entre o i -elétron e A -núcleos é $r_{iA} = |r_{iA}| = |r_i - R_A|$; a distância entre os elétrons i -ésimo e j -ésimo é $r_{ij} = |r_i - r_j|$, e a distância entre A núcleos e B núcleos é $R_{AB} = |R_A - R_B|$.

Temos na equação (1):

- M_A é a razão da massa do núcleo A pela massa do elétron;
- Z_A é o número atômico do núcleo A ;

- Os operadores Laplacianos, ∇_i^2 e ∇_A^2 envolvem diferenciações em relação às coordenadas do i-ésimo elétron e A-ésimo núcleo, respectivamente;
- o primeiro e segundo termo representam as energias cinéticas dos elétrons e núcleos, respectivamente;
- o terceiro termo representa a energia de interação entre núcleos e elétrons (atração de Coulomb) e o quarto e quinto termos a energia de repulsão entre os elétrons e entre os núcleos, respectivamente.

Uma molécula é um sistema de muitos corpos, para o qual a equação de Schrödinger independente do tempo não possui uma solução analítica. Algumas aproximações devem ser impostas para tornar a equação solúvel, diminuindo o número de graus de liberdade, isto é, o número de variáveis. O número de variáveis da equação pode ser reduzido, por exemplo, considerando somente os elétrons da camada de valência, tomando os demais como se formassem com o núcleo um “caroço iônico”, e então resolver a Equação de Schrödinger para os elétrons de valência se movendo no campo potencial do "caroço" iônico. A justificativa para essa aproximação é o fato de que os elétrons de valência em geral contribuem para a formação das ligações químicas que originam os sistemas moleculares.

Existem dois modelos principais para cálculos de estrutura molecular: a teoria do orbital molecular e a teoria de ligações de valência, sendo que a primeira tem sofrido um desenvolvimento maior do que a segunda. Assim, apresentaremos neste capítulo o tratamento de Hartree-Fock, baseado na teoria do orbital molecular.

A primeira aproximação em cálculos de estrutura molecular é a **aproximação de Born-Oppenheimer**, que separa o problema eletrônico do problema nuclear, admitindo que isto pode ser feito pelo fato de a massa nuclear ser muito maior que a

eletrônica. Assim, os elétrons se moveriam no campo produzido pelos núcleos em cada instante, de forma que os núcleos poderiam ser tomados como estacionários para uma análise do movimento eletrônico. Assim, os termos de energia potencial da interação que misturam as variáveis dos núcleos e elétrons, dificultando a separação da equação, são eliminados. O problema é então simplificado, pois desta forma eliminam-se na expressão do hamiltoniano os termos de energia cinética dos núcleos e a energia potencial da interação entre os núcleos se torna constante (segundo e quinto termo na equação 1). Por outro lado, a Equação de Schrödinger deve ser resolvida para várias posições do núcleo, num processo iterativo. O conjunto de soluções assim obtido permite a construção de uma superfície de energia potencial molecular e a identificação das conformações de equilíbrio da molécula (nos pontos de menor energia). A aproximação de Born-Oppenheimer é bastante realista para estados eletrônicos fundamentais, sendo menos realista para estados excitados e algumas outras situações, onde outras aproximações são necessárias¹².

Embora a aproximação de Born-Oppenheimer seja uma simplificação considerável, persiste uma dificuldade fundamental, que é característica de sistemas polieletrônicos: a correlação eletrônica, ou seja, as interações mútuas entre os elétrons, que afetam os seus movimentos, sendo necessária outra aproximação para tornar o problema aproximadamente solúvel. Para o estudo de moléculas é comum a utilização da **aproximação de Hartree-Fock**. Este é um modelo de partículas independentes, no qual cada elétron é tratado como se estivesse se movendo isolado num potencial resultante que descreve a média de suas interações coulombianas com os núcleos e os demais $N - 1$ elétrons. Esta energia potencial efetiva é determinada a partir de uma função de onda, proposta inicialmente, como

uma expansão num conjunto de funções monoelétrônicas (spin – orbitais). Resolvendo a Equação de Schrodinger para cada elétron, considerando o potencial efetivo, o spin-orbital para este elétron é calculado numa nova aproximação. O processo é repetido para todos os elétrons que compõem o sistema molecular, separadamente. Uma vez determinados todos os novos spin-orbitais, repete-se o ciclo re-otimizando-os até que o valor do potencial obtido sofra uma mudança suficientemente pequena se comparada com o ciclo anterior. Por isso, o método de Hartree-Fock é dito autoconsistente.

A função de onda inicial para o cálculo é escrita como um produto anti-simetrizado de spin-orbitais, $\chi(x)$, ou seja, uma função que descreverá tanto a função espacial como a de spin. O spin do elétron é contado através de duas funções ortonormais, $\alpha(\omega)$ e $\beta(\omega)$ (spin +1/2 e -1/2, respectivamente).

Teremos:

$$\chi_{2i-1}(x) = \Psi_i(r)\alpha(\omega) \quad (2)$$

$$\chi_{2i}(x) = \Psi_i(r)\beta(\omega), \quad \text{onde } i = 1, 2, \dots, k. \quad (3)$$

Os spin-orbitais devem satisfazer à condição de ortonormalidade:

$$\langle \chi_i | \chi_j \rangle = \delta_{ij}. \quad (4)$$

Com as funções de onda de um elétron, podemos escrever a função de onda de um sistema de elétrons, supostamente, não interagentes. Um hamiltoniano H_1 deste sistema pode ser descrito como uma somatória de hamiltonianos de um elétron $h(i)$. Estes serão constituídos apenas de um termo para a energia cinética e outro para a energia potencial do i-ésimo elétron, já que inicialmente desprezaremos a repulsão elétron-elétron. As autofunções dos hamiltonianos $h(i)$ serão os spin-orbitais $\chi(x)$, ou seja, teremos:

$$h(i)\chi_j(x) = \varepsilon_j\chi_j(x) \quad (5)$$

A função de onda multieletrônica será dada pelo produto de spin-orbitais:

$$\Psi^{HP}(x_1, x_2, \dots, x_n) = \chi_i(x_1)\chi_j(x_2)\dots\chi_k(x_n), \quad (6)$$

que é uma autofunção de H_1 , portanto teremos:

$$H_1\Psi^{HP} = E_1\Psi^{HP}, \quad (7)$$

onde E_1 será a soma dos autovalores dos spin-orbitais:

$$E_1 = \varepsilon_i + \varepsilon_j + \varepsilon_k. \quad (8)$$

A função de onda aproximada Ψ^{HP} é chamada de produto de Hartree e podemos observar dois aspectos da mesma: a) não inclui de forma completa os efeitos de repulsão elétron-elétron; e, b) não respeita a antissimetria requerida para uma autofunção eletrônica, ou seja, para a troca de coordenadas entre dois elétrons não há a necessária troca de sinal para a função de onda para férmions.

A idéia é utilizar Ψ^{HP} como função de onda aproximada para o sistema de elétrons interagentes. Para antissimetrizá-la corretamente utilizaremos o chamado determinante de Slater, onde colocaremos os spin-orbitais alinhados em linhas e colunas, cada coluna referindo-se a um mesmo spin-orbital, enquanto que cada linha refere-se a um mesmo elétron, ou vice-versa. Com isso, a troca de duas linhas ou colunas (coordenadas de dois elétrons) causará a troca de sinal do determinante (da função de onda). De forma compacta podemos escrever a função de onda do estado fundamental como o seguinte determinante de Slater:

$$|\Psi_0\rangle = |\chi_1\chi_2\dots\chi_n\rangle. \quad (9)$$

Com o hamiltoniano eletrônico total sendo dado por

$$H_{el} = -\sum_{i=1}^n \frac{1}{2} \nabla_i^2 - \sum_{i=1}^n \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^n \sum_{j>i}^n \frac{1}{r_{ij}}, \quad (10)$$

e a função de onda total descrita acima, podemos escrever a expressão para a energia total do sistema. Necessitamos ainda da definição analítica de um determinante para escrever:

$$|\Psi_0\rangle = \sum_{i=1}^{n!} (-1)^{P_i} P_i \{\chi_m(1)\chi_n(2)\dots\}, \quad (11)$$

onde P_i é o operador que permuta o índice da coluna e p_i é o número de transposições necessárias para restaurar uma dada permutação $i_1, i_2 \dots i_n$, para uma ordem natural 1, 2, 3, ..., n.

Após algumas manipulações algébricas podemos chegar, através destas expressões, à expressão para a energia do estado fundamental¹:

$$E_0 = \langle \Psi_0 | H | \Psi_0 \rangle = \sum_i^n \langle i | h | i \rangle + \frac{1}{2} \sum_i^n \sum_j^n \langle ii | jj \rangle - \langle ij | ji \rangle. \quad (12)$$

Onde foram usadas as seguintes definições:

$$\langle i | h | j \rangle = \zeta dx_1 \chi_i^*(x_1) h(x_1) \chi_j(x_1), \text{ as integrais de um elétron e,}$$

$$\langle ij | kl \rangle = \zeta dx_1 dx_2 \chi_i^*(x_1) \chi_j(x_1) \frac{1}{r_{1,2}} \chi_k^*(x_2) \chi_l(x_2), \text{ as integrais de dois elétrons.}$$

O caminho que nos leva às equações de Hartree-Fock consiste em minimizar a energia do sistema E_0 ($\{\chi_i\}$) em relação aos spin-orbitais, sendo que a ortonormalidade dos mesmos funcionará como condição de vínculo. Escrevemos então a função de Lagrange L ($\{\chi_i\}$) para a energia do estado fundamental:

$$L(\{\chi_i\}) = E_0(\{\chi_i\}) - \sum_i^n \sum_j^n \lambda_{ij} (\langle i | j \rangle - \delta_{ij}), \quad (13)$$

com λ_{ij} constituindo um conjunto de multiplicadores de Lagrange.

Utilizaremos ainda duas definições:

Operador de Troca:
$$K_j(1) \chi_i(1) = [\zeta dx_2 \chi_j^*(2) \frac{1}{r_{12}} \chi_i(2)] \chi_j(1), \quad (14)$$

Operador de Coulomb:
$$J_j(1)\chi_i(1) = [\zeta dx_2 \chi_j^*(2) \frac{1}{r_{12}} \chi_j(2)]\chi_i(1). \quad (15)$$

O Operador de Coulomb $J(1)$ representa a interação entre o elétron 1 e a densidade eletrônica dada por $|\chi_j(2)|^2$; já o Operador de Troca não possui uma interpretação física similar. Este termo surge como manifestação da contribuição para a energia do sistema, devido à antissimetria das funções de spin-orbitais. Podemos observar pela expressão 14, que quando $K_j(1)$ opera sobre $\chi_i(1)$, ocorre a troca entre o elétron 1 e o elétron 2 do lado direito de $\frac{1}{r_{12}}$. O resultado desta operação depende do valor do spin-orbital $\chi_i(1)$ em todo o espaço.

O princípio variacional consiste em tomar variações arbitrárias nas funções χ , obedecendo ao vínculo de ortonormalidade, e impor que as mesmas levem a uma variação nula na energia.

Assim, chegamos a **Forma não Canônica das Equações de Hartree-Fock**, por não ser uma equação de autovalores na forma mais comum:

$$[h(1) + \sum_{j=1}^n J_j(1) - K_j(1)]\chi_i(1) = \sum_{j=1}^n \lambda_{ij} \chi_j(1), \quad (16)$$

com $i = 1, 2, 3, \dots, n$. Definindo o **operador de Fock** como,

$F = [h(1) + \sum_{j=1}^n J_j(1) - K_j(1)]$, podemos reescrever a equação 16 como:

$$F|\chi_i\rangle = \sum_{j=1}^n \lambda_{ij} |\chi_j\rangle. \quad (17)$$

¹Para as moléculas de camada fechada consideramos um par de spin orbitais para cada orbital espacial (RHF-Restricted Hartree-Fock). Para as moléculas de camada aberta, os spin orbitais podem diferir tanto na função espacial quanto na função de spin (UHF-Unrestricted Hartree-Fock).

Temos agora que resolver um conjunto de equações diferenciais de um elétron. Se todos os λ_{ij} com $i \neq j$ fossem iguais a zero, teríamos uma equação de autovalores. Isso pode ser conseguido mudando o conjunto de base, através de uma transformação unitária. Através da aplicação de uma transformação unitária aos spin-orbitais $\{\chi_i\}$ podemos encontrar um novo conjunto $\{\chi'_i\}$ que será também um conjunto ortonormal. Essa transformação, uma rotação no sistema de coordenadas, não afeta o significado físico da função de onda, nem a forma dos operadores de Troca e de Coulomb. Do ponto de vista dos multiplicadores de Lagrange, λ_{ij} , uma expressão matricial envolvendo a nova e a antiga matriz pode ser escrita: $\lambda' = U^+ \lambda U$; e podemos encontrar U, existente e única, tal que a transformação diagonalize λ . E então, sem perda de generalidade, a **Forma Canônica das Equações de Hartree-Fock** pode ser escrita como:

$$F|\chi_i\rangle = \lambda_i|\chi_i\rangle \quad (18)$$

O princípio destas equações consiste em substituir o problema multieletrônico por vários problemas de um elétron, de forma que a repulsão eletrônica possa ser tratada através de um campo médio. O operador de Fock é um operador hermitiano, todos seus autovalores são reais e as autofunções pertencentes a diferentes autovalores são mutuamente ortogonais. Com a solução da equação 18 teremos um conjunto de soluções em que os n autovalores devem ser os de menor valor para o estado fundamental. As n funções χ'_i correspondentes são chamadas de *orbitais moleculares ocupados* do estado fundamental, as demais autofunções de F são denominadas de *estados virtuais*. Ainda, como o operador de Fock depende dos χ'_i , a Equação Canônica de Hartree-Fock deve ser resolvida de forma iterativa até a autoconsistência.

Existem diferentes formas de se resolver a equação de Hartree-Fock. Uma delas, proposta por Roothaan em 1951, consiste em expandir a função espacial em termos de um conjunto finito de funções de base monoelétrônicas. Na teoria dos orbitais moleculares as funções de onda moleculares são determinadas via combinação linear de orbitais atômicos, denominada **MO-LCAO** (Molecular Orbital-Linear Combination of Atomic Orbitals):

$$\chi'_i = \sum_{\mu=1}^{\xi} C_{\mu i} \phi_{\mu}, \text{ com } i = 1, 2, \dots, \xi \quad (19)$$

(supondo uma base completa e de dimensão finita ξ para que a função de onda seja descrita perfeitamente).

Chegamos às equações de Roothaan substituindo a expansão 19 na Forma Canônica da Equação de Hartree:

$$\sum_{\nu} F_{\mu\nu} C_{\nu i} = \varepsilon_i \sum_{\nu} S_{\mu\nu} C_{\nu i}, \text{ ou} \quad (20)$$

$$\sum_{\nu=1} C_{\nu i} \zeta d^3 r_1 \phi_{\mu}^*(1) F(1) \phi_{\nu}(1) = \varepsilon_i \sum_{\nu=1} C_{\nu i} \zeta d^3 r_1 \phi_{\mu}^*(1) \phi_{\nu}(1) \quad (21)$$

com $i = 1, 2, \dots, \xi$.

Onde foram utilizadas as seguintes definições:

* **integral / matriz de recobrimento (overlap):** $S_{\mu\nu} \equiv \zeta d^3 r_1 \phi_{\mu}^*(1) \phi_{\nu}(1)$ ou

$$S_{\mu\nu} \equiv \langle \phi_{\mu} | \phi_{\nu} \rangle;$$

* **integral / matriz de Fock:** $F_{\mu\nu} \equiv \zeta d^3 r_1 \phi_{\mu}^*(1) F(1) \phi_{\nu}(1)$ ou $F_{\mu\nu} \equiv \langle \phi_{\mu} | F | \phi_{\nu} \rangle$.

Então na forma matricial a equação de Hartree-Fock Roothaan pode ser escrita como:

$$FC = SC\varepsilon \quad (22)$$

Para que as equações tenham soluções diferentes da trivial, a seguinte condição deve ser satisfeita: $\det(F_{\mu\nu} - \varepsilon_i S_{\mu\nu}) = 0$. As raízes desta equação secular dão as energias dos orbitais moleculares ε_i . As equações são resolvidas por um processo iterativo, e o ciclo é repetido até que o critério de convergência seja alcançado.

Métodos de cálculos *ab initio* ou *de primeiros princípios* calculam explicitamente as integrais envolvendo os orbitais atômicos, usando apenas constantes físicas fundamentais, sem utilizar nenhum outro parâmetro experimental. Os cálculos *ab initio* apresentam resultados tanto melhores quanto melhor for o conjunto de base usado na expansão das funções de onda, sendo que o número finito de funções de base requerido para tornar os cálculos exequíveis também limita a qualidade dos resultados obtidos. Quanto maior o número e mais adequadas para os átomos que compõem a molécula forem as funções utilizadas, mais precisos serão os resultados e, também será maior a exigência computacional dos cálculos. Já está sendo usada como alternativa dentro dos métodos *ab-initio* a fim de reduzir o custo computacional dos cálculos a Teoria do Funcional Densidade ^[4,5].

Os métodos semi-empíricos foram desenvolvidos devido às dificuldades em se aplicar um cálculo *ab initio* para sistemas moleculares, que em geral possuem um grande número de átomos, o que torna o número de integrais de dois elétrons a serem resolvidas muito elevado. Assim, esses métodos utilizam resultados experimentais para parametrizar algumas integrais que envolvem os dois elétrons.

Na próxima seção apresentaremos de forma breve alguns métodos semi-empíricos desenvolvidos para o estudo da estrutura eletrônica de moléculas, mencionando suas principais diferenças e simplificações adicionais envolvidas em cada caso.

2.2. Métodos e aproximações Semi-empíricas [1, 6-12]

Em princípio, cálculos utilizando as equações de Hartree-Fock-Roothaan ficariam limitados a sistemas mais simples, pois o tempo computacional necessário cresce com k^4 , onde k é o número de funções de base ou orbitais atômicos, associado ao grande número de integrais de dois elétrons a serem resolvidas. Para equacionar estas dificuldades uma das linhas de pesquisa é a dos métodos semi-empíricos, que mantêm o formalismo autoconsistente da teoria de orbitais moleculares usando a expansão em orbitais atômicos e busca solucionar as dificuldades inerentes ao método utilizando hamiltonianos mais simples, desprezando algumas integrais decorrentes da formulação de Schrodinger ou ainda, introduzindo parâmetros obtidos de resultados experimentais. As aproximações variam de acordo com o método adotado.

O método semi-empírico monoelétrônico mais simples para moléculas não planares é o **método de Huckel Extendido (EHT – Extended Huckel Theory)**. Não é um método autoconsistente, é rápido e qualitativamente bom para investigar a estrutura de orbitais de sistemas moleculares, mas não é utilizado em procedimentos de otimização de geometria.

O método EHT utiliza o modelo de elétron independente, admitindo a descrição do hamiltoniano eletrônico como a soma de hamiltonianos efetivos de um elétron:

$$H_{el} = \sum_i H_i^{eff}$$

Construímos assim uma matriz H , onde os orbitais moleculares χ_i e suas energias orbitais ε_i são soluções do problema de autovalores padrão $HC = SC\varepsilon$,

demonstrável pelo método da variação linear de funções. A matriz C contém os coeficientes da combinação linear que descreve os orbitais moleculares em termos

de orbitais atômicos (aproximação LCAO): $\chi_i = \sum_{\mu=1} C_{\mu} \phi_{\mu}$, e ε é uma matriz diagonal contendo as energias orbitais.

2.2.1. Aproximações NDO

Muitos métodos semi-empíricos autoconsistentes, baseados na teoria Hartree-Fock, foram desenvolvidos. Um conjunto de métodos apresenta a aproximação **NDO** (Neglect of Differential Overlap), caracterizada por desprezar os efeitos do recobrimento entre os orbitais atômicos nos cálculos.

A aproximação mais simples despreza completamente o recobrimento diferencial - **CNDO** (Complete Neglect of Differential Overlap): $\phi_{\mu} \phi_{\nu} = \delta_{\mu\nu} \phi_{\mu} \phi_{\mu}$. O método (**CNDO** – Complete Neglect of Differential Overlap) foi proposto por Pople, Santry e Segal em 1965, sendo o método semi-empírico autoconsistente mais simples, que utiliza como base um conjunto de orbitais atômicos do tipo Slater para os elétrons de valência e a aproximação **ZDO** (Zero Differential Overlap), que despreza as integrais de repulsão eletrônicas de orbitais atômicos diferentes, centrados no mesmo átomo. As conseqüências da aproximação **ZDO** são: i) a matriz de recobrimento **S** é reduzida à matriz unidade ($S_{\mu\nu} = \delta_{\mu\nu}$); ii) as integrais de um elétron envolvendo três centros (dois das funções de base e uma do operador) são zero; iii) todas as integrais de três e quatro centros, que são as integrais mais numerosas, são desprezadas, ou seja, a teoria considera apenas integrais de um e de dois centros.

Todos os métodos semi-empíricos são baseados na aproximação **ZDO**. Os métodos de aproximação diferem, principalmente, no grau com que a **ZDO** é aplicada para as integrais de repulsão eletrônica.

A aproximação **ZDO** não é invariante para uma rotação do sistema de coordenadas. Para manter a invariância o método **CNDO** faz uma aproximação adicional: faz com que as integrais de repulsão eletrônica de dois elétrons dependam sobre qual átomo o orbital atômico está centrado e não dependa da natureza dos orbitais.

No método **CNDO** a exclusão de integrais de troca de um centro é uma das maiores limitações, o que, freqüentemente, torna a teoria incapaz de distinguir estados provenientes de uma mesma configuração. Essa dificuldade não aparece numa segunda aproximação, chamada **NDDO** (Neglect of Diatomic Differential Overlap), também introduzida por Pople, Santry e Segal em 1965, que consiste em desprezar o recobrimento diferencial diatômico: $\phi_{\mu}^A \phi_{\nu}^B = \delta_{AB} \phi_{\mu}^A \phi_{\nu}^A$. Nesta aproximação, apenas o recobrimento entre orbitais atômicos em átomos diferentes são negligenciadas nos cálculos das integrais. Porém, esse método requer o cálculo de um número muito grande de integrais de dois centros. A solução, então, é encontrada no método intermediário chamado **INDO** ((Intermediate Neglect of Differential Overlap).

O método **INDO** surgiu em 1967 com Pople, Beveridge e Doboser. No **INDO**, os recobrimentos entre orbitais atômicos do mesmo átomo não são desprezados nas integrais de repulsão de um centro, mas ainda o são nas integrais de repulsão de dois centros.

Nos métodos semi-empíricos há parâmetros que necessitam ser determinados por algum processo independente, não sendo única a forma de

determiná-los, e constitui o que é denominado de *parametrização* do método. Quanto melhor esses parâmetros maior a eficácia do método na obtenção de propriedades moleculares.

Entre os métodos mais recentes estão o **MNDO** (Modified NDO), o **AM1** (Austin Method 1) e o **PM3** (Parametric Method 3) parametrizados para a aproximação **NDDO**. Ainda mais recentes, estão os métodos **PM5** (Parametric Method 5), **PM6** (Parametric Method 6), do Stewart.

2.2.2. Aproximações MNDO, AM1 e PM3

O **MNDO** é um método semi-empírico, sofisticado, autoconsistente e parametrizado para o estudo de geometrias moleculares e caminhos de reação.

Os orbitais moleculares da camada de valência (Ψ_i) [χ_i] são representados por combinações lineares de um conjunto de base mínima de camada de valência (ϕ_v) na aproximação **LCAO**. Os coeficientes da expansão são obtidos pelas equações de Roothaan, que na aproximação **NDDO** é escrita como:

$$\sum_v (F_{\mu v} - E_i \delta_{\mu v}) C_{vi} = 0,$$

onde E_i , é o autovalor associado ao orbital molecular Ψ_i e $\delta_{\mu v}$ é a delta de Kronecker.

A energia eletrônica E_{el} será dada por:

$$E_{el} = \frac{1}{2} \sum_{\mu} \sum_{\nu} P_{\mu\nu} (H_{\mu\nu} + F_{\mu\nu}),$$

onde $P_{\mu\nu}$ são os elementos de matriz de ordem de ligação e $H_{\mu\nu}$ do hamiltoniano de caroço.

Utilizando-se a notação ϕ_μ e ϕ_ν para os orbitais atômicos centrados em um átomo A e ϕ_λ e ϕ_σ para aqueles centrados no átomo B (A \neq B genéricos), os elementos da matriz de Fock no formalismo **MNDO** serão dados por:

$$F_{\mu\mu} = U_{\mu\mu} + \sum_B V_{\mu\mu,B} + \sum_\nu^A P_{\nu\nu} (\langle \mu\mu | \nu\nu \rangle - \frac{1}{2} \langle \mu\nu | \mu\nu \rangle) + \sum_B \sum_{\lambda,\sigma}^B P_{\lambda\sigma} \langle \mu\mu | \lambda\sigma \rangle,$$

para os termos da diagonal e

$$F_{\mu\nu} = \sum_B V_{\mu\nu,B} + \frac{1}{2} \sum_\nu P_{\mu\nu} (3 \langle \mu\nu | \mu\nu \rangle - \langle \mu\mu | \nu\nu \rangle) + \sum_B \sum_{\lambda,\sigma}^B P_{\lambda\sigma} \langle \mu\nu | \lambda\sigma \rangle,$$

para os termos fora da diagonal. Temos ainda:

$$F_{\mu\lambda} = \beta_{\mu\lambda} - \frac{1}{2} \sum_\nu^A \sum_\sigma^B P_{\nu\sigma} \langle \mu\nu | \lambda\sigma \rangle,$$

onde:

- $P_{\mu\nu}$ é a matriz de ordem de ligação (*bond order*);
- $U_{\mu\mu}$ são as energias monoelétrônicas de um centro;
- $\langle \mu\mu | \nu\nu \rangle$ e $\langle \mu\nu | \mu\nu \rangle$ são as integrais de Coulomb e de troca, respectivamente;
- $\beta_{\mu\lambda}$ são as integrais de ressonância de dois centros;
- $V_{\mu\nu,B}$ representa o termo de atração eletrostática;
- $\langle \mu\nu | \lambda\sigma \rangle$ são as integrais de repulsão de dois centros.

A energia total E_{tot}^{mol} é escrita como a soma da energia eletrônica e as repulsões $E_{AB}^{caroço}$ entre os “caroços” dos átomos A e B:

$$E_{tot}^{mol} = E_{el} + \sum_{A < B} \sum E_{AB}^{caroço}.$$

O calor de formação molecular C_F será dado por:

$$C_F^{mol} = E_{tot}^{mol} - \sum_A E_{el}^A + \sum_A C_F^A,$$

onde E_{tot}^{mol} é a energia total da molécula, E_{el}^A é a energia eletrônica e C_F^A os calores de formação experimentais para os átomos da mesma.

Na aproximação **MNDO**, os vários termos da matriz de Fock e as repulsões $E_{AB}^{caroço}$ não são determinadas analiticamente, mas sim determinados através de dados experimentais ou de expressões semi-empíricas que contêm parâmetros ajustáveis aos dados experimentais. A introdução destes parâmetros compensará, em princípio, as limitações da função de onda total (que despreza a correlação eletrônica) e erros inerentes às simplificações implementadas pelo método.

Os termos de um centro são obtidos através de vários estados de valência do átomo e seus íons, a partir de valores espectroscópicos correspondentes. Por outro lado, as integrais de repulsão de dois centros representam a energia de interação de carga ($e\phi_\mu\phi_\nu$) no átomo A e ($e\phi_\lambda\phi_\sigma$) no átomo B, e seria inconsistente tentar calculá-las analiticamente devido às aproximações básicas do **MNDO**. Do ponto de vista de mecânica clássica, estas integrais serão iguais à soma sobre todas as interações de momentos de multipolo (M_{lm} com l e m especificando a ordem e orientação dos mesmos) de duas distribuições de carga. Desta forma, as integrais de repulsão de dois centros são expandidas em termos das interações de multipolo-multipolo:

$$\langle \mu\nu | \lambda\sigma \rangle = \sum_{l_1} \sum_{l_2} \sum_m [M_{l_1 m}^A, M_{l_2 m}^B].$$

Cada multipolo M_{lm} é representado através de uma configuração apropriada $[M_{lm}]$ de 2^l pontos de carga de magnitude $\frac{e}{2^l}$, com separações D_l . A interação $[M_{l_1 m}^A, M_{l_2 m}^B]$ entre dois multipolos é, então, calculada pela aplicação de uma fórmula semi-empírica apropriada para cada uma das interações entre os pontos de carga nas duas configurações, e através da soma sobre todos eles. Denota-se a distância

entre os pontos de carga i e j nas configurações interagentes dos átomos A e B como R_{ij} .

Os termos de atração elétron-caroço ($V_{\mu\mu,B}$) e repulsão caroço-caroço ($E_{AB}^{caroço}$) também serão parametrizados através de parâmetros ajustáveis.

O **MNDO** retém todos os termos de dois centros envolvendo recobrimento diferencial monoatômico, incluindo interações de multipolos de ordem superior. O método, porém, falha na descrição das ligações de hidrogênio e superestima as repulsões entre átomos não ligados. Desta forma foi necessário corrigir a forma de repulsão caroço-caroço, dando origem ao método denominado **AM1**. A função de repulsão de caroço (CRF) do **AM1** será dada por:

$$CRF(AB) = Z_A Z_B \gamma_{SS} [1 + F(A) + F(B)],$$

onde:

$$F(A) = \exp(-\alpha_A R_{AB}) + \sum_i K_{Ai} \exp[L_{Ai} (R_{AB} - M_{Ai})^2],$$

$$F(B) = \exp(-\alpha_B R_{AB}) + \sum_j K_{Bj} \exp[L_{Bj} (R_{AB} - M_{Bj})^2],$$

- Z_A e Z_B são os números atômicos dos núcleos A e B, respectivamente.
- L representa parâmetros que determinam as larguras das funções Gaussianas utilizadas na parametrização. M e K são parâmetros que foram otimizados.

O formalismo usado no **AM1** é essencialmente o mesmo do **MNDO**, com exceção do termo **CRF**. As integrais de repulsão eletrônicas de um centro permanecem inalteradas. Os parâmetros K, α e M_x estão contidos no **CRF**. O **AM1** mostrou-se superior ao **MNDO** devido ao fato de as geometrias de equilíbrio estarem em maior concordância com os resultados experimentais.

O método **PM3**, introduzido por James J. P. Stewart, é na verdade o **AM1** com nova parametrização, definida através do estudo de um número muito maior de dados experimentais do que os considerados para os métodos anteriores. Alguns trabalhos mostram um melhor acordo do **PM3** com resultados experimentais, tanto para parâmetros estruturais quanto para outras propriedades de algumas famílias de moléculas orgânicas [15,16,17,19].

Podemos acrescentar que as vantagens dos métodos semi-empíricos são a rapidez no cálculo e a precisão. Tanto mais preciso é o método quanto maior e mais diversificado é o conjunto de moléculas utilizadas para parametrização.

O método semi-empírico **PM3** (assim como o MNDO e o AM1) é baseado na aproximação Hartree-Fock e não inclui explicitamente a correlação eletrônica, tendendo a superestimar a separação entre os orbitais moleculares ocupados e desocupados e portanto superestimar as energias de excitação eletrônica, não sendo adequado para simular o espectro de absorção associado às transições eletrônicas moleculares. Zerner e colaboradores prepararam uma parametrização do **INDO (ZINDO/S)**[13] para a simulação dos espectros, utilizando interação de configurações (**CI**)[14], que permite incluir nos cálculos os efeitos da correlação eletrônica, ainda que de forma aproximada. Estes espectros são particularmente importantes já que uma das famílias estudadas, as furanocumarinas, podem ser ativadas usando-se luz ultravioleta (fotoquimioterapia/terapia fotodinâmica).

Para os cálculos realizados neste trabalho foram usados métodos semi-empíricos, que são econômicos, quando comparamos o tempo gasto na execução dos cálculos em relação aos métodos *ab initio* e, já mostraram em trabalhos anteriores que produzem boa qualidade de resultados. Esse aspecto é essencial, tendo em vista o grande número de estruturas moleculares investigadas e o

propósito de testar a M.I.E.[15], em princípio aplicável a conjuntos de moléculas contendo centenas ou milhares de derivados produzidos por métodos combinatórios de síntese de possíveis novos fármacos.

2.3 – Metodologia de Índices Eletrônicos – MIE [15-25]

A **MIE** foi desenvolvida, inicialmente, para investigar a ação carcinogênica dos Hidrocarbonetos Policíclicos Aromáticos (HAP's), com o objetivo de distinguir qualitativamente HAP's ativos de inativos. A MIE envolve parâmetros puramente eletrônicos e está baseada no conceito de densidade local de estados (LDOS – Local Density of States), e em valores críticos da diferença de energia eletrônica entre os níveis dos orbitais moleculares de fronteira.

A densidade de estados (DOS - *Density Of States*) é definida como o número de estados eletrônicos por unidade de energia. Para um conjunto de níveis discretos não degenerados (E_i) de um dado sistema molecular, teríamos, por exemplo: $DOS(E) = \delta_{E,E_i}$.

O conceito correlato da LDOS, isto é, a DOS calculada sobre uma região molecular específica, nos permite mapear a distribuição de um orbital molecular, ocupado ou virtual, sobre um esqueleto molecular. Dessa forma, a LDOS é uma indicação de como um átomo, ou um conjunto de átomos, contribui para a formação de um orbital molecular específico.

O conceito de LDOS é introduzido para que se possa descrever a distribuição espacial dos estados sobre o sistema que se está estudando. A contribuição de cada átomo para um dado nível eletrônico pode ser ponderada de diversas formas. A definição de “peso” mais usada pelos químicos leva em conta a densidade de probabilidade calculada no sítio em questão, com o fator sendo dado, por:

$$W_{\mu}(E_i) = \sum_{\nu} S_{\mu\nu} c_i^{*\mu} c_i^{\nu},$$

onde a soma deve se estender sobre os sítios delimitados pelos orbitais atômicos envolvidos.

Para hamiltonianos baseados na aproximação da combinação linear dos orbitais atômicos (LCAO), como os utilizados nesse trabalho, a contribuição de cada átomo para a formação de um nível eletrônico é dada pelo coeficiente do orbital atômico elevado ao quadrado, isto é, pela densidade de probabilidade correspondente ao nível, naquele sítio. Para LDOS a soma é feita sobre todos os orbitais atômicos correspondentes aos átomos de uma região que se deseja

$$\text{investigar } (n_i \text{ a } n_f): LDOS(E_i) = 2 \sum_{m=n_i}^{n_f} |c_{mi}|^2 .$$

O número 2 na equação acima vem do princípio de exclusão de Pauli (máximo de 2 elétrons por nível eletrônico).

A MIE utiliza dois descritores: η e Δ . O primeiro parâmetro, η , é definido como uma medida da diferença da contribuição dos átomos de uma determinada região molecular para a formação de dois níveis moleculares:

$$\eta = LDOS_{nível1} - LDOS_{nível2}, \text{ ou seja,}$$

$$\eta = 2 \sum_{m=n_i}^{n_f} (|c_{m,nível1}|^2 - |c_{m,nível2}|^2) .$$

O segundo parâmetro fundamental da MIE, Δ , é definido como a diferença de energia entre os mesmos dois níveis moleculares da equação acima:

$$\Delta = E_{nível1} - E_{nível2}$$

Δ contém uma informação global da molécula (envolvendo os autovalores moleculares) enquanto η contém informações locais que refletem a distribuição de cargas sobre a molécula, ou seja, pode ser interpretado como uma medida indireta de carga local e, conseqüentemente, reatividade química. Os parâmetros η e Δ são utilizados então para construir regras que possam classificar, qualitativamente, compostos ativos e inativos. De forma geral, as regras obtidas envolvem a

comparação com valores críticos: η_c e Δ_c . As regras obtidas podem apresentar a forma:

se $\eta > \eta_c$ e $\Delta > \Delta_c$, a molécula será ativa, caso contrário será inativa.

Os valores críticos, η_c e Δ_c , são determinados através de uma análise sistemática dos parâmetros teóricos frente à classificação de atividade dos compostos dada pelos índices experimentais. Estes valores são definidos como o valor limite de η e Δ através dos quais os compostos ativos podem ser separados dos inativos.

Os orbitais moleculares de interesse na investigação com a MIE, são os orbitais de fronteira HOMO (Highest Occupied Molecular Orbital), HOMO-1, LUMO (Lowest Unoccupied Molecular Orbital) e LUMO+1. O destaque dado a estes orbitais está relacionado com sua importância na reatividade dos compostos, uma vez que estes são os orbitais envolvidos nas reações químicas entre dois compostos e nas excitações eletrônicas de mais baixa energia.

Durante a busca dos valores críticos, η_c e Δ_c , são determinados quais os orbitais moleculares [(HOMO e HOMO-1) ou (LUMO e LUMO+1)] que melhor representam a atividade biológica dos compostos - a base para a construção das regras de seleção.

Desde seu desenvolvimento em 1996, a MIE já foi aplicada no estudo de atividade de compostos de famílias distintas com aplicações das mais diversificadas: carcinogênico, anti-tumoral, antibiótico, anti-inflamatório, anti-malarial e com atividade progestacional. É intrigante que com apenas dois parâmetros quânticos comuns seja possível identificar, pelo menos em nível qualitativo, compostos ativos e inativos de classes de moléculas com atividades biológicas distintas. Além disso, as regras se mantiveram essencialmente inalteradas nos casos em que diferentes

hamiltonianos foram utilizados. A MIE se mostra assim competitiva com técnicas tradicionais mais usadas no estudo de relações estrutura-atividade e em quimiometria, como PCA e HCA . PCA e HCA são métodos muito utilizados em pesquisa de reconhecimento de padrões, e serão descritas nas próximas seções.

2.4 – Análise de Componentes Principais – PCA [26-29]

PCA é uma ferramenta exploratória que mapeia o espaço das amostras (que em nosso estudo são os compostos estudados - moléculas) através de escores e as variáveis individuais por “loadings” em um novo espaço vetorial definido pelas componentes principais (PCs). As componentes principais são obtidas em ordem decrescente de importância, pois a primeira componente é a combinação linear das variáveis com maior amplitude de variância nos dados originais. Os escores permitem identificar claramente se as moléculas são similares ou não. Dos “loadings” as variáveis (que são os parâmetros eletrônicos, neste trabalho) mais importantes podem ser facilmente identificadas, bem como a correlação entre elas.

O método de análise PCA é considerado um ótimo método de seleção das variáveis relevantes na separação de compostos químicos quanto à uma atividade experimental específica.

PCA é uma metodologia classificada dentro da chamada análise estatística multivariada de reconhecimento de padrões, que abrange quantas variáveis se desejar, e cujo objetivo é reduzir o número dessas variáveis para um conjunto menor de parâmetros, eliminando as variáveis originais correlacionadas. Assim as moléculas são separadas em subconjuntos observando suas semelhanças, o que permite a classificação qualitativa da atividade biológica de diversificadas famílias de fármacos.

De forma geral, cada amostra (molécula) pode ser representada como um ponto no espaço N-dimensional das variáveis consideradas (parâmetros eletrônicos), onde cada variável equivale a um eixo. Este espaço N-dimensional é um espaço de variáveis correlacionadas, ou seja, onde informações semelhantes são dadas por

diferentes variáveis (**espaço N**). Queremos reduzir a dimensionalidade deste espaço, e o papel da PCA é construir um novo espaço (**espaço C**) de variáveis descorrelacionadas, que é o espaço de componentes principais (PCs).

Antes da análise PCA, as variáveis devem ser escalonadas, para que possam ser comparadas numa mesma escala, apesar de possuírem unidades diferentes. Utilizamos o auto-escalonamento que transforma as variáveis tal que sua média é zero e a variância deve estar entre 0 e 1. Na construção do **espaço C** alguns vínculos devem ser considerados, sendo os mais significativos a necessidade de que a distribuição estatística dos pontos (dispersão) no **espaço N** seja mantida e que os novos eixos sejam descorrelacionados (ortogonais).

A distribuição estatística das amostras é quantificada pela variância da distribuição dos pontos e as componentes principais são construídas maximizando-se esta variância. Desta forma a primeira componente principal (PC1) é a que contém a maior informação estatística do sistema inicial, ou seja, a primeira PC é gerada de maneira que responda pela maior parte da variância, seguida pela PC2, e assim sucessivamente até a PCn. O que se obtém como resultado é um novo espaço onde as primeiras PC's são responsáveis pela maior parte da informação relevante da distribuição e da similaridade entre as amostras. Matematicamente a construção dos PC's equivale a uma rotação de eixos de forma que os PC's são apenas combinações lineares das variáveis originais e são ortogonais entre si.

Os resultados de uma análise PCA são apresentados por um conjunto de amostras (gráfico de score) e um conjunto de descritores (gráfico de loading) representados em planos xy (planos das componentes principais de maior variância) (Figuras 2.1 e 2.2). As distribuições dos pontos nestes dois gráficos são correlacionadas, e para um melhor entendimento dos resultados obtidos com a

metodologia devem ser analisadas em conjunto. As amostras mais à direita no gráfico de score apresentam um alto valor para as variáveis mais à direita no gráfico de loading. O mesmo ocorrendo para amostras e descritores localizados mais à esquerda, na parte superior ou inferior dos gráficos.

Durante a aplicação da PCA, grande parte dos descritores inicialmente considerados são descartados, restando apenas os poucos capazes de separar as amostras em subgrupos de acordo com os dados experimentais de controle. A seleção dos “melhores” descritores é feita visualmente através dos resultados obtidos para os gráficos de score.

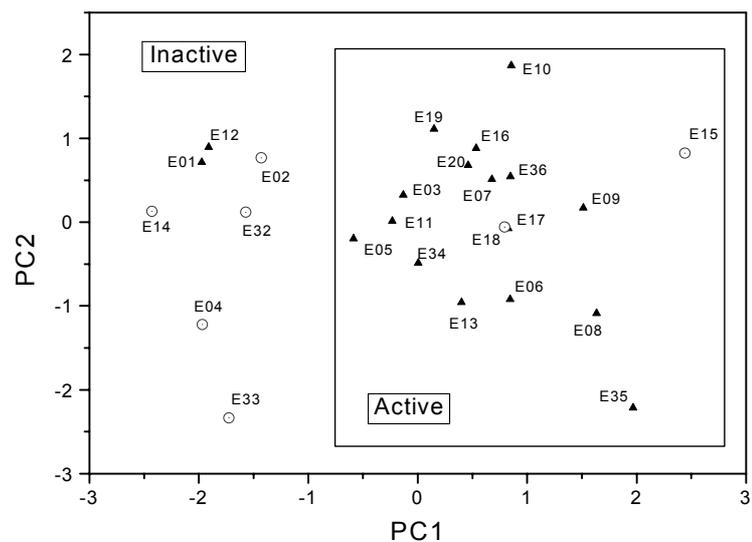


Figura 2. 1 – Gráfico de escores. Distribuição de amostras agrupadas pelo PCA.

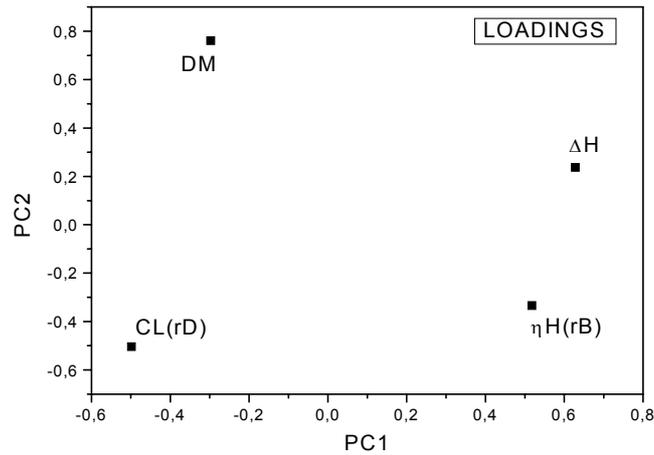


Figura 2.2 – Gráfico de loadings. Distribuição dos descritores responsáveis pela separação das amostras na figura 2.1.

2.4.1 – Obtendo as Componentes Principais (PC's)

Antes da obtenção das componentes principais iremos definir alguns itens que serão necessários. Para representar uma variável aleatória p-dimensional utilizaremos o vetor aleatório, representado por \mathbf{X} : $X^T = [X_1, \dots, X_p]$, onde X_1, \dots, X_p , são as variáveis aleatórias. Os vetores serão sempre representados sem índices baixos.

O vetor média $\mu^T = [\mu_1, \dots, \mu_p]$ tem como componentes:

$$\mu^T = \langle X_i \rangle = \int_{-\infty}^{\infty} x f_i(x) dx, \quad (1)$$

onde $f_i(x)$ é uma função peso, essas componentes são as médias das componentes de X . A variância da i-ésima componente é dada por:

$$\text{Var}(X_i) = \langle (X_i - \mu_i)^2 \rangle = \langle X_i^2 \rangle - \mu_i^2 \quad (2)$$

No caso univariado, a variância é usualmente denotada por σ_i^2 , mas para corresponder à notação da covariância que será definida a seguir, iremos representá-la por σ_{ii} no caso multivariado.

A covariância de duas variáveis X_i e X_j é definida por,

$$Cov(X_i, X_j) = \langle (X_i - \mu_i)(X_j - \mu_j) \rangle \quad (3)$$

dessa forma a covariância é o produto do desvio de duas variáveis em relação às suas respectivas médias. Em particular, se $i = j$ a covariância nada mais é do que a variância da variável i . Assim não temos necessidade de definir a variância, uma vez que ela é um caso particular da covariância. A covariância de X_i e X_j é usualmente representada por σ_{ij} ; se $i = j$, a variância de X_i é representada por σ_{ii} , como já dito.

A equação (3) é muitas vezes escrita em forma equivalente,

$$\sigma_{ij} = \langle X_i, X_j \rangle - \mu_i \mu_j \quad (4)$$

Com p variáveis, existem p variâncias e $1/2p(p-1)$ covariâncias. É útil apresentar estas quantidades na forma matricial, representada por Σ , cujo elemento (i,j) é σ_{ij} .

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$$

Essa matriz é muitas vezes chamada matriz de dispersão, matriz variância-covariância ou simplesmente matriz covariância. Os termos da diagonal são as variâncias, enquanto os termos fora da diagonal são as covariâncias. Como $\sigma_{ij} = \sigma_{ji}$ a matriz Σ é simétrica.

A matriz Σ pode ser representada em relação as variáveis e as respectivas médias, utilizando as equações (3) e (4).

$$\Sigma = \langle \langle X - \mu \rangle \langle X - \mu \rangle^T \rangle \quad (5)$$

$$\Sigma = \langle XX^T \rangle - \mu\mu^T.$$

Ilustraremos o uso das covariâncias na procura da variância de qualquer combinação linear das componentes de X . Consideremos a seguinte combinação linear $Y = a^T X$, onde $a^T = [a_1, \dots, a_p]$ é um vetor de constantes. Então Y_i é uma variável simples aleatória, cuja média é dada por:

$$\langle Y \rangle = a^T \mu, \quad (6)$$

enquanto a variância é obtida por $Var(Y) = \langle [a^T (X - \mu)]^2 \rangle$.

Como $a^T(X - \mu)$ é um escalar e todo escalar é igual a sua transposta, podemos expressar $Var(Y)$ em termos de Σ , usando (5):

$$Var(Y) = \langle a^T (X - \mu)(X - \mu)^T a \rangle = a^T \langle (X - \mu)(X - \mu)^T \rangle a = a^T \Sigma a \quad (7)$$

Consideremos o caso $p = 2$, onde $a^T X = a_1 X_1 + a_2 X_2$. Podemos escrever:

$$\begin{aligned} Var(a_1 X_1 + a_2 X_2) &= Var(a_1 X_1) + Var(a_2 X_2) + Cov(a_1 X_1, a_2 X_2) \\ &= a_1^2 Var(X_1) + a_2^2 Var(X_2) + 2a_1 a_2 Cov(X_1, X_2) \end{aligned}$$

ou,

$$\begin{bmatrix} a_1 & a_2 \end{bmatrix} \begin{bmatrix} Var(X_1) & Cov(X_1, X_2) \\ Cov(X_1, X_2) & Var(X_2) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$

As equações (6) e (7) podem ser generalizadas pela notação que se A é qualquer matriz ($p \times m$) de constantes, então pode facilmente ser mostrado que $A^T X$ possui vetor médio e a matriz de covariância é dada por:

$$\langle A^T X \rangle = A^T \mu \quad (8)$$

$$Var(A^T X) = A^T \Sigma A$$

Embora as covariâncias sejam úteis para muitas propostas matemáticas, elas são raramente utilizadas como estatísticas descritivas. Se duas variáveis estão linearmente relacionadas, então a covariância irá ser positiva ou negativa dependendo da inclinação da reta. E ainda, os coeficientes são de difícil interpretação por dependerem das unidades nas quais as duas variáveis foram medidas. Deste modo introduzimos um elemento, o coeficiente de correlação, que é dado pela covariância normalizada pelo produto dos desvios padrões das duas variáveis. A correlação de X_i e X_j será denotada ρ_{ij} :

$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j} \quad (9)$$

onde σ_i é o desvio padrão de X_i .

Com p variáveis, existem p variâncias e $p(p-1)/2$ distintas correlações. É muito útil representar as correlações por uma matriz ($p \times p$) cujos elementos são os coeficientes de correlação. Essa matriz será denotada por P , os termos da diagonal de P são unitários e os fora da diagonal são tal que a matriz é simétrica.

Para relacionarmos a matriz P com a matriz Σ introduziremos a matriz D , diagonal de ordem ($p \times p$), cujos termos da diagonal são os desvios padrões das componentes de X .

$$D = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_p \end{bmatrix}$$

As matrizes de covariância e correlação ficam relacionadas por:

$$\Sigma = DPD$$

ou

$$P = D^{-1}\Sigma D^{-1}$$

(10)

Onde os termos a diagonal de D^{-1} são recíprocos dos respectivos desvios padrões.

Voltemos as componentes principais.

Suponhamos que $X^T = [X_1, \dots, X_p]$ seja uma variável aleatória com média μ e matriz de covariância Σ . O objetivo é determinar, por meio de combinações lineares das variáveis originais, um novo conjunto de variáveis não correlacionadas, cujas variâncias diminuem a partir da primeira para a última variável. Identifiquemos as novas variáveis ou coordenadas por Y_1, Y_2, \dots, Y_p que são combinações lineares das variáveis originais:

$$Y_j = a_{1j}X_1 + a_{2j}X_2 + \dots + a_{pj}X_p = a_j^T X \quad (11)$$

onde $a_j^T = [a_{1j}, \dots, a_{pj}]$ é um vetor constituído por constantes. As novas variáveis possuem unidade arbitrária. Impomos a restrição que $a_j^T a_j = \sum_{k=1}^p a_{kj}^2 = 1$ de modo que este procedimento particular de normalização assegura que toda transformação é ortogonal, ou, em outras palavras, que a distância no espaço p são preservadas.

Para encontrarmos a primeira componente principal devemos determinar a_1 de tal modo a maximizar a variância $(a_1^T X)$, com a restrição de $a_1^T a_1 = 1$. A segunda componente principal é encontrada determinando a_2 , tal que $Y_2(Var(a_2^T X))$ tenha a maior variância possível e seja a segunda maior para o problema. Além disso, Y_2 deve ser não correlacionada com Y_1 . Seguimos este processo de forma sucessiva para encontrar as componentes principais.

Para encontra Y_1 , temos:

$$Var(Y_1) = Var(a_1^T X) = a_1^T \Sigma a_1 \quad (12)$$

assim nossa tarefa se resume a maximizar $a_1^T \Sigma a_1$.

Para maximizar esta função devemos utilizar o método dos multiplicadores de Lagrange, visto que a função ser maximizada depende de mais de uma variável, e possui uma ou mais restrições. Com somente uma restrição (ou vínculo), este método utiliza o fato de que os pontos estacionários de uma função diferenciável de p variáveis, $f(x_1, x_2, \dots, x_p)$, sujeita a uma restrição $g(x_1, x_2, \dots, x_p)=c$, são tais que existe um número λ , chamado de multiplicador de Lagrange que satisfaça:

$$\frac{\partial f}{\partial x_i} - \lambda \frac{\partial g}{\partial x_i} = 0 \quad i=1,2,\dots,p \quad (13)$$

no ponto estacionário.

Essas p equações, aliadas às restrições, são suficientes para determinar as coordenadas dos pontos estacionários. É também necessário verificar se um ponto estacionário é Máximo, mínimo ou ponto de sela. Para isso usamos a função $L(x)$, tal que:

$$L(x) = f(x) - \lambda[g(x) - c]$$

onde o termo entre colchetes é nulo. Então os conjuntos de equações obtidas em

(13) podem ser escritos simplesmente como: $\frac{\partial L}{\partial x} = 0$

No nosso caso, queremos maximizar a função $\Sigma = DPD$ com a restrição $a_1^T a_1 - 1 = 0$, e a nova função pode ser escrita:

$$L(a_1) = a_1^T \Sigma a_1 - \lambda[a_1^T a_1 - 1] \quad (14)$$

Derivando em relação à a_1 ,

$$\frac{\partial L}{\partial a_1} = 2\Sigma a_1 - 2\lambda a_1$$

Fazendo esta igual a zero,

$$(\Sigma - \lambda I)a_1 = 0 \quad (15)$$

A matriz unitária I foi inserida na equação (15) de modo que o termo entre colchetes é da mesma ordem, isto é $(p \times p)$. Agora temos um passo crucial no argumento. Se a equação (15) possui solução não trivial para a_1 , então $(\Sigma - \lambda I)$ deve ser uma matriz singular. Deste modo λ deve ser escolhido de forma que:

$$|\Sigma - \lambda I| = 0 \quad (16)$$

Assim, uma solução não nula para a equação (15) existe, se e só se, λ é um autovalor de Σ . De modo geral Σ tem p autovalores, que deverão ser todos positivos, já que Σ é semidefinida positiva (ver eq. 3,4 e 5). Se os autovalores são $\lambda_1, \lambda_2, \dots, \lambda_p$, e se são distintos e obedecem $\lambda_1 > \lambda_2 > \dots > \lambda_p$. Então escolheremos um desses valores para determinar a primeira componente principal, efetuando o cálculo da variância:

$$Var(a_1^T X) = a_1^T \Sigma a_1 = a_1^T \lambda a_1 = \lambda$$

Como nós queremos maximizar esta variância, escolhemos λ_1 . Então de posse da equação (15), a componente principal, a_1 , que nós estávamos procurando, deve ser o autovetor correspondente de Σ para o maior autovalor.

A segunda componente principal, isto é $Y_2 = a_2^T X$, é obtida utilizando-se os mesmos procedimentos, adicionando a restrição $a_2^T a_2 - 1 = 0$, desta forma asseguramos que Y_2 será não correlacionada com Y_1 . Assim:

$$Cov(Y_1 Y_2) = Cov(a_1^T X, a_2^T X) = \langle a_2^T (X - \mu)(X - \mu)^T a_1 \rangle = a_2^T \Sigma a_1 \quad (17)$$

Devemos exigir que esta seja nula, mas desde que $\Sigma a_1 = \lambda a_1$, o que é equivalente a impor que a_1 e a_2 sejam ortogonais.

Com o objetivo de maximizar a variância Y_2 , isto é $a_2^T a_1$, estamos sujeitos a duas restrições. Então será necessário o uso de dois multiplicadores de Lagrange, que iremos escrever como λ e δ , considerando:

$$L(a_2) = a_2^T \Sigma a_2 - \lambda [a_2^T a_2 - 1] - \delta a_2^T a_1 \quad (18)$$

e seguindo o mesmo raciocínio para Y_1 encontramos Y_2 . No ponto estacionário, teremos.

$$\frac{\partial L}{\partial a_2} = 2(\Sigma - \lambda I)a_2 - \delta a_1 \quad (19)$$

multiplicando por a_1^T chegamos a: $2a_1^T \Sigma a_2 - \delta = 0$, pois $a_1^T a_2 = 0$. Por (17), vemos que $a_1^T \Sigma a_2 = 0$, tal que δ se anule no ponto estacionário. Assim (19) se torna $(\Sigma - \lambda I)a_2 = 0$.

E desta vez devemos escolher λ_2 , o segundo maior autovalor. Seguindo os mesmos argumentos podemos encontrar a j -ésima componente associada ao j -ésimo maior autovalor.

2.5 – Análise Hierárquica de Clusters – HCA [26-29]

O principal objetivo da HCA é enfatizar o agrupamento natural de amostras similares baseado em sua proximidade no espaço multidimensional formado pelas variáveis utilizadas. É também uma ferramenta exploratória usada para validar o agrupamento previamente verificado pela PCA. Os resultados são apresentados na forma de um dendograma, permitindo a visualização de agrupamentos e a correlação entre as amostras.

No HCA, as distâncias (euclidianas) entre as amostras são calculadas e transformadas em uma matriz de similaridade, cujos elementos são índices de similaridade na faixa de zero até um; uma distância menor significa um maior índice. Distâncias pequenas são indicadores de que as amostras são relativamente similares. Em princípio, nós podemos esperar que pontos representando compostos ativos se agruparão em uma região limitada do espaço no gráfico dos escores das PCs, enquanto os pontos representando compostos inativos se agruparão em outra região.

A análise HCA sempre tem início com o cálculo da distância entre as amostras duas a duas. O par mais próximo é selecionado e transformado em um grupo que passa a ser identificado no gráfico de similaridade como uma amostra única. Isso é feito sucessivamente unindo novas amostras aos pares ou aos grupos já formados. A maneira de se considerar a medida da proximidade entre amostras ou grupos de amostras no método HCA não é única. Apresentamos abaixo de forma bem rápida as mais comuns.

No método de vizinhos mais próximos a distância entre um grupo que acaba de ser formado e as demais amostras (ou outro grupo já formado anteriormente) é

dada pela menor distância entre a amostra e as amostras do grupo (calculadas separadamente). O método de vizinhos mais distantes baseia-se na mesma idéia, mas a distância considerada na análise de similaridade é a maior distância entre a amostra e as amostras do grupo.

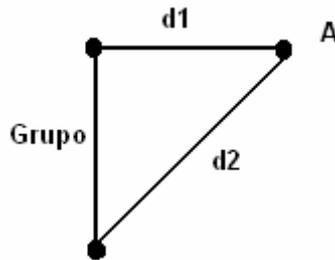


Figura 2.3 – Ilustração do cálculo da distância entre um grupo de duas amostras e uma amostra A.

Na ilustração da Figura 2.3, para a análise de vizinhos mais próximos d_1 corresponde à medida de proximidade entre eles, já na análise de vizinhos mais distantes a proximidade é dada por d_2 .



Figura 2.4 - Cálculo da distância entre grupos para o método dos centróides.

No método de centróides, ilustrado na figura 2.4, cada par de amostras passa a ser identificado por um único ponto localizado no centro do grupo e a medida da proximidade entre grupos neste caso é dada pela distância entre os centros.

As distâncias entre as amostras são calculadas e transformadas em uma matriz **S**, cujos elementos são índices de similaridade. A similaridade é calculada de acordo com a expressão:

$$S_{pq} = 1 - \frac{d_{pq}}{d_{máx}}$$

onde S_{pq} é um elemento da matriz de similaridade **S**, $d_{máx}$ é a maior distância entre quaisquer das amostras da distribuição e d_{pq} é a menor distância entre duas amostras ou grupo de amostras.

Na figura 2.5 é mostrado um dendograma de similaridade no qual as amostras são agrupadas.

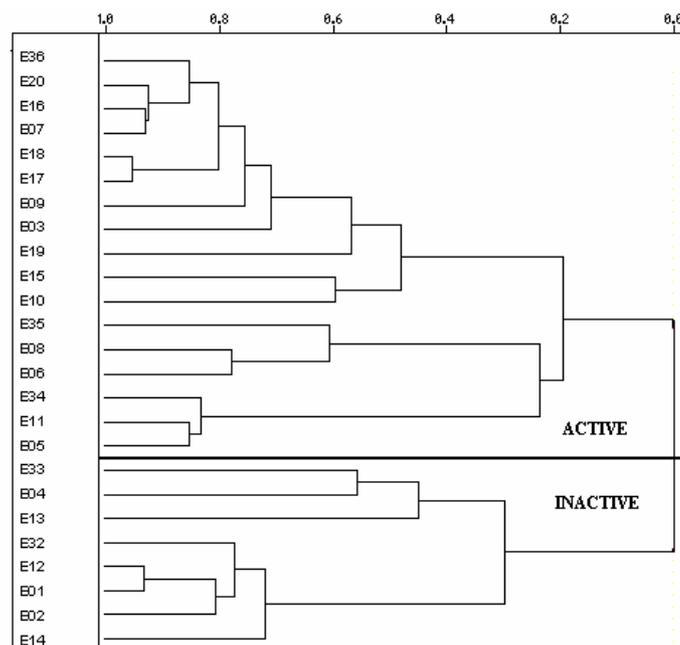


Figura 2.5 – Representação de um dendograma obtido pelo método HCA.

No dendograma as linhas horizontais representam as amostras e as linhas verticais indicam a similaridade entre elas. A similaridade varia de 0 a 1. Quanto mais próxima de zero menor a distância entre as amostras e mais semelhantes são

as mesmas. O resultado final de um dendograma nos permite verificar como as amostras se aglomeram e se há uma relação de semelhança entre os vários subgrupos formados.

2.6 – Referências

- [1] P.W. Atkins and R.S. Friedman, *Molecular Quantum Mechanics*, Oxford University Press Inc., 3^a Edição, 1997.
- [2] I. N. Levine, *Quantum Chemistry*, Prentice Hall, New Jersey, 4^a edição, 1991.
- [3] Andrew Streitwieser, John Wiley & Sons, *Molecular Orbital Theory for Organic Chemists*, New York, 1961.
- [4] R. O. Jones e O. Gunnarsson, *Revs. Mod. Phys.* 61, 689 (1989).
- [5] P. Ballone, M. Marchi, *J. Phys. Chem. A* **103**, 3097-3102 (1999).
- [6] Hoffman, R., "An Extended Huckel Theory", *J. Chem. Phys.* 1963, 39(6), 1397-1412,
- [7] J.A. Pople e D. L. Beveridge, *Approximate Molecular Orbital Theory*, McGraw-Hill, New York, 1970.
- [8] G. A. Segal in *Semiempirical Methods of Electronic Structure Theory*, Plenum, New York, 1977.
- [9] R. C. Bingham, M. J. S. Dewar et al., *J. Am. Chem. Soc.* 97, 1285, 1294, 1302, 1307 (1975), e referências citadas.
- [10] Dewar et. Al., *J. Am. Chem Soc.* 97, 1311 (1975); *Science* 187, 1037 (1975).
- [11] J. J. P. Stewart, *J. Comp. Chem.* 10 (1991) 209; 10 (1991) 221; 11 (1990) 543.
- [12] J. D. M. Vianna, A. Fazzio e S. Canuto, *Teoria Quântica de Moléculas e Sólidos*, Editora Livraria da Física, 1^a Edição, 2004.
- [13] W. D. Edwards e M.C. Zerner, *Theor. Chem. Acta* 72, 347 (1987).
- [14] R. McWeeny in S. Wilson e G. H. F. Dierksen (Eds), *Methods in Computational Molecular Physics*, Plenum Press, N.Y., 1992.
- [15] P. M. V. B. Barone, A. Camilo Jr., D. S. Galvão, *Phys. Rev. Lett.* **77**, 1186 (1996).
- [16] P. M. V. B. Barone, A. Camilo Jr., D. S. Galvão, *J Mol. Struct. (Theochem)* **505**, 55 (2000).
- [17] R. S. Braga, P. M. V. B. Barone, D. S. Galvão, *J. Mol. Struct. (Theochem)* **464**, 257 (1999).
- [18] I. N. Levine, *Quantum Chemistry*, 4th ed., Prentice-Hall, Inc., Englewood Cliffs: New Jersey, 1991.
- [19] S. F. Braga, D. S. Galvão, *J. Chem. Inf. Comput. Sci.* **43**, 699 – 706 (2003).

- [20] R. Vendrame, R. S. Braga, Y. Takahata, D. S. Galvão, *J. Chem. Inf. Comput. Sci.* **39**, 1094 (1999).
- [21] R. Vendrame, R. S. Braga, Y. Takahata, D. S. Galvão, *J. Mol. Struct (Theochem)* **529**, 253 (2001).
- [22] M. Cyrillo, D. S. Galvão, *J. Mol. Struct. (Theochem)* **464**, 267 (1999).
- [23] R. Vendrame, V. R. Coluci, R. S. Braga e D. S. Galvão *J. Mol. Struct (Theochem)* **619**, 195 (2002).
- [24] R. S. Braga, R. Vendrame, D. S. Galvão, *J Chem Inf. Comput. Sci.* **40**, 1377(2000).
- [25] L. C. de Melo, S. F. Braga e P. M. V. B. Barone; *Journal of Molecular Graphics and Modelling.* **25**, 912-920 (2007).
- [26] Chatfield C. and Collins, A. J.; "Introduction to Multivariate Analysis", Cambridge (1980).
- [27] D. L. Massart, B. G. M. Vandeginste, S. N. Deming, Y. Michotte, L. Kaufman, *Chemometrics: a textbook 2.* Elsevier.
- [28] J. M. Deane, *Multivariate Pattern Recognition in Chemometrics, illustrated by case studies*, R. G. Brereton (Editor), Elsevier, Amsterdam, 1992.
- [29] J. C. Miller e J. N. Miller, *Statistics for Analytical Chemistry*, Ellis Horwood Ptr Prentice Hall, New York, 1994 ,3^a edição.

CAPÍTULO 3

FAMÍLIAS DE COMPOSTOS ESTUDADOS

3.1. Elipticinas

A elipticina (5,11-dimetil-6H-pyrido[4,3]carbazole), molécula orgânica planar aromática, foi isolada pela primeira vez em 1959 de folhas da planta *Ochrosia elliptica* Labell (*Apocynaceae*) que é cultivada na Oceania, tendo sua atividade biológica, assim como a de alguns derivados, evidenciada em 1967 durante estudos sobre atividade anticancerígena de algumas plantas. Nesta oportunidade, a elipticina e seu derivado 9-metoxielipticina foram obtidos a partir do extrato de dois outros arbustos, *Ochrosia Moorei* e *Escavatia coccinea*, podendo ainda ser a elipticina isolada do caule da *Ochrosia Acuminata* [1,2,3,4]. O alcalóide elipticina é um composto orgânico que possui uma estrutura planar simples, cujo núcleo é composto da metade de um carbazol ligado a um anel de piridina (Figura 3.1) [5]. Sua atividade no tratamento quimioterápico contra vários tipos de câncer como a leucemia mieloblástica, adenocarcinoma, câncer de mama e alguns tumores sólidos, associada aos efeitos colaterais limitados e à ausência de toxicidade hematológica, reforçam o interesse por suas propriedades biológicas [6,7,8, 9-14].

A estrutura básica dos 40 derivados da elipticina e compostos análogos olivacina e isoelipticina incluídos neste trabalho é apresentada na **Figura 3.1** e na **Tabela 3.1**. Olivacinas, E32 a E37, diferem da elipticina somente pela troca do grupo metil da posição R_{11} para a R_1 . Isoelipticinas, E38 a E40, são derivados com um átomo de nitrogênio em uma posição diferente no anel piridínico – ver **Tabela 3.1**.

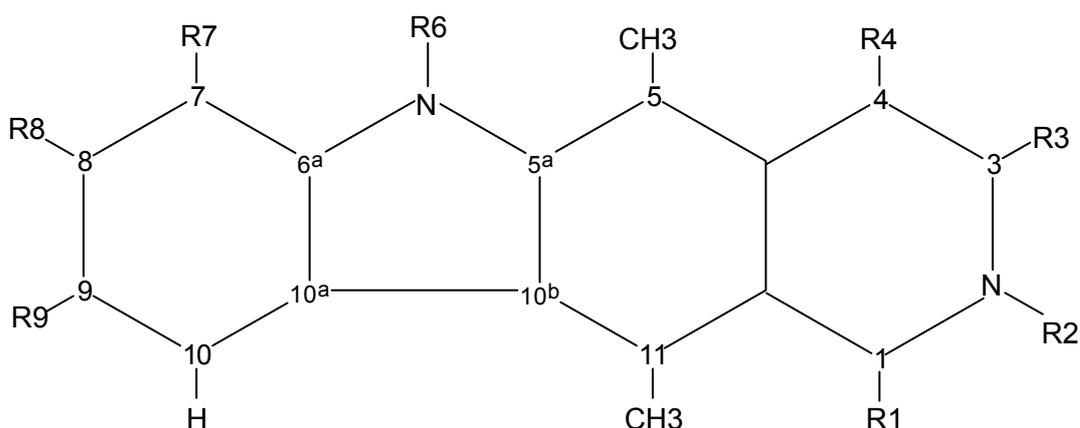


Figura 3.1 – Estrutura básica da Elipticina.

Para nossos estudos dividimos o conjunto de moléculas em dois grupos: **G1** e **G2**. O grupo **G1** reúne os 25 derivados para os quais a atividade biológica é conhecida experimentalmente (moléculas 1-20 e 32-36 na Tabela 3.1). O grupo **G2** reúne os 15 derivados para os quais não há informações experimentais disponíveis e estes serão considerados para resultados preditivos (moléculas 21-31 e 37-40 na Tabela 3.1). Como não há dados experimentais avaliados sob as mesmas condições para todos os compostos estudados, nós utilizamos para o grupo **G1** a mesma aproximação proposta por Villemin et al. [15], classificando os compostos simplesmente em ativos (**A**) e inativos (**I**). Para as moléculas E01 a E10 consideramos a escala experimental de atividade proposta por Cavaliere et. al. [16] para definir os compostos ativos e inativos, e para as demais moléculas o índice experimental IC_{50} [2]. Nós consideramos a molécula como ativa se o valor do IC_{50} é menor do que $1 \mu\text{M}$. O índice IC_{50} é determinado in vitro, e indica a concentração inibidora que reduz em 50% a taxa de crescimento das células após 24 horas de exposição à droga, em relação às células não expostas. Desta forma, quanto menor o valor de IC_{50} , maior a atividade da droga testada.

n°	Molécula	R_1	R_2	R_4	R_6	R_7	R_8	R_9	R_{10}	A.
E01	Elíptica (E)	H	-	H	H	H	H	H	H	A
E02	9E-aminoE	H	-	H	H	H	H	NH ₂	H	I
E03	9-hidroxiE	H	-	H	H	H	H	OH	H	A
E04	7-hidroxiE	H	-	H	H	OH	H	H	H	I
E05	9-metoxiE	H	-	H	H	H	H	OCH ₃	H	A
E06*	Elípticinium (EM)	H	CH ₃	H	H	H	H	H	H	A
E07*	2N-metil, 9-hidroxiE	H	CH ₃	H	H	H	H	OH	H	A
E08*	2N-metil, 1-metil, E	CH ₃	CH ₃	H	H	H	H	H	H	A
E09*	2N-metil, 6N-metil, E	H	CH ₃	H	CH ₃	H	H	H	H	A
E10*	2N-metil, 6N-metil, 9-hidroxiE	H	CH ₃	H	CH ₃	H	H	OH	H	A
E11	9-hidroxi-6-metilE	H	-	H	CH ₃	H	H	OH	H	A
E12	6-metilE	H	-	H	CH ₃	H	H	H	H	A
E13	9-metoxi-1-aminoE	NH ₂	-	H	H	H	H	OCH ₃	H	A
E14	9-bromoE	H	-	H	H	H	H	Br	H	I
E15*	7-hidroxi-2-metilEM	H	CH ₃	H	H	OH	H	H	H	I
E16*	9-metoxi-2-metilEM	H	CH ₃	H	H	H	H	OCH ₃	H	A
E17*	9-hidroxi 7-metilEM	H	CH ₃	H	H	CH ₃	H	OH	H	A
E18*	9-hidroxi 8,10-dimetilEM	H	CH ₃	H	H	H	CH ₃	OH	CH ₃	I
E19*	9-hidroxi-2-(2etilamino)etilEM	H	DEAEt	H	H	H	H	OH	H	A
E20*	2-etil-9-hidroxiE	H	CH ₂ CH ₃	H	H	H	H	OH	H	A
E21	9-metoxi-6-metilE	H	-	H	CH ₃	H	H	OCH ₃	H	-
E22	11-demetil-1-CHO-9-metoxiE	CH O	-	H	H	H	H	OCH ₃	H	-
E23	11-demetil-1-CH ₂ OH-9-metoxiE	CH ₂ OH	-	H	H	H	H	OCH ₃	H	-
E24	6-etilE	H	-	H	CH ₂ CH ₃	H	H	H	H	-
E25	9-metilE	H	-	H	H	H	H	CH ₃	H	-
E26	7-metilE	H	-	H	H	CH ₃	H	H	H	-
E27	11-demetil-9-hidroxiE	H	-	H	H	H	H	OH	H	-
E28*	4-HO-9-metoxi-2-metilEM	H	CH ₃	HO	H	H	H	OCH ₃	H	-
E29*	4-CO ₂ CH ₃ -9-metoxi-2-metilEM	H	CH ₃	CO ₂ CH ₃	H	H	H	OCH ₃	H	-
E30*	9-metil-2-metilEM	H	CH ₃	H	H	H	H	CH ₃	H	-
E31*	2-CH ₂ OCH ₃ E	H	CH ₂ OCH ₃	H	H	H	H	H	H	-
E32	Olivacina (O)	CH ₃	-	H	H	H	H	H	H	I
E33	7-hidroxiO	CH ₃	-	H	H	OH	H	H	H	I
E34	9-hidroxiO	CH ₃	-	H	H	H	H	OH	H	A
E35*	9-hidroxi-2-metilOM	CH ₃	CH ₃	H	H	H	H	OH	H	A
E36*	9-hidroxi-2,8,10-trimetilOM	CH ₃	CH ₃	H	H	H	CH ₃	OH	CH ₃	A
E37	9-metoxiO	CH ₃	H	H	H	H	H	OCH ₃	H	-
E38	3N-isoE **	H	H	H	H	H	H	H	H	-
E39	1N-11demetil-2,4dimetilE **	H	CH ₃	CH ₃	H	H	H	H	H	-
E40	4N-E **	H	H	-	H	H	H	H	H	-

Tabela 3.1 – Elíptica e derivados estudados. Os radicais numerados referem-se a estrutura mostrada na Figura 3.1. *moléculas de camada eletrônica aberta; ** o número que precede o símbolo N indica a posição do átomo de nitrogênio no anel piridínico. A e I refere-se a ativo e inativo, E e EM refere-se a elíptica e elípticinium, e O e OM a olivacina e olivacinium. Para as moléculas de camada fechada foi utilizado o RHF-Restricted Hartree-Fock (onde consideramos um par de spin orbitais para cada orbital espacial). Para as moléculas de camada aberta foi utilizado o UHF-Unrestricted Hartree-Fock (os spin orbitais podem diferir tanto na função espacial quanto na função de spin).

3.2. Furanocumarinas

As furanocumarinas são compostos poli-heterocíclicos derivados da cumarina pela adição de um anel de furano nas posições 6,7 (psoraleno - **Fig. 3.2**) ou 7,8 (pseudopsoraleno ou angelicina), possuindo classes diversificadas de derivados naturais e sintéticos que exibem atividade fotobiológica e fototerapêutica. Em sistemas vivos, seu alvo principal é o DNA, ao qual se liga covalentemente após irradiação por radiação UVA (320 – 400 nm). Derivados das furanocumarinas têm sido investigados e utilizados no tratamento de várias doenças caracterizadas por condições hiperproliferativas e, mais recentemente, em fotoquimioterapia extracorporal [17,18, 19-20].

Os psoralenos são furanocumarinas que exibem atividade fotobiológica e fototerapêutica. O composto psoraleno foi primeiro isolado de sementes de *Psoralea corylifolia*, e posteriormente outros psoralenos foram isolados de extratos da planta *Ammi majus*, encontrada no Vale do Nilo, no Egito e do Severn, na Inglaterra [21].

As angelicinas são furanocumarinas que também possuem propriedades fotobiológicas interessantes. Alguns efeitos já considerados são: efeitos antiproliferativos, indução de hemólise em eritrócitos, inativação de microorganismos procariotes e eucariotes e algumas viroses, além de fototoxicidade de pele e mutagênese. A angelicina foi primeiro isolada de raízes da *Angelica archangelica*, e depois observada em outras plantas da família *Umbelliferae*. Alguns derivados metoxilados também foram isolados de fontes naturais, os vegetais *Pimpinella saxifraga* e *Heracleum sphondylium* [22]. Vários derivados metilados da angelicina têm sido investigados e tem sido mostrado experimental [23,24] e teoricamente [25] que a presença dos grupos metil na angelicina influencia a orientação da molécula

dentro da cavidade de intercalação do DNA, aumentando a atividade antiproliferativa, dependendo da posição destes substituintes [26-28].

Neste trabalho estudamos derivados do psoraleno, e todas as moléculas estudadas têm camada fechada (número par de elétrons), possuindo estruturas semelhantes. Algumas moléculas que apresentam estrutura um pouco diferente são também mostradas na figura 3.2 (derivados do psoraleno que diferem pela posição do anel de furano ligado à cumarina e pela adição de um grupo benzeno). Os grupos laterais que compõem cada molécula que possuem estruturas semelhantes, mas diferem no número, tipo e posição de ligação do radical, são explicitados na tabela 3.2.

O conjunto investigado é formado por 52 compostos, 31 para os quais a atividade fotodinâmica [29], determinada através de testes com irradiação de luz com comprimento de onda na faixa do ultravioleta (365,5 nm) sobre pele humana, foi relatada (moléculas P, P13, P14 e P25 a P52, denominado grupo **G1**) e 21 compostos para os quais não foram relatados testes experimentais (moléculas P1 a P12 e P15 a P24, denominado grupo **G2**). Deixamos de incluir as angelicinas no estudo, devido às poucas informações experimentais obtidas que permitissem um estudo mais consistente.

Estes grupos foram estudados separadamente, sendo o trabalho dividido em duas etapas: a primeira dedicada a estabelecer correlações entre a atividade biológica relatada e seus descritores teóricos para o **G1**, e a segunda dedicada a propor a atividade qualitativa para os compostos do **G2**, aplicando as regras e padrões observados na primeira parte do estudo.

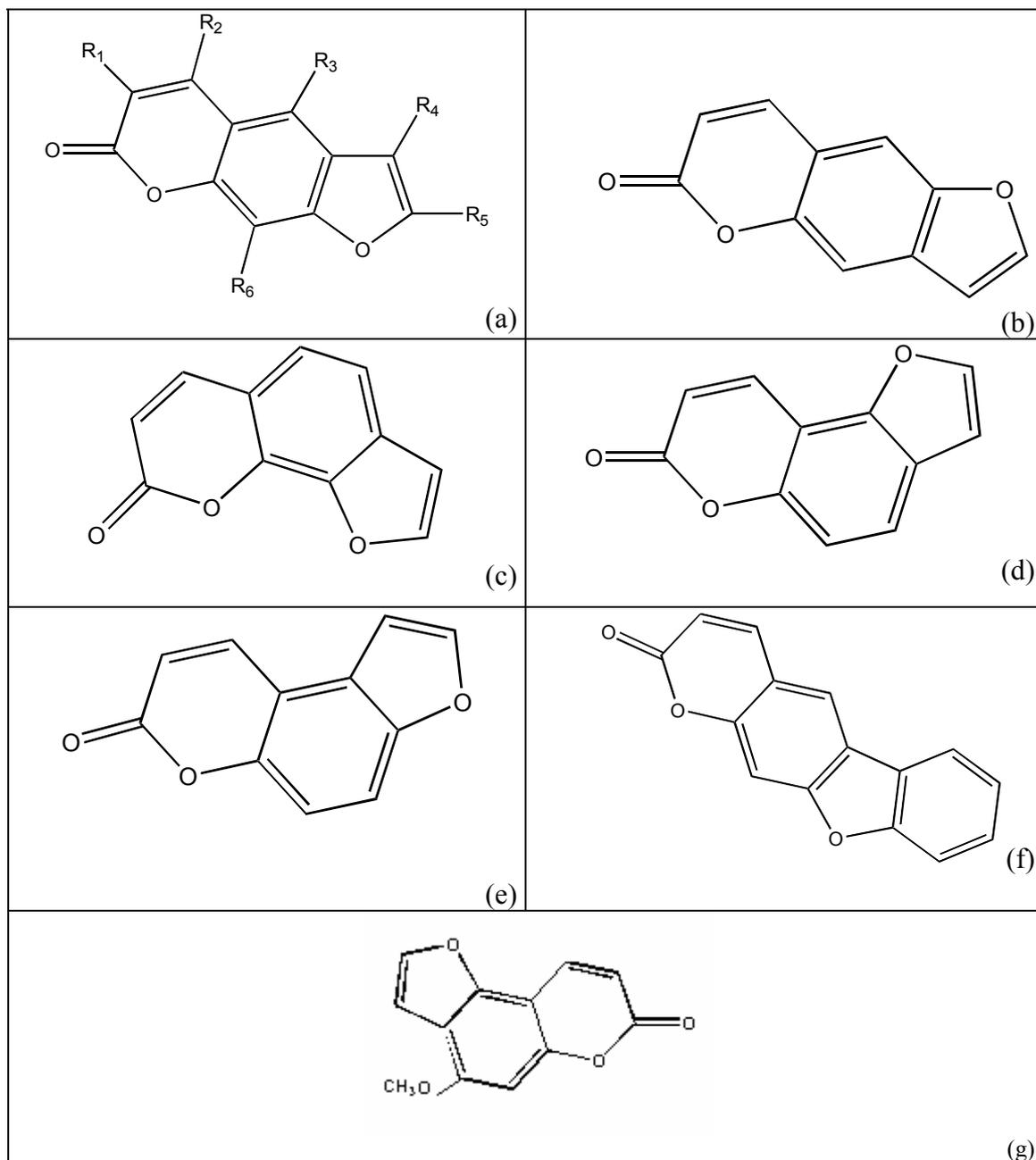


Figura 3.2: Estrutura Molecular: a) psoraleno ($R_i = H$); b) psoraleno1(P_1); c) psoraleno 2 (P_2) Pseudoisopsoraleno; d) psoraleno 3 (P_3) Allopsoraleno; e) psoraleno 4 (P_4) Isopseudoisopsoraleno; f) benzopsoraleno (P_{24}); g) allobergapten.

Nº.	Molécula	R ₁	R ₂	R ₃	R ₄	R ₅	R ₆	A _f
P	Psoraleno	H	H	H	H	H	H	A
P ₅	4,5',8 Trimetilpsoraleno	H	CH ₃	H	H	CH ₃	CH ₃	-
P ₆	4 Metilpsoraleno	H	CH ₃	H	H	H	H	-
P ₇	4,4' Dimetilpsoraleno	H	CH ₃	H	CH ₃	H	H	-
P ₈	4,5' Dimetilpsoraleno	H	CH ₃	H	H	CH ₃	H	-
P ₉	4,8 Dimetilpsoraleno	H	CH ₃	H	H	H	CH ₃	-
P ₁₀	3 Metilpsoraleno	CH ₃	H	H	H	H	H	-
P ₁₁	3,4',8 Trimetilpsoraleno	CH ₃	H	H	CH ₃	H	CH ₃	-
P ₁₂	3,4' Dimetilpsoraleno	CH ₃	H	H	CH ₃	H	H	-
P ₁₃	5 Metoxy psoraleno (5 Mop ou Bergapten)	H	H	OCH ₃	H	H	H	A
P ₁₄	8 Metoxy psoraleno (8 Mop ou Xanthotoxin)	H	H	H	H	H	OCH ₃	A
P ₁₅	8 Mop 1	H	H	O ₂ C ₄ H ₉	H	H	OCH ₃	-
P ₁₆	8 Mop 2	H	H	OC ₄ H ₉	H	H	OCH ₃	-
P ₁₇	8 Mop 3	H	H	OCH	H	H	OCH ₃	-
P ₁₈	5,8 Dimetoxypsoraleno	H	H	OCH ₃	H	H	OCH ₃	-
P ₁₉	3,4' Dimetil 8 Mop	CH ₃	H	H	CH ₃	H	OCH ₃	-
P ₂₀	4,4' Dimetil 8 Mop	H	CH ₃	H	CH ₃	H	OCH ₃	-
P ₂₁	4,4',5' Trimetil 8 Mop	H	CH ₃	H	CH ₃	CH ₃	OCH ₃	-
P ₂₂	3,4',5' Trimetil 8 Mop	CH ₃	H	H	CH ₃	CH ₃	OCH ₃	-
P ₂₃	3,4,4' Trimetil 8 Mop	CH ₃	CH ₃	H	CH ₃	H	OCH ₃	-
P ₂₅	8-hidroxy psoraleno (xanthotoxin)	H	H	H	H	H	OH	I
P ₂₆	5-hidroxy psoraleno (bergaptol)	H	H	OH	H	H	H	I
P ₂₇	4',5' dihidropsoraleno	H	H	H	H ₂	H ₂	H	I
P ₂₈	3,4 dihidropsoraleno	H ₂	H ₂	H	H	H	H	I
P ₂₉	4' metilpsoralenoo	H	H	H	CH ₃	H	H	A
P ₃₀	4' phenyl – 4 metilpsoraleno	H	CH ₃	H	phenyl 	H	H	I
P ₃₁	3,4 dihidro xanthotoxin	H ₂	H ₂	H	H	H	OCH ₃	I
P ₃₂	3 metil 8MOP	CH ₃	H	H	H	H	OCH ₃	A
P ₃₃	4 metil 8MOP	H	CH ₃	H	H	H	OCH ₃	A
P ₃₄	4' metil 8MOP	H	H	H	CH ₃	H	OCH ₃	A
P ₃₅	5', 4 dimetil 8MOP	H	CH ₃	H	H	CH ₃	OCH ₃	A
P ₃₆	5' phenyl – 4 metil 8MOP	H	CH ₃	H	H	phenyl	OCH ₃	I
P ₃₇	5 cloro 8MOP	H	H	Cl	H	H	OCH ₃	A
P ₃₈	5 nitro 8MOP	H	H	NO ₂	H	H	OCH ₃	I
P ₃₉	5 amino 8MOP	H	H	NH ₂	H	H	OCH ₃	I
P ₄₀	5 acetilamino 8MOP	H	H	acetila mino	H	H	OCH ₃	I
P ₄₁	4', 5' dihidro bergapten	H	H	OCH ₃	H ₂	H ₂	H	I
P ₄₂	4 metil 5MOP	H	CH ₃	OCH ₃	H	H	H	A
P ₄₃	4 metil – 4', 5' dihidro 5MOP	H	CH ₃	OCH ₃	H ₂	H ₂	H	I
P ₄₄	Allobergapten	Fig.						A
P ₄₅	4', 5' dihidro allobergapten	H	H	H	H ₂	H ₂	H	I
P ₄₆	4', 5', 3, 4 tetrahidro allo 5MOP	H ₂	H ₂	OCH ₃	H ₂	H ₂	H	I
P ₄₇	4 metil allobergapten	H	CH ₃	H	H	H	H	A
P ₄₈	4 metil 4', 5' dihidro allo 5MOP	H	CH ₃	OCH ₃	H ₂	H ₂	H	I
P ₄₉	8 nitro 5MOP	H	H	OCH ₃	H	H	NO ₂	I
P ₅₀	8 amino 5MOP	H	H	OCH ₃	H	H	NH ₂	I
P ₅₁	8 acetilamino 5MOP	H	H	OCH ₃	H	H	acetila mino	I
P ₅₂	5,8 dihidroxy psoraleno	H	H	OH	H	H	OH	I

Tabela 3.2 – Psoraleno e derivados estudados. Os radicais numerados referem-se a estrutura

mostrada na Figura 3.2a. A e I refere-se a ativo e inativo.

3.3. Referências

- [1] Goodwin, S., Smith, A F. and Horning, E. C.. *J. Am. Chem. Soc.* **1959**, *81*, 1903-1908; Woodward, R. B., Iacobucci, G. A. and Hochstein, F. A.. *J. Am. Chem. Soc.* **1959**, *81*, 4434-4435.
- [2] Gribble, G. W., in *The Alkaloids*, (Academic Press, 1990) Vol. 39, p.239.
- [3] E. M. Acton, V. L. Narayanan, P.A. Risbood, R. H. Shoemaker, D. T. Vistica and M. R. Boyd, *J. Med. Chem* **37**, 2185-2189 (1994).
- [4] M. Sainsburg in *The Chemistry of Antitumor Agents*, D. E. V. Wilman (Ed.), Chapman and Hall, N. Y., 1990, p.410.
- [5] C. Auclair, *Archives of Biochemmistry Biophysics* **259**, 1-14 (1987).
- [6] Auclair, C.. *Arch. Biochem. Biophys.* **1987**, *259*, 1.
- [7] Anderson, W. K., Gopalsamy, A. and Reddy, P. S.. *J. Med. Chem.* **1994**, *37*, 1955-1963; Jurayj, J., Haugwitz, R. D., Varma, R. K., Paull, K. D., Barret, J. F and Cushman, M.. *J. Med. Chem.* **1994**, *37*, 2190-2197.
- [8] Paoletti, C., Cros, S., Dat-Xuong, N., Lecointe, P., and Moisand, A.. *Chem. Biol. Interact.* **1979**, *25*, 45-58.
- [9] Chabane, H., Lamazzi, C., Thiéry, V., Guillaumetb, G. and Bessona, T.. *Tetrahedron Letters* **2002**, *43*, 2483–2486.
- [10] Tran-Thi, H. A., Nguyen-Thi, T., Michel, S., Tillequin, F., KOCH, M., Pfeiffer, B., Pierré, A., Trinh-Van-Dufat, H.. *Chem. Pharm. Bull.* **2004**, *52*, 540-545.
- [11] Ergun, Y., Patir, S., Okay, G.. *Synthetic Communications.* **2004**, *34*, 435-442.
- [12] Ishikura, M., Hino, A., Yaginuma, T., Ágata, I. and Katagiri, N.. *Tetrahedron* **2000**, *56*, 193–207.
- [13] Diaz, M., Cobas, A., Guitian, E., Castelo, L.. *European Journal of Organic Chemistry* **2001**, *23*, 4543-4549.
- [14] Ishikura, M., Hino, A., Katagiri, N.. *Heterocycles* **2000**, *53*, 11-14.
- [15] Villemin, D.; Cherqaoui, D.; Mesbah, A.. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1288-1293.
- [16] Cavaliere, E. L.; Rogan, E. G.; Roth, R. W.; Saugier, R. K.; Hakam, A.. *Chem. Biol. Interact.* **1983**, *47*, 87-109.
- [17] Richard L. Edelson, *Light Activated Drugs*, Scientific American, 1988.

- [18]** J.Ladik and W.Forner, *The Beginnings of cancer in the cell* (Springer-Verlag, Berlin, Heidelberg, 1994).
- [19]** Boehncke WH, Konig K., Ruck A., Kaufmann R., Sterry W.. *Acta Derm Venereol*, 74 (3):201-5, 1994.
- [20]** Rook A.H., Suchin K.R., KAO D.M.F., et al.. *J. Invest. Derm. Symp P*, 4: (1) 85-90, 1999.
- [21]** Gillenvater, A., et al.. *Arch. Otolaryngol Head Neck Surg.*, 124: 1251-1258, 1998.
- [22]** Bordin, F., Dall 'Acqua, F. and Guiotto, A.. *Pharmac. Ther.*, Vol. 52, pp 331-363, 1991.
- [23]** Caffieri S., V. Lucchini, P. Rodighiero, G. Miolo and F. Dall 'Acqua. *Photochem. Photobiol.*, 48, 573-577, 1988.
- [24]** Miolo G., V. Lucchini, P. Rodighiero, F. Dall 'Acqua and S. Caffieri. *Photochem Photobiol*, 59, 277-283, 1994
- [25]** Demaret J., J.P. Ballini and P. Vigny. *J. Comput. Aided Mol. Des.* 7, 683-698, 1993.
- [26]** Ebermann, R., Alth G., Kreitner M. and Kubin A.. *Journal of Photochemistry and Photobiology B-Biology*, 36: (2) 95-97, 1996.
- [27]** Gia O., Uriarte E., Zagotto G., Baccichetti F., Antonello C., Marcianimagno S.. *Journal of Photochemistry and Photobiology B-Biology*, 14: (1-2) 95-104, 1992.
- [28]** Potapenko A.Y.. *Journal Photoch. Photobiol B*, 9: (1) 1-33, 1991.
- [29]** L. Musajo and G. Rodighiero, *Experientia*, Vol. XVIII – Fasc. 4, 153-200, 15.IV.1962.

CAPÍTULO 4

ELIPTICINAS

4.1 – Introdução

A correlação da resposta biológica de compostos químicos com parâmetros eletrônicos moleculares foi sugerida há mais de 50 anos [1], com o argumento de que a relação entre a atividade biológica e as propriedades eletrônicas das drogas poderia, associado ao processo experimental, auxiliar o processo de síntese de novos compostos promissores para o tratamento quimioterápico [2-7]. O grande número de derivados da elipticina sintetizados, estudados e mapeados do ponto de vista das relações estrutura-atividade, representam amostras úteis para estes propósitos [8-15].

Em trabalhos anteriores foi investigado um pequeno grupo de derivados da elipticina, através dos métodos semi-empíricos AM1 (Austin Method 1) [16], INDO/CI (Intermediate Neglect of Differential Overlap/Configuration Interaction), PM3 (Parametric Method 3) [17] e ZINDO/CI (Zerner's Intermediate Neglect of Differential Overlap with Configuration Interaction) [18]. Diversas propriedades eletrônicas calculadas, como estabilidade relativa, energias dos orbitais ou as transições eletrônicas do espectro de absorção não mostraram estar ligadas à atividade biológica das drogas. Notou-se, no entanto, uma correlação entre a atividade biológica e o momento de dipolo destas moléculas [19-21]. A direção do vetor momento de dipolo foi também analisada, mas não mostrou nenhuma correlação com a atividade das drogas. Por outro lado, existem diferenças na distribuição de carga da molécula devido à presença de grupos laterais, originando diferenças no momento de dipolo e no padrão do potencial eletrostático molecular (distribuição dos sítios ativos sobre o esqueleto molecular), o que fornece um critério para classificação das moléculas de acordo com sua atividade biológica.

Nesse trabalho, investigamos sistematicamente um conjunto maior de derivados da elipticina, com o objetivo de investigar propriedades eletrônicas às quais a atividade biológica poderia estar relacionada. Embora os processos envolvidos na interação das drogas com o meio biológico e em seus mecanismos de ação possam ser muito complexos, as propriedades moleculares são fatores importantes para a definição de sua atividade biológica.

Conforme descrito em detalhes no capítulo 3, o conjunto de compostos pertencentes à família das elipticinas foi dividido em dois grupos: **G1** e **G2**, o primeiro reunindo os 25 derivados para os quais a atividade biológica é conhecida experimentalmente, e o segundo, os 15 derivados para os quais não há informações experimentais disponíveis que serão considerados para resultados preditivos.

Apresentamos nas próximas seções, separadamente, a análise conformacional, análise de propriedades eletrônicas e óticas e a investigação teórica de relações estrutura- atividade.

Finalizando o capítulo comparamos os resultados de diferentes metodologias utilizadas para a investigação da correlação entre a atividade biológica e grandezas eletrônicas.

4.2 Resultados

4.2.1 – Análise Conformacional

Iniciamos nossa investigação com a realização de uma busca conformacional para obtenção da estrutura geométrica de mínimo global de cada um dos 40 compostos estudados (**Tabela 3.1**), assumindo que a conformação de menor energia em vácuo é a fonte primária de informações teóricas para uma investigação de correlação com a atividade biológica, cientes de que nem sempre a conformação mínima é necessariamente a estrutura ativa.

A utilização de métodos semi-empíricos foi a alternativa adotada, por produzirem bons resultados na investigação geométrica de muitos compostos orgânicos. Confirmando essa hipótese, comparamos a geometria otimizada através de métodos semi-empíricos com dados estruturais obtidos com raio-X, encontrados na literatura para alguns compostos [22]. Na comparação entre geometrias experimentais de raios-X e geometrias teoricamente obtidas em vácuo, é esperada uma discordância devido ao efeito do empacotamento do sistema cristalino. Neste caso as maiores distorções deverão ocorrer para os valores dos ângulos diedrais, seguidos dos valores dos ângulos de ligação. Na **Tabela 4.1** ilustramos a comparação da geometria molecular para o derivado E12 (6-metileipticina) calculado usando o método semi-empírico PM3 e os dados experimentais de raios-X. Os erros máximos encontrados para os parâmetros estruturais foram de 4,2% para os comprimentos e de 5,5% para os ângulos de ligação. O desvio quadrático médio é de 0,024Å para os comprimentos de ligação e 0,58° para os ângulos de ligação. Observamos que o maior acordo entre a análise teórica e dados de raios-X

ocorre para os comprimentos de ligação. Essa análise corrobora resultados anteriores, em que o método PM3 também apresentou melhor eficiência que os demais métodos utilizados em cálculos com E1 e E6 [21]. Com base nestes resultados, que mostram o bom desempenho do método PM3 para a descrição das propriedades geométricas da molécula E12, selecionamos esse método para investigar as demais moléculas.

Para um estudo detalhado da geometria varremos o espaço de configurações formado pelas conformações geométricas, construídas através da variação dos ângulos diedrais dos grupos laterais flexíveis. Para uma busca sistemática, variamos simultaneamente estes ângulos diedrais dentro do intervalo de 360 graus, em passos de 20 graus (em um total de 324 conformações). As geometrias obtidas foram então otimizadas mantendo-se fixos os diedrais variados. Os resultados foram obtidos considerando um sistema gasoso. A localização do mínimo global dentro do espaço de conformações está indicada por uma pequena esfera.

Estes cálculos foram realizados utilizando o programa Mopac2000 integrado ao pacote computacional CAChe [27,29], o qual inclui o método PM3 aplicado na busca conformacional. Este programa facilita o estudo sistemático de geometria permitindo uma variação simultânea de até dois diedrais flexíveis, o que nos permite mapear a distribuição de conformações estáveis que os compostos possuem, assim como localizar seu mínimo global. Na **Figura 4.1** ilustramos a superfície do calor de formação (Kcal/mol), obtida para as estruturas otimizadas como função da variação simultânea dos ângulos diedrais e na **Figura 4.2** para a variação de apenas um ângulo diedral.

Comprimento ligação (Å)			Ângulo de ligação (°)		
Átomos	Raio-X-ray	PM3	Átomos	Raio-X	PM3
8 - 7	1.3690	1.3930	8 - 7 - 6a	116.2	117.8
7 - 6a	1.4030	1.3900	7 - 6a - 10a	122.8	121.3
6a - 6	1.3900	1.4490	7 - 6a - 6	127.5	128.6
6 - 5a	1.4060	1.4520	10a - 6a - 6	109.7	110.1
5a - 5	1.3840	1.3800	6a - 6 - 5a	109.8	105.8
5 - R5	1.5020	1.4860	6 - 5a - 10b	107.1	109.0
5 - 4a	1.4050	1.4290	10b - 5a - 5	122.9	121.7
4a - 4	1.4250	1.4270	5a - 5 - 4a	115.0	117.3
4 - 3	1.3480	1.3680	R5 - 5 - 4a	121.0	120.4
3 - 2	1.3430	1.3820	5 - 4a - 11a	122.8	120.8
2 - 1	1.3210	1.3230	11a - 4a - 4	115.3	117.5
1 - 1a	1.4220	1.4300	4a - 4 - 3	119.9	120.6
1a - 11	1.4160	1.4260	4 - 3 - 2	125.5	121.6
11 - R11	1.5270	1.4890	3 - 2 - 1	116.6	119.4
11 - 10b	1.3580	1.3750	2 - 1 - 11a	123.4	122.8
10b - 10a	1.4700	1.4580	1 - 11a - 4a	119.2	118.2
10a - 10	1.3840	1.3890	4a - 11a - 11	120.8	121.1
10 - 9	1.3850	1.3920	11a - 11 - R11	122.7	120.3
9 - 8	1.3610	1.3940	11a - 11 - 10b	117.4	117.6
10a - 6a	1.3940	1.4160	R11 - 11 - 10b	119.9	122.0
10b - 5a	1.4500	1.4420	11 - 10b - 5a	121.1	121.4
11a - 4a	1.4150	1.4140	5a - 10b - 10a	106.3	107.2
			10b - 10a - 6a	107.1	107.7
			10b - 10a - 10	134.7	132.2
			6a - 10a - 10	118.2	120.0
			10a - 10 - 9	119.3	118.6
			10 - 9 - 8	121.0	121.0
			9 - 8 - 7	122.5	121.2

Tabela 4.1 – Comparação entre os valores experimentais e teóricos calculados com método semi-empírico PM3 para a molécula E12 (6-metileipticina).

Analisamos as conformações possíveis para todas as cadeias laterais variando seus diedrais flexíveis e analisamos então a estabilidade relativa do conjunto de geometrias obtido. A partir da localização da conformação de menor energia, iniciamos a obtenção dos descritores quânticos necessários para a exploração teórica de estrutura-atividade dos compostos estudados.

Tendo sido obtidas as geometrias de mínimo global das 40 moléculas em vácuo, passamos à obtenção das propriedades eletrônicas associadas a cada elipticina e à busca da correlação com a atividade biológica dos compostos. Adotamos a hipótese de que a conformação de menor energia obtida para um composto em vácuo é a primeira e melhor candidata para análise em estudos teóricos de correlação com a atividade biológica.

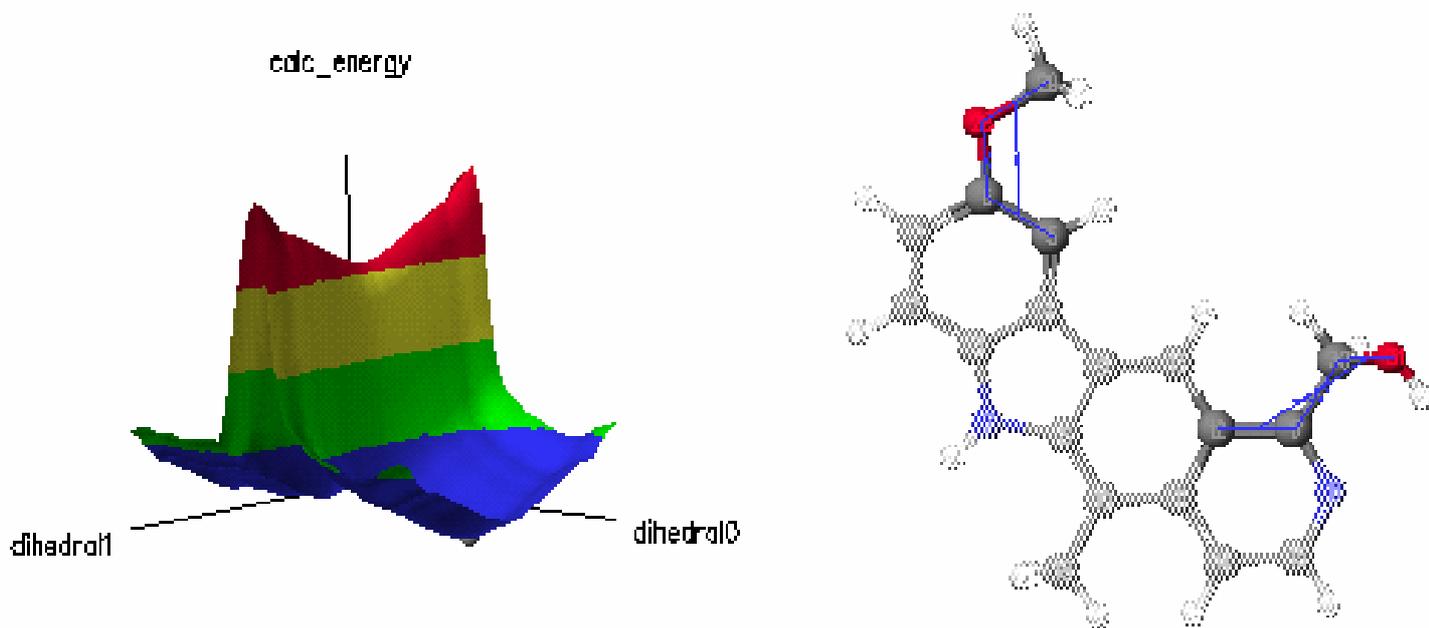


Figura 4.1 – Busca sistemática da conformação mais estável da molécula elipticina E23 no v acuo. (à direita) Superfície do Calor de Formação (Kcal/mol) como uma função da variação dos ângulos diedrais (em destaque na figura) em passos de 20 graus. (à esquerda).

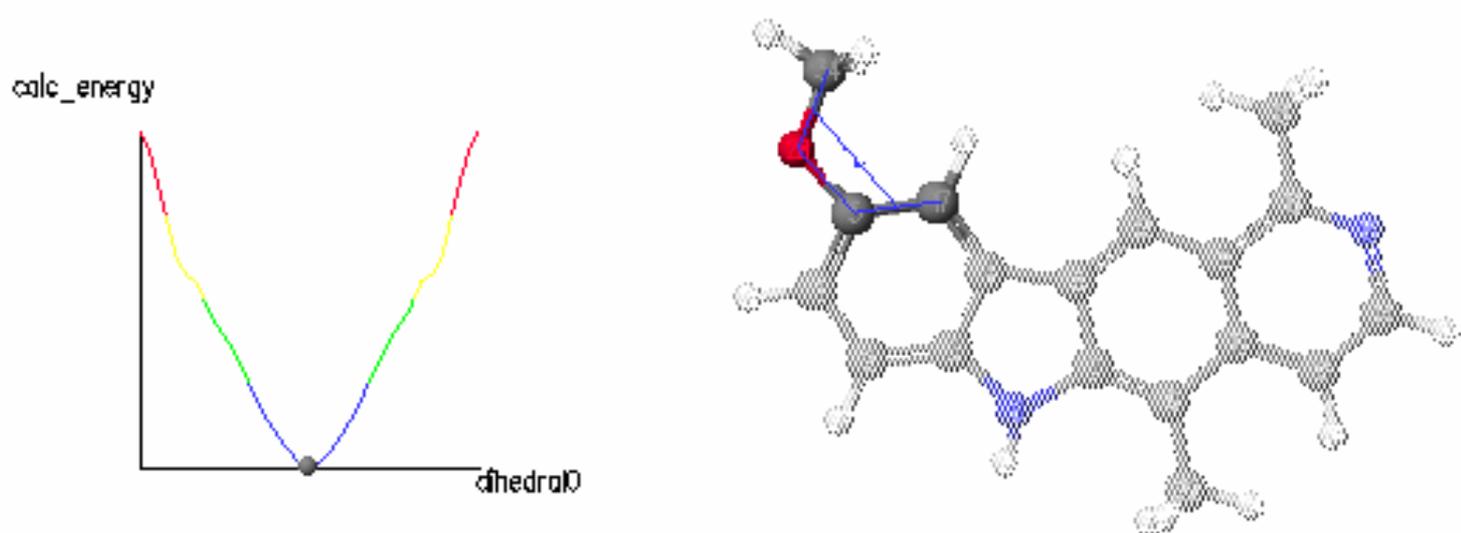


Figura 4.2 – Busca sistemática da conformação mais estável da molécula elipticina E37 no v acuo. (à direita) Superfície do Calor de Formação (Kcal/mol) como uma função da variação do ângulo diedral (em destaque na figura) em passos de 20 graus. (à esquerda).

4.2.2 – Análise de Propriedades Eletrônicas e Óticas

Concluída a investigação estrutural dos compostos, estudamos as mudanças ocorridas na distribuição dos orbitais de fronteira, através da densidade local de estados (LDOS), para as conformações mínimas. A análise da LDOS é uma forma prática e rápida de se determinar a distribuição espacial dos orbitais moleculares. Para a obtenção da LDOS a molécula foi dividida em regiões. O critério de seleção destas regiões considera que elas devem ser comuns a todos os compostos da classe investigada (esqueleto básico), e devem ser reconhecidamente importantes do ponto de vista químico e/ou biológico.

A LDOS foi analisada em detalhes para os orbitais de fronteira HOMO (Highest Occupied Molecular Orbital), HOMO-1, LUMO (Lowest Unoccupied Molecular Orbital) e LUMO+1 sobre sete regiões moleculares específicas: o anel benzênico possuindo a maior ordem de ligação (região **rA**); os sítios atômicos N2, N6 e C₉ (regiões **rB**, **rC** e **rD**, respectivamente); o anel pentagonal (região **rF**), o anel contendo o carbono C9 (região **rN**), e os sítios atômicos 1-4 do anel pirídínico (região **rP**) (veja **Figura 4.3**). Cada região foi escolhida a partir de uma motivação em particular. A região **rA** é susceptível ao ataque eletrofílico, em decorrência da densidade eletrônica associada à ligação com maior ordem de ligação. As regiões **rB** e **rC** têm sua densidade de carga associada com o índice de afinidade ao DNA (K_{app}) correlacionada a atividade citotóxica e antitumoral [25]. **rD** representa o sítio onde as substituições são mais relevantes para a atividade das elipticinas. Experimentalmente verifica-se que a presença de um grupo OH na posição R9 aumenta consideravelmente sua atividade em relação a um derivado ou análogo sem a presença do grupo hidroxila [6]. O anel pentagonal – região **rF** - contém o

heteroátomo nitrogênio (N) e é preservado em todas as estruturas. A região **rN** representa um dos anéis mais expostos do esqueleto molecular (o outro é o que envolve a região **rA**) onde reações bioquímicas são mais susceptíveis de ocorrer, e **rP** corresponde à região que na análise do potencial eletrostático molecular aparece como diferenciadora para as elipticinas inativas biologicamente [26].

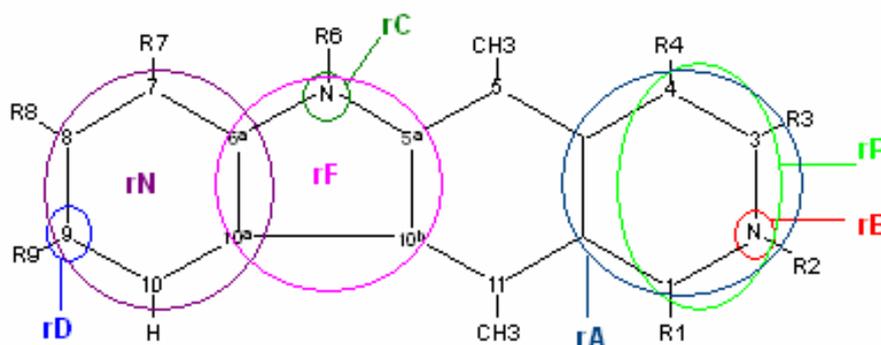


Figura 4.3 – Regiões sobre o esqueleto molecular das elipticinas selecionadas para os cálculos de LDOS.

Todos os parâmetros eletrônicos foram calculados com o método semi-empírico PM3, através do pacote computacional MOPAC6, integrado ao Chem2Pac [46]. O pacote Spartan [24] foi utilizado para a visualização da orientação do vetor momento de dipolo elétrico das moléculas otimizadas e da distribuição espacial dos orbitais de fronteira das moléculas (ilustração abaixo – **Figura 4.4**). O método UHF foi usado para os cálculos envolvendo as moléculas de camada aberta e o método RHF para as moléculas de camada fechada.

A orientação do momento de dipolo, assim como a distribuição dos orbitais LUMO e HOMO, segue um padrão idêntico para todos os compostos com dados

experimentais (grupo **G1**), não apresentando características particulares que diferenciem o grupo de compostos ativos do grupo de inativos.

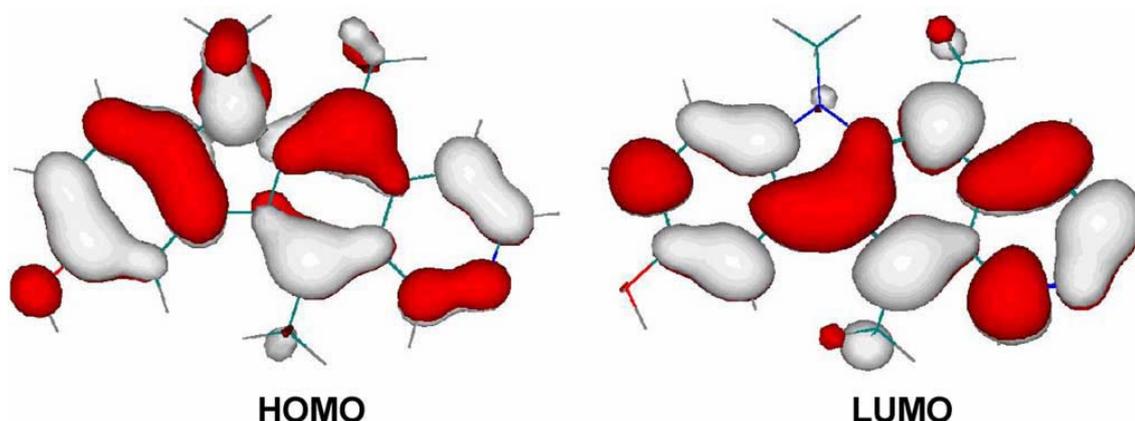


Figura 4.4 – Distribuição Espacial para o HOMO e LUMO (Isosuperfície da função de onda molecular).

Realizamos também um estudo considerando o comportamento de excitação eletrônica dos compostos. Parâmetros como os descritos anteriormente são bem descritos pelo método PM3, porém grandezas como energias de transições eletrônicas são superestimadas por este método, uma vez que utilizam a aproximação ZDO (Zero Differential Overlap) desconsiderando a correlação eletrônica [28]. Assim, para uma simulação mais realística dos espectros de absorção ótica, o método ZINDO/S-CI (Zerner`s Intermediate Neglect of Differential Overlap/Spectroscopic-Configuration Interaction) foi utilizado o método que apresenta uma versão especialmente calibrada para o estudo de compostos orgânicos [29,30].

Os espectros foram simulados considerando as geometrias obtidas dos cálculos PM3 e utilizando um conjunto médio de 200 configurações para o CI (singlete/triplete). A aproximação de Mataga-Nishimoto foi utilizada para as integrais gama com os fatores de interação usuais de 1.267 e 0.585 [31]. Os erros médios cometidos pelo método ZINDO/S-CI são de apenas 1000 - 2000 cm^{-1} (~0.1-0.2 eV) para todas as bandas abaixo de 45.000 cm^{-1} [32]. Os espectros de absorção gerados através destes cálculos são representados graficamente usando-se uma envoltória Lorentziana com largura média de 0.05 eV para as energias de transição cujos pesos são dados pelos valores da força de oscilador.

Os espectros são muito semelhantes para todas as moléculas, apresentando o limiar de absorção na região de 3.4-3.9 eV, e o pico de absorção máxima na região de 4.5-4.9 eV. Este comportamento está de acordo com a regularidade observada para as diferenças de energia LUMO-HOMO destas moléculas, já que essa diferença fornece uma aproximação (embora superestimada) para o valor em energia da primeira transição eletrônica. Portanto, na análise dos espectros das moléculas do grupo **G1**, não observamos nenhuma especificidade nas características espectroscópicas que pudesse ser correlacionada à atividade biológica.

Ilustramos na **Figura 4.5** as simulações dos espectros de dois compostos ativos (E11 e E12) e dois inativos (E32 e E33). Na **Tabela 4.2**, referente à Figura 4.5, estão as principais contribuições das configurações eletrônicas para a primeira transição e para a transição mais forte. As configurações predominantes para o limiar de absorção, assim como para o pico de absorção máxima, são formadas da transição envolvendo elétrons retirados dos orbitais HOMO e HOMO-1 e colocados nos orbitais LUMO e LUMO+1, padrão observado para todas as moléculas. A

notação utilizada, $|A \Rightarrow B \rangle$ indica que uma configuração foi formada retirando-se um elétron do orbital molecular A e colocando-o no orbital molecular B.

#	Limiar de absorção			Absorção máxima		
	Energia (eV)	Força de oscilador	Principais contribuições	Energia (eV)	Força de oscilador	Principais contribuições
E32	3.680	0.179	-0.53 $H \Rightarrow L+1 \rangle$ 0.33 $H-1 \Rightarrow L \rangle$	4.799	5.799	-0.37 $H \Rightarrow L+1 \rangle$ 0.54 $H-1 \Rightarrow L \rangle$
E33	3.702	0.116	-0.52 $H \Rightarrow L+1 \rangle$ 0.33 $H-1 \Rightarrow L \rangle$	4.707	5.842	-0.38 $H \Rightarrow L+1 \rangle$ 0.49 $H-1 \Rightarrow L \rangle$
E11	3.66	0.414	-0.56 $H \Rightarrow L+1 \rangle$ 0.32 $H-1 \Rightarrow L \rangle$	4.67	5.193	-0.35 $H \Rightarrow L+1 \rangle$ 0.51 $H-1 \Rightarrow L \rangle$
E12	3.64	0.288	0.54 $H \Rightarrow L+1 \rangle$	4.66	5.613	-0.36 $H \Rightarrow L+1 \rangle$ 0.51 $H-1 \Rightarrow L \rangle$

Tabela 4.2 – Energia, força de oscilador e principais contribuições para o limiar de absorção (primeira transição) e pico máximo (máxima absorção) para espectros simulados pelo método ZINDO. A notação $|A \Rightarrow B \rangle$ indica uma configuração gerada pela remoção de um elétron do orbital molecular A para o orbital molecular B.

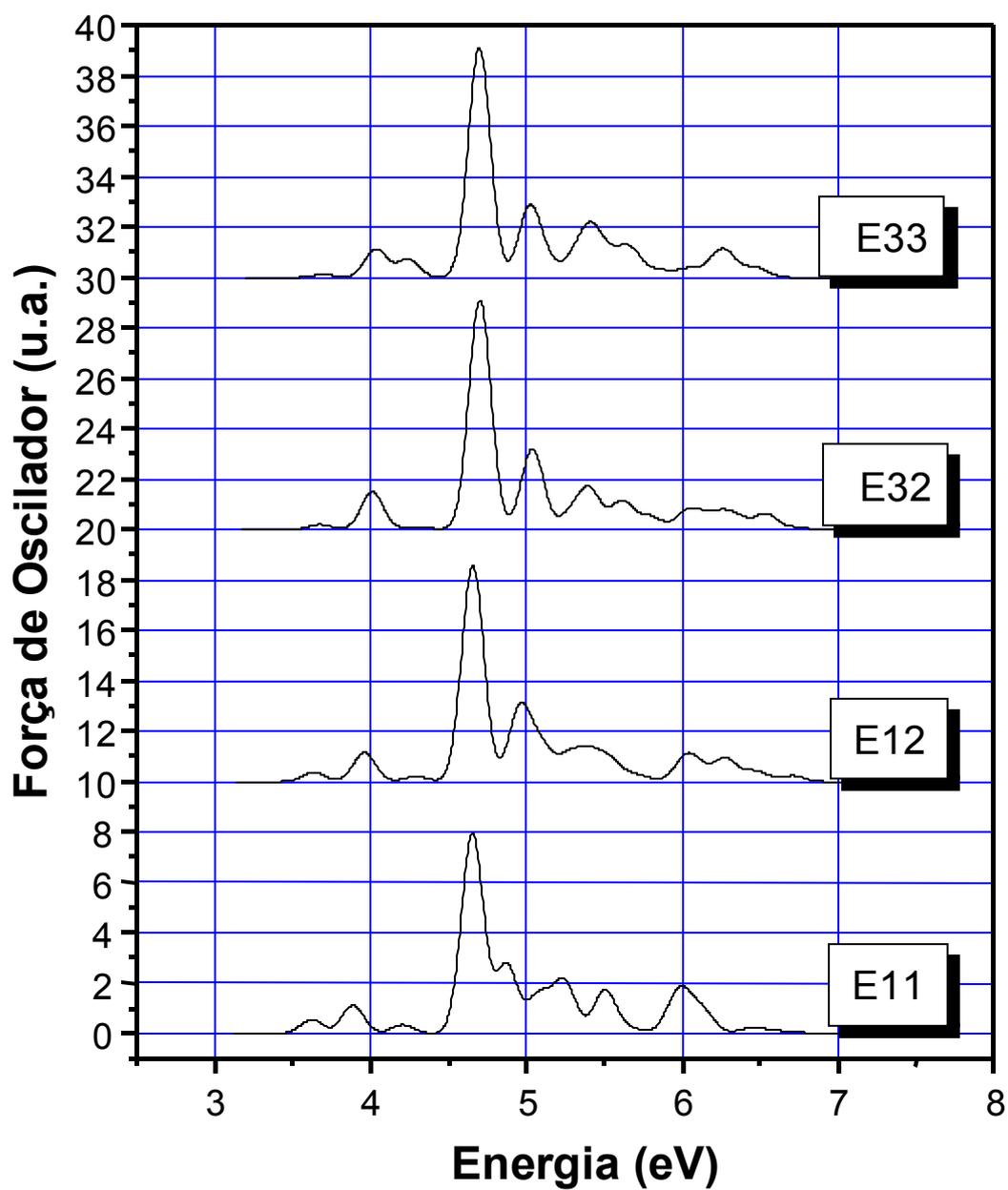


Figura 4.5 – Simulação dos espectros de absorção para derivados da elipticina ativos (E11 e E12) e inativos (E32 e E33).

As propriedades eletrônicas das drogas podem ainda ser interpretadas através do potencial eletrostático molecular (MEP) [33,34], que tem sido usado como indicador de reatividade e da importância de determinados sítios moleculares para as interações com as biomoléculas [28,35,36]. Para investigar as regiões de maior atividade nas elipticinas, determinamos o potencial eletrostático molecular considerando as cargas obtidas nos cálculos PM3 como pontuais, localizadas nos sítios atômicos (aproximação ZDO – Zero Differential Overlap). O potencial eletrostático foi calculado sobre pontos do plano do esqueleto molecular, com passos de 0,01Å nas direções x e y.

Os cálculos de potencial eletrostático podem indicar os sítios mais ativos no esqueleto molecular, através das regiões onde este é mais intenso.

Quase todas as moléculas apresentam um pico no potencial eletrostático associado ao sítio 9, identificando este como um sítio ativo. Experimentalmente substituições na posição C9 são identificadas como relevantes para a atividade das elipticinas [37]. Observamos também que as moléculas que apresentam sítios ativos no anel de piridina são biologicamente inativas (sítios atômicos 1-4 na Figura 3.1). Assim os cálculos de MEP indicam uma espécie de “competição” entre a densidade de carga nas vizinhanças da posição C9 e nos sítios sobre o anel de piridina como uma possível condição para a atividade destes compostos. No entanto, uma análise baseada somente nos sítios ativos não é o bastante para distinguir regras para uma diferenciação clara dos compostos em ativos e inativos.

Por outro lado, se associamos a análise dos sítios ativos e dos valores do momento de dipolo conseguimos identificar os compostos de acordo com sua atividade biológica.

Assim, obtemos uma regra para discriminação das moléculas ativas e inativas:

- ◆ *se a molécula tem $DM < 2.6$ Debye e não possui sítio ativo sobre o anel de piridina, ela é ativa. Se a molécula não obedece a esse critério, ela é inativa.*

Essa regra permite identificar corretamente 20 das 25 moléculas do grupo G1. Somente as moléculas ativas E01, E10, E12, E19 e E36 são erroneamente classificadas como inativas.

Esse resultado sugere que novos derivados da elipticina com promissora atividade biológica poderiam ser sintetizados com substituições seletivas no esqueleto molecular de forma que o DM e o MEP sobre o anel de piridina sejam reduzidos.

Na **Tabela 4.3** apresentamos os valores do momento de dipolo e sítios mais ativos para as moléculas do grupo **G1**, assim como as informações de atividade para demonstrar os resultados discutidos acima.

Usando a regra acima, propomos a classificação das moléculas do grupo **G2** (moléculas sem informação experimental de atividade biológica) em ativas ou inativas – **Tabela 4.4**. Para este grupo de compostos o critério baseado no valor do DM e MEP indica que oito moléculas seriam inativas e sete ativas.

Os resultados encontrados neste estudo preliminar, utilizando parâmetros eletrônicos, são uma boa indicação da aplicabilidade destes parâmetros como descritores capazes de nortear a análise da atividade biológica.

Nas seções seguintes, 4.2.3 e 4.2.4, apresentaremos os resultados obtidos com a metodologia de índices eletrônicos (MIE) e com a análise de componentes principais (PCA) e análise hierárquica de clusters (HCA), envolvendo os demais parâmetros eletrônicos.

#	DM (Debye)	Sítios Ativos		#	DM (Debye)	Sítios Ativos	
E01	2,843	r ₁ , r ₄ , N ₂	A	E14	2,852	r ₈ , r ₁₀ , C ₉ , r ₇	I
E02	2,884	r ₁₀ , N ₂ , N ₆	I	E15	2,580	r ₇ , C ₅ , C ₈ , r ₁ , C ₄	I
E03	2,342	r ₇ , r ₉ , N ₆	A	E16	2,631	r ₁₀ , r ₇ , r ₉ , R ₈ , C ₆	A
E04	2,608	r ₉ , N ₂ , N ₆	I	E17	2,267	r ₉ , r ₈ , r ₁₀ , r ₇	A
E05	1,895	r ₇ , N ₆	A	E18	2,222	r ₇ , r ₁ , C ₈	I
E06	1,984	r ₇ , N ₆	A	E19	3,10	r ₇ , r ₈ , C ₁₀ , R ₉	A
E07	2,456	r ₇ , r ₁₀ , N ₆	A	E20	2,618	r ₈ , r ₇ , r ₁₀ , R ₉ , N ₆	A
E08	1,862	r ₄ , r ₇ , r ₉	A	E32	2,731	r ₈ , r ₁₀ , r ₉ , r ₇	I
E09	2,003	r ₄ , r ₉ , N ₆	A	E33	1,728	C ₄ , C ₇ , r ₈ , C ₅	I
E10	2,868	r ₇ , N ₂ , N ₆	A	E34	1,723	r ₇ , r ₈ , r ₉ , r ₁₀ , R ₅	A
E11	1,769	r ₈ , r ₉ , R ₇	A	E35	0,979	r ₈ , r ₉ , r ₇ , r ₁₀	A
E12	2,840	R ₈ , r ₇ , r ₉	A	E36	2,322	r ₉ , r ₁₁ , r ₇ , C ₈ , R ₁	A
E13	2,225	r ₉ , r ₁₀ , r ₇ , R ₈	A				

Tabela 4.3 – Resultados para o momento de dipolo (DM) e sítios ativos das moléculas do grupo G1. Sítios, radicais e regiões mais ativas. C_n: carbono n, R_n: grupo lateral no carbono n e r_n: vizinhança do carbono n (1Å raio). A e I, refere-se, respectivamente, a composto ativo e inativo, dentro da aproximação proposta por Villemin et al.

#	D.M. (Debye)	Sítios ativos	Atividade proposta
E21	1.852	r ₇ , r ₈ , r ₁₀ , r ₉	A
E22	2.492	r ₇ , r ₉ , r ₁₀ , r ₈ , R ₁₁	A
E23	2.223	r ₇ , R ₁ , r ₉ , r ₁₀ , r ₈	I
E24	3.134	r ₇ , r ₈ , r ₉ , R ₅ , R ₁₀	I
E25	2.932	R ₇ , r ₈ , r ₁₀ , C ₉	I
E26	3.038	r ₇ , r ₉ , r ₈ , r ₁₀	I
E27	2.012	r ₇ , r ₈ , r ₁₀ , r ₉	A
E28	2.918	r ₁₀ , r ₈ , r ₇ , r ₉ , C ₄ , N ₆	I
E29	2.461	r ₉ , r ₁₀ , r ₇ , r ₈ , C ₁₁ , r ₁ , N ₂	I
E30	1.221	r ₇ , r ₈ , r ₁₀ , r ₉	A
E31	2.153	R ₈ , r ₇ , r ₉ , r ₁₀	A
E37	3.751	r ₉ , C ₇ , r ₈ , r ₁₀	I
E38	2.676	r ₈ , r ₇ , r ₉ , r ₁₀	I
E39	2.433	r ₈ , r ₁₀ , r ₉ , r ₇	A
E40	1.691	r ₈ , r ₉ , r ₇ , r ₁₀	A

Tabela 4.4 – Resultado para o valor do momento de dipolo e sítios ativos para as moléculas do grupo G2. Na última coluna a atividade proposta para o grupo. C_n: carbono n, R_n: grupo lateral no carbono n e r_n: vizinhança do carbono n (1Å raio).

4.2.3 – Investigação da Atividade Biológica através da Metodologia de Índices Eletrônicos – MIE

Nesta seção investigamos a correlação entre descritores teóricos e a atividade biológica através da metodologia de índices eletrônicos (MIE) [38,39]. O objetivo da MIE é separar moléculas ativas e inativas (com relação a uma classe específica de atividade biológica) utilizando, preferencialmente, os parâmetros eletrônicos, Δ e η . A MIE relaciona a atividade biológica com propriedades eletrônicas através de regras baseadas em valores críticos destes parâmetros. Δ fornece uma informação global da molécula através da diferença de energia entre dois estados de fronteira vizinhos, e η uma informação local através da medida da diferença na contribuição dos átomos de uma região molecular específica para a formação de dois estados de fronteira vizinhos.

Como no estudo da MIE a divisão das moléculas em regiões é essencial, o esqueleto molecular das elipticinas foi dividido em sete regiões, conforme detalhamento na seção anterior (Figura 4.3).

Considerando as regiões acima, analisamos os valores dos seguintes parâmetros eletrônicos: a energia do HOMO (Highest Occupied Molecular Orbital) e HOMO-1; a diferença de energia entre HOMO e HOMO-1 (ΔH); a energia do LUMO (Lowest Unoccupied Molecular Orbital energy) e LUMO+1; a diferença de energia entre LUMO e LUMO+1 (ΔL); a contribuição para a densidade local de estados (LDOS) para as regiões selecionadas, na formação dos orbitais HOMO e HOMO-1 (**CH** e **CH-1**) e suas diferenças (ηH); a contribuição para a densidade local de estados (LDOS) para as regiões selecionadas, na formação dos orbitais LUMO e LUMO+1 (**CL** e **CL+1**) e suas diferenças (ηL).

Através de uma busca exploratória sistemática, considerando os descritores calculados, analisamos os parâmetros Δ e η para as sete regiões e para os orbitais de fronteira. Verificamos então quais os valores críticos e as regras de seleção que melhor classificam as elípticas quanto ao seu comportamento ativo ou inativo. A melhor classificação de atividade foi obtida para a diferença de energia entre os orbitais HOMO e HOMO-1 (ΔH) e para a contribuição dos átomos da região **rD** para a formação do orbital LUMO (**CL(rD)**).

Analisando os valores dos descritores frente à classificação experimental de atividade resumida pela análise qualitativa dispostos na Tabela 3.1, podemos estabelecer a seguinte regra de seleção de atividade:

- ◆ se $\Delta H \leq 0,85$ eV e $CL(rD) \geq 0,022$ eV, a molécula será inativa, caso contrário será ativa.

Com a regra acima, podemos classificar 23 (vinte e três) das 25 (vinte e cinco) moléculas pertencentes ao grupo **G1** (moléculas 1-20 e 32-36), o que corresponde a um acerto de 92% na comparação com os dados experimentais. Este é um resultado muito bom já que somente dois descritores são considerados na análise. Apenas as moléculas E15 e E18 são incorretamente classificadas como ativas.

Na **Tabela 4.5** estão resumidos os valores das grandezas eletrônicas relacionadas com os descritores que deram origem à regra de separação de atividade.

molécula	ΔH (eV)	CL(rD)	A. Q.	molécula	ΔH (eV)	CL(rD)	A. Q.
E01	0,658	0,021	A	E14	0,692	0,029	I
E02	0,808	0,023	I	E15	1,119	0,009	I
E03	0,952	0,019	A	E16	0,980	0,015	A
E04	0,669	0,034	I	E17	0,956	0,017	A
E05	0,802	0,018	A	E18	0,979	0,017	I
E06	1,131	0,025	A	E19	1,055	0,021	A
E07	0,983	0,015	A	E20	1,004	0,017	A
E08	1,087	0,020	A	E32	0,614	0,023	I
E09	1,102	0,013	A	E33	0,402	0,028	I
E10	0,967	0,008	A	E34	0,758	0,015	A
E11	0,785	0,013	A	E35	1,094	0,020	A
E12	0,636	0,020	A	E36	1,061	0,015	A
E13	0,683	0,018	A				

Tabela 4.5 - Descritores ΔH e CL(rD) calculados com o PM3. A. Q. indica a atividade biológica experimental apresentada de forma qualitativa.

Aplicando a regra acima propomos uma classificação de atividade para as moléculas do grupo **G2**. Apresentamos na **Tabela 4.6** as grandezas eletrônicas e a proposição de atividade, observando que das 15 moléculas não testadas a regra estabelecida na MIE prevê que 11 seriam ativas.

Molécula	ΔH (eV)	CL(rD)	A. P.
E21	0,779	0,015	A
E22	0,795	0,008	A
E23	0,776	0,015	A
E24	0,669	0,022	I
E25	0,684	0,025	I
E26	0,603	0,022	I
E27	0,774	0,015	A
E28	0,922	0,005	A
E29	0,844	0,008	A
E30	1,085	0,023	A
E31	1,121	0,024	A
E37	0,763	0,017	A
E38	0,756	0,022	I
E39	0,548	0,020	A
E40	0,668	0,021	A

Tabela 4.6 - Descritores ΔH e CL(rD) calculados com o PM3. A. P. indica a atividade biológica qualitativa proposta através da metodologia MIE.

A fim de obtermos uma avaliação estatisticamente independente dos resultados encontrados, realizamos também uma análise teórica aplicando metodologias convencionalmente utilizadas nos estudos de reconhecimento de padrões muito utilizadas em investigações químicas de fármacos e de alimentos, a “Análise de Componentes Principais” (PCA) e “Análise Hierárquica de Clusters” (HCA).

4.2.4 – Investigação da Atividade Biológica através da Análise de Componentes Principais (PCA) e da Análise Hierárquica de Clusters (HCA)

A PCA utiliza um grande número de descritores simultaneamente no estudo de separação de amostras. De um modo geral, a metodologia consiste em gerar novas variáveis não correlacionadas (componentes principais – PC) através da combinação linear dos descritores moleculares inicialmente considerados na análise. Os compostos avaliados podem, a princípio, ser separados em subgrupos de acordo com sua atividade experimental. Esta separação, realizada no novo espaço de componentes principais, é obtida como fruto da seleção de apenas alguns dos descritores considerados no início da investigação. Os demais, descorrelacionados com a atividade, vão sendo, durante a análise, descartados.

Para o estudo realizado com as metodologias PCA e HCA, consideramos os parâmetros calculados com o método PM3 dentro do programa Einsigth [47].

Para a aplicação da técnica de PCA selecionamos inicialmente 56 descritores. Os descritores considerados envolvem entre outros:

- a energia do HOMO (Highest Occupied Molecular Orbital) e HOMO-1;
- a diferença de energia entre HOMO e HOMO-1 (ΔH);
- a energia do LUMO (Lowest Unoccupied Molecular Orbital energy) e LUMO+1;
- a diferença de energia entre LUMO e LUMO+1 (ΔL);
- a contribuição para a densidade local de estados (LDOS) para as regiões selecionadas, na formação dos orbitais HOMO e HOMO-1 (**CH** e **CH-1**) e suas diferenças (ηH);

- a contribuição para a densidade local de estados (LDOS) para as regiões selecionadas, na formação dos orbitais LUMO e LUMO+1 (**CL** e **CL+1**) e suas diferenças ($\eta\mathbf{L}$);
- o valor do momento de dipolo (**DM**);
- o calor de formação (**HF**);
- refratividade (**R**),
- polarizabilidade (**Po**),
- massa (**M**),
- energia de hidratação (**HE**);
- energia de solvatação (**SE**),
- coeficiente de partição molecular octanol-água ($\log \mathbf{P}$).

A energia do HOMO (relacionada ao potencial de ionização) e energia do LUMO (relacionada a afinidade eletrônica) são medidas da susceptibilidade da molécula perder um par de elétrons para um agente eletrofílico ou aceitar um par de elétrons de um agente nucleofílico. A refratividade e polarizabilidade são propriedades que medem a susceptibilidade da molécula tornar-se polarizada. O momento de dipolo elétrico representa a força de interação do tipo polar. E o coeficiente de partição é o principal descritor que caracteriza a afinidade relativa de uma droga entre o meio aquoso e lipídico, propriedade correlacionada com a atividade da droga devido á relação direta com os fenômenos de transporte, absorção, distribuição e metabolismo.

Dos 56 descritores iniciais, os parâmetros refratividade, polarizabilidade e massa foram descartados em uma análise preliminar. De forma similar, outros descritores foram descartados por se mostrarem inábeis na distinção entre moléculas ativas e inativas.

A separação das moléculas em subgrupos de atividades distintas (ativos e inativos), foi alcançada utilizando-se uma combinação final dos seguintes descritores:

- ΔH - a diferença de energia entre HOMO e HOMO-1;
- $CL(rD)$ - a contribuição dos átomos da região rD para a formação do orbital LUMO;
- DM – o valor do momento de dipolo; e
- $\eta H(rB)$ – diferença entre LDOS para os orbitais de fronteira HOMO e HOMO-1 (CH and $CH-1$) na região rB .

Três parâmetros selecionados e responsáveis pelos melhores resultados na investigação da PCA são parâmetros eletrônicos relacionados com os descritores da MIE. Durante as investigações com a PCA, testes considerando a não utilização dos descritores da MIE mostraram uma queda na porcentagem de acerto das predições. Este resultado é uma nova evidência da importância dos parâmetros eletrônicos na classificação de atividades biológicas.

A **Figura 4.6** e **Figura 4.7** apresentam para o grupo **G1** resultados obtidos para a metodologia PCA. As moléculas estão representadas por pontos em um plano de duas componentes principais (PC1 e PC2). Os descritores utilizados foram auto-escalados antes da análise. Os compostos estão separados em dois subgrupos de atividade (ativos e inativos) horizontalmente opostos, indicando que o eixo PC1 é o principal responsável por esta separação. O grupo de moléculas ativas está localizado na parte direita do gráfico e o grupo de inativas está distribuído na parte esquerda do gráfico. O plano PC1 x PC2 conserva 74% da variância total dos dados originais.

Os componentes PC1 e PC2 são escritos em termos dos descritores inicialmente considerados no cálculo segundo as equações:

$$\mathbf{PC1} = - 0,30 \mathbf{DM} + 0,63 \Delta\mathbf{H} - 0,50 \mathbf{CL(rD)} + 0,52 \eta\mathbf{H(rB)} \quad (1)$$

$$\mathbf{PC2} = 0,76 \mathbf{DM} + 0,24 \Delta\mathbf{H} - 0,50 \mathbf{CL(rD)} - 0,33 \eta\mathbf{H(rB)} \quad (2)$$

Estas equações mostram que a contribuição do descritor **DM** é menor que as dos demais descritores na PC1, e maior na PC2.

Estes novos resultados, obtidos por uma metodologia classicamente utilizada no estudo de reconhecimento de padrões, confirmam a importância das variáveis eletrônicas na investigação de atividade das elipticinas. O **DM** mostra correlação com a atividade das elipticinas [25,26,40], e MEP indica sítios ativos sobre a região **D**. Ao mesmo tempo, investigações experimentais identificam a região **D** como relevante para a atividade biológica das elipticinas. Por outro lado, $\Delta\mathbf{H}$, **CL(rD)** e $\eta\mathbf{H(rB)}$ não foram considerados em estudos anteriores e são novos parâmetros os quais podemos considerar em futuras investigações.

Na análise com a PCA, as moléculas E01 e E12 são incorretamente classificadas como inativas e as moléculas E15 e E18 incorretamente classificadas como ativas. Isto corresponde a classificação correta de 21 moléculas e, portanto a 84% de acerto.

Analisando em conjunto os gráficos das **Figuras 4.6 e 4.7**, verificamos que os descritores **CL(rD)/DM** e $\Delta\mathbf{H}/\eta\mathbf{H(rB)}$ têm papéis similares na separação das moléculas ativas e inativas em regiões horizontalmente distintas no gráfico SCORE. Se a esta análise agregarmos as informações contidas na equação 1, podemos inferir que para que novos derivados da elipticina ofereçam índices de atividade eles

devem apresentar valores positivos para as variáveis **DM** e **CL(rD)** , combinados com valores negativos para as variáveis ΔH e $\eta H(rB)$.

Na **Figura 4.8** apresentamos o resultado da HCA para o conjunto de variáveis apresentadas no resultado PCA acima. Os descritores foram auto-escalados. As moléculas ativas foram separadas das inativas e agrupadas em dois subgrupos distintos no gráfico. A formação de dois grupos distintos, um para as moléculas ativas e outro para as inativas, demonstra que os descritores adotados na classificação de atividade foram capazes de avaliar corretamente a similaridade que deve existir entre todos os compostos ativos e entre todos os inativos. Os dois clusters ativo e inativo (separados por uma linha horizontal entre as moléculas E05 e E33) apresentam uma similaridade igual a zero, o que demonstra que as moléculas ativas e inativas são bem separadas no espaço quadridimensional da PCA. Nesta análise HCA temos as moléculas E01, E12, E15, E18 (incorretamente classificadas na PCA) e E13 em clusters distintos de sua atividade, o que corresponde a um acerto de 80% na classificação.

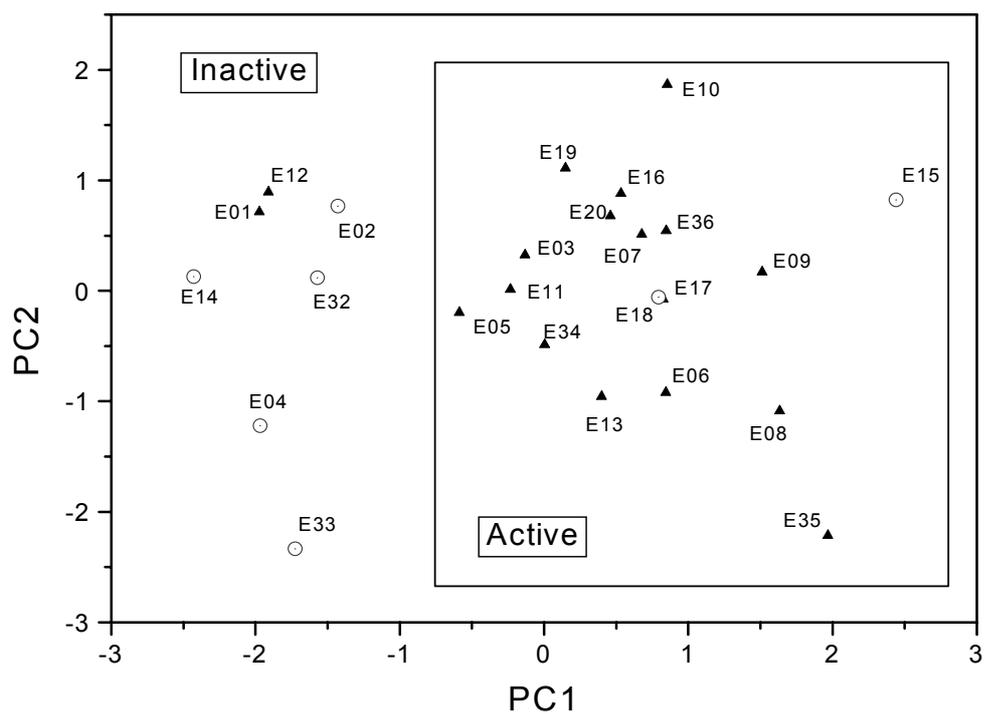


Figura 4.6 - Gráfico SCORE do Grupo G1– distribuição dos compostos no plano das componentes principais PC1 x PC2. Triângulo e círculo representam moléculas ativas e inativas, respectivamente.

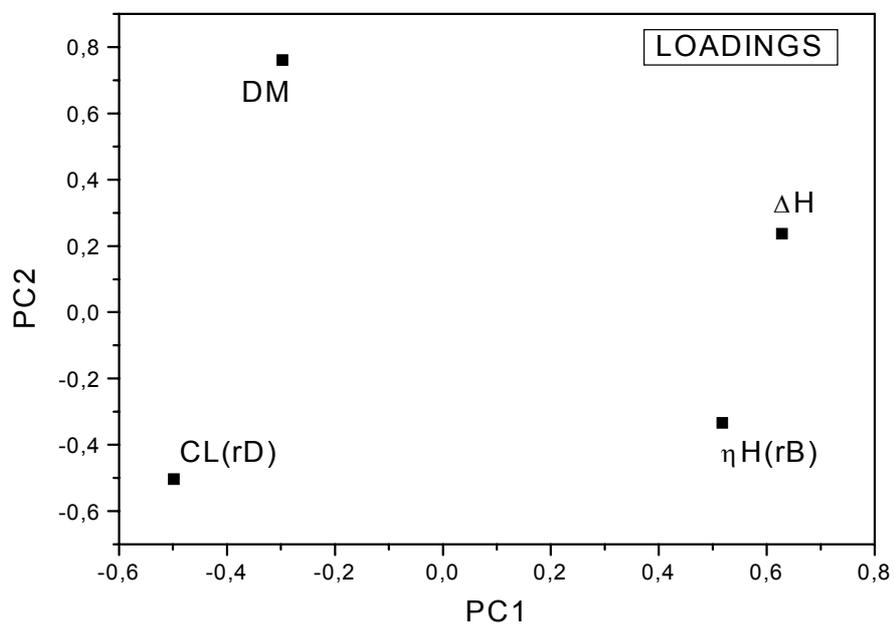


Figura 4.7 - Gráfico LOADINGS para grupo G1 – distribuição dos descritores responsáveis pela separação dos compostos em grupos distintos de atividade.

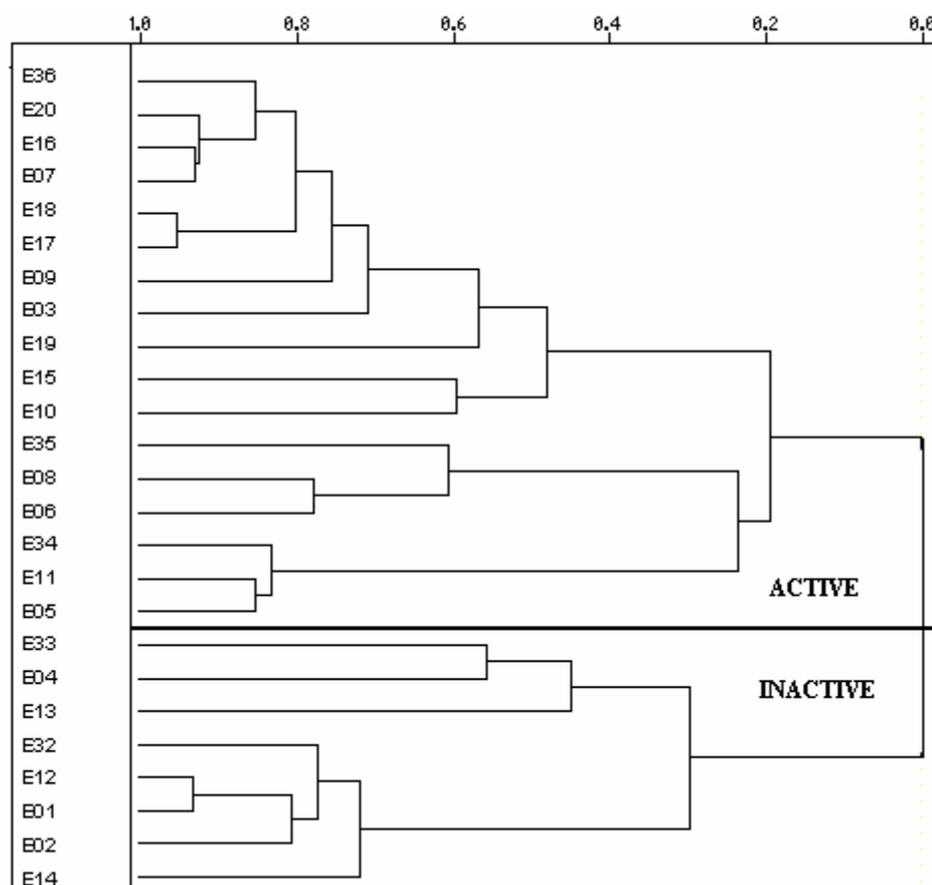


Figura 4.8 - Dendrograma da distribuição hierárquica do grupo G1. Os compostos ativos e inativos estão separados em grupos distintos e apresentam uma similaridade nula entre si.

Estas combinações de parâmetros foram selecionadas entre dezenas de outras por apresentarem os melhores resultados. Podemos observar que incluem parâmetros eletrônicos e que estes muitas vezes são isoladamente os que mais contribuem para a formação dos eixos PC1 e PC2, o que é mais uma indicação da relevância dos descritores eletrônicos na classificação, ainda que qualitativa, da atividade biológica dos compostos.

Os padrões de separação de atividades construídos serão agora utilizados para a proposição de atividade biológica dos compostos do grupo **G2**.

A metodologia PCA é classificada dentro da quimiometria como uma técnica exploratória e apresenta limitações em estudos de previsão. A aplicação da metodologia nos estudos de simulação de atividade se resume em decidir se novos compostos possuem de um modo geral as características de outro grupo de compostos já conhecido.

No caso específico do nosso estudo gostaríamos de comprovar a semelhança em atividade das moléculas do grupo **G2** e do grupo **G1**. Para tal, os compostos do grupo **G2** devem ser projetados no espaço de componentes principais PC1 x PC2 construído para os parâmetros selecionados pelo grupo **G1**.

Projetamos os compostos nos resultados PCA apresentados nesta seção. Os resultados das projeções das novas moléculas no espaço das componentes principais são apresentados na **Figura 4.9**. Projetamos no gráfico os 40 compostos (25 do grupo **G1** e 15 do grupo **G2**) no espaço de componentes principais construído como combinação das variáveis que foram capazes de fazer a separação em atividade do grupo **G1** isoladamente.

Observamos na **Figura 4.9** que os compostos do grupo **G2** se misturam com os do grupo **G1**, mostrando afinidade tanto com as moléculas ativas quanto com as inativas. Temos nove das quinze moléculas ocupando a região das moléculas ativas, sendo elas a E21, E22, E27, E28, E29, E30, E31, E39 e E40. As outras seis - E23, E24, E25, E26, E37 e E38 - ocupam a região das inativas.

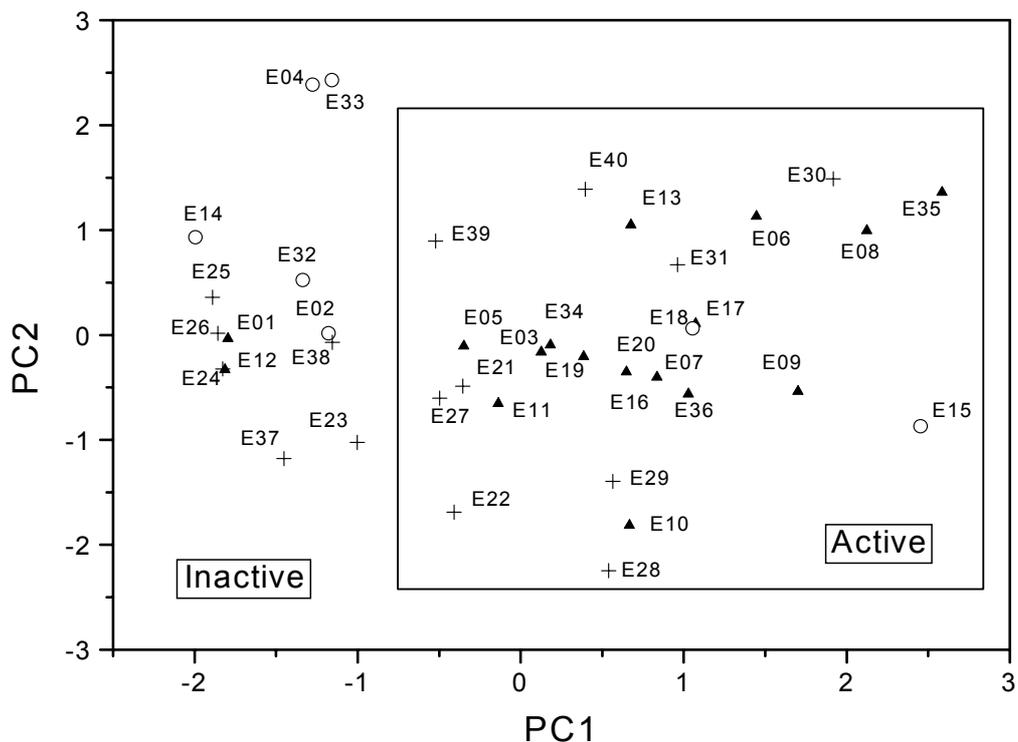


Figura 4.9 - Resultados da análise de PCA. Gráfico de separação dos compostos do grupo G1 e proposição de atividade para os compostos do grupo G2. Os símbolos triângulo, círculo e cruz representam ativo, inativo e não testado, respectivamente.

Os eixos PC1 e PC2 acima são construídos de acordo com as equações:

$$\text{PC1} = -0,44 \text{ DM} + 0,64 \Delta\text{H} - 0,30 \text{ CL}(\text{rD}) + 0,55 \eta\text{H}(\text{rB}) \quad (3)$$

$$\text{PC2} = -0,40 \text{ DM} - 0,22 \Delta\text{H} + 0,81 \text{ CL}(\text{rD}) + 0,37 \eta\text{H}(\text{rB}) \quad (4)$$

As equações mantêm as mesmas variáveis investigadas no início da seção, mas as contribuições de cada variável para a formação dos componentes principais foi alterada pela presença das novas moléculas. Isto ocorre uma vez que o sistema é

autoescalado após a inclusão de novos compostos de forma a se uniformizar os valores dos diferentes descritores, evitando uma análise que favoreça algum descritor em particular.

Na **Figura 4.10** apresentamos o resultado da HCA para o conjunto de variáveis apresentadas no resultado PCA para o grupo de 40 compostos. Na HCA somente as moléculas E39 e E40 estão inseridas em uma classe diferente de atividade proposta pela PCA.

Para concluirmos na **Tabela 4.7** estão sumarizadas todas as informações de previsão de atividade obtidas para o **G2**. Os índices utilizados na tabela seguem o padrão adotado no estudo, **A** para composto ativo, e **I** para inativo. Analisando os resultados de todas as metodologias observamos que: a) os compostos E23 e E37 obtiveram uma indicação de atividade e serão então considerados inativos; b) os compostos E28, E29, E39 e E40 obtiveram uma indicação de inatividade e serão então considerados ativos. Considerando então uma classificação por votos, sendo uma pontuação de no mínimo 3 indicações necessária para classificar o composto quanto à atividade, nossos resultados indicam que dos 15 compostos do grupo **G2**, 9 (nove) deverão ser biologicamente ativos e 6 (seis) inativos.

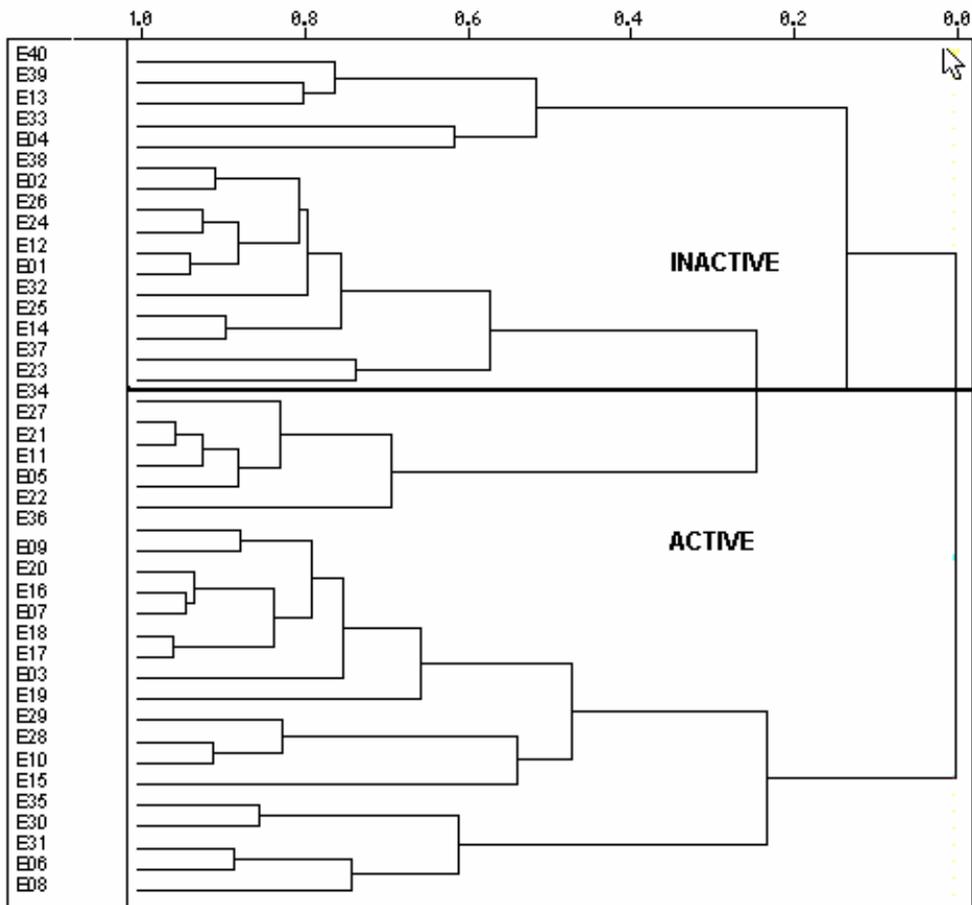


Figura 4.10 - Dendrograma da distribuição hierárquica do conjunto de moléculas (G1 + G2).

Molécula	PCA	HCA	MIE	Seção 4.2.2
E21	A	A	A	A
E22	A	A	A	A
E23	I	I	A	I
E24	I	I	I	I
E25	I	I	I	I
E26	I	I	I	I
E27	A	A	A	A
E28	A	A	A	I
E29	A	A	A	I
E30	A	A	A	A
E31	A	A	A	A
E37	I	I	A	I
E38	I	I	I	I
E39	A	I	A	A
E40	A	I	A	A

Tabela 4.7: Sumário da atividade proposta para os compostos do grupo G2 das elipticinas segundo diferentes metodologias.

4.3 – Referências

- [1] L. P. Hammett, *Physical Organic Chemistry* (McGraw-Hill, New York, 1940).
- [2] Dalton, L. K., Demerac, S. Elmes, B. C., Loder, J. W., Swan, J. M., and Teitei, J... *Aust. J. Chem.* 1967, 20, 2715-2727.
- [3] JaszoldHoworko, R. *et al.* *J. Med Chem.* 1994, 37, 2445-2452.
- [4] Anderson, W. K., Gopalsamy, A. and Reddy, P. S.. *J. Med. Chem.* 1994, 37, 1955-1963; Jurayj, J., Haugwitz, R. D., Varma, R. K., Paull, K. D., Barret, J. F and Cushman, M.. *J. Med. Chem.* 1994, 37, 2190-2197.
- [5] Mathe, G., Pontiggia, P., Bourut, C., Chenu, E. and Orbach-Arbouys, S.. *Biom. and Pharmacother.* 1994, 48, 51-53.
- [6] Auclair, C.. *Arch. Biochem. Biophys.* 1987, 259, 1.
- [7] Kohn, K. W., Ross, W. E. and Glaubinger, D., in F. E. Hahn (Ed.) *Antibiotics*, vol.2, Springer-Verlag, Berlin, 1979, p. 195
- [8] Le Pecq, J. B., Dat-Xuong, N., Gosse, C., and Paoletti, C. A.. *Proc. Natl. Acad. Sci. USA* 1974, 71, 5078-5082.
- [9] Paoletti, C., Cros, S., Dat-Xuong, N., Lecointe, P., and Moisand, A.. *Chem. Biol. Interact.* 1979, 25, 45-58.
- [10] Chabane, H., Lamazzi, C., Thiéry, V., Guillaumetb, G.. *Tetrahedron Letters* 2002, 43, 2483–2486.
- [11] Tran-Thi, H. A., Nguyen-Thi, T., Michel, S., Tillequin, F., KOCH, M., Pfeiffer, B., Pierré, A., Trinh-Van-Dufat, H.. *Chem. Pharm. Bull.* 2004, 52, 540-545.
- [12] Ergun, Y., Patir, S., Okay, G.. *Synthetic Communications.* 2004, 34, 435-442.
- [13] Ishikura, M., Hino, A., Yaginuma, T., Ágata, I. and Katagiri, N.. *Tetrahedron* 2000, 56, 193–207.

- [14] Diaz, M., Cobas, A., Guitian, E., Castelo, L.. *European Journal of Organic Chemistry* 2001, 23, 4543-4549.
- [15] Ishikura, M., Hino, A., Katagiri, N.. *Heterocycles* 2000, 53, 11-14.
- [16] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart, *J. Am. Chem. Soc.* 107, 3902 (1985).
- [17] J. J. P. Stewart, *J. Comp. Chem.* 10, 209 (1991); 10, 221 (1991).
- [18] J. Ridley, and M. C. Zerner *Theor. Chim. Acta*, 72, 347 (1987).
- [19] S. O. Dantas, F. C. Lavarda, B. Laks, and D. S. Galvão, *J. Mol. Struct. (Theochem)* 253, 319 (1992).
- [20] S. O. Dantas, and D. S. Galvão, *J. Mol. Struct. (Theochem)* 257, 37 (1992).
- [21] P.M.V.B.Barone, S.O.Dantas, and D.S.Galvão, *J. Mol. Struct. (THEOCHEM)* 465, 219 (1999).
- [22] R.G. Harvey, *American Scientist* 70, 386 (1982).
- [23] P. M. V. B. Barone, A. Camilo Jr., and D. S. Galvão, *Synthetic Metals* 102, 1410 (1999).
- [24] SPARTAN program, version 4.1, Wavefunction Inc., Irvine, CA92715, USA.
- [25] Dantas, S. O., Lavarda, F. C., Laks, B., and Galvão, D. S.. *J. Mol. Struct. (THEOCHEM)* 1992,253,319-332.
- [26] Braga, S. F., de Melo, L. C. and Barone, P. M. V. B.. *J. Mol. Struct. (THEOCHEM)*. 2004, 710, 51-59.
- [27] MOPAC program, version 6.0, Quantum Chemistry Exchange No. 455.
- [28] G. H. Loew, D. Berkowitz, H. Weinstein, and S. Srebrenik, in *Molecular Quantum Phamacology* (Reidel Dordrecht, 1974), E. D. Bergmann and B. Pullman (Eds.), pp. 355.

- [29]** CACHE 5.0, Fujitsu Limited, Chiba City, Chiba 2618588, Japan (www.CACheSoftware.com)
- [30]** C. A. Reynolds, G. G. Ferenczy, and W. G. Richards, *J. Mol. Struct. (THEOCHEM)* 256, 249 (1992).
- [31]** W. D. Edwards, M. C. Zerner, *Theor. Chim. Acta*, 72, 347 (1987).
- [32]** Zerner, M. C., D. B. Boyd (Eds.), *Reviews in Computational Chemistry*, VCH, New York, 1991, pp. 313-365.
- [33]** E. Scrocco and J. Tomasi, *Top. Curr. Chem.*, 42, 95 (1973).
- [34]** S. Guha. D. Majumdar and A. K. Bhattacharjee, *J. Mol. Struct. (THEOCHEM)* 256, 61 (1992).
- [35]** E. Scrocco and J. Tomasi, in *Advances in Quantum Chemistry* (Academic Press, N. Y., 1978), P. Löwdin (Ed.), Vol.11pp. 115.
- [36]** R. Bonaccorsi, E. Scrocco and J. Tomasi, *J. Chem. Phys.* 52, 5270 (1970).
- [37]** Barone, P. M. V. B., Camilo, A., Jr., Galvão, D. S.. *Phys. Rev. Lett.* 1996, 77, 1186-1189.
- [38]** Braga, R. S., Barone, P. M. V. B., Galvão, D. S.. *J. Mol. Struct. (THEOCHEM)* 1999, 464, 257-266.
- [39]** Barone, P. M. V. B., Dantas, S. O., and Galvão, D. S... *J. Mol. Struct. (THEOCHEM)*. 1999, 465, 219-229.
- [40]** Barone, P. M. V. B., Braga, R. S., Camilo, A., Galvão, D. S.. *J. Mol. Struct. (THEOCHEM)* 2000, 505, 55-66.
- [41]** Vendrame, R., Braga, R. S., Takahata, Y., Galvão, D. S.. *J. Chem. Inf. Comput. Sci.* 1999, 39, 1094-1104.
- [42]** Santo, L.L.D., Galvão, D. S.. *J. Mol. Struct. (THEOCHEM)* 1999, 464, 273-279.
- [43]** Braga, S. F., Galvão, D. S.. *J. Chem. Inf. Comput. Sci.* 2003, 43, 699-706.

[44] Vendrame, R., Braga, R. S., Galvão, D. S.. *J. Mol. Struct. (THEOCHEM)* 2002, 619, 195-205.

[45] Braga, R. S., Vendrame, R., Galvão, D. S.. *J. Chem. Inf. Comput. Sci.* 2000, 40, 1377-1385.

[46] Cyrillo, M., Galvão, D. S.. *J. Mol. Struct. (THEOCHEM)* 1999, 464, 267-272;

Cyrillo, M., Galvão, D. S., Chem2Pac: a Computational Chemistry Integrator for Windows. *EPA Newsletter* 1999, 67, 31-38.

<http://www.ifi.unicamp.br/gsonm/chem2pac>.

[47] Einsight 3.0. Infometrix, Inc. 2200 Sixth Ave, Suite 833, Seattle WA, 98121, 1991.

CAPÍTULO 5

FURANOCUMARINAS

5.1 – Introdução

Entre as variadas linhas de pesquisa que buscam formas terapêuticas menos agressivas para o tratamento de tumores, se encontra a investigação de drogas que são ativadas por luz. Uma vez administradas ao paciente, essas drogas não exibem nenhum efeito citotóxico até serem expostas à radiação em comprimento de onda específico [1,2]. Os cientistas têm então se dedicado a pesquisa de agentes fotosensibilizadores com baixos efeitos fototóxicos e mutagênicos, tomando como modelo a estrutura tricíclica de drogas fotoquimioterapêuticas naturais conhecidas.

O mecanismo de ação citotóxica das furanocumarinas envolve inicialmente intercalação entre dois pares de bases nitrogenadas do DNA, formando um complexo molecular furanocumarina-DNA, que a princípio se apresenta inerte. Após ser irradiado com luz na região UV do espectro, ocorre uma reação fotoquímica entre a droga intercalada e um par de bases nitrogenadas pirimidínicas (citosina e timina), na qual se forma um ciclo-aduto. A timina é o nucleotídeo de preferência para formação de aduto. A fotorreação ocorre com a participação da dupla ligação entre os carbonos 4',5' do anel de furano ou 3,4 da cumarina com a dupla ligação entre os carbonos 5,6 nas bases pirimidínicas. Absorvendo um segundo fóton e, sendo a configuração geométrica favorável, poderá formar-se uma nova ligação com a base pertencente à fita oposta através da outra dupla ligação fotorreativa, estabelecendo uma ligação cruzada entre as duas fitas do DNA. Estudos demonstram que os psoralenos, devido à sua estrutura linear, são capazes de formar o *cross-link* entre as fitas de DNA impedindo a duplicação das hélices e a reprodução celular, matando o tumor, enquanto que as angelicinas, devido à sua estrutura angular em geral não formam *links* entre as duas fitas da macromolécula.

Isso poderia explicar a menor ou nenhuma atividade das angelicinas, apesar de possuírem maior índice de fotoligação com o DNA. Este comportamento pode estar relacionado à diferente geometria de intercalação exibida pelas furanocumarinas angulares e conseqüente modificações no alinhamento das duplas ligações fotorreativas da droga com respeito à base do DNA [2-8]. No entanto, os psoralenos apresentam uma maior fototoxicidade de pele do que as angelicinas, motivo pelo qual se pesquisa derivados de angelicinas que possam apresentar maior reatividade.

Portanto, correlacionar índices eletrônicos com atividade para esta família de moléculas pode não ser tão simples, principalmente devido à diversidade de aplicações clínicas e às poucas informações quantitativas sobre a atividade biológica.

Neste capítulo apresentamos os resultados para o estudo da relação entre descritores teóricos e a atividade dos psoralenos. Iniciamos o estudo buscando as estruturas geométricas referentes ao estado de menor energia e determinando as propriedades eletrônicas dos compostos. Nas seções seguintes apresentaremos os resultados da investigação usando a MIE, a PCA e a HCA para discriminar as moléculas quanto à atividade.

5.2 – Resultados

5.2.1 – Análise Conformacional

As furanocumarinas estudadas possuem, em sua maioria, uma estrutura molecular pouco flexível, com grupos laterais pequenos ligados ao esqueleto molecular, o que facilita a obtenção das geometrias dos compostos.

Para o grupo de compostos estudados neste capítulo não foram encontrados na literatura dados de geometria de raios-X que validassem a escolha de um método com base em resultados comparativos entre geometrias experimentais e geometrias obtidas teoricamente. Com base no histórico de resultados obtidos com métodos semi-empíricos, inclusive já utilizados no estudo das elipticinas (Capítulo 4), o método PM3 foi selecionado para a investigação da estrutura geométrica dos compostos estudados também neste capítulo.

Assumindo que a conformação de menor energia é a melhor estrutura para o estudo de propriedades eletrônicas, otimizamos todas as estruturas em vácuo, a partir de uma conformação inicial construída com o programa Hyperchem [9], realizando busca conformacional apenas para aqueles compostos que apresentam grupos laterais flexíveis. Nesse caso, como no estudo das elipticinas, variamos simultaneamente os ângulos diedrais dentro do intervalo de 360 graus, em passos de 20 graus (em um total de 324 conformações). As geometrias assim obtidas foram então otimizadas mantendo-se fixos os diedrais variados.

Estes cálculos foram realizados utilizando-se o programa Mopac2000 integrado ao pacote computacional Cache WorkSystem PRO [10,11], o qual inclui o método PM3 aplicado na otimização das geometrias moleculares. Na **Figura 5.1**

ilustramos a superfície do calor de formação (Kcal/mol), obtida para as estruturas otimizadas como função da variação de um ângulo diedral e variação simultânea dos ângulos diedrais.

Analisamos as conformações possíveis para todas as cadeias laterais variando seus diedrais flexíveis e analisamos então a estabilidade do conjunto de geometrias obtido. A partir da localização da conformação de menor energia, iniciamos a obtenção dos descritores quânticos necessários para a exploração teórica de estrutura-atividade dos compostos estudados.

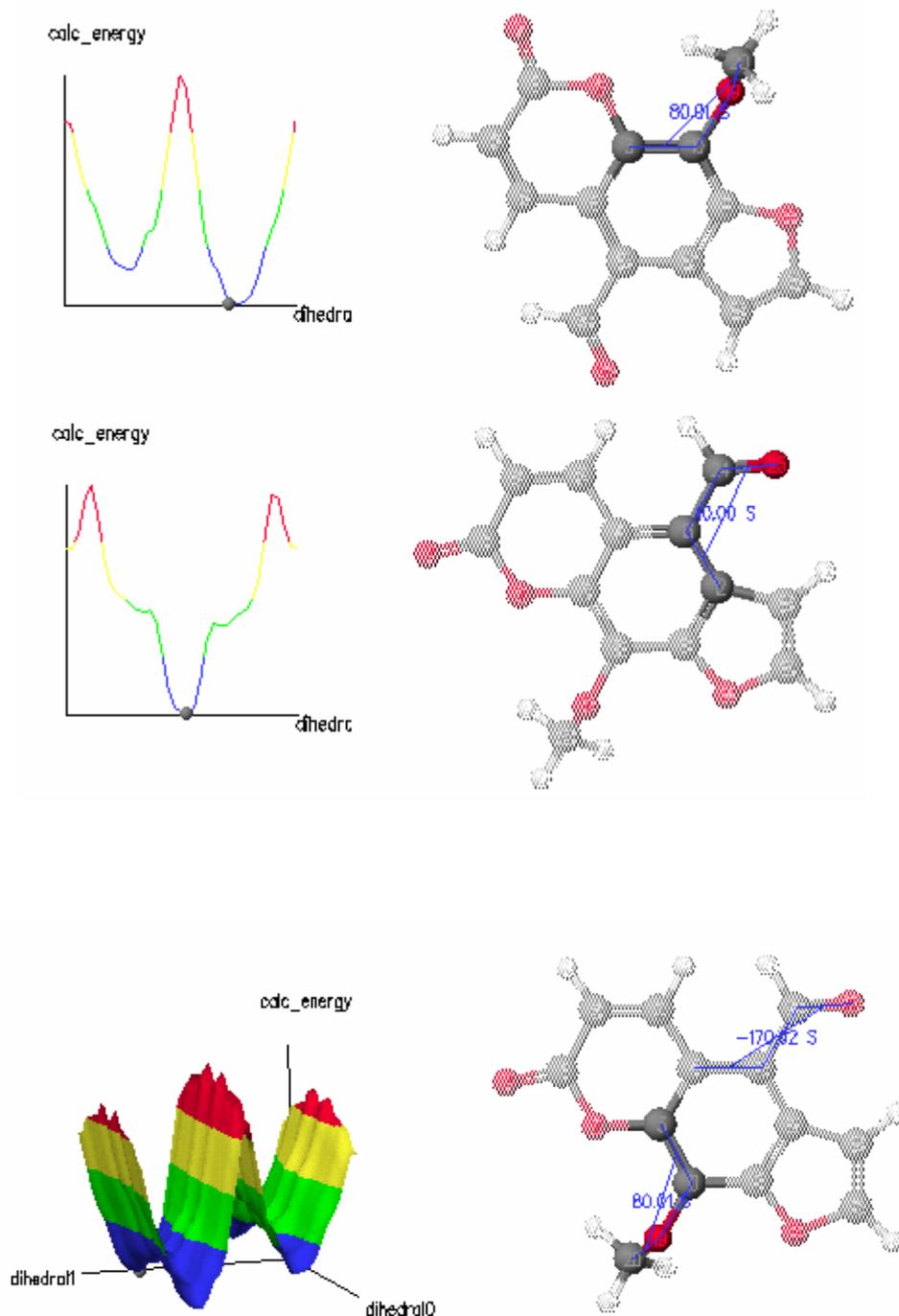


Figura 5.1 – Busca sistemática da conformação mais estável da molécula psoraleno P₁₇ no vácuo. (à direita) Superfície do Calor de Formação (Kcal/mol) como uma função da variação do ângulo diedral (em destaque na figura) em passos de 20 graus. (à esquerda).

5.2.2– Investigação da Atividade Biológica através da Metodologia de Índices Eletrônicos – MIE

Nesta seção investigamos a correlação entre descritores teóricos e a atividade fotodinâmica através da metodologia de índices eletrônicos (MIE).

A partir das geometrias obtidas na seção 5.2.1, calculamos os parâmetros eletrônicos Δ e η , assim como os parâmetros $C(n)r_i$. Esses cálculos podem também ser executados de forma automática usando o programa Chem2Pac [12].

O esqueleto molecular foi dividido em cinco regiões específicas (**Figura 5.2**) para o estudo da contribuição dos átomos na formação dos orbitais moleculares:

- ◆ anel benzênico externo que apresenta a maior ordem de ligação em todos os compostos (região **rA**);
- ◆ o anel de furano (região **rB**);
- ◆ os sítios fotoreativos 3,4 e 4',5' (região **rC**);
- ◆ os sítios fotoreativos 3 e 4 do anel benzênico externo (região **rD**);
- ◆ os sítios fotoreativos 4' e 5' do anel de furano (região **rE**).

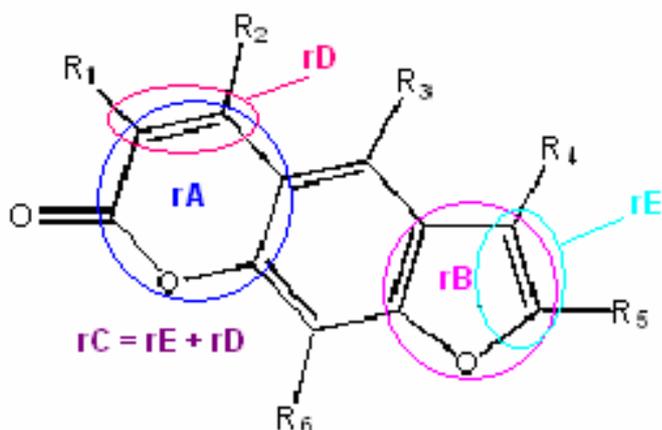


Figura 5.2: Regiões sobre o esqueleto molecular dos psoralenos selecionadas para os cálculos de LDOS.

Considerando as cinco regiões acima, calculamos os parâmetros eletrônicos utilizados na MIE:

- a energia do HOMO (Highest Occupied Molecular Orbital) e LUMO (Lowest Unoccupied Molecular Orbital energy) e de seus primeiros vizinhos HOMO-1 e LUMO+1;
- a diferença de energia entre os orbitais: HOMO e HOMO-1 (ΔH) e LUMO e LUMO+1 (ΔL);
- a contribuição para a densidade local de estados (LDOS) para as cinco regiões moleculares selecionadas, na formação dos orbitais HOMO e HOMO-1 ($C(H)$ e $C(H-1)$) e suas diferenças ($\eta H = C(H) - C(H-1)$); e,
- a contribuição para a densidade local de estados (LDOS) para as cinco regiões moleculares selecionadas, na formação dos orbitais LUMO e LUMO+1 ($C(L)$ e $C(L+1)$) e suas diferenças ($\eta L = C(L+1) - C(L)$).

Uma análise preliminar destes parâmetros mostrou que aqueles relacionados com os orbitais HOMO, HOMO - 1, LUMO e LUMO+1 apresentavam maior

correlação com os dados experimentais de atividade fotodinâmica. A melhor classificação de atividade foi obtida para a diferença de energia entre os orbitais LUMO e LUMO+1 (η_L) e para a contribuição dos átomos da região **rB** para a formação do orbital HOMO (**C(H)**).

Analisando os valores dos descritores frente à classificação experimental de atividade resumida pela análise qualitativa dispostos na Tabela 3.2, podemos estabelecer a seguinte regra de seleção de atividade:

- ◆ se $\eta_L > 0,5$ e $CH(rB) > 0,65$, a molécula será ativa, caso contrário será inativa.

Com a combinação binária descrita acima, que define a região na qual as moléculas são ativas, podemos classificar 28 (vinte e oito) das 31 (trinta e uma) moléculas pertencentes ao grupo **G1**, o que corresponde a um acerto de 90% na comparação com os dados experimentais. Temos um resultado muito bom, já que somente dois parâmetros são envolvidos na análise. Apenas as moléculas inativas P26, P30 e P40 são incorretamente classificadas como ativas.

Na **tabelas 5.1** estão os valores das grandezas eletrônicas relacionadas com os descritores da MIE que deram origem à regra de separação de atividade.

Nº.	C(H)B	η_{LB}	A.Q.	Nº.	C(H)B	η_{LB}	A.Q.
P	0,984	0,620	A	P38	0,720	0,080	I
P13	0,850	0,602	A	P39	0,541	0,568	I
P14	0,867	0,568	A	P40	0,668	0,552	I
P25	0,594	0,685	I	P41	0,675	0,238	I
P26	0,807	0,604	I	P42	0,899	0,616	A
P27	0,742	0,254	I	P43	0,651	0,278	I
P28	1,049	-0,078	I	P44	0,659	0,881	A
P29	1,054	0,609	A	P45	0,248	0,492	I
P30	0,946	0,544	I	P46	0,854	0,293	I
P31	0,890	0,030	I	P47	0,632	0,873	A
P32	0,872	0,571	A	P48	0,234	0,483	I
P33	0,861	0,584	A	P49	0,996	0,345	I
P34	0,934	0,556	A	P50	0,360	0,645	I
P35	0,927	0,590	A	P51	0,552	0,474	I
P36	0,876	0,321	I	P52	0,491	0,620	I
P37	0,715	0,551	A				

Tabela 5.1: Descritores η_L e CH(rB) calculados com o PM3. A.Q. indica a atividade biológica experimental apresentada de forma qualitativa.

Nos propomos agora a utilizar a regra de classificação de atividade construída na primeira parte para estudar o grupo **G2**, o qual é composto pelas moléculas que ainda não foram testadas, ou cuja informação experimental de atividade fotodinâmica não foi relatada na literatura.

Apresentamos na **tabela 5.2** os valores dos parâmetros eletrônicos para as moléculas do grupo **G2** e a proposição de atividade, notando que das 22 (vinte e duas) moléculas não testadas a regra prevê que 15 (quinze) seriam ativas.

Nº.	η_{LB}	C(H)B	A.P.	Nº.	η_{LB}	C(H)B	A.P.
P1	0,107	0,907	I	P12	0,613	1,023	A
P2	-0,152	0,611	I	P15	0,470	0,782	I
P3	0,869	0,803	A	P16	0,560	0,796	A
P4	0,550	0,428	I	P17	0,445	0,882	I
P5	0,625	0,920	A	P18	0,559	0,698	A
P6	0,635	0,965	A	P19	0,564	0,936	A
P7	0,625	1,038	A	P20	0,572	0,923	A
P8	0,441	1,017	I	P21	0,578	0,984	A
P9	0,619	0,855	A	P22	0,566	0,994	A
P10	0,624	0,926	A	P23	0,575	0,924	A
P11	0,596	0,926	A	P24	0,341	0,748	I

Tabela 5.2: Descritores η_L e CH(rB) calculados com o PM3. A.P. indica a atividade biológica qualitativa proposta através da metodologia MIE.

Na próxima seção investigamos a correlação da atividade biológica experimental e parâmetros teóricos através das técnicas de análise de componentes principais (PCA) e análise hierárquica de clusters (HCA).

5.2.3 – Investigação da Atividade Biológica através da Análise de Componentes Principais (PCA) e da Análise Hierárquica de Clusters (HCA)

Nesta seção, reunindo dentro da mesma análise um número maior de parâmetros, utilizamos as metodologias PCA e HCA para o estudo da correlação entre a atividade biológica e descritores teóricos. Dentro da quimiometria a metodologia PCA é classificada como uma técnica exploratória e apresenta limitações em estudos de previsão, porém iremos utilizar essa técnica para investigar o comportamento de atividade do grupo **G2** frente às características do grupo de compostos já conhecido.

Em nosso estudo temos o intuito de avaliar a semelhança em atividade das moléculas do grupo **G1** e do grupo **G2**, projetando no espaço de componentes principais PC1 x PC2 construído para os parâmetros selecionados pelo **G1** os compostos do grupo **G2**.

Inicialmente selecionamos 42 descritores, que incluem os descritores da MIE para as cinco regiões do esqueleto molecular estudadas. Entre outros, os descritores envolvem:

- a energia do HOMO (Highest Occupied Molecular Orbital) e HOMO-1;
- a diferença de energia entre HOMO e HOMO-1 (ΔH);
- a energia do LUMO (Lowest Unoccupied Molecular Orbital energy) e LUMO+1;
- a diferença de energia entre LUMO e LUMO+1 (ΔL);

- a contribuição para a densidade local de estados (LDOS) para as regiões selecionadas, na formação dos orbitais HOMO e HOMO-1 (**CH** e **CH-1**) e suas diferenças ($\eta\mathbf{H}$);
- a contribuição para a densidade local de estados (LDOS) para as regiões selecionadas, na formação dos orbitais LUMO e LUMO+1 (**CL** e **CL+1**) e suas diferenças ($\eta\mathbf{L}$);
- o valor do momento de dipolo (**DM**);
- o calor de formação (**HF**);
- refratividade (**R**),
- polarizabilidade (**Po**),
- massa (**M**),
- energia de hidratação (**HE**);
- energia de solvatação (**SE**),
- coeficiente de partição octanol-água (**log P**).

Durante a aplicação do PCA, grande parte dos descritores inicialmente considerados são descartados, restando apenas os poucos capazes de separar as amostras em subgrupos de acordo com os dados experimentais de controle. A seleção dos “melhores” descritores é feita visualmente através dos resultados obtidos para os gráficos de escore.

Considerando a atividade fotodinâmica dos compostos, e inúmeras combinações dos parâmetros acima, foi possível separar os psoralenos em um subconjunto na avaliação da PCA. De todas as combinações de descritores obtidas na classificação dos compostos, apresentamos o conjunto que exemplifica os nossos resultados, o qual denominamos “Conjunto A”:

▲ **Conjunto A** - inclui os descritores:

- **CH(rB)** - a contribuição dos átomos da região **rB** para a formação do orbital HOMO;
- **CL(rB)** – a contribuição dos átomos da região **rB** para a formação do orbital LUMO;
- $\eta\mathbf{H}(\mathbf{rB})$ – diferença entre LDOS para os orbitais de fronteira HOMO e HOMO-1 (**CH** and **CH-1**) na região **rB**; e,
- $\eta\mathbf{L}(\mathbf{rB})$ – diferença entre LDOS para os orbitais de fronteira LUMO e LUMO + 1 (**CL** and **CL+1**) na região **rB**.

Observamos que entre os parâmetros responsáveis pelos melhores resultados na PCA estão também os descritores eletrônicos da MIE que produziram a melhor classificação de atividade dando origem a regra de separação obtida na seção 5.2.2.

A **Figura 5.3** apresenta para o grupo **G1** o resultado obtido para a metodologia PCA, onde mostramos o gráfico de separação dos compostos quanto à atividade biológica, segundo as variáveis do conjunto A. A notação que iremos usar em todas as figuras envolvendo a metodologia PCA serão os símbolos triângulo (**▲**), círculo (**○**) e cruz (**+**) representando ativo, inativo e não testado/sem informação, respectivamente.

As moléculas estão representadas por pontos em um plano de duas componentes principais (PC1 e PC2). Os descritores utilizados foram auto-escalados antes da análise. Para o conjunto A o grupo de moléculas ativas está localizado na parte esquerda superior do gráfico e o grupo de inativas está distribuído de forma mais esparsa pelo gráfico. De forma geral, ambos os eixos PC1 e PC2 são responsáveis pela separação, já que não há uma separação em grupos que se distanciam em linhas bem definidas (vertical ou horizontal).

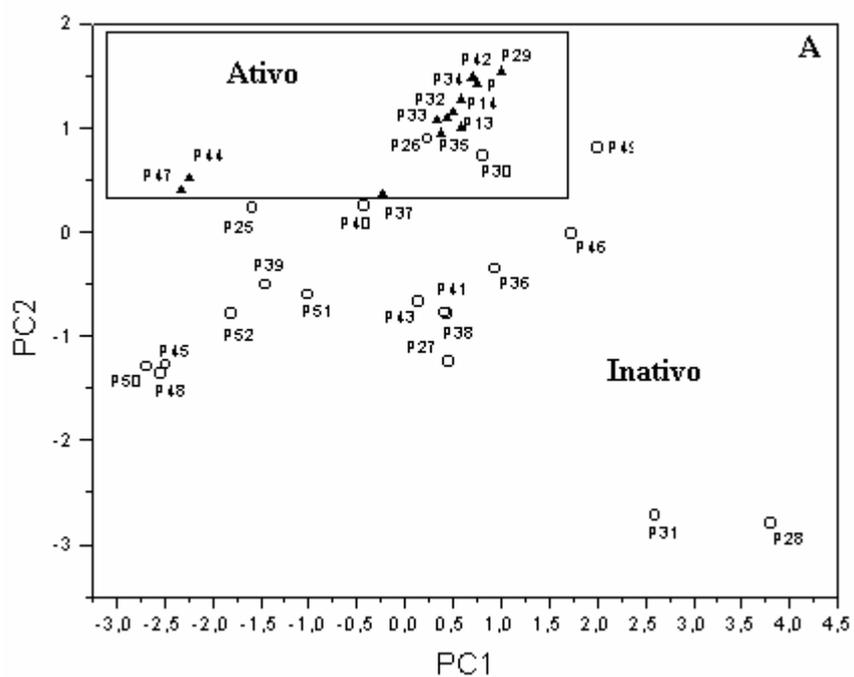


Figura 5.3 – Resultado da análise PCA. Gráfico de distribuição dos compostos no plano PC1 x PC2, segundo as variáveis do conjunto A.

Para o resultado apresentado, os eixos PC1 e PC2, que acumulam 92% da variância do sistema, são escritos em termos dos descritores teóricos segundo as equações:

▲ **Conjunto A:**

$$\text{PC1} = 0,52 \eta\mathbf{H}(\mathbf{rB}) - 0,44 \eta\mathbf{L}(\mathbf{rB}) + 0,55 \mathbf{CH}(\mathbf{rB}) + 0,48 \mathbf{CL}(\mathbf{rB}) \quad (1)$$

$$\text{PC2} = 0,49 \eta\mathbf{H}(\mathbf{rB}) + 0,56 \eta\mathbf{L}(\mathbf{rB}) + 0,44 \mathbf{CH}(\mathbf{rB}) - 0,50 \mathbf{CL}(\mathbf{rB}) \quad (2)$$

As equações mostram que há uma maior contribuição de **CH(rB)** para PC1 e $\eta\mathbf{L}(\mathbf{rB})$ para PC2.

Esta nova apresentação de resultados, obtidos por uma metodologia classicamente utilizada no estudo de reconhecimento de padrões, e que inclui as variáveis eletrônicas na investigação de atividade dos psoralenos, indica novamente a relevância dos parâmetros eletrônicos teóricos. Observamos nos resultados o envolvimento da região **B** que segundo as investigações experimentais está, de modo geral, envolvida na foto reação que ocorre com a participação da dupla ligação entre os carbonos 4',5' do anel de furano.

Na análise com a PCA, as moléculas P26 e P30 representam erros na classificação, sendo incorretamente classificadas como ativas, o que representa um índice de acerto de 93% (noventa e três por cento). Notemos que estas moléculas também apareceram incorretamente classificadas na análise com a MIE.

Na **Figura 5.4** apresentamos o resultado da HCA, construído com base na distribuição dos compostos no espaço das componentes principais, que nos mostra como estes grupos se aglutinam e a similaridade existente entre eles.

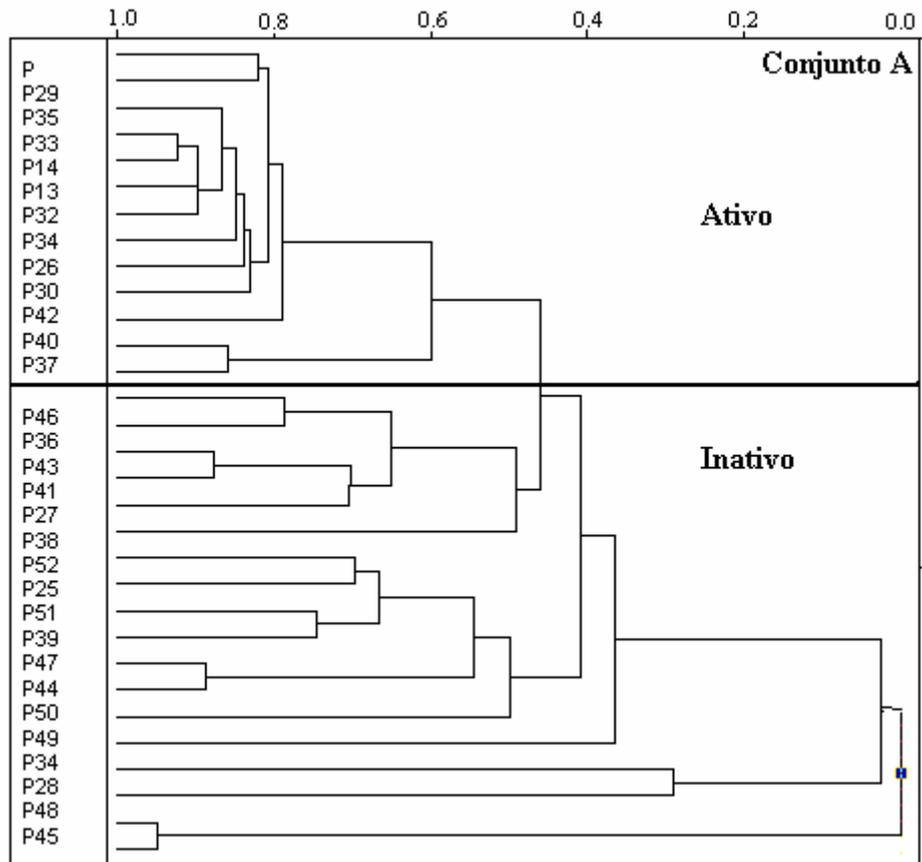


Figura 5.4 – Resultado da análise HCA. Gráfico de separação dos compostos em agrupamentos, segundo as variáveis do conjunto A.

As moléculas ativas foram separadas das inativas e agrupadas em um subgrupo distinto no gráfico. A formação de dois grupos distintos, um para as moléculas ativas e outro para as inativas, demonstra que os descritores adotados na classificação de atividade foram capazes de avaliar corretamente a similaridade que deve existir entre os compostos ativos e entre os inativos. Os dois clusters ativo e inativo estão separados por uma linha horizontal. Na análise HCA temos, de acordo com o conjunto avaliado, as seguintes moléculas em clusters distintos de sua atividade experimental: P26, P30, P40, P44 e P47.

Apresentaremos agora o resultado da metodologia PCA, para os compostos do grupo **G2**, quando os projetamos no espaço construído para os parâmetros selecionados pelo grupo **G1**. Gostaríamos neste estudo de avaliar a semelhança em atividade das moléculas do grupo **G1** e do grupo **G2**.

Projetamos os novos compostos no conjunto A de resultados PCA e apresentamos na **Figura 5.5** o resultado da projeção das 22 novas moléculas nos espaços das componentes principais construído como combinação das variáveis que foram capazes de fazer a separação em atividade do grupo **G1** isoladamente.

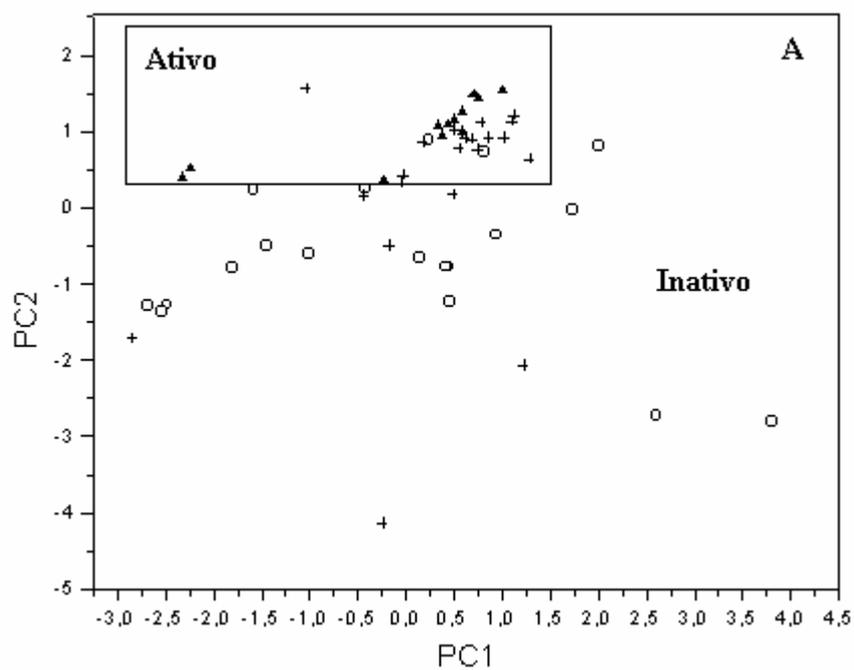


Figura 5.5 - Resultados da análise de PCA. Gráfico de incluindo todos os compostos e proposição de atividade. Onde ▲ representa ativo, ○ representa inativo e + representando não testado/sem informação.

Observamos da figura acima que os compostos do grupo **G2** se misturam com os do grupo **G1**, mostrando uma maior afinidade com o grupo das moléculas ativas, região do gráfico onde se concentram em média 15 das 22 moléculas das quais não temos informação experimental. Aparecem como inativas as moléculas P1, P2, P4, P17, P18 e P24. As demais ocupam sempre a região das ativas. Sendo que alguns compostos estão no limite entre as duas regiões.

Os eixos PC1 e PC2 acima são construídos de acordo com as equações:

▲ **Conjunto A:**

$$\mathbf{PC1} = 0,64 \eta\mathbf{H}(\mathbf{rB}) - 0,26 \eta\mathbf{L}(\mathbf{rB}) + 0,66 \mathbf{CH}(\mathbf{rB}) + 0,29 \mathbf{CL}(\mathbf{rB}) \quad (7)$$

$$\mathbf{PC2} = 0,31 \eta\mathbf{H}(\mathbf{rB}) + 0,66 \eta\mathbf{L}(\mathbf{rB}) + 0,24 \mathbf{CH}(\mathbf{rB}) - 0,64 \mathbf{CL}(\mathbf{rB}) \quad (8)$$

Observamos que há alteração das contribuições de cada variável para a formação dos componentes principais com a presença de novas moléculas, isto ocorre uma vez que o sistema é autoescalado após a inclusão de novos compostos de forma a se uniformizar os valores dos diferentes descritores, evitando uma análise que favoreça algum descritor em particular. No entanto, temos ainda maior contribuição de **CH(rB)** para PC1 e $\eta\mathbf{L}(\mathbf{rB})$ para PC2, observamos ainda que a contribuição das variáveis que envolvem LUMO e LUMO + 1 decrescem em PC1 e crescem em PC2.

5.3– Referências

- [1] Potapenko A.Y.. *Journal Photoch. Photobiol B*, 9: (1) 1-33, 1991.
- [2] Bordin, F., Dall 'Acqua, F. and Guiotto, A... *Pharmac. Ther.*, Vol. 52, pp 331-363, 1991.
- [3] Towers G.H.N., Page J.E., Hudson J.B.. *Current Organic Chemistry*; 1: (4) 395-414, 1997.
- [4] Ebermann, R., Alth G., Kreitner M. and Kubin A.. *Journal of Photochemistry and Photobiology B-Biology*, 36: (2) 95-97, 1996.
- [5] Edenharder R., Speth C., Decker M., Kolodziej H., Kayser O., Platt K.L.. *Mutation Research-Genetic Toxicology*; 345: (1-2) 57-71, 1995.
- [6] Neal J.J., WU D.. *Pesticide Biochemistry and Physiology*, 50: (1) 43-50, 1994.
- [7] Miolo G., Tomanin R., Derossi A., Dall 'Acqua F., Zacchello F., Scarpa M.. *Journal of Photochemistry and Photobiology B-Biology*, 26: (3) 241-247, 1994.
- [8] Zhao J.F., Zhang Y.J., Jin X.H., Athar M., Santella R.M., Bickers D.R., Wang Z.Y.. *Journal of Investigative Dermatology*, 113: (6) 1070-1075, 1999.
- [9] Hyperchem TM 5.1. 1997 Hypercube.
- [10] MOPAC program, version 6.0, Quantum Chemistry Exchange No. 455.
- [11] CACHE 5.0, Fujitsu Limited, Chiba City, Chiba 2618588, Japan (www.CACheSoftware.com)
- [12] Cyrillo, M., Galvão, D. S.. *J. Mol. Struct. (THEOCHEM)* 1999, 464, 267-272;
Cyrillo, M., Galvão, D. S., Chem2Pac: a Computational Chemistry Integrator for Windows. *EPA Newsletter* 1999, 67, 31-38.
<http://www.ifi.unicamp.br/qsonm/chem2pac>.

CAPÍTULO 6

CONCLUSÕES FINAIS

6.1. Conclusões finais

Neste trabalho, nos dedicamos a investigar sistematicamente propriedades teóricas (eletrônicas) de conjuntos de compostos pertencentes às famílias das elipticinas e das furanocumarinas com o objetivo de determinar regras e padrões indicadores de uma correlação entre a atividade biológica destes compostos e características intrínsecas de sua estrutura molecular. A importância deste estudo é evidenciada pela observação de que pequenas variações na estrutura química de compostos de uma família são responsáveis por grandes mudanças em sua atividade biológica. Apesar da complexidade envolvida no processo de interação composto-alvo biológico e em sua rota metabólica no organismo, as propriedades moleculares são fundamentais para a definição do caráter de atividade que ele apresenta. Por este motivo, investigações de atividade biológica, mesmo que qualitativa, baseada em fatores que dependam diretamente da estrutura e dos átomos constituintes de um composto, representam um grande avanço na formulação/seleção de novos compostos teste.

Estudos de relações qualitativas de atividade baseados em parâmetros teóricos são freqüentemente identificados como estudos de estrutura-atividade (SAR - *Structure-Activity Relationship*) e consideram a atividade biológica através de índices experimentais de atividade. Para nosso estudo, consideramos uma análise “binária” de atividade, onde ativo ou inativo são as classificações experimentais básicas para a investigação.

Dentre os vários métodos de investigação SAR conhecidos, aplicamos em nossas análises a Metodologia de Índices Eletrônicos (MIE), a Análise de

Componentes Principais (PCA - *Principal Component Analysis*), a Análise Hierárquica de Clusters (HCA – *Hierachical Cluster Analysis*).

As famílias estudadas foram divididas em dois sub-grupos, o primeiro com os compostos para os quais estão disponíveis informações experimentais sobre a atividade biológica (**G1**), e o segundo, com compostos não testados ou sem informação experimental disponível (**G2**), até a realização deste trabalho.

Para o grupo **G1** das elipticinas, nossos resultados mostram que os descritores eletrônicos ΔH (diferença de energia entre os orbitais de fronteira HOMO e HOMO-1), **CL(rD)** e **CH(rB)** (relacionados com a distribuição dos orbitais LUMO e HOMO sobre átomos isolados do anel exposto do esqueleto molecular) aplicados na análise da MIE mostraram maior capacidade de seleção, reproduzindo com 92% de acerto a classificação experimental de atividade do conjunto das 25 moléculas estudadas, enquanto que o momento de dipolo (**DM**) e o potencial eletrostático molecular (MEP) reproduziram as mesmas informações com 80% de acerto.

A importância dos descritores eletrônicos é reforçada pela análise através das metodologias de reconhecimento de padrões Análise de Componentes Principais (PCA) e Análise Hierárquica de Agrupamentos (HCA). Mostramos que a PCA seleciona os descritores ΔH (diferença de energia entre os orbitais de fronteira HOMO e HOMO-1), **CL(rD)** e **CH(rB)** (relacionados com a distribuição dos orbitais LUMO e HOMO sobre átomos isolados do anel exposto do esqueleto molecular), derivados da Metodologia de Índices Eletrônicos(MIE) e o **DM** selecionado na primeira parte deste estudo e em trabalhos anteriores. As técnicas PCA e HCA classificam corretamente 84% e 80% das moléculas, respectivamente.

Com as regras e padrões obtidos para o grupo **G1** sugerimos o comportamento de atividade biológica para os compostos pertencentes ao grupo **G2**.

Combinando o resultado dos três diferentes métodos propomos que 9 das 15 moléculas sejam ativas. Embora a PCA e HCA não sejam métodos classificatórios, eles são úteis para uma análise independente dos resultados da MIE para as moléculas não testadas. Além disso, resultados preliminares para o grupo das moléculas não testadas obtidos com análise de Redes Neurais Artificiais, dão suporte às conclusões do presente trabalho.

Para o grupo **G1** dos psoralenos, nossos resultados mostram que os descritores eletrônicos η_L (diferença de energia entre os orbitais LUMO e LUMO+1) e **CH(rB)** (contribuição dos átomos da região **rB** para a formação do orbital HOMO), também aplicados na análise da MIE, mostraram maior capacidade de seleção definindo a região na qual as moléculas são ativas, reproduzindo com 90% de acerto a classificação experimental de atividade do conjunto das 31 moléculas estudadas nesse grupo.

A investigação com os métodos PCA e HCA novamente mostraram a relevância dos resultados obtidos com a MIE, selecionando dentre dezenas de descritores os eletrônicos como importantes na análise de atividade biológica.

A PCA selecionou os descritores **CH(rB)** (contribuição dos átomos da região **rB** para a formação do orbital HOMO), **CL(rB)** (contribuição dos átomos da região **rB** para a formação do orbital LUMO), $\eta_H(rB)$ (diferença entre LDOS para os orbitais de fronteira HOMO e HOMO-1 na região **rB**), e, $\eta_L(rB)$ (diferença entre LDOS para os orbitais de fronteira LUMO e LUMO + 1 na região **rB**) com 93% de acerto na classificação.

Com as regras e padrões obtidos para o **G1** estudamos e sugerimos o comportamento de atividade biológica para os compostos que compõem o **G2**. Combinando o resultado da MIE e PCA propomos que 15 das 22 moléculas sejam ativas.

Como conclusão desta investigação nós propomos que a MIE, com suas regras simplificadas envolvendo apenas 2 parâmetros, pode ser efetivamente aplicada como guia “pré-síntese” nas investigações de novos compostos. Isto é, as informações fornecidas pelos parâmetros da MIE podem ser utilizadas orientando a escolha de novos compostos a serem sintetizados. Os novos compostos podem ser selecionados priorizando-se substituições químicas específicas, que mantenham/elevem as características eletrônicas requeridas pelos padrões (das variáveis eletrônicas) observados para as moléculas ativas.

A utilização conjunta dos resultados obtidos com as metodologias MIE, PCA e HCA tornam-se uma ferramenta ainda mais poderosa no campo do planejamento computacional. A investigação com os métodos PCA e HCA mais uma vez refletiram a confiabilidade dos resultados obtidos com a MIE, apresentando ótimos índices de acerto na reprodução dos resultados experimentais e selecionando dentre dezenas de descritores os eletrônicos como importantes na análise de atividade biológica.