

Universidade Estadual de Campinas
Instituto de Física Gleb Wataghin
Departamento de Física Aplicada

Dissertação de Mestrado

**ESTUDO ESTRUTURA-ATIVIDADE DA
COMBRETASTATINA E DERIVADOS**

Marcelo Amorim Marchiori

Orientador: Prof. Dr. Douglas Soares Galvão

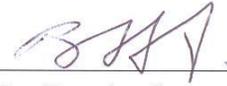
Co-Orientadora: Dra. Scheila Furtado Braga Llanes

Campinas, Setembro de 2007

Estudo Estrutura-Atividade da Combretastatina e Derivados

Este exemplar corresponde à redação final da tese devidamente corrigida e defendida por **Marcelo Amorim Marchiori** e aprovada pela comissão julgadora.

Campinas, 17 de março de 2008.



Prof. Dr. Douglas Soares Galvão

Banca examinadora:

Prof. Dr. Douglas Soares Galvão IFGW/UNICAMP.

Prof. Dr. Profa. Iris Torriani IFGW/UNICAMP.

Prof. Dr. David Lima Azevedo DEP. FÍSICA/UFMA.

Dissertação apresentada ao Instituto de Física Gleb Wataghin, UNICAMP como requisito parcial para obtenção do título de **Mestre em Física**.

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO IFGW - UNICAMP

M332e

Marchiori, Marcelo Amorim
Estudo estrutura-atividade da combretastatina
e derivados / Marcelo Amorim Marchiori. --
Campinas, SP : [s.n.], 2007.

Orientador: Douglas Soares Galvão.
Dissertação (mestrado) - Universidade Estadual de
Campinas, Instituto de Física "Gleb Wataghin".

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

- **Título em inglês:** Structure-activity study of combretastatin and derivatives
- **Palavras-chave em inglês (Keywords):**
 1. Structure-activity relationship (Biochemistry)
 2. Quantum chemistry
 3. Electronic structure
- **Área de concentração:** Estrutura Eletrônica de Átomos e Moléculas; Teoria
- **Titulação:** Mestre em física
- **Banca examinadora:**
Prof. Douglas Soares Galvão
Prof. David Lima Azevedo
Prof^a. Íris Concepción Linares de Torriani
- **Data da defesa:** 11.09.2007



MEMBROS DA COMISSÃO JULGADORA DA TESE DE MESTRADO DE **MARCELO AMORIM MARCHIORI – RA 049323** APRESENTADA E APROVADA AO INSTITUTO DE FÍSICA “GLEB WATAGHIN”, DA UNIVERSIDADE ESTADUAL DE CAMPINAS, EM 11 / 09 / 2007.

COMISSÃO JULGADORA:

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

Prof. Dr. David Lima Azevedo
DF/UFMA

Profa. Dra. Iris Concepción Linares de Torriani
DFMC/IFGW/UNICAMP

Não pude sofrer por ti, amigo,
mas o traço que te dedico
tem a marca do meu semblante,
do meu pezaroso sorriso,
da dor que insiste em me derrubar.
Se, em minhas mãos,
coubessem infinitas palavras,
existiria, ainda, nesses versos uma lacuna,
uma vazia preenchido pelo silêncio
e pela falta que você nos faz.
O tempo agora é seu
e teus olhos miram o que te faz maior.
Se o vento te alcançar com estes versos, saiba:
A bondade tem seu nome
e a saudade o nosso, sempre.

*Para ti, Edinaldo.
(in memoriam)*

Agradecimentos

Agradeço

ao professor Douglas, pela oportunidade e confiança;

à minha Co-orientadora Scheila, por ter sido muito mais que isto;

aos meus pais, Heloísio e Josiane, e aos meus irmãos Daniel e Ricardo, pelo carinho, apoio e todos os sacrifícios feitos;

à minha esposa Gláucia, minha amiga e amada;

a todos da minha família que me apoiaram e ajudaram nas minhas escolhas;

à minha nova família, Sr. Geraldo (*in memoriam*), D. Maria José, Sr. João, minha cunhada Lela e minha sobrinha Ariele (Eu vou te pegar, hein!!!), por me fazerem sentir em casa;

aos amigos do peito e, ou melhor, irmãos por opção, Bruno Mendonça, André, Paulinho, Henrique e Ygor, por tanto tempo juntos;

aos amigos do GSONM, Sato, Gustavo, Pedro, Vitor, Karla e Bruno (ex membro), pela ajuda e discussões;

aos meus eternos amigos do curso de Física da Universidade Federal de Viçosa;

aos novos amigos do IFGW e, em especial, ao pessoal da minha sala: Sérgio, Vladimir, André e Diego;

e a todos que, de alguma forma contribuíram para o cumprimento de mais esta etapa.

Meus sinceros agradecimentos aos membros da Banca Examinadora, Prof. David Lima Azevedo e Prof. Iris Torriani pela disponibilidade e boa vontade de avaliação deste trabalho; A todo mundo da CPG pelo apoio fundamental dado aos alunos da pós-graduação do IFGW; e à Capes pelo apoio financeiro.

Resumo

A Combretastatina é um estilbeno isolado na década de 80, e vem sendo amplamente estudada pela indústria farmacêutica devido à sua promissora ação anticarcinogênica. Como fármaco anticarcinogênico, age interrompendo o ciclo de polimerização e despolimerização dos microtúbulos, componente celular extremamente importante para a motilidade, manutenção estrutural e mitose celular. Sua principal forma de atuação consiste em despolimerizar os microtúbulos estáveis das células endoteliais da vasculatura tumoral, levando ao bloqueio do fluxo sanguíneo que alimenta os tumores cancerígenos. Uma das grandes vantagens da Combretastatina, em relação aos demais medicamentos antineoplásicos, é o fato de não levar à resistência medicamentosa no tratamento quimioterápico.

Investigamos a estrutura da Combretastatina e 17 derivados por meio de métodos semiempíricos e estudamos a relação entre as propriedades teóricas e a atividade experimental destes compostos, utilizando três metodologias de reconhecimento de padrões: a Metodologia de Índices Eletrônicos (MIE), a Análise de Componentes Principais (PCA) e a Análise Hierárquica de Clusters (HCA).

Para cada metodologia construímos regras e padrões, permitindo a classificação dos compostos em ativos e inativos, a partir das propriedades calculadas teoricamente. Os resultados das três metodologias confirmam a aplicabilidade da MIE e reforçam a importância das variáveis eletrônicas para a classificação da atividade biológica das Combretastatinas.

Abstract

Combretastatin, a stilbene isolated in 80's, has been widely studied by the pharmaceutical industry due to its promising anticarcinogenic action. As an antineoplastic agent it acts interrupting the polymerization-depolymerization cycle of the microtubules, an important cellular component to motility, structural maintenance and cellular mitosis. Its main feature consists in dissociate the microtubules in endothelial cells of the tumoral vascular system, leading to disruption of the blood flow that feeds the carcinomas. One of the great advantages of Combretastatin, when compared with others compounds, is the fact that it does not lead to drug resistance in chemotherapy treatments.

We investigated the structure of Combretastatin and 17 derivatives using semiempirical methods. We performed the study of the relationship between theoretical properties and experimental activity of these molecules using three pattern recognition methodologies: Electronic Index Methodology (EIM), Principal Component Analysis (PCA) and Hierarchical Clusters Analysis (HCA).

For each methodology we found rules and patterns capable of classifying our molecules into active or inactive, using the properties theoretically calculated. The results obtained from the three methodologies confirm the applicability of the EIM and reinforce the importance of the electronic variables for the classification of the biological activity of Combretastatins.

Índice

1	Introdução	1
1.1	O Câncer	2
1.2	Divisão Celular	3
1.3	A Combretastatina	7
1.3.1	Combretastatina e derivados	10
2	Metodologia	13
2.1	Teoria de Campo Auto-Consistente	14
2.1.1	Aproximação de Hartree-Fock-Roothaan	14
2.1.2	Equações de Roothaan	21
2.1.3	A densidade de carga	22
2.2	Métodos Semi-empíricos	26
2.2.1	Aproximações NDO	26
2.2.2	Aproximação MNDO	27
2.2.3	Austin Method 1- AM1	31
2.2.4	Parametric Method 3-PM3	31
2.3	Métodologias de Análise Multivariada	32
2.3.1	Análise de Componentes Principais	33
2.3.2	Resumo sobre a utilização do PCA	39
2.3.3	Análise Hierárquica de Clusters- HCA	40
2.3.4	Esboço sobre o algoritmo do HCA	42
2.4	A Metodologia dos Índices Eletrônicos	43

3	Resultados sobre Investigação Estrutura-Atividade da Combretastatina e derivados	49
3.1	Estudo da geometria molecular	49
3.1.1	Dados experimentais para um derivado da Combretastatina	50
3.1.2	O cristal de C_{rx}	51
3.1.3	Avaliação dos métodos semi-empíricos.	52
3.2	Metodologia de Otimização das moléculas	55
3.2.1	Procedimentos computacionais	56
3.3	Estudo da estrutura-atividade	58
3.3.1	A MIE aplicada a Combretastatina e derivados	58
3.3.2	Investigação da Atividade Biológica da Combretastatina e Derivados através das técnicas: PCA e HCA	62
3.4	Conclusões e perspectivas	67

Capítulo 1

Introdução

Os grandes avanços tecnológicos dos dias atuais propiciam uma visão esperançosa no que diz respeito à possibilidade de tratamentos mais eficientes e, até mesmo, de cura de muitas doenças que afetam a humanidade. Essa esperança baseia-se no êxito da síntese de novos fármacos, que atuam cada vez com maior eficiência no combate à patogenia. Aliado a isto, os estudos teóricos sobre as propriedades físicas destes compostos têm dado luz a questões fundamentais para a compreensão dos mecanismos físicos, químicos e bioquímicos que conferem eficiência no combate a muitas doenças. Na verdade, os estudos teóricos são fundamentais também do ponto de vista econômico. Se analisarmos o processo de pesquisa de novos fármacos, veremos que os testes biológicos, *in vivo* e *in vitro*, são extremamente custosos, tanto, no que diz respeito ao tempo despendido, como em termos de recursos alocados nos protocolos de testes. Para a realização destes testes, é necessário, de modo geral, que, primeiramente, tanto o composto a ser utilizado como fármaco, como a causa da patologia, sejam isolados. Geralmente, a obtenção destes compostos é feita a partir de trabalhos de campo (composto natural), ou da síntese em laboratório (composto sintético). Daí, verifica-se experimentalmente, a evolução da doença sob a influência do medicamento teste. Para que os resultados sejam confiáveis esta verificação experimental deve ser repetida até que os dados estatísticos sejam satisfatórios para determinação de um resultado seguro sobre a ação da droga. O alto valor dos novos medicamentos, reflete o custo referente às pesquisas dos laboratórios farmacêuticos. Logo, a compreensão dos mecanismos responsáveis pela eficácia de um determinado composto, pode introduzir nestas pesquisas um caráter preditivo, uma espécie de regra de seleção de candidatos a bons medicamentos. No caso específico dos cânceres, os comentários se aplicam. O processo de pesquisa, desde a obtenção dos compostos,

até os testes biológicos, geram custos bastante elevados. Além disso, devemos ressaltar que o grande número de diferentes tumores representa uma dificuldade extra para a elaboração de novos medicamentos. O fato de que uma determinada droga, por exemplo, não inibe o crescimento de um tumor de mama, não significa que ela não seja eficiente no combate a um tumor de próstata. Assim, custos extras podem ser gerados e um poder preditivo sobre a eficiência de novas drogas seria de grande valia tanto para a indústria farmacêutica como para a população em geral, pois, como resultado da redução dos custos de pesquisa para os laboratórios, surgiria a redução do valor final do medicamento no mercado. Portanto, estabelecer uma correlação da atividade biológica com indicadores que permitam selecionar, *a priori*, quais os compostos mais promissores e, também, descartar os mais improváveis, seria uma contribuição importante para o desenvolvimento de novos, e melhores, compostos para uso farmacológico.

1.1 O Câncer

Câncer, neoplasia ou tumor maligno são os nomes comumente atribuídos a uma classe de doenças caracterizadas, de modo geral, por um descontrole na divisão de células aberrantes dentro do organismo. Não é possível determinar em que período da história da humanidade o câncer surgiu, mas acredita-se que esta seja uma das doenças mais antigas a acompanhar o homem. A deficiência nos diagnósticos em períodos passados, impossibilita a determinação das origens históricas da doença. Nos dias atuais, com a melhoria dos métodos de diagnóstico, verifica-se que boa parcela da população mundial sofre com algum tipo de câncer. Por exemplo, segundo dados do Instituto Nacional de Câncer (INCA), estima-se que em 2006, aproximadamente 470.000 pessoas no território brasileiro tenham desenvolvido algum tipo de neoplasia [1]. Diante de dados tão alarmantes o estudo de novas drogas e métodos que possibilitem uma maior eficácia no controle ou na eliminação destes tumores é extremamente relevante.

A comunidade científica ainda não identificou os mecanismos de surgimento da doença, mas há um consenso de que além de fatores genéticos hereditários, a exposição excessiva à radiação, o tabagismo, a alimentação com excesso de gordura, entre outros, contribuem significativamente para o surgimento de um câncer[1, 2]. A diferenciação entre os tipos de câncer dá-se através do tipo de célula afetada. Desta forma, um mesmo órgão pode sofrer com diversos tipos de tumores se este for formado por mais de uma célula. Outras características que diferenciam os diversos tipos de câncer são a velocidade de multiplicação e seu poder de invadir órgãos e tecidos vizinhos

ou distantes (metástase)[3].

Alguma alteração, que ainda não se sabe a natureza, ocorre no DNA celular e a partir daí ela passa a não responder aos comandos do organismo, dividindo-se indiscriminadamente. As células resultantes dessa divisão descontrolada, em muitos dos casos, aglomeram-se formando uma massa de células aberrantes, o tumor. O tumor cria uma vasculatura própria e se alimenta dos nutrientes destinados, inicialmente, às células saudáveis, debilitando significativamente o organismo. Em linhas gerais, as células de um câncer são parasitas vorazes. O grande risco, ocorre quando estas células atingem a corrente sanguínea, disseminando-se por todo o corpo e debilitando ainda mais o organismo.



Figura 1.1: Células cancerígenas que não respondem aos comandos que finalizam o processo de mitose [4].

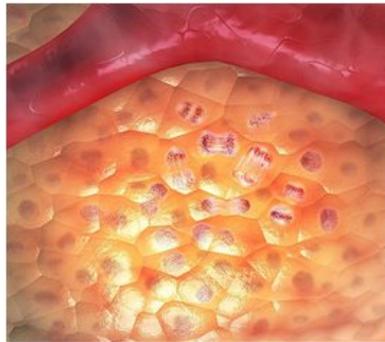


Figura 1.2: A divisão celular torna-se descontrolada[4].

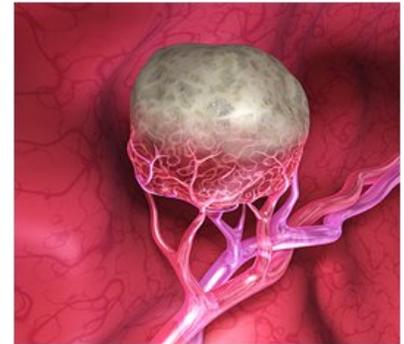


Figura 1.3: Tumor sólido formado por aglomeração celular[4].

Para que possamos compreender o processo de surgimento de um câncer e os mecanismos de atuação da Combretastatina, analisaremos com mais detalhes o ciclo de divisão celular.

1.2 Divisão Celular

A maioria das células eucariotas provém de uma série de eventos ordenados que constituem o **ciclo celular**, durante o qual seus cromossomos são duplicados e separados para gerar duas novas células. A regulação desse ciclo é extremamente importante para o desenvolvimento normal de um organismo multicelular. A perda de controle pode, no fim das contas, levar ao surgimento de um câncer. Podemos dizer que são, basicamente, duas as fases críticas no processo de divisão celular: a duplicação do DNA e a mitose.

Simplificadamente, podemos dividir a seqüência de eventos que conduzem à mitose em quatro fases, como mostra a Figura 1.4.

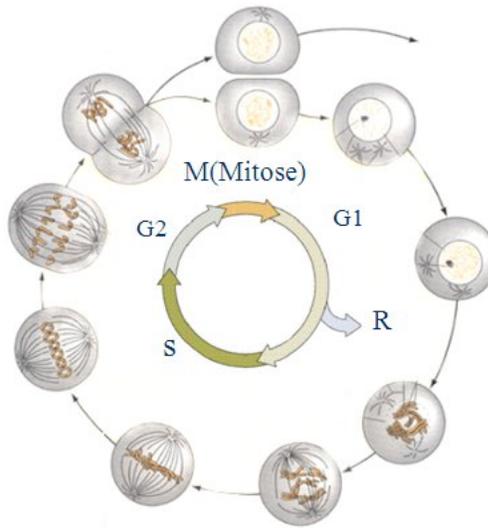


Figura 1.4: Etapas da divisão celular.

A divisão celular inicia-se no estágio **G1**, fase na qual são emitidos sinais (“Sim” ou “Não”) para que o processo continue ou seja interrompido. Estes sinais são manifestações de determinadas proteínas e é, através destes sinais, emitidos no sub-estágio de averiguação **R**, que a célula atesta se está apta para passar à etapa seguinte, a fase **S**, onde o DNA é duplicado. Para garantir que não haja danos na duplicação, a célula passa pela fase **G2**, na qual é feita a análise de todo o material genético. Se a cópia não for exata, são liberadas proteínas supressoras que bloqueiam a passagem ao estágio final, a mitose **M**, fase em que duas células idênticas são formadas[3].

Quando as etapas de verificação não ocorrem, células defeituosas podem surgir do processo de divisão celular. Se este processo acontecer sistematicamente, estas células passam a se dividir sem que o organismo necessite, aglomerando-se numa estrutura massiva compostas de células aberrantes, o tumor. As células de um tumor benigno, não se espalham por outras partes do corpo e, por isso, raramente significam uma ameaça à vida. Já as células dos tumores malignos apresentam anomalias em seus cromossomos e se dividem descontroladamente, invadindo tecidos e órgãos vizinhos e, em alguns casos, alcançando a corrente sanguínea e, enfim, disseminando-se por todo o corpo, num processo denominado metástase.

Lancemos, neste ponto, um olhar um pouco mais cuidadoso sobre um componente celular extremamente importante no processo da mitose. Note, na Figura 1.4, que, no início da fase **G1** alguns filamentos mostram-se de forma mais pronunciada. Estes filamentos são conhecidos como microtúbulos e tem função estrutural, ou seja, dão forma às células, impedindo o colapso de sua membrana celular sobre suas organelas e núcleo, além de desempenhar papel fundamental no processo de divisão celular.

Os Microtúbulos

O microtúbulo (figura 1.5) é um polímero formado por sub-unidades globulares denominadas tubulina, arranjadas numa estrutura cilíndrica com aproximadamente 24 nm de diâmetro[3]. A tubulina (figura 1.6), por sua vez, é um hetero-dímero composto por dois monômeros: a α -tubulina e a β -tubulina.

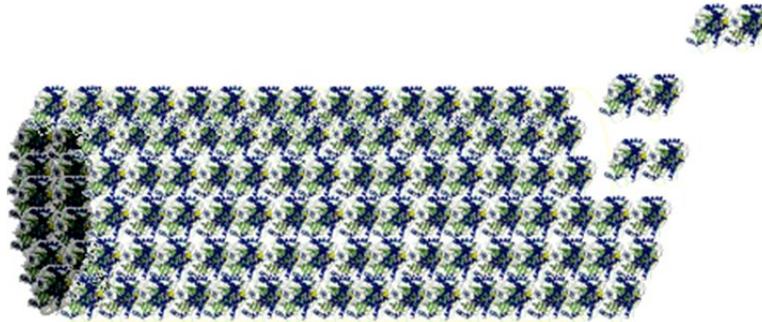


Figura 1.5: Estrutura esquemática de um microtúbulo

A interação entre a α -tubulina e a β -tubulina é bastante forte, tanto que, sob condições normais, esses monômeros raramente se dissociam.

Existem dois tipos de microtúbulos:

- Estáveis, com longo tempo de vida;
- Instáveis, com curto tempo de vida.

Os microtúbulos estáveis são responsáveis pela manutenção da estrutura celular, enquanto que os instáveis, são importantes para os diversos tipos de movimento celulares, como por exemplo, o descarte de organelas, a absorção de alimentos (fagocitose) e a segregação física do material celular no processo de divisão [3].

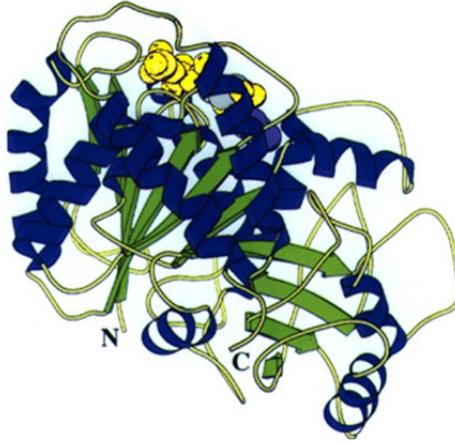


Figura 1.6: Estrutura da proteína tubulina

A indústria farmacêutica tem dedicado muito de suas pesquisas ao estudo de drogas que têm como alvo primário os microtúbulos. Cada uma destas drogas atua em um sítio específico da tubulina. São três os sítios mais importantes, até agora identificados: os dos taxóides, dos *Vinca* alcalóides e o da colchicina. Os taxóides, composto pelo Taxol e derivados, são uma classe de agentes quimioterápicos amplamente estudados e caracterizados pela alta eficiência na ação anticarcinogênica, quebrando o ciclo de polimerização-despolimerização da tubulina em microtúbulos[5, 7, 6, 8]. Os taxóides ligam-se a determinado um sítio da β -tubulina e estimulam a polimerização dos heterodímeros, estabilizando e inibindo sua posterior despolimerização[6]. Este processo bloqueia o ciclo de reprodução celular na fase da mitose, onde os cromossomos duplicados para a formação de duas células irmãs são separados em pólos distintos da célula original. A polimerização estimulada pelos taxóides ocorre, mesmo em condições pouco favoráveis, como baixa temperatura e ausência de proteínas associadas aos microtúbulos. Já, tanto os *Vinca* alcalóides, como a Colchicina e derivados, tem como alvo sítios distintos da β -tubulina. O primeiro grupo, têm como principal problema terapêutico, a resistência à droga de muitas células cancerígenas, enquanto que o segundo tem sérios efeitos colaterais como: anemia aguda e redução significativa de leucócitos do sangue[7].

A característica comum dos três compostos descritos acima é o fato de todos eles atuarem interrompendo o ciclo de polimerização - despolimerização da tubulina. Compostos cuja ação antineoplásica está baseada somente na inibição do ciclo mitótico, não têm atuação eficiente em células que não estão se dividindo quando a droga alcança o seu sítio terapêutico. Estas células, portanto, podem sobreviver e adquirir resistência à droga. Assim, a interrupção do supri-

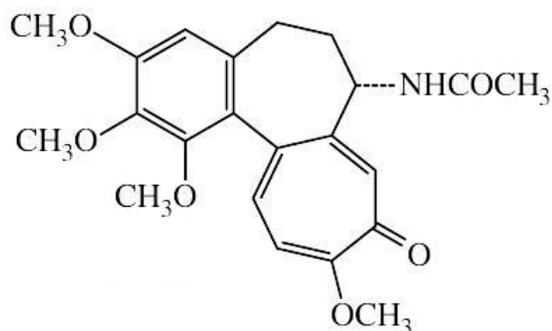


Figura 1.7: Estrutura molecular da Colchicina.

mento de sangue dos tumores é uma via alternativa importante, uma vez que a sua atuação resulta na morte da célula, independentemente de sua fase no ciclo celular. Entretanto, a atuação isolada de agentes antivasculares causa a morte do tumor central e não possui efeito significativo em tumores periféricos. Estes medicamentos são mais efetivos quando administrados em conjunto com a radioterapia e medicamentos citotóxicos [5]. Desta forma, a atividade dual (antivascular e citotóxica) de compostos pode ser bastante vantajosa. A Colchicina diferencia-se dos taxóides e dos *Vinca* alcalóides, porque também consegue agir na vasculatura que irriga o tumor. No entanto, ela não é uma droga muito eficiente, uma vez que possui índice terapêutico baixo, ou seja, deve ser administrada em concentrações próximas dos limites aceitáveis para o organismo[7]. A grande esperança está no conjunto de novos compostos estruturalmente relacionados, que atuam no mesmo sítio da Colchicina. Dentre estes, podemos citar a Combretastatina e derivados, objeto de estudo deste trabalho.

1.3 A Combretastatina

A Combretastatina é uma pequena molécula orgânica, um estilbeno, encontrada e isolada em 1982 pelo pesquisador George R. Pettit [12] num pequeno salgueiro africano cujo nome científico é *Combretum caffrum*. Sua estrutura é bastante simples, possui dois anéis de benzeno poli-oxigenados ligados por uma ponte de etileno.

Devido à rigidez das ligações duplas entre átomos de carbono, a estrutura não apresenta muita liberdade ao redor da ponte que liga os anéis aromáticos, tornando-a estruturalmente muito semelhante à Colchicina (ver figura 1.7), o que é considerado crucial na sua atividade anticarcinogênica. No entanto, apresenta baixa solubilidade em água, o que a tornaria pouco

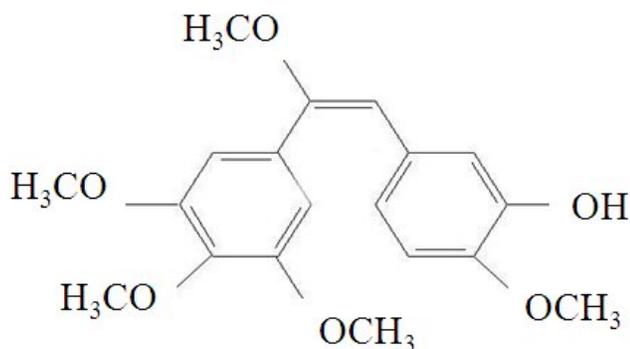


Figura 1.8: Estrutura molecular da Combretastatina CA-4. Cada vértice das estruturas geométricas possui um átomo de carbono.

eficiente na sua administração dentro de organismos vivos. Na tentativa de se obterem compostos mais ativos e também mais solúveis, modificações foram consideradas na estrutura inicial da Combretastatina. Estas modificações ocorrem através de mudanças nos grupos ligados aos dois anéis aromáticos e da ponte.

Como já mencionamos anteriormente, a Combretastatina possui importantes vantagens sobre os demais medicamentos que atuam na tubulina. Ela possui dupla função antineoplásica, pois tem ação citotóxica e antivascular. A ação antivascular pode ser dividida em dois mecanismos primários: o primeiro seria impedir a formação de novos vasos que alimentem o tumor e o segundo, a atuação em vasos já estabelecidos. Apesar de também possuir ação antiangiogênica, a Combretastatina e derivados, destacam-se pela eficiência na interrupção do fluxo sanguíneo que alimenta o tumor. Ao atingirem as células endoteliais da vasculatura tumoral, ela atua sobre a tubulina, despolimerizando os microtúbulos. Conseqüentemente, as células endoteliais perdem sua forma, bloqueando o fluxo de sangue até o tumor. Sem alimento e oxigênio, o tumor acaba morrendo. Um esquema desta atuação está apresentado nas figuras 1.9 a 1.16.

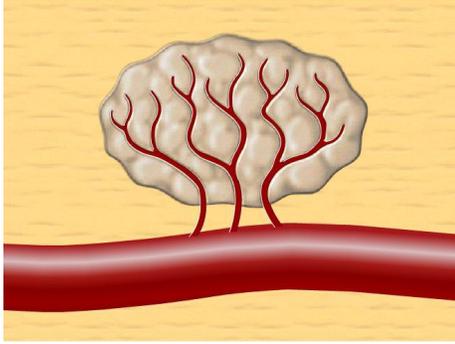


Figura 1.9: Vasos sanguíneos que alimentam o tumor [11].

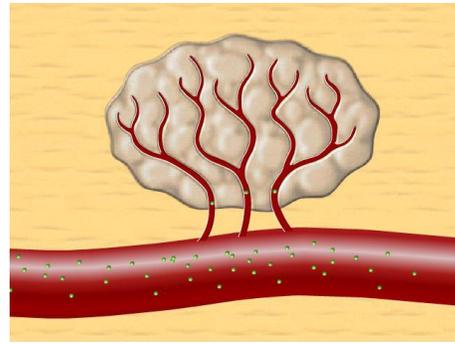


Figura 1.10: A Combretastatina atinge a vasculatura tumoral [11].

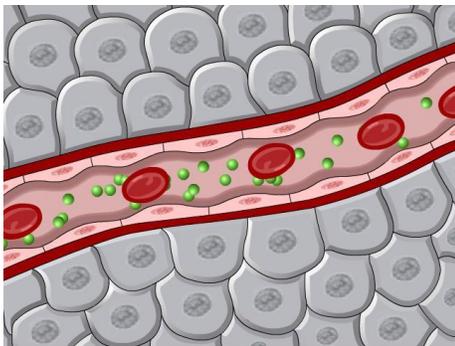


Figura 1.11: A Combretastatina dentro da vasos sanguíneos irrigam o tumor [11].

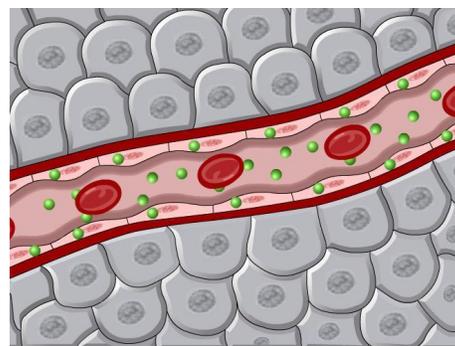


Figura 1.12: Ela penetra nas células endoteliais [11].

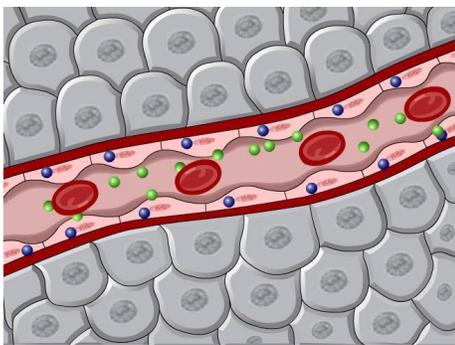


Figura 1.13: Ligando-se à tubulina [11].

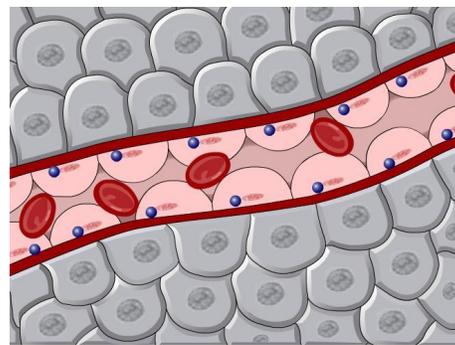


Figura 1.14: As células endoteliais perdem sua forma [11].

Outra importante característica da Combretastatina é a sua eficiência, mesmo no combate a células cancerígenas que conseguem, com facilidade, adquirir resistência à droga[9, 10].

Fica evidente, portanto, a importância da Combretastatina no que diz respeito ao combate

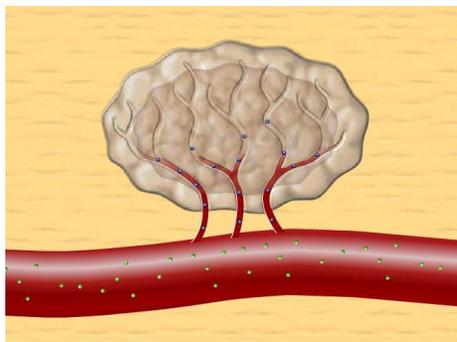


Figura 1.15: Aos poucos, o fluxo sanguíneo diminui [11].

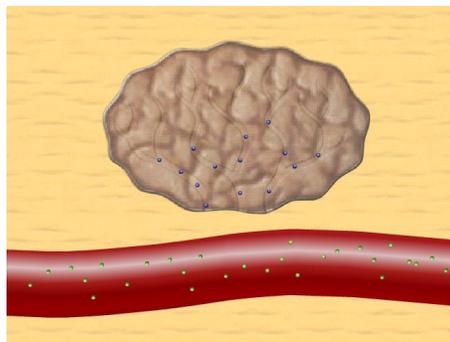


Figura 1.16: Até que o o tumor fica sem alimento e morre [11].

ao câncer. O estudo da sua relação estrutura-atividade, estabelecendo uma correlação entre sua atividade biológica com indicadores teóricos, é o objetivo deste trabalho. Introduziremos agora, o conjunto de moléculas estudadas nesta dissertação ¹.

1.3.1 Combretastatina e derivados

Neste projeto estudamos um conjunto de dezoito moléculas, obtido do trabalho experimental de S. Ducki *et al* [5]. Nele está a Combretastatina e mais 17 compostos estruturalmente relacionados pelo esqueleto básico comum, dado pela figura 1.17.

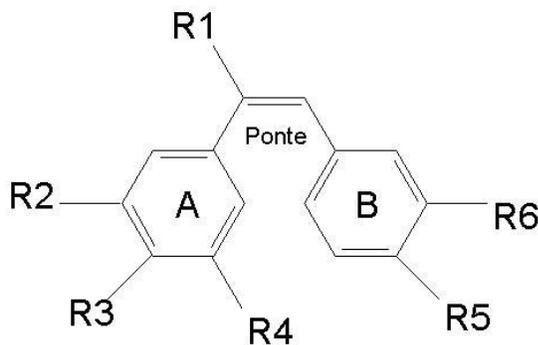


Figura 1.17: Esqueleto da Combretastatina e derivados estudados.

¹Algumas das propriedades mencionadas sobre a Combretastatina e outras que não foram mencionadas aqui, podem ser encontradas nas referências de [5] a [24]

As moléculas diferem entre si porque possuem diferentes substituintes laterais. A tabela 1.1 mostra cada uma das dezoito moléculas com seus respectivos grupos laterais.

	R1	R2	R3	R4	R5	R6	IC₅₀[5]
C1	H	CH ₃	CH ₃	CH ₃	OCH ₃	OH	0.12
C2	CH ₃	OCH ₃	OCH ₃	OCH ₃	OCH ₃	OH	0.13
C3	Etil	OCH ₃	OCH ₃	OCH ₃	OCH ₃	OH	0.13
C4	H	OCH ₃	OCH ₃	OCH ₃	OCH ₃	OH	0.18
C5	H	OCH ₃	OCH ₃	OCH ₃	OCH ₃	H	0.20
C6	H	OEtil	OEtil	OEtil	OCH ₃	OH	0.50
C7	H	CH ₃	CH ₃	CH ₃	OCH ₃	H	0.65
C8	H	CH ₃	CH ₃	CH ₃	OCH ₃	F	0.70
C9	H	OEtil	OEtil	OEtil	OCH ₃	F	1.20
C10	H	CH ₃	CH ₃	OCH ₃	OCH ₃	F	1.25
C11	CH ₃	OCH ₃	OCH ₃	OCH ₃	OCH ₃	H	1.30
C12	H	OCH ₃	OCH ₃	OCH ₃	CH ₃	CH ₃	2.0
C13	H	H	H	H	OCH ₃	OH	2.5
C14	H	CH ₃	CH ₃	H	OCH ₃	OH	3.1
C15	H	OCH ₃	OCH ₃	OCH ₃	Etil	H	3.4
C16	H	F	F	F	OCH ₃	OH	4.5
C17	H	F	F	F	OCH ₃	F	5.5
C18	H	CH ₃	CH ₃	H	OCH ₃	F	6.6

Tabela 1.1: Conjunto de moléculas estudado. R_n indica os grupos laterais do esqueleto da figura 1.17 e IC_{50} o índice experimental de atividade biológica.

Na última coluna estão dipostos os valores do IC_{50} , índice obtido experimentalmente que avalia a concentração de cada uma das 18 moléculas estudadas para que ocorra 50% da inibição da polimerização da tubulina. Desta forma, quanto menor o índice, menor a dose do fármaco a ser utilizada para inibição da polimerização da tubulina, ou seja, mais eficiente é o composto. Os dados experimentais foram obtidos com referência ao valor experimental para a Combretastatina.

Utilizamos o seguinte critério de classificação com respeito a atividade das moléculas da tabela 1.1:

- Se $IC_{50} < 1$ então a molécula será considerada ativa.

- Se $IC_{50} \geq 1$ então a molécula será considerada inativa.

A tabela 1.2, abaixo, separa nossas moléculas segundo o critério de classificação acima.

Ativas	C01	C02	C03	C04	C05	C06	C07	C08		
Inativas	C09	C10	C11	C12	C13	C14	C15	C16	C17	C18

Tabela 1.2: Classificação de atividade a partir do IC_{50} .

Estabelecido o critério de atividade experimental, estamos aptos para o estudo estrutura-atividade do nosso conjunto de moléculas. Antes, porém introduziremos os aspectos mais importantes sobre as metodologias de estudo utilizadas nesta dissertação.

Capítulo 2

Metodologia

Lidaremos, nas seções seguintes deste capítulo com dois dos principais métodos computacionais utilizados para o estudo de sistemas moleculares: os métodos *ab initio* e os semi-empíricos. Os métodos *ab initio* são puramente teóricos, ou seja, dada a configuração do sistema, todos os cálculos são feitos sem a inclusão de qualquer parâmetro auxiliar. No entanto, estes métodos são, na média, mais lentos e requerem um alto custo computacional, mesmo para sistemas relativamente pequenos, como sistemas com poucas dezenas de átomos. Por outro lado, os métodos semi-empíricos utilizam parâmetros obtidos empiricamente e a acuracidade destes métodos está associada a acuracidade dos dados experimentais usados na obtenção dos parâmetros. Entretanto, os métodos semi-empíricos são rápidos e precisos, o suficiente, para lidar com sistemas em que o número de átomos é bastante grande. De forma resumida, o poder dos métodos semi-empíricos está, não no seu rigor teórico, mas na sua capacidade de reprodução de propriedades químicas a partir dos parâmetros ajustáveis.

Qualquer método semi-empírico é composto por uma parte teórica e por um conjunto de parâmetros pré-estabelecidos. Podemos, em linhas gerais, dizer que algumas das integrais associadas a propriedades físico-químicas de algum átomo que deveriam ser calculadas no método Hartree-Fock, são substituídas por parâmetros empíricos correspondentes às respectivas propriedades. Assim, ganha-se tempo computacional, pois se o parâmetro associado a integral que deve ser calculada pertencer ao banco de dados do método utilizado, não é necessário que os cálculos sejam realizados. A parte de maior consumo de tempo nos métodos *ab initio* é o cálculo do grande número de integrais de repulsão eletrônica. Para obtermos maiores detalhes sobre os métodos semi-empíricos, introduziremos, de forma qualitativamente, agora, as conhecidas

equações de Hartree-Fock-Roothaan.

2.1 Teoria de Campo Auto-Consistente

Um problema de configuração relativamente simples como o problema de três corpos, não possui solução analítica e, à medida que o número de componentes do sistema aumenta, aumenta, também, a dificuldade da análise matemática do problema. No domínio quântico, à exceção de sistemas simples como o íon H_2^+ , todos os outros sistemas moleculares possuem um número maior de núcleos e elétrons que interagem entre si. Ou seja, a maioria absoluta dos sistemas moleculares são suficientemente complexos para não admitirem solução analítica. Diversos métodos foram, e ainda vem sendo desenvolvidos, na tentativa de contornar este problema. Os principais objetivos destes métodos são encontrar, com segurança, as propriedades moleculares fundamentais, além de determinar o estado fundamental para estes sistemas.

2.1.1 Aproximação de Hartree-Fock-Roothaan

Para se encontrar a função de onda que descreve o estado fundamental de uma molécula, é necessário resolver a equação de Schrödinger independente do tempo

$$H|\Psi_0\rangle = E_0|\Psi_0\rangle, \quad (2.1)$$

onde H é o Hamiltoniano do sistema molecular e $|\Psi_0\rangle$ é a função de onda do estado fundamental com energia E_0 . Se desprezarmos efeitos relativísticos, o Hamiltoniano para este tipo de sistema, com M núcleos e N elétrons, em unidades atômicas, é dado por

$$H = -\sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} + \sum_{A=1}^M \frac{1}{2M_A} \nabla_A^2 + \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{R_{AB}}. \quad (2.2)$$

onde:

- Z_A é o número atômico do núcleo A;
- M_A é a razão entre a massa do núcleo A e a massa do elétron;
- os operadores laplacianos ∇_i^2 e ∇_A^2 são calculados em termos das coordenadas do i -ésimo elétron e do A -ésimo núcleo, respectivamente;

- o primeiro e o quarto termos representam as energias cinéticas dos elétrons e núcleos, respectivamente;
- o segundo termo está associado à energia de interação coulombiana entre núcleos e elétrons e o terceiro à energia de repulsão entre os elétrons;
- o quinto termo representa o potencial de repulsão entre os núcleos.

No entanto, o Hamiltoniano (2.2), como é de se esperar, impossibilita uma solução analítica para a equação (2.1). Devemos, portanto, utilizar alguma aproximação que simplifique a equação de Schrödinger para sistemas moleculares.

Aproximação de Bohr-Oppenheimer

A aproximação de Bohr-Oppenheimer é fundamental para o cálculo quântico de sistemas moleculares. Como podemos ver na equação (2.2), existem dois termos de origem puramente eletrônica, outros dois de origem nuclear e um de interação entre elétrons e núcleos. A aproximação de Bohr-Oppenheimer nos fornece uma maneira de estudar os problemas eletrônico e nuclear separadamente.

Como a massa dos núcleos é muito maior que a massa eletrônica, podemos admitir que, comparativamente com movimento eletrônico, eles se movem muito mais lentamente. Como conseqüência, a energia cinética dos núcleos é muito pequena comparada à energia cinética dos elétrons, o que significa dizer que o quarto termo da equação (2.2) é desprezível e pode ser desconsiderado. Assim, uma boa aproximação seria considerar que os elétrons se movem num campo gerado por núcleos fixos, logo o quinto termo da equação (2.2) seria, na verdade, uma constante. Como sabemos, uma constante adicionada ao Hamiltoniano não altera suas autofunções, logo, podemos reescrever a equação (2.2) como

$$H_{el}|\Psi_{el}\rangle = \left[-\sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} \right] |\Psi_{el}\rangle = E_{el}|\Psi_{el}\rangle, \quad (2.3)$$

onde H_{el} e $|\Psi_{el}\rangle$ so o Hamiltoniano e a autofunção do problema eletrônico aproximado.

De forma análoga, como os elétrons movem-se muito mais rápido do que os núcleos, é razoável substituir os termos que envolvem as coordenadas eletrônicas de (2.2), pelas suas médias com a relação à função de onda dos elétrons, Ψ_{el} . Ou seja,

$$H_N = \sum_{A=1}^M \frac{1}{2M_A} \nabla_A^2 + \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{R_{AB}} + \left\langle - \sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} \right\rangle. \quad (2.4)$$

Desta forma, construímos um Hamiltoniano para o movimento do núcleo sob a ação de um campo médio gerado pela distribuição dos elétrons¹. A função de onda para os núcleos é obtida resolvendo a equação

$$H_N |\Psi_N\rangle = E_N |\Psi_N\rangle \quad (2.5)$$

e a solução geral é dada por

$$|\Psi_0\rangle = |\Psi_{el}\rangle |\Psi_N\rangle \quad (2.6)$$

Neste texto focaremos somente a metodologia de solução da equação (2.3). Com este intuito, o primeiro passo a ser dado é considerar que os elétrons não interagem entre si, ou seja, a energia de cada elétron é dada pelo Hamiltoniano

$$h(i) = \frac{1}{2} \nabla_i^2 + \sum_A \frac{Z_A}{r_{iA}}, \quad (2.7)$$

logo o Hamiltoniano do sistema será

$$H_1 = \sum_{i=1}^N h(i). \quad (2.8)$$

A equação de Schrödinger, com o Hamiltoniano acima, é separável, logo, dizer que os elétrons não interagem entre si, significa dizer que podemos escrever a função de onda multieletrônica $|\Psi_{el}\rangle$ como o produto de funções de onda de um elétron. Neste ponto, o spin eletrônico deve ser levado em consideração. Para isso, utilizam-se funções ortonormais $\alpha(\omega)$ e $\beta(\omega)$ que representam os spins $+1/2$ e $-1/2$, respectivamente. Escreveremos, portanto, uma função de onda que contenha, simultaneamente, a dependência espacial e de spin. A este tipo de função dá-se o nome de spin-orbital e escreve-se

$$\chi_{2i-1}(x) = \psi_i(r) \alpha(\omega) \quad \text{e} \quad \chi_{2i}(x) = \psi_i(r) \beta(\omega). \quad (2.9)$$

Deve ficar claro que está subentendida a dependência do spin em $x = (\vec{r}, \omega)$. Devido às condições de ortonormalidade, os spins-orbitais, devem obedecer à relação

¹Para mais detalhes ver referências [25], [26], [27], [28].

$$\int \chi_i^*(x) \chi_j(x) dx = \langle \psi_i | \psi_j \rangle = \delta_{ij}. \quad (2.10)$$

Considera-se como autofunções dos hamiltonianos $h(i)$ os spins-orbitais $\chi(x)$, logo,

$$h(i) \chi_i(x) = \epsilon_i \chi_i(x), \quad (2.11)$$

desta forma, a função de onda multieletrônica será dada pelo produto dos spins-orbitais

$$\Psi_H(x_1, x_2, \dots, x_N) = \chi_i(x_1), \chi_j(x_2), \dots, \chi_k(x_N), \quad (2.12)$$

uma vez que Ψ_H é autofunção de H_1 , ou seja,

$$H_1 \Psi_H = E_1 \Psi_H \quad (2.13)$$

onde E_1 é dado por

$$E_1 = \epsilon_1 + \epsilon_2 + \dots + \epsilon_N. \quad (2.14)$$

A função de onda Ψ_H é conhecida como função de Hartree e devemos destacar dois pontos importantes sobre ela:

1. ela não inclui efeitos de repulsão inter-eletrônicas;
2. ela não é antissimétrica com relação à troca de coordenadas, logo não pode representar férmions.

No entanto, é possível encontrar uma função antissimétrica partindo de Ψ_H . Para isto, basta utilizarmos o determinante de Slater

$$\Psi_0 = \begin{vmatrix} \chi_1(\mathbf{x}_1) & \chi_1(\mathbf{x}_2) & \cdots & \chi_1(\mathbf{x}_N) \\ \chi_2(\mathbf{x}_1) & \chi_2(\mathbf{x}_2) & \cdots & \chi_2(\mathbf{x}_N) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_N(\mathbf{x}_1) & \chi_N(\mathbf{x}_2) & \cdots & \chi_N(\mathbf{x}_N) \end{vmatrix}. \quad (2.15)$$

Destá forma, a troca de coordenadas de duas partículas, o que significa a troca de duas colunas na matriz acima, implicaria na alteração do sinal do determinante e, conseqüentemente, o sinal da função de onda. Baseado na definição da função determinante, podemos escrever

$$|\psi_0\rangle = \sum_{i=1}^{N!} (-1)^{p_i} P_i \{\chi_m(1), \chi_n(2), \dots\}, \quad (2.16)$$

onde P_i é operador que permuta o índice das colunas 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 $1, 2, \dots, N$.

Utilizando as equações (2.2) e (2.16), podemos obter a energia do estado fundamental, pois

$$E_0 = \langle \psi_0 | H | \psi_0 \rangle = \sum_{i=1}^N \langle i | h | i \rangle + \frac{1}{2} \sum_N^i \sum_N^j (\langle ii | jj \rangle - \langle ij | ji \rangle). \quad (2.17)$$

onde

$$\langle i | h | j \rangle = \langle i | h(1) | j \rangle = \int \chi_i^*(x_1) h(r_1) \chi_j(x_1) dx_1 \quad (2.18)$$

está relacionado com o hamiltoniano de 1-elétron (integrais de 1-elétron) e

$$\langle ij | kl \rangle = \int \chi_i^*(x_1) \chi_j(x_1) \frac{1}{r_{12}} \chi_k^*(x_2) \chi_l(x_2) dx_1 dx_2 \quad (2.19)$$

está relacionado com os termos de interação (integrais de 2-elétrons).

A expressão (2.17) nos fornece a energia do estado fundamental. Precisamos agora encontrar que condições deve satisfazer o conjunto $\{\chi_i\}$ para que E_0 seja mínima, respeitando sempre a condição de ortonormalidade para os spins-orbitais,

$$\langle \chi_i | \chi_j \rangle = \delta_{ij}.$$

Note que este, é um problema de obtenção de extremos de uma função restrita a vínculos. Uma das principais técnicas de resolução deste tipo de problema é a dos Multiplicadores de Lagrange. Para isto, escrevemos o funcional

$$L(\{\chi_i\}) = E_0(\{\chi_i\}) - \sum_i^N \sum_j^N \lambda_{ij} (\langle i | j \rangle - \delta_{ij}), \quad (2.20)$$

em termos do conjunto de variáveis $\{\chi_i\}$. Daí efetuamos a derivada funcional de $L(\{\chi_i\})$ com respeito ao conjunto $\{\chi_i\}$ e encontramos que condições devem satisfazer o conjunto de multiplicadores para que a relação

$$\delta [L(\{\chi_i\})] = 0 \quad (2.21)$$

seja satisfeita. Se utilizarmos as definições abaixo

$$K_j(1) \chi_i(1) = \left[\int \chi_j^*(2) \frac{1}{r_{12}} \chi_i(2) dx_2 \right] \chi_j(1), \quad (2.22)$$

e

$$J_j(1) \chi_i(1) = \left[\int \chi_j^*(2) \frac{1}{r_{12}} \chi_j(2) dx_2 \right] \chi_i(1) \quad (2.23)$$

conhecidos como Operador de Troca e Operador de Coulomb, respectivamente, então, para que a condição (2.21) seja satisfeita, os multiplicadores de Lagrange $\{\lambda_i\}$ devem obedecer à seguinte relação

$$\left[h(1) + \sum_{j=1}^N J_j(1) - K_j(1) \right] \chi_i(1) = \sum_{j=i}^N \lambda_{ij} \chi_j(1), \quad i = 1, 2, 3, \dots, N, \quad (2.24)$$

comumente chamada de **Forma Não-Canônica das Equações de Hartree-Fock**. No entanto, note que se definirmos

$$\mathbf{F} = \left[h(1) + \sum_{j=1}^N J_j(1) - K_j(1) \right], \quad (2.25)$$

podemos reescrever (2.24) como

$$\mathbf{F}|\chi_i\rangle = \sum_{j=1}^N \lambda_{ij} |\chi_j\rangle \quad (2.26)$$

que, a princípio pode conter termos não diagonais. Entretanto, como o conjunto $\{\chi_i\}$ é ortogonal, podemos construir um novo conjunto de orbitais $\{\chi'_i\}$ a partir de uma transformação unitária aplicada ao conjunto $\{\chi_i\}$, ou seja, podemos escrever

$$\chi'_i = \sum_j U_{ij} \chi_j. \quad (2.27)$$

Todas as informações físicas do sistema ainda estão contidas no conjunto construído utilizando a equação (2.27), uma vez que uma transformação unitária pode ser vista como, apenas, uma rotação do sistema de coordenadas, digamos por um ângulo ϕ . De fato, pode-se mostrar que a função de onda em termos dos novos orbitais é escrita como

$$|\Psi'_0\rangle = e^{i\phi} |\Psi_0\rangle. \quad (2.28)$$

Outro ponto que deve ser destacado é o fato de que tanto o operador de Troca, como o de Coulomb, não têm sua forma afetada por uma transformação unitária.

Se voltarmos nosso olhar para os multiplicadores de Lagrange, veremos as vantagens de se escrever a função de onda em termos dos orbitais (2.27), pois podemos escrever um novo conjunto de multiplicadores a partir da expressão matricial

$$\lambda' = \mathbf{U}^\dagger \lambda \mathbf{U}. \quad (2.29)$$

Pode-se verificar que sempre é possível encontrar uma transformação unitária \mathbf{U} tal que a transformação (2.29), acima, diagonaliza λ e que, além disso, ela é única². Logo justifica-se a introdução de uma nova base de orbitais, uma vez que, em termos desta base, podemos reescrever a equação (2.26) em sua forma diagonal

$$\mathbf{F}|\chi_i\rangle = \lambda_i|\chi_i\rangle, \quad (2.30)$$

que também é conhecida por **Forma Canônica das Equações de Hartree-Fock**. Note que o complicado problema multieletrônico foi reduzido ao estudo da equação de Schrödinger para um elétron sob a influência de um campo médio gerado pela distribuição de cargas devida aos demais elétrons do sistema.

Vamos restringir nosso problema a sistemas de N elétrons (N um número par), com todos eles emparelhados dois a dois em cada um dos $n = N/2$ orbitais espaciais. Ou seja, restringiremos a discussão a sistemas no estado fundamental e de camada fechada. O formalismo para os estados fundamentais de camada aberta é bem distinto deste e não será tratado aqui, uma vez que a essência do método é a mesma.

O intuito, agora, é deixar de representar a função de onda em termos dos spins-orbitais, para passar a uma representação puramente espacial. Portanto, se integrarmos sobre as funções de spin e considerarmos a condição de ortonormalidade obtemos

$$\mathbf{F}(1)\Psi_j(1) = \epsilon_j\Psi_j(1), \quad (2.31)$$

com o operador de Fock para camada fechada sendo escrito como

²Ver referência [29]

$$\mathbf{F}(1) = \mathbf{h}(1) + \sum_{i=1}^{N/2} [2J_i - K_i] \quad (2.32)$$

onde as definições dos operadores de Coulomb (J) e de Troca (K) são totalmente análogas às aquelas derivadas para os spins orbitais

$$J_j(1) \Psi_i(1) = \left[\int \Psi_j^*(2) \frac{1}{r_{12}} \Psi_j(2) dr_2 \right] \Psi_i(1), \quad [\text{Operador de Coulomb}] \quad (2.33)$$

$$K_j(1) \Psi_i(1) = \left[\int \Psi_j^*(2) \frac{1}{r_{12}} \Psi_i(2) dr_2 \right] \Psi_j(1), \quad [\text{Operador de Troca}]. \quad (2.34)$$

Além disso, o fator dois multiplicando o Operador de Coulomb na equação (2.32) é devido à degenerescência do spin.

Uma vez eliminada a dependência do spin de nossas equações, o cálculo de orbitais moleculares torna-se equivalente ao problema de resolver a equação (2.31). Esta equação, é na verdade, uma equação íntegro-diferencial, cuja solução não é de modo algum, trivial. Veremos, a seguir, uma forma de contornar este problema.

2.1.2 Equações de Roothan

Ao invés de resolver a equação (2.31), Roothaan³ introduziu um conjunto de funções de base espaciais capaz de transformar um problema íntegro-diferencial num problema matricial. De fato, se a função de onda for descrita por uma base completa de dimensão finita (l) temos

$$\Psi_i = \sum_{\mu=1}^l C_{\mu i} \varphi_{\mu} \quad (2.35)$$

com $i = 1, 2, \dots, l$. Esta aproximação é comumente conhecida como LCAO (*Linear Combination of Atomic Orbitals*).

Substituindo a equação (2.35) na equação (2.31) encontramos

$$\sum_{\nu=1}^l C_{\nu i} \int \varphi_{\mu}^*(1) \mathbf{F}(1) \varphi_{\nu}(1) d^3r_1 = \epsilon_i \sum_{\nu=i}^l C_{\nu i} \int \varphi_{\mu}^*(1) \varphi_{\nu}(1) d^3r_1, \quad (2.36)$$

ou

$$\sum_{\nu} C_{\nu i} F_{\mu\nu} = \epsilon_i \sum_{\nu} C_{\nu i} S_{\mu\nu}, \quad i = 1, 2, \dots, l \quad (2.37)$$

³O leitor interessado no trabalho original de Roothaan deve consultar a referência [30]

onde definimos

$$F_{\mu\nu} = \int \varphi_{\mu}^*(1) \mathbf{F}(1) \varphi_{\nu}(1) d^3r_1, \quad (2.38)$$

$$S_{\mu\nu} = \int \varphi_{\mu}^*(1) \varphi_{\nu}(1) d^3r_1. \quad (2.39)$$

Logo, resolver (2.31) é equivalente a encontrar a solução de

$$\mathbf{FC} = \mathbf{SC}\epsilon, \quad (2.40)$$

ou seja, precisamos resolver um problema matricial. A equação (2.40) é conhecida como **Equação de Roothaan**.

Para determinarmos o conjunto de orbitais moleculares espaciais $\{\Psi_i\}$ de Hartree-Fock e as energias orbitais ϵ_i , devemos diagonalizar (2.40), eliminando a matriz \mathbf{S} . Antes, porém, introduziremos um conceito importante que nos auxiliará bastante: a densidade de carga.

2.1.3 A densidade de carga

Imagine um elétron descrito pela função de onda $\Psi_i(\mathbf{r})$. A probabilidade de se encontrar este elétron num elemento de volume $d\mathbf{r}$ localizado no ponto \mathbf{r} é dado por $|\Psi_i(\mathbf{r})|^2 d\mathbf{r}$, ou seja, a função de distribuição de probabilidade (densidade de carga) é $|\Psi_i(\mathbf{r})|^2$. Se, ao invés de um elétron, tivermos uma molécula de camada fechada descrita por uma função de onda de um único determinante, com cada um dos orbitais moleculares Ψ_i contendo dois elétrons, então a densidade de carga será dada por

$$\rho(\mathbf{r}) = 2 \sum_i^{N/2} |\Psi_i(\mathbf{r})|^2. \quad (2.41)$$

Escrevendo Ψ_i em termos da expansão (2.35) encontramos

$$\rho(\mathbf{r}) = \sum_{\mu\nu} P_{\mu\nu} \varphi_{\mu}(\mathbf{r}) \varphi_{\nu}^*(\mathbf{r}) \quad (2.42)$$

onde

$$P_{\mu\nu} = 2 \sum_i^{N/2} C_{\mu i} C_{\nu i}^* \quad (2.43)$$

é conhecido como a matriz densidade. Deve ser ressaltado que a matriz \mathbf{P} determina univocamente a densidade de carga $\rho(\mathbf{r})$. Além disso ela está diretamente associada a matriz dos

coeficientes de expansão \mathbf{C} , como pode ser verificado a partir da equação (2.43). Logo, podemos escrever os resultados dos cálculos Hartree-Fock para camada fechada, tanto em termos de $C_{\mu i}$ como de $P_{\mu\nu}$. Utilizando as equações (2.32),(2.33) e (2.34), reescrevemos o operador de Fock para camada fechada em termos da matriz densidade

$$\begin{aligned} \mathbf{F}(\mathbf{r}_1) &= \int \varphi_\mu^*(\mathbf{r}_1) h(\mathbf{r}_1) \varphi_\nu(\mathbf{r}_1) d\mathbf{r}_1 + \sum_{\lambda\sigma} P_{\lambda\sigma} \left[\langle \mu\nu | \sigma\lambda \rangle - \frac{1}{2} \langle \mu\nu | \sigma\nu \rangle \right] \\ &= H_{\mu\nu}^{\text{Caroço}} + \sum_{\lambda\sigma} P_{\lambda\sigma} \left[\langle \mu\nu | \sigma\lambda \rangle - \frac{1}{2} \langle \mu\nu | \sigma\nu \rangle \right] \end{aligned} \quad (2.44)$$

onde

$$H_{\mu\nu}^{\text{Caroço}} = \int \varphi_\mu^*(\mathbf{r}_1) h(\mathbf{r}_1) \varphi_\nu(\mathbf{r}_1) d\mathbf{r}_1.$$

A equação (2.44) pode ser reescrita explicitando os operadores de Coulomb e de troca. De fato, se definirmos

$$T_{\mu\nu} \equiv \int \varphi_\mu^*(1) \left[-\frac{1}{2} \nabla_i^2 \right] \varphi_\nu(1) d^3r_1 \quad (2.45)$$

$$V_{\mu\nu} \equiv \int \varphi_\mu^*(1) \left[-\sum_j \frac{Z_j}{r_{1j}} \nabla_i^2 \right] \varphi_\nu(1) d^3r_1 \quad (2.46)$$

$$G_{\mu\nu} \equiv \sum_{\lambda\sigma} P_{\lambda\sigma} \left[\langle \mu\nu | \sigma\lambda \rangle - \frac{1}{2} \langle \mu\nu | \sigma\nu \rangle \right] \quad (2.47)$$

com

$$\langle \mu\nu | \sigma\lambda \rangle = \int \varphi_\mu^*(1) \varphi_\nu(1) \frac{1}{r_{12}} \varphi_\lambda^*(2) \varphi_\sigma(2) d^3r_1 d^3r_2$$

temos

$$F_{\mu\nu} = (T_{\mu\nu} + V_{\mu\nu}) + G_{\mu\nu}. \quad (2.48)$$

Analisando a equação (2.48) vemos que os dois primeiros termos envolvem apenas um elétron e são invariantes para um dado conjunto de funções de base $\{\varphi_\mu\}$. Entretanto a expressão para $G_{\mu\nu}$ envolve dois elétrons e depende da matriz densidade \mathbf{P} , além de um conjunto de integrais de dois elétrons. É, justamente, o cálculo deste termo que constitui a maior dificuldade no método Hartree-Fock, ou seja, que necessita de mais recurso computacional e dispende mais tempo.

Com o intuito de eliminar \mathbf{S} da equação (2.40) devemos encontrar uma transformação \mathbf{X} , não unitária, que ortogonalize o conjunto $\{\varphi\}$, que em geral, não é ortogonal. Matematicamente significa dizer que o novo conjunto de orbitais, obtidos pela aplicação da transformação \mathbf{X} , dados por

$$\varphi'_\mu = \sum_\nu \mathbf{X}_{\nu\mu} \varphi_\nu, \quad \mu = 1, 2, \dots, N \quad (2.49)$$

devem satisfazer a condição

$$\int \varphi'^*_\mu(\mathbf{r}) \varphi'_\nu(\mathbf{r}) = \delta_{\mu\nu}. \quad (2.50)$$

Substituindo a equação (2.49) em (2.50) obtemos

$$\sum_\lambda \sum_\sigma X_{\lambda\mu}^* S_{\lambda\sigma} X_{\sigma\nu} = \delta_{\mu\nu} \quad (2.51)$$

ou, na forma matricial

$$\mathbf{X}^\dagger \mathbf{S} \mathbf{X} = \mathbf{1} \quad (2.52)$$

Além disso, como \mathbf{S} é uma matriz Hermitiana, podemos encontrar uma transformação unitária \mathbf{U} que diagonaliza \mathbf{S} , ou seja,

$$\mathbf{U}^\dagger \mathbf{S} \mathbf{U} = \mathbf{s} \quad (2.53)$$

com \mathbf{s} diagonal. A partir das equações (2.52) e (2.53) pode-se mostrar que

$$\mathbf{X} = \mathbf{S}^{-1/2} = \mathbf{U} \mathbf{s}^{-1/2} \mathbf{U}^\dagger.$$

Além disso, se escrevermos

$$\mathbf{C}' = \mathbf{X}^{-1} \mathbf{C} \text{ ou } \mathbf{C} = \mathbf{X} \mathbf{C}' \quad (2.54)$$

então

$$(\mathbf{X}^\dagger \mathbf{F} \mathbf{X}) \mathbf{C}' = (\mathbf{X}^\dagger \mathbf{S} \mathbf{X}) \mathbf{C}' \epsilon, \quad (2.55)$$

ou seja,

$$\mathbf{F}'\mathbf{C}' = \mathbf{C}'\epsilon \quad (2.56)$$

onde

$$\mathbf{F}' = \mathbf{X}^\dagger \mathbf{F} \mathbf{X}. \quad (2.57)$$

é a nova matriz de Fock.

Uma vez que a matriz \mathbf{S} é diagonalizada utilizando a transformação unitária \mathbf{U} , a matriz diagonal \mathbf{s} será formada pelos autovalores de \mathbf{S} . Logo o método de Hartree-Fock-Roothaan resume-se em diagonalizar a matriz \mathbf{F}' dada pela equação (2.56).

A seguir mostraremos um esquema do algoritmo computacional utilizado para o cálculo do método de Hartree-Fock-Roothaan a partir da implementação numérica.

Procedimento SCF (Hartree-Fock-Roothaan)

1. Inserir o conjunto de átomos, as distâncias interatômicas e o conjunto de funções de base;
2. Calcular os elementos $S_{\mu\nu}$, $H_{\mu\nu}$ e $\langle \mu\nu | \sigma\lambda \rangle$;
3. Diagonalizar \mathbf{S} e obter \mathbf{X} ;
4. Calcular a matriz densidade \mathbf{P} ;
5. Encontrar a matriz \mathbf{G}
6. Determinar a matriz de Fock \mathbf{F} e em seguida a nova matriz de Fock \mathbf{F}' ;
7. Diagonalizar \mathbf{F}' obtendo \mathbf{C}' e ϵ ;
8. Calcular $\mathbf{C} = \mathbf{X}\mathbf{C}'$;
9. Determinar uma nova matriz \mathbf{P} ;
10. Critério de convergência para os elementos de \mathbf{P} ;
11. Se o critério não for satisfeito, retornar ao item 5, se sim, construir as matrizes \mathbf{C} , \mathbf{P} , \mathbf{F} etc.
12. Cálculo dos valores esperados.

2.2 Métodos Semi-empíricos

Mesmo com o crescente aumento dos recursos computacionais nos dias atuais, o cálculo *ab initio* (primeiro princípios) descrito anteriormente, é muito custoso do ponto de vista computacional quando aplicado a moléculas de médio e grande porte (dezenas e centenas de átomos, respectivamente). O número de integrais de dois elétrons (elementos de $G_{\mu\nu}$) torna-se muito grande e o seu cálculo, muitas vezes, inviável.

Parece óbvio que a forma de se contornar este problema é aplicar mais aproximações ao método de Hartree-Fock-Roothaan. Poderíamos nos dedicar ao estudo de diversas destas aproximações, mas focaremos, somente naquelas em que existe a incorporação de parâmetros empíricos com o intuito de substituir algumas das integrais que deveriam ser resolvidas. Este tipo de procedimento deu origem aos denominados métodos semi-empíricos, bastante utilizados no cálculo das estruturas geométricas de sistemas moleculares com um número significativo de átomos. A seguir apresentaremos alguns dos métodos semi-empíricos mais utilizados na literatura científica.

2.2.1 Aproximações NDO

Nos cálculos *ab initio*, onde todas as equações e integrais desenvolvidas na seção anterior são resolvidas, a qualidade dos resultados obtidos está diretamente ligada à base utilizada na expansão da função de onda. Para que os cálculos sejam, ao menos praticáveis, o número de funções do conjunto de base deve ser finito e a acuracidade dos resultados depende, tanto do número, como das funções escolhidas para a expansão. Portanto, mesmo que a escolha das funções seja a mais apropriada para os átomos envolvidos, existe um limite na descrição imposto pelo número finito de funções que compõem a base. Se tentarmos contornar este problema, aumentando a quantidade de funções, outro problema surge: a grande quantidade de integrais de dois elétrons a serem calculadas. Suponhamos que nosso conjunto de funções de base contenha M orbitais atômicos. O número de integrais de dois elétrons é da ordem de M^4 . Estamos, claramente, diante de um impasse: melhorar a qualidade dos resultados aumentando o conjunto de funções de base ou reduzir o custo computacional, reduzindo a quantidade de integrais de dois elétrons a ser calculada. Uma forma de se contornar este impasse seria aplicar algum tipo de aproximação a estas integrais. As aproximações mais simples que podem ser efetuadas, inicialmente, são:

- utilizar a Aproximação do Carço, considerando apenas os elétrons de valência;
- desprezar completamente o recobrimento diferencial [31, 32, 33, 34]

$$\varphi_\mu\varphi_\nu = \delta_{\mu\nu}\varphi_\mu\varphi_\mu. \quad (2.58)$$

Esta aproximação é conhecida como CNDO (*Complete Neglect of Differential Overlap*). Uma das maiores limitações da CNDO é a exclusão das integrais de troca de 1-centro, que impossibilita, por exemplo, o cálculo da densidade de estados de alguns orbitais para determinadas estruturas [35].

Uma segunda aproximação seria desconsiderar o recobrimento diferencial entre átomos distintos

$$\varphi_\mu^i\varphi_\nu^j = \delta_{\mu\nu}\varphi_\mu^i\varphi_\mu^i, \quad (2.59)$$

comumente conhecido como NDDO (*Neglect of Diatomic Differential Overlap*). A desvantagem deste método é o grande número de integrais de 2-centros a serem calculadas. Uma forma de se minimizar os problemas da CNDO e da NDDO, é manter o produto $\varphi_\mu\varphi_\nu$, envolvendo orbitais atômicos diferentes (φ_μ e φ_ν), somente nas integrais de 1-centro. Por ser um caso intermediário entre a CNDO e a NDDO, esta aproximação ganhou o nome de INDO (*Intermediate Neglect Differential Overlap*). Este tipo de negligenciamento de recobrimento diferencial foi utilizado em diversos outros métodos como ZINDO (*Zerner INDO*), MNDO (*Modified NDO*), AM1 (*Austin Method 1*), PM3 (*Parametric Method 3*), entre outros. Na verdade, existem diversos tipos de parametrizações para o método INDO e, uma das mais populares deu origem ao MNDO, e este, após algumas melhorias na forma de parametrização de algumas integrais e nos parâmetros utilizados, deu origem aos métodos AM1 e PM3. Discutiremos, agora, com um pouco mais de detalhes estes três últimos métodos.

2.2.2 Aproximação MNDO

Um dos primeiros métodos semi-empíricos autoconsistentes desenvolvido, foi o MNDO [36, 37, 38]. Neste método, o tratamento é restrito à moléculas de camada fechada e aos seus elétrons de valência, os quais assume-se moverem num campo gerado pelo carço composto pelo núcleo e pelos elétrons de camadas internas (aproximação de carço). Os orbitais moleculares da camada

de valência Ψ_i , como já vimos na equação (2.85), são representados por combinações lineares de um conjunto de base mínima de orbitais atômicos de valência $\{\varphi_\nu\}$ na aproximação LCAO. Os coeficientes das funções nesta expansão são obtidos a partir das equações de Hartree-Fock-Roothaan na aproximação NDDO, ou seja, a partir da equação

$$\sum_{\nu} (F_{\mu\nu} - E_i \delta_{\mu\nu}) C_{\nu i} = 0, \quad (2.60)$$

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

A energia eletrônica E_{el} será dada por

$$E_{el} = \frac{1}{2} \sum_{\mu} \sum_{\nu} P_{\mu\nu} (H_{\mu\nu} + F_{\mu\nu}), \quad (2.61)$$

onde $P_{\mu\nu}$ são os elementos da matriz densidade (ordem de ligação) e $H_{\mu\nu}$ do Hamiltoniano de caroço.

A partir de agora utilizaremos a seguinte notação :

- φ_{μ} e φ_{ν} representam os orbitais centrados no átomo A ;
- φ_{λ} e φ_{σ} representam os orbitais centrados no átomo B ($B \neq A$).

Desta forma os elementos da matriz de Fock no formalismo MNDO serão dados por

$$F_{\mu\mu} = U_{\mu\mu} + \sum_B V_{\mu\mu}^B + \sum_{\nu}^A P_{\nu\nu} \left(\langle \mu\mu|\nu\nu \rangle - \frac{1}{2} \langle \mu\nu|\mu\nu \rangle \right) + \sum_B \sum_{\lambda,\sigma} P_{\lambda\sigma} \langle \mu\mu|\lambda\sigma \rangle, \quad (2.62)$$

para os termos da diagonal e

$$F_{\mu\nu} = \sum_B V_{\mu\nu}^B + \frac{1}{2} \sum_{\nu}^A P_{\mu\nu} (\langle \mu\nu|\mu\nu \rangle - \langle \mu\mu|\mu\nu \rangle) + \sum_B \sum_{\lambda,\sigma} P_{\lambda\sigma} \langle \mu\nu|\lambda\sigma \rangle, \quad (2.63)$$

para os termos fora da diagonal.

Nas equações (2.62) e (2.63) utilizamos as seguintes definições :

- $U_{\mu\mu}$ são as energias monoelétrônicas de um centro;
- $P_{\mu\nu}$ são os elementos da matriz de ordem de ligação ;
- $\langle \mu\mu|\nu\nu \rangle$ e $\langle \mu\nu|\mu\nu \rangle$ são as integrais de Coulomb e de troca de dois centros, respectivamente;
- $V_{\mu\nu}^B$ representa o termo de atração eletrostática de um centro;
- $\langle \mu\nu|\lambda\sigma \rangle$ são as integrais de repulsão de dois centros.

Podemos, neste ponto, escrever a energia molecular total como a soma da energia eletrônica e a energia de repulsão entre os “caroços” dos átomos A e B

$$E_{total} = E_{el} + \sum_A \sum_{A < B} E_{AB}^{caroço}. \quad (2.64)$$

No MNDO, ao invés do termo coulombiano $\frac{Z_A Z_B}{R_{AB}}$, onde R_{AB} é a distância entre os caroços, utiliza-se a expressão

$$E_{AB}^{MNDO} = Z_A Z_B \langle s_A s_A | s_B s_B \rangle [1 + \exp(-\alpha_A R_{AB}) + \exp(-\alpha_B R_{AB})] \quad (2.65)$$

onde s_A e s_B representam funções atômicas do tipo s e α_A e α_B são parâmetros ajustáveis do modelo.

Já o calor de formação C_F pode ser escrito como

$$C_F = E_{total} - \sum_A E_{el}^A + \sum_A C_F^A, \quad (2.66)$$

onde E_{total} é a energia total da molécula, E_{el}^A e C_F^A são, respectivamente, as energia eletrônica e calor de formação obtidos experimentalmente para cada um dos átomos da molécula.

Os termos da matriz de Fock e as energias de repulsão $E_{AB}^{caroço}$, ao invés de determinados analiticamente, são determinados através de dados experimentais ou de expressões semi-empíricas que contêm parâmetros ajustáveis aos dados obtidos no experimento. A princípio, a introdução destes parâmetros compensará deficiências básicas da função de onda total (é bom lembrar que ela não leva em consideração a correlação eletrônica) e erros causados, principalmente, pelas simplificações do método.

Para alguns átomos, as energias teóricas de vários estados de valência e seus íons são ajustados de acordo com os valores espectroscópicos experimentais, ajustando, desta forma, os termos de um centro. Estes valores são diferentes dos obtidos analiticamente por outros métodos. Já as integrais de repulsão de dois centros, que representam a energia de interação entre cargas ($e\varphi_\mu\varphi_\nu$) no átomo A e carga ($e\varphi_\lambda\varphi_\sigma$) no átomo B , não podem ser calculadas analiticamente por conta das aproximações básicas do MNDO. Mas existe uma forma de se contornar isso, pois estas integrais podem ser vistas, classicamente, como a soma sobre todas as interações de momentos de multipolo M_{lm} (l e m especificam a ordem e a orientação dos multipolos, respectivamente.) de

duas distribuições de carga. Isto significa dizer que estas integrais de repulsão de dois centros, podem ser expandidas em termos das interações multipolo-multipolo. Matematicamente,

$$\langle \mu\nu | \lambda\sigma \rangle = \sum_{l_1} \sum_{l_2} \sum_m \left[M_{l_1 m}^A, M_{l_2 m}^B \right]. \quad (2.67)$$

Cabe ressaltar que cada multipolo M_{lm} é representado a partir de uma configuração $[M_{lm}]$ de 2^l pontos de carga $E/2^L$ com separações D_l . A interação entre dois multipolos ($[M_{l_1 m}^A, M_{l_2 m}^B]$), é, então, calculada pela aplicação de uma fórmula semi-empírica para cada uma das interações entre os pontos de carga nas duas configurações e, através da soma sobre todos eles. Portanto, se chamarmos de R_{ij} a distância entre os pontos de carga i e j nas configurações de interação entre os átomos A e B , respectivamente, teremos

$$\left[M_{l_1 m}^A, M_{l_2 m}^B \right] = \frac{e^2}{2^{l_1+l_2}} \sum_{i=1}^{2^{l_1}} \sum_{j=1}^{2^{l_2}} f_l(R_{ij}), \quad (2.68)$$

onde $f_l(R_{ij})$ é uma expressão semi-empírica apropriada. De forma análoga e, em linhas gerais, os termos $V_{\mu\mu}^B$, atração elétron-carço, e $E_{AB}^{caroço}$, repulsão entre caroços, também são parametrizados a partir de parâmetros ajustáveis.

Uma vantagem do MNDO é o fato de que ele considera todos os termos de dois centros envolvendo o recobrimento diferencial monoatômico, incluindo interações de multipolos de ordem superior. Portanto, na aproximação MNDO, existe uma dependência angular para os termos de dois centros devido às possíveis orientações dos multipolos. Em outras palavras, se estivermos lidando com sistemas moleculares, podemos dizer que existe uma dependência angular para a repulsão elétron-elétron e a atração caroço-elétron dentro da molécula. Comparado aos seus antecessores, cuja aproximação considera apenas interações de monopolo para a distribuição de cargas e, cuja repulsão elétron-elétron e a atração caroço-elétron são dadas a partir de uma média esfericamente simétrica, o MNDO, mostra-se superior no que diz respeito às propriedades moleculares, uma vez que os efeitos direcionais desempenham papel fundamental neste tipo de sistemas. Em contrapartida, o método possui falhas no que diz respeito à descrição das ligações de hidrogênio e superestima as repulsões entre átomos não ligados quando estes estão mais afastados que a distância de van der Waals. Uma das formas de se corrigir este problema seria modificar a forma de repulsão caroço-carço, o que é feito no seu método sucessor, o AM1.

2.2.3 Austin Method 1- AM1

Como já mencionamos, acima, a principal diferença entre MNDO e AM1 está na forma do termo de repulsão de caroço. Ao invés de utilizar a equação (2.65) utiliza-se a expressão

$$E_{AB}^{AM1} = Z_A Z_B \langle s_A s_A | s_B s_B \rangle [1 + \chi(A) + \chi(B)] \quad (2.69)$$

com

$$\chi(A) = \exp(-\alpha_A R_{AB}) + \sum_i \kappa_{Ai} \exp[L_{Ai} (R_{AB} - M_{Ai})^2] \quad (2.70)$$

$$\chi(B) = \exp(-\alpha_B R_{AB}) + \sum_j \kappa_{Bj} \exp[L_{Bj} (R_{AB} - M_{Bj})^2] \quad (2.71)$$

onde α_A , α_B , κ_{Ai} , κ_{Bj} , L_{Ai} , L_{Bj} , M_{Ai} e M_{Bj} são os parâmetros ajustáveis.

O formalismo usado no AM1 é, essencialmente, o mesmo do MNDO, com exceção, é claro, do termo para a energia de repulsão caroço-caroço. As integrais de repulsão eletrônica de um centro permanecem inalteradas. No entanto, o AM1 mostrou-se superior ao MNDO, principalmente, porque a geometrias de equilíbrio, obtidas do método, estão em maior concordância com os resultados experimentais.

2.2.4 Parametric Method 3-PM3

Como vimos, os métodos MNDO e AM1 são bastante similares, suas estruturas teóricas diferem somente na forma do termo de repulsão caroço-caroço. Já o método PM3 difere do AM1 somente nos conjuntos de valores experimentais e funções de referência utilizados. Funções de referência são quantidades a-dimensionais que podem ser identificadas com algum fenômeno observável experimentalmente. Propriedades moleculares são utilizadas para definir estas funções. Dentro do PM3, são basicamente três os tipos de funções de referência:

- **Funções de referência experimentais:** grande parte delas são derivadas de dados experimentais. Entretanto, em alguns casos, os resultados obtidos a partir de cálculos *ab initio* bastante refinados são utilizados onde os dados experimentais são ausentes ou de baixa qualidade.

- **Funções de referência calculadas:** os valores para propriedades moleculares calculadas usando o conjunto de parâmetro disponíveis também são utilizados para se construir funções de referência.
- **Funções de referência simplificadas:** existem sérias limitações para a utilização dos parâmetros do método dentro de uma equação. É impossível se prever o resultado da modificação de um determinado parâmetro numa função de referência calculada sem que seja efetuado um cálculo completo de MNDO, ou, até mesmo uma otimização geométrica. Logo, é extremamente útil representar as equações do MNDO numa forma mais simples possível, de maneira que possa ser utilizada em qualquer tipo de cálculo que envolve modificação dos parâmetros. Geralmente escreve-se a dependência da função de referência em termos dos parâmetros usados na descrição do método, como um simples somatório

$$q_i(\Delta P) = c_i + \sum_j \sum_k a_{ijk} \Delta P_j^k \quad \text{Função de Ref. Simplificada} \quad (2.72)$$

onde c_i é o calor de formação calculado e ΔP_j no parâmetro j entre o ponto no qual q_i foi avaliado, P_j^0 , usando MNDO e o valor, P_j , utilizado no método, isto é, $\Delta P_j = (P_j^0 - P_j)$.

Não entraremos em detalhes sobre as características destas funções⁴. O que deve ser destacado é o fato de que as funções mais adequadas para a descrição do sistema são obtidas a partir de um processo de otimização e os parâmetros são ajustados de modo a minimizar o quadrado da diferença entre os valores calculados e as funções experimentais de referência.

Existe um consenso de que tanto o AM1, como o PM3, são ambos melhores que o MNDO. Entretanto, quando comparamos o AM1 e o PM3 não é possível determinar, de forma absoluta, qual o melhor. No capítulo 3 utilizaremos um trabalho experimental para avaliar qual dos dois métodos é o mais apropriado para a determinação das estruturas geométricas de menor energia para o conjunto de moléculas deste trabalho.

2.3 Metodologias de Análise Multivariada

Nesta seção faremos uma breve introdução sobre dois dos principais métodos de análise multivariada, o PCA (*Principal Component Analysis*) e o HCA (*Hierarchical Cluster Analysis*).

⁴Para mais detalhes [43, 44].

O intuito aqui, não é introduzir formalmente estes dois métodos de análise, mas sim, delinear as estruturas matemáticas que dão forma aos dois⁵.

O PCA é um método largamente utilizado em problemas de reconhecimento de padrões, desde a quimiometria aplicada à classificação da atividade biológica de fármacos, até à seleção de qualidade de alimentos. Imaginemos que um determinado grupo de objetos possa ser descrito por um conjunto de propriedades comuns a todos eles. Podemos utilizar o PCA como uma ferramenta para seleção das propriedades que mais contribuem, tanto para a semelhança, como para a distinção entre estes objetos. Já o HCA é uma metodologia que determina agrupamentos (*clusters*) a partir do nível de similaridade entre os objetos.

2.3.1 Análise de Componentes Principais

Suponhamos que dispomos de um conjunto de n moléculas e que, além disso, de alguma forma, calculamos, ou medimos, q propriedades para cada uma delas. O que podemos concluir a respeito do nosso conjunto de moléculas a partir, somente, de suas propriedades calculadas? Para tentar responder esta pergunta, podemos, inicialmente, organizar os dados disponíveis numa tabela, como segue

	P ₁	P ₂	⋯	P _q
M ₁	x_{11}	x_{12}	⋯	x_{1q}
M ₂	x_{21}	x_{22}	⋯	x_{2q}
⋮	⋮	⋮	⋱	⋮
M _n	x_{n1}	x_{n2}	⋯	x_{nq}

Tabela 2.1: Tabela com as propriedades (P_{*i*}) para cada uma das moléculas (M_{*i*}).

Sem qualquer perda de informação, ao invés de escrevermos a tabela 2.1, podemos escrever o conjunto de propriedades calculadas utilizando notação matricial,

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1q} \\ x_{21} & x_{22} & \cdots & x_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nq} \end{pmatrix} \quad (2.73)$$

Existem duas formas de se interpretar a matriz acima:

⁵O leitor interessado numa discussão mais detalhada e formal deve consultar [39].

1. Cada um das linhas seria um vetor de q coordenadas $x_i^T = (x_{i1}, x_{i2}, \dots, x_{iq})$ num espaço q -dimensional,

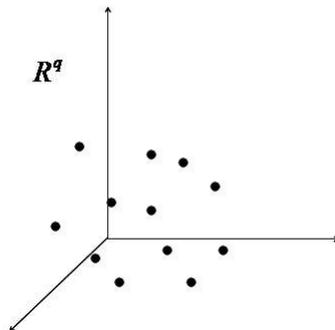


Figura 2.1: Conjunto de n pontos num espaço q -dimensional

ou seja, a matriz (2.1) pode ser vista como um conjunto de n pontos num espaço q -dimensional.

2. Cada um das colunas seria um vetor de n coordenadas $x_j = (x_{1j}, x_{2j}, \dots, x_{nj})^T$ num espaço n -dimensional,

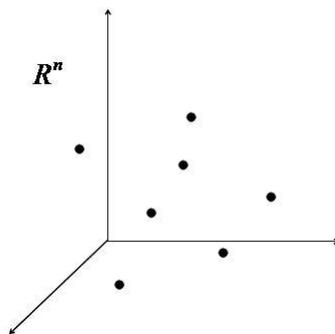


Figura 2.2: Conjunto de q pontos num espaço n -dimensional

ou seja, a matriz (2.1) pode ser vista como um conjunto de q pontos num espaço n -dimensional.

Adotaremos, neste texto, a primeira das interpretações. Note que, se n ou q forem maiores do que 3, perdemos a capacidade de construir representações gráficas visualmente representáveis destes conjuntos. Entretanto, esse mesmo conjunto de pontos pode ser representado a partir

de uma projeção em espaços de dimensão, convenientemente, menor. De fato, a expressão matemática para a projeção de pontos x_i numa reta L_1 é dada por

$$p_{x_i} = x_i^T \frac{u_1}{\|u_1\|} = x_i^T u_1, \quad (2.74)$$

onde u_1 é o vetor unitário na direção de L_1 .

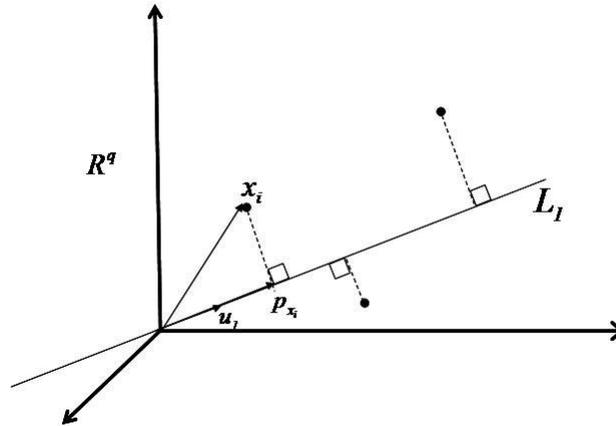


Figura 2.3: Projeção dos pontos na reta L_1

Mas qual será a melhor reta para efetuar esta projeção? Em outras palavras: qual reta melhor representa a distribuição dos pontos em \mathbb{R}^q ? A resposta para esta pergunta pode ser dada da seguinte forma: a melhor reta para a projeção dos pontos x_i é dada pela reta que minimiza a soma do quadrado da distância ortogonal de cada ponto x_i à reta L_1 , ou seja, a expressão

$$\sum_{i=1}^n \|x_i - P_{x_i}\|^2 = \sum_{i=1}^n (\|x_i\|^2 - \|P_{x_i}\|^2), \quad (2.75)$$

deve ser mínima, ou, equivalentemente,

$$\sum_{i=1}^n \|P_{x_i}\|^2 \quad (2.76)$$

deve ser máxima. Logo, o que devemos fazer é encontrar $u_1 \in \mathbb{R}^q$ tal que $\sum_{i=1}^n \|P_{x_i}\|^2$ seja máxima, respeitando o vínculo $\|u_1\| = 1$. É possível encontrar u_1 utilizando o seguinte resultado:

Teorema 2.1 *O vetor u_1 , tal que $\|u_1\| = 1$, que minimiza $\sum_{i=1}^n \|x_i - P_{x_i}\|^2$ é o autovetor de $X^T X$ associado ao maior dos autovalores (λ_1) de $X^T X$, onde X é dada por .*

O vetor u_1 é chamado, neste caso, de *primeiro eixo fatorial* e $Xu_1 = z_1$ a *primeira variável fatorial*.

Analogamente, se desejarmos efetuar uma projeção em um plano, pode ser verificado que este novo espaço bidimensional que otimiza a distribuição de pontos, deve conter o vetor u_1 obtido a partir do Teorema 2.1. O melhor plano para projeção será aquele gerado pelos dois primeiros eixos fatoriais. Para encontrar o *segundo eixo fatorial*, u_2 , podemos utilizar o seguinte resultado:

Teorema 2.2 *O segundo eixo fatorial, u_2 é o autovetor de $X^T X$ associado ao segundo maior autovalor, λ_2 , de $X^T X$.*

Um resultado importante é o fato de que $u_1^T u_2 = 0$, ou seja, os dois primeiros eixos principais são ortogonais. O plano gerado pelos vetores u_1 e u_2 terá como novos eixos as duas primeiras variáveis fatoriais, $z_1 = Xu_1$ e $z_2 = Xu_2$, como mostra a próxima figura.

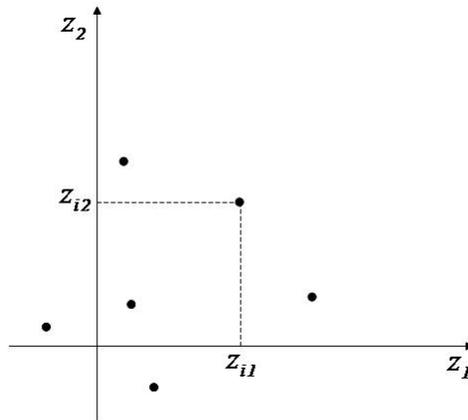


Figura 2.4: Conjunto de pontos projetados no plano gerado pelas duas primeiras variáveis fatoriais.

Podemos dizer, que z_1 e z_2 são duas novas variáveis completamente independentes, assim como dois eixos ortogonais.

De forma inteiramente análoga, se a projeção for num espaço de dimensão s , com $s \leq q$, devemos encontrar os s -primeiros vetores unitários $\{u_1, u_2, \dots, u_s\}$, ou seja, os autovetores de $X^T X$ associados aos s -primeiros autovalores $\{\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_s\}$. Devemos ter em mente que estes vetores unitários são mutuamente ortonormalizados e os novos eixos do espaço s -dimensional serão construídos a partir das novas variáveis vetoriais $z_k = Xu_k = (z_{1k}, z_{2k}, \dots, z_{nk})$ que, na verdade, são combinações lineares das variáveis originais $x_{[i]}$, que, como já vimos, correspondem

à i -ésima linha da matriz X . Matematicamente,

$$z_{ik} = \sum_{m=1}^q x_{im}u_{mk}. \quad (2.77)$$

Portanto é possível projetarmos a distribuição de pontos original, propriamente contida em R^q , em novos eixos, construídos em termos das variáveis antigas (eq. (2.77)), de forma que esta projeção contenha o máximo de informação sobre a distribuição original. No entanto, o verdadeiro intuito do PCA é selecionar as variáveis que mais contribuem, tanto para a semelhança, como para as diferenças entre as moléculas da Tabela 2.1. Logo, o PCA também é uma técnica de redução dimensional.

No caso de n moléculas com q variáveis utilizadas para sua descrição, não sabemos, *a priori* que tipo informação cada uma destas variáveis fornecem dentro deste espaço multidimensional. É possível que duas, ou mais, destas variáveis, contenham o mesmo tipo de informação para distribuição estatística dos pontos. Suponha, por exemplo, que, para cada molécula (M_i), tenhamos medido uma propriedade arbitrária e, que o resultado da medida está dado abaixo,

$$\begin{array}{l} M_1 \rightarrow 1 \\ M_2 \rightarrow 2 \\ \vdots \quad \vdots \quad \vdots \\ M_n \rightarrow n \end{array}$$

Suponha, também, que para uma outra propriedade arbitrária, tenham sido encontrados os seguintes valores

$$\begin{array}{l} M_1 \rightarrow 2 \\ M_2 \rightarrow 4 \\ \vdots \quad \vdots \quad \vdots \\ M_n \rightarrow 2n \end{array}$$

Podemos notar que estes dois conjuntos de variáveis contribuem da mesma forma para a distribuição dos pontos no espaço, uma vez que, são, na verdade, múltiplos um do outro. Logo, a inclusão destes dois conjuntos de dados tem o mesmo efeito que a inclusão de apenas um, o que significa, que uma delas pode ser descartada. Este é o espírito do PCA. Primeiro procura-se que variáveis estão correlacionadas, ou seja, as que contribuem aproximadamente da mesma forma,

para que uma delas seja eliminada e, assim, consigamos uma mesma distribuição de pontos, mas com um número menor de variáveis.

Imagine que saibamos, de antemão o grau de importância de cada uma das variáveis da matriz X . Matematicamente, poderíamos evidenciar esta importância através de uma média ponderada

$$\delta^T X = \sum_{j=1}^q \delta_j X_j \quad \text{tal que} \quad \sum_{j=1}^q \delta_j^2 = 1. \quad (2.78)$$

Mas se não soubermos, inicialmente, qual o peso de cada uma das variáveis, como podemos determinar a importância de cada uma delas? A resposta é: otimizando o vetor $\delta = (\delta_1, \delta_2, \dots, \delta_q)^T$. Na verdade, busca-se o vetor δ que maximiza a variância da projeção $\delta^T X$, ou seja, encontramos δ tal que

$$\max_{\{\delta: \|\delta\|=1\}} \text{Var}(\delta^T X) = \max_{\{\delta: \|\delta\|=1\}} \delta^T \text{Var}(X) \delta. \quad (2.79)$$

A matriz $\text{Var}(X)$ é dada por

$$\text{Var}(X) = \Sigma = \Gamma \Lambda \Gamma^T, \quad (2.80)$$

onde

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_q \end{pmatrix} \quad (2.81)$$

é a matriz dos autovalores de $X^T X$ e

$$\Gamma = \begin{pmatrix} \gamma_1 & \gamma_2 & \cdots & \gamma_q \end{pmatrix} = \begin{pmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1q} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{n1} & \gamma_{n2} & \cdots & \gamma_{nq} \end{pmatrix} \quad (2.82)$$

é a matriz cujas colunas são formadas pelos autovetores de $X^T X$. Logo, podemos reescrever a expressão (2.80) como

$$\Gamma \Lambda \Gamma^T = \sum_{j=1}^q \lambda_j \gamma_j \gamma_j^T. \quad (2.83)$$

A maximização de $Var(X)$ pode ser entendida como uma forma de melhor separar as variáveis do sistema, ou seja, a otimização de δ nos fornece a melhor forma de se projetar as variáveis numa linha reta e, esta reta, tem a mesma orientação de δ . O procedimento para se encontrar δ é análogo ao já descrito para o caso dos eixos fatoriais, pois, na verdade, ele é o autovetor de $\Gamma\Lambda\Gamma^T$ associado ao maior de seus autovalores. Se, ao invés de projetarmos numa reta, desejarmos uma projeção num plano, devemos encontrar o autovetor ξ associado ao segundo maior autovalor de $\Gamma\Lambda\Gamma^T$. Os eixos formados a partir de δ e ξ são comumente chamados de $PC1$ e $PC2$, pois representam, respectivamente, as novas variáveis que contribuem para a maior separação dos pontos no espaço. Devemos ter em mente que, tanto $PC1$ quanto $PC2$, são novos eixos descritos em termos de uma combinação linear das variáveis antigas, como uma espécie de média ponderada destas variáveis, ou seja, o quadrado dos coeficientes da combinação linear são uma estimativa do grau de importância de suas respectivas variáveis para a projeção dos pontos no novo plano. Desta forma, variáveis com maiores coeficientes são mais importantes para a distribuição dos pontos. Outro ponto importante é o fato de que, quando visualizamos a projeção destes pontos no plano formado pelos dois vetores de maior contribuição para a maximização da variância, estamos olhando para distribuição dos pontos onde há, também, a maximização da informação sobre o sistema.

No entanto, o descarte de variáveis não é um processo automatizado. É necessário que identifiquemos estas variáveis e as descartemos. A seguir será apresentado um resumo sobre a técnica do PCA, bem como os procedimentos utilizados para a busca dos descritores de maior relevância, ou seja os procedimentos que fazem do PCA um técnica de redução dimensional.

2.3.2 Resumo sobre a utilização do PCA

Se para cada uma das n amostras houver um conjunto de q -variáveis, cada uma destas amostras pode ser representada por um ponto num espaço q -dimensional, onde cada uma das variáveis está associada a uma dimensão. Não podemos garantir que todas estas variáveis estão decorrelacionadas, ou seja, que os eixos destes espaço q -dimensional sejam mutuamente ortogonais. Variáveis que não são "ortogonais" entre si, possuem informações semelhantes, que conseqüentemente contribuem de forma semelhante para a distribuição de pontos no espaço. O principal objetivo do PCA é construir novos eixos em termos das variáveis antigas, que sejam ortogonais entre si e que forneçam o máximo possível de informação sobre as amostras.

A princípio, as variáveis que descrevem as amostras têm variações de unidades muito diferentes. Por isso é necessário que todas elas sejam transformadas, de maneira que possuam o mesmo domínio de variação. Na mais comum das transformações, todas as variáveis possuem valores entre 0 e 1. Feito isso, o PCA busca os novos eixos que melhor representam o sistema. Alguns vínculos devem ser respeitados nesta busca: a dispersão deve ser mantida e como já vimos os novos eixos devem ser descorrelacionados (ortogonais). A distribuição estatística dos pontos está relacionada com a variância da distribuição dos pontos e novos eixos (componentes principais) são obtidos a partir da maximização da variância. O resultado final é a obtenção de q novas variáveis (escritas como combinação linear das variáveis antigas) que contêm em ordem decrescente, a informação estatística sobre o sistema inicial, isto é, a componente principal PC1 contém maior parte da informação estatística, seguida pela componente PC2 e, assim, sucessivamente até a PC q . A visualização é comumente feita na projeção dos n pontos no plano gerado pelas duas primeiras componentes principais.

Entretanto o método não é capaz de, automaticamente, selecionar as variáveis mais importantes, aquelas que contribuem mais para a separação das amostras. Os resultados do PCA são apresentados pelo conjunto de amostras (*score*) e um conjunto de variáveis, também conhecido como descritores (*loading*), representados no plano formado por PC1 e PC2. A busca pelas variáveis mais importantes deve ser feita utilizando conjuntamente os gráficos de *score* e de *loading*. Por exemplo, as variáveis que estão mais à direita no gráfico de *loading*, apresentam grandes valores nas amostras mais à direita no gráfico de *score*. O mesmo vale para as variáveis localizadas à esquerda no gráfico de *loading*. Em contrapartida as variáveis que estão no centro do gráfico de *loading* contribuem muito pouco para a distinção das propriedades das amostras, ou seja, elas contribuem muito pouco para a distribuição dos pontos no plano $PC1 \times PC2$. Estas variáveis, portanto, podem ser descartadas, o que significa uma redução no número de variáveis necessárias para caracterização do sistema e uma conseqüente redução do espaço de representação. O descarte de variáveis deve cessar quando percebe-se que cada uma das variáveis restantes é importante para a separação das amostras no gráfico de *score*.

2.3.3 Análise Hierárquica de Clusters- HCA

O HCA- *Hierarchical Cluster Analysis* é um conjunto de técnicas de construção de grupos (*clusters*) a partir de um conjunto de dados multivariados. Na verdade, o objetivo é construir grupos

com propriedades homogêneas a partir de um grande conjunto de dados, aparentemente, heterogêneo. Os grupos, ou clusters, devem ser o mais homogêneo possíveis e as diferenças entre os diversos clusters deve ser a maior possível. Podemos dividir o HCA em duas etapas fundamentais:

1. Escolha da medida de proximidade (similaridade): Checa-se para cada par de objetos a similaridade entre os seus valores. Esta similaridade é definida como uma medida da proximidade dos objetos. Quanto maior a similaridade, mais próximos eles estão e, mais homogêneos serão.
2. Escolha do algoritmo de construção dos grupos: O algoritmo deve ser capaz de separar os objetos em grupos, de tal forma que as diferenças entre grupos distintos e a semelhança entre membros de um mesmo grupo, sejam as maiores possíveis.

A similaridade é definida a partir da menor distância entre duas amostras (d_{ij}) (ou grupos de amostras) e da maior distância (d_{max}) entre quaisquer das amostras da distribuição, de acordo com a equação:

$$S_{ij} = 1 - \frac{d_{ij}}{d_{max}}. \quad (2.84)$$

A distância entre dois pontos é calculada segundo a métrica Euclidiana⁶ pela expressão

$$d_{ij} = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \dots + (x_{pi} - x_{pj})^2},$$

logo quanto menor a distância entre dois pontos, maior a similaridade entre eles. O resultado do HCA, geralmente, é dado a partir de um dendograma como o mostrado na figura 2.5

⁶Na verdade, outros tipos de Métrica podem ser utilizadas. Para mais detalhes ver referência [39].

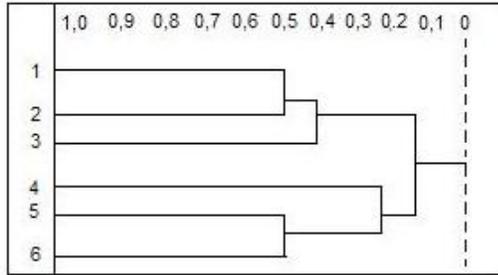


Figura 2.5: Representação típica de um dendrograma obtido pelo método HCA.

Em linhas gerais, o algoritmo de busca dos grupos pode ser resumido a partir do roteiro abaixo.

2.3.4 Esboço sobre o algoritmo do HCA

1. O primeiro passo é calcular as distâncias entre todos os pares de pontos, ou seja, para cada ponto p_i , calcula-se a distância d_{ij} , com $i \neq j$, entre o ponto p_i e cada um dos pontos p_j .
2. Cada um dos pares p_{ij} de menor distância formará um grupo, que será representado por um novo ponto p_k , localizado na metade da distância entre cada um dos pares de pontos.
3. Calcula-se a distância entre os pares de grupos, formados pelos novos pontos p_k (grupos), ou seja, d_{kl} , com $l \neq k$.
4. Cada um dos pares de grupos de menor distância formará um novo grupo (ponto p_m), localizado na metade da distância entre grupos p_k e p_l encontrados no item 2.
5. Repetem-se os mesmos procedimentos descritos nos itens 3 e 4 até que todos os pontos sejam inseridos em um dos grupos.

Desta forma, encontramos como resultado final, grupos, que na verdade são representados por pontos no espaço, cuja similaridade entre amostras e entre grupos é dada pela expressão (2.84). Logo, a similaridade dos pontos pertencentes a grupos distintos é dada pela similaridade entre os grupos. Em contrapartida, elementos de um mesmo grupo podem possuir similaridades distintas entre si, uma vez que, as distâncias entre si, em \mathbb{R}^q , podem ser diferentes.

Uma nova Metodologia

Objetivo deste trabalho é investigar sistematicamente propriedades (eletrônicas e físico-químicas) de um conjunto de moléculas utilizados como fármacos, com o intuito de determinar regras e padrões indicadores de uma correlação entre atividade biológica destes fármacos e características intrínsecas de sua estrutura molecular.

A evidência experimental de que pequenas alterações na estrutura química de compostos de uma determinada família são responsáveis por grande mudanças em sua atividade biológica, deixa claro que este tipo de investigação é extremamente importante. Sistemas biológicos, são extremamente complexos, caracterizados, principalmente por interações de moléculas orgânicas. Apesar da complexidade envolvida no processo de interação, no caso específico, composto-alvo biológico e sua rota metabólica no organismo, as propriedades moleculares dos fármacos são fundamentais para a definição do caráter de atividade que ele desempenha. Portanto, um estudo, mesmo que qualitativo, fundamentado em parâmetros que dependam diretamente da conformação e dos átomos constituintes de um composto, representa um passo rumo a uma melhoria no processo de formação e seleção de novos fármacos.

2.4 A Metodologia dos Índices Eletrônicos

De maneira geral, os métodos que estudam a relação estrutura-atividade diferem nos parâmetros moleculares utilizados para a investigação da atividade biológica. No entanto, o objetivo comum de qualquer um destes métodos, é encontrar regras que possibilitem fazer previsões confiáveis sobre a atividade de compostos não sintetizados ou não testados, auxiliando, desta forma, na escolha entre vários candidatos a serem sintetizados e principalmente, na construção de novas estruturas.

Mas dentre tantos parâmetros possíveis, quais são aqueles capazes de nos fornecer com segurança uma regra que diferencie os compostos estudados segundo sua atividade biológica?

A Metodologia de Índices Eletrônicos (MIE), proposta por Barone, Camilo e Galvão [41], foi desenvolvida, inicialmente, com o objetivo de distinguir qualitativamente os Hidrocarbonetos Policíclicos Aromáticos em termos de sua atividade carcinogênica (causadores ou não). Empregando regras extremamente simples, que envolvem parâmetros puramente eletrônicos,



Figura 2.6: Diante de tantas propriedades, quais são as mais importantes?

a MIE surge como uma alternativa extremamente eficiente⁷ para o estudo da relação estrutura-atividade de fármacos em geral.

Muito utilizado na química quando se lida com a interação entre dois compostos, o conceito de Chave-Fechadura, pode ser, de certo modo, aplicado dentro da MIE. Como veremos, a Metodologia de Índices Eletrônicos leva em consideração as diferentes contribuições dadas por cada uma das principais regiões moleculares na correlação de sua estrutura com sua atividade biológica, respeitando, de certa forma o caráter local (depende da região molecular) das interações intermoleculares e traz, diferentemente das técnicas estatísticas de análise multivariada, um caráter intuitivo sobre os mecanismos básicos de interação entre dois compostos.

É razoável pensar que os orbitais de fronteira de duas moléculas interagentes desempenhem papel extremamente importante no que diz respeito à afinidade entre elas. Pensando nisso, a MIE foi construída baseada nos conceitos de Densidade Local de Estados (*LDOS-Local Density of States*) e nos valores críwticos da diferença de energia 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. A LDOS, é justamente, o conceito que nos permite estudar a contribuição de propriedades eletrônicas locais para o estudo da relação estrutura-atividade de um composto, pois é a densidade de estados, calculada sobre uma região molecular específica. A LDOS nos permite mapear a distribuição de um orbital molecular, ocupado ou virtual, sobre as diversas porções do esqueleto molecular. Em outras palavras, a LDOS é uma indicação de como um átomo, ou um conjunto de átomos, contribui para a formação de um

⁷Ver o resumo dos resultados obtidos na tabela 2.4.

determinado orbital molecular.

Para Hamiltonianos baseados na aproximação de combinação linear de orbitais atômicos (LCAO),

$$\psi_m = \sum_i c_{mi} \varphi_i \quad (2.85)$$

os cálculos da LDOS envolvem a contribuição de cada átomo para um nível eletrônico específico medida pelo coeficiente do orbital molecular elevado ao quadrado, ou seja, pela densidade de probabilidade correspondente ao nível, naquele sítio molecular. Para cálculos envolvendo mais de um orbital ou várias regiões moleculares a LDOS é obtida somando-se as contribuições dos orbitais atômicos selecionados

$$LDOS(E_i) = 2 \sum_{m=n_i}^{n_f} |c_{mi}|^2. \quad (2.86)$$

Na equação acima, o fator dois que multiplica o somatório é consequência direta do Princípio de Exclusão de Pauli. A obtenção da LDOS exige a divisão da molécula em regiões. No entanto essa divisão deve ser criteriosa, considerando estruturas comuns a todos os compostos do conjunto estudado (estruturas que pertençam ao esqueleto básico) e que desempenhem um papel reconhecidamente importante do ponto de vista químico e, ou, biológico.

Anteriormente surgiu a questão: Quais variáveis contribuem mais significativamente para a classificação de uma molécula como ativa ou inativa para uma determinada atividade biológica? O que a MIE propõe é:

Apenas dois parâmetros são capazes de nos fornecer, com eficiência, instrumentos que nos permitam classificar um determinado conjunto de moléculas estruturalmente semelhantes, segundo sua atividade biológica.

Essas duas variáveis são:

$$\begin{aligned} \eta &= LDOS_{nível\ 1} - LDOS_{nível\ 2} \\ &= 2 \sum_{m=n_i}^{n_f} (|c_{m\ nível\ 1}|^2 - |c_{m\ nível\ 2}|^2), \end{aligned} \quad (2.87)$$

uma medida da diferença da contribuição dos átomos de uma determinada região molecular para a formação de dois níveis moleculares;

$$\Delta = E_2 - E_1 \quad (2.88)$$

a separação, em termos da energia, dos dois níveis moleculares da expressão (2.87).

É importante notar que, enquanto o parâmetro Δ contém informação global sobre a molécula em termos das auto-energias moleculares, o parâmetro η contém informações locais que refletem a distribuição de cargas sobre as diversas regiões moleculares.

Os orbitais moleculares de interesse na investigação utilizando 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 (principalmente HOMO e LUMO) está relacionada com sua importância na reatividade dos compostos, uma vez que estes são envolvidos nas reações químicas entre moléculas. A figura 2.7 ilustra como os parâmetros Δ e η podem ser obtidos diretamente dos gráficos para DOS e LDOS. Note que cada uma das linhas verticais representa um nível de energia. O parâmetro η é dado pela diferença de altura entre as linhas que descrevem o HOMO-1 e HOMO e LUMO e LUMO+1. Já o parâmetro Δ pode ser encontrado a partir da distância (horizontal) entre os níveis consecutivos.

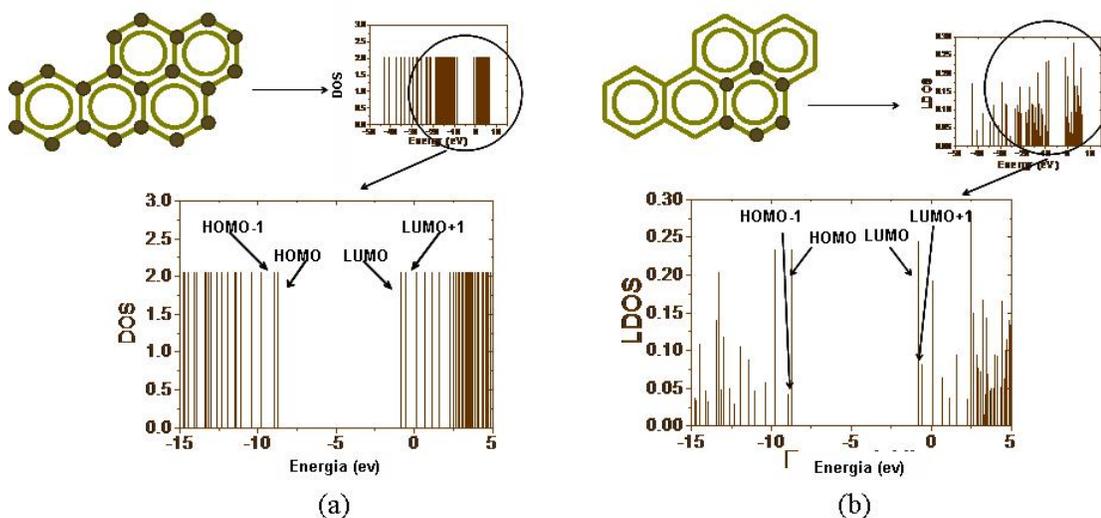


Figura 2.7: Espectro eletrônico através de cálculos de (a) DOS e (b) LDOS realizados nas regiões indicadas.

Os parâmetros Δ e η são utilizados, então, para a construir as regras de classificação de atividade biológica dos compostos. Como já mencionamos a regra encontrada é extremamente simples e, tipicamente pode ser escrita como:

Regra: Se $\Delta > \Delta_c$ e $\eta < \eta_c$, a molécula será ativa; caso contrário, será inativa.

Temos que Δ_c e η_c são valores críticos para cada classe específica de moléculas. É importante notar que os valores para η_c podem ser distintos em diferentes regiões, uma vez que cada região contribui de uma forma para a distribuição de cargas. Logo, deve ficar claro que a regra encontrada, está associada a uma das regiões estudadas da molécula. As regras podem ser melhor sumarizadas com o emprego de Tabelas Booleanas. Por exemplo, a regra descrita acima pode ser resumida na tabela 2.2, onde os sinais + e - indicam se a condição para o parâmetro, especificado na regra, é satisfeita (+) ou não (-).

Δ	η	Atividade prevista pela MIE
+	+	Ativa
+	-	Inativa
-	+	Inativa
-	-	Inativa

Tabela 2.2: Tabela Booleana para classificação de atividade dos compostos.

Para ilustrar a forma de atuação da MIE, imaginemos que estamos estudando um conjunto de dez moléculas $\{M_1, M_2, \dots, M_{10}\}$, cuja atividade biológica foi determinada experimentalmente. Suponhamos que, baseados nos dados experimentais, possamos separar nosso conjunto de moléculas da seguinte forma

Atividade	Moléculas
Ativas	$M_1 M_2 M_3 M_4 M_5$
Inativas	$M_6 M_7 M_8 M_9 M_{10}$

Tabela 2.3: Classificação hipotética do conjunto de dez moléculas.

Para utilizar a MIE devemos, primeiramente, dividir as moléculas, que por hipótese possuem esqueleto comum, em regiões a partir de critérios que devem ser definidos caso a caso. Definimos os orbitais de fronteira mais relevantes para o problema. Feito isso calculamos os parâmetros Δ e η para cada uma das regiões em cada uma das moléculas. Com isso obtemos tabelas que relacionam as moléculas com seus respectivos valores de Δ e η , calculados sobre as regiões pré-determinadas. O que falta, agora, é determinar os valores críticos Δ_c e η_c que reproduzam, com máxima fidelidade, a classificação obtida a partir dos dados experimentais (tabela 2.3). Deve

ficar claro que, para diferentes regiões, em geral, obteremos diferentes valores críticos e diferentes percentuais de acerto com respeito à classificação da tabela 2.3. A região com maior percentual fornecerá a regra da MIE.

A Metodologia de Índices Eletrônicos vem sendo aplicada no estudo de diferentes classes de compostos orgânicos com atividades biológicas distintas. O sucesso da utilização da MIE pode ser verificada a partir da tabela 2.4. Note que o índice de acerto é bastante elevado, principalmente, se destacarmos que são utilizados apenas dois parâmetros para a classificação .

Classe de Compostos	Ação Biológica	Número de Moléculas	PCA (%)	MIE (%)
HAP	Carcinogênico	81	83.3	87.2
Lapachol	Anti-tumoral Antibiótico Anti-malarial	25	-	96.0
Ácidos Naftóicos	Inibidor da Integrase no HIV	6	-	100
Talidomidas	Anti-inflamatório	32	-	90.0
Esteróides	Afinidade de ligação da CBG	31	100	100
Acetoxi-progesteronas	Atividade Progesteronal	21	85.7	90.5
Taxóides	Anti-tumoral	20	100	85.0
Benzo[c]quinolizin-3-onas	Inibidor da 5alpha-reductase	23	-	100
Quinolinas	Inibidor da Integrase no HIV	32	93.75	93.8
Tetraciclinas	Antibiótico	14	100	100

Tabela 2.4: Comparação entre os resultados obtidos pelo PCA e MIE para diferentes classes de compostos [40].

É importante ressaltar que a obtenção dos parâmetros Δ e η deve ser feita somente após termos garantias de que todas as estruturas moleculares do conjunto a ser estudado correspondem às estruturas advindas de algum critério estabelecido de antemão. No nosso caso, como veremos adiante, exigimos que a geometria final de cada uma das moléculas estudadas neste trabalho fosse um mínimo da energia conformacional.

Capítulo 3

Resultados sobre Investigação Estrutura-Atividade da Combretastatina e derivados

3.1 Estudo da geometria molecular

Introduzimos no capítulo anterior, toda a metodologia que será utilizada no estudo da relação estrutura-atividade do nosso conjunto de moléculas. A próxima etapa é utilizar algum método mecânico-quântico para encontrar a geometria mais estável (de menor energia) da Combretastatina e seus derivados. Denominaremos este processo de otimização da geometria. Dado o número de átomos da Combretastatina e a necessidade de uma busca conformacional refinada, a utilização de métodos *ab initio* seria inviável devido ao custo computacional destes métodos. Uma alternativa seria a utilização de algum método semi-empírico. Na seção 2.2.2, destacamos que os métodos mais utilizados para cálculo de geometrias são o AM1 e o PM3, mas, não possuímos garantias, *a priori*, sobre o melhor desempenho de um ou outro, para o nosso conjunto de moléculas. Devemos buscar critérios seguros que garantam confiabilidade à comparação entre os resultados do AM1 e do PM3 e, uma das formas de se realizar isso, é utilizar dados coletados experimentalmente para moléculas, pelo menos, similares às que devem ser otimizadas. Ou seja, utilizar dados experimentais como parâmetros de comparação entre os resultados obtidos pelos métodos semi-empíricos.

3.1.1 Dados experimentais para um derivado da Combretastatina

Apesar de não haver disponível na literatura, trabalhos experimentais sobre a geometria de nenhuma das moléculas estudadas neste trabalho, dados experimentais para compostos bastante similares foram encontrados [14]. Para avaliar o poder de reprodutibilidade da geometria experimental dos métodos AM1 e PM3, selecionamos a molécula C_{rx} , apresentada na figura 3.1. Esta molécula, apesar de não estar incluída em nosso grupo de análise, possui o mesmo esqueleto molecular central (caroço) que a Combretastatina e seus derivados.

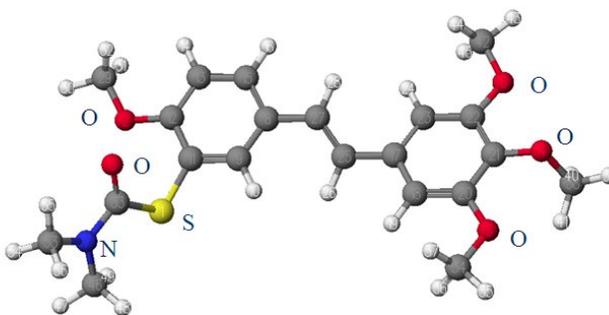


Figura 3.1: Molécula C_{rx} cuja geometria foi obtida através de análise de raio-x[14].

Otimizamos C_{rx} utilizando cada um dos dois métodos semi-empíricos e colhemos os resultados para os seguintes parâmetros estruturais:

- Comprimento de ligação ;
- Ângulo de ligação ;
- Ângulo diedral.

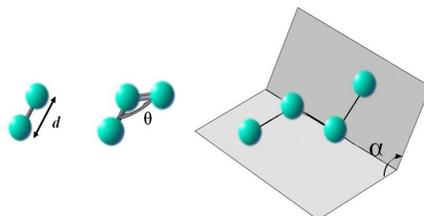


Figura 3.2: Comprimento de ligação, ângulo de ligação e o ângulo diedral formado por dois planos que se intersectam.

Na figura 3.3, abaixo, definiremos a nomenclatura de classificação dos átomos constituintes de C_{rx} .

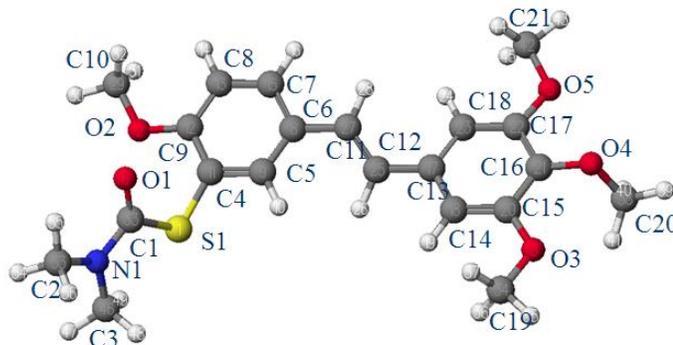


Figura 3.3: Numeração dos átomos que constituem C_{rx} .

Com a nomenclatura utilizada acima e, baseados na figura 3.2, definiremos:

- C_i-O_j , como o comprimento de ligação entre o carbono- i e o oxigênio- j da Figura 3.3;
- $C_i-C_j-C_k$, como o ângulo pelas linhas retas que ligam C_i e C_k a C_j ;
- $C_i-C_j-C_k-C_n$, como o ângulo diedral formado pela interseção dos planos que contém $C_i-C_j-C_k$ e $C_j-C_k-C_n$, respectivamente.

3.1.2 O cristal de C_{rx} .

É de se esperar que haja algumas discordâncias na comparação entre uma geometria obtida experimentalmente e uma geometria obtida a partir de cálculos teóricos num ambiente de vácuo, principalmente, devido ao empacotamento do sistema cristalino.

No entanto, em cristais moleculares, as forças intermoleculares (principalmente, a força de van der Waals) são muito menores do que as forças intra-moleculares. Ou seja, a rigidez da molécula, na maioria das vezes, é suficientemente grande para evitar distorções significativas na sua geometria. Ainda assim, o empacotamento molecular pode provocar distorções substanciais na geometria dos compostos. No caso do cristal estudado em [14] (ver figura 3.4), a disposição das moléculas C_{rx} dentro da célula cristalina contribui para o relaxamento, principalmente dos ângulos diedrais. Se, ao invés de empacotadas como na figura 3.4, as moléculas estivessem

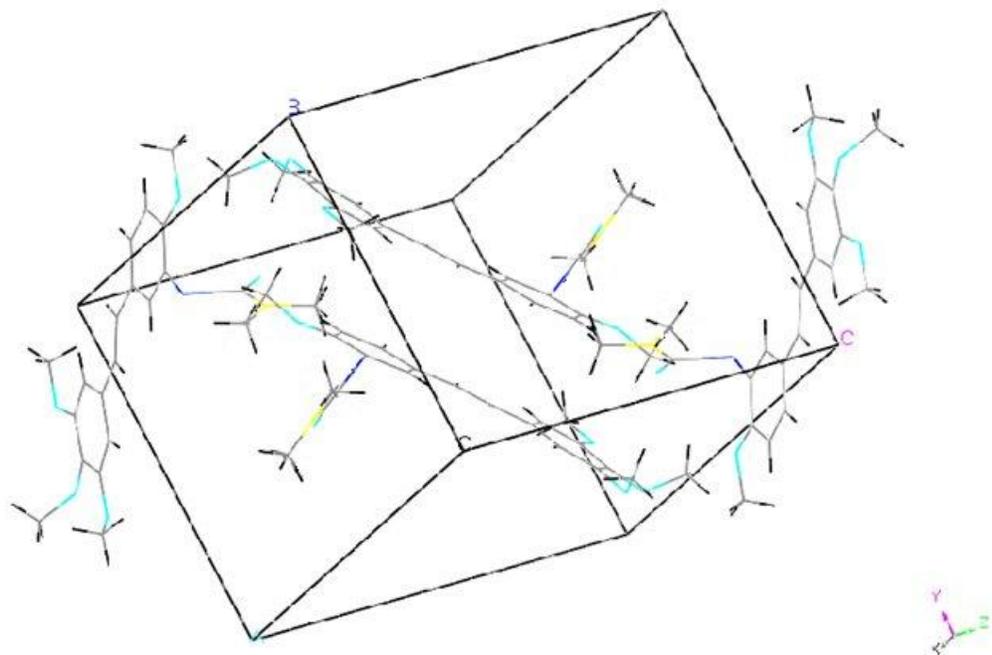


Figura 3.4: Célula do cristal de C_{rx} analisado.

todas sobrepostas dentro do cristal, este tipo de empacotamento, certamente, contribuiria para a planarização destes ângulos e, conseqüentemente para significativos desvios dos valores calculados teoricamente.

Justificamos a utilização dos dados obtidos através do estudo de raio-x do cristal de C_{rx} como fontes de comparação dos resultados teóricos provenientes dos métodos semi-empíricos AM1 e PM3. A próxima seção ocupa-se da análise destes dados com o intuito de garantir quais destes métodos nos fornece os melhores dados estruturais para a descrição da geometria do nosso conjunto de moléculas.

3.1.3 Avaliação dos métodos semi-empíricos.

A seguir apresentaremos, nas tabelas 3.1, 3.3 e 3.2, os resultados obtidos, experimentalmente e após otimização geométrica, para os comprimentos de ligação, ângulos de ligação e ângulos diedrais mais relevantes para a geometria de C_{rx} .

Comprimento de ligação (Å).

	R-X	PM3	AM1		R-X	PM3	AM1
S1-C4	1,771	1,769	1,703	C8-C7	1,386	1,385	1,388
O1-C1	1,212	1,218	1,244	C5-C4	1,373	1,395	1,398
N1-C3	1,448	1,479	1,434	C7-C6	1,387	1,399	1,406
S1-C1	1,793	1,835	1,769	C6-C11	1,466	1,456	1,452
N1-C2	1,457	1,481	1,436	C11-C12	1,313	1,342	1,343
C9-O2	1,360	1,374	1,375	C12-C13	1,464	1,458	1,455
O2-C10	1,428	1,407	1,425	C13-C18	1,390	1,397	1,399
C15-O3	1,375	1,381	1,381	C18-C17	1,379	1,397	1,400
O3-C19	1,386	1,408	1,424	C17-C16	1,394	1,414	1,411
O5-C21	1,414	1,406	1,423	C16-C15	1,386	1,410	1,412
O4-C20	1,402	1,412	1,429	C15-C14	1,382	1,399	1,398
C4-C9	1,397	1,411	1,411	C14-C13	1,390	1,395	1,402
C9-C8	1,384	1,404	1,405	C5-C6	1,393	1,396	1,399

Tabela 3.1: Valores para o comprimento de ligação obtidos experimentalmente e através dos métodos semi-empíricos PM3 e AM1.

Ângulo Diedral (Graus).

	R-X	PM3	AM1		R-X	PM3	AM1
C7-C6-C11-C12	182,22	179,91	164,46	C13-C18-C17-C16	-1,30	0,23	-1,14
C6-C11-C12-C13	183,16	179,94	180,08	C9-C4-S1-C1	-81,89	-73,73	-70,46
C12-C11-C6-C5	2,27	0,21	16,47	C4-S1-C1-N1	179,27	171,09	175,23
C11-C12-C13-C18	5,51	0,29	-26,62	C8-C9-O2-C10	9,39	-6,05	-9,50
C9-C4-C5-C6	0,07	-0,74	-1,19	C14-C15-O3-C19	32,58	18,86	-9,69
C9-C8-C7-C6	0,33	-0,16	-0,04	C17-C16-O4-C20	-109,17	-95,01	-109,92
C11-C12-C13-C14	172,00	179,77	154,32	C16-C17-O5-C21	-177,21	-175,38	-171,17
C13-C14-C15-C16	-0,70	0,24	-0,14				

Tabela 3.2: Valores para o ângulos diedrais obtidos experimentalmente e através dos métodos PM3 e AM1.

Ângulo de ligação (Graus).

	R-X	PM3	AM1		R-X	PM3	AM1
C4-S1-C1	99,84	104,64	105,37	C7-C6-C5	116,45	119,11	118,54
C1-N1-C3	124,70	118,37	120,8	C6-C5-C4	122,37	121,61	121,74
C3-N1-C2	116,80	113,57	117,48	C5-C4-C9	120,03	118,49	118,58
O1-C1-S1	122,00	123,12	124,84	C5-C6-C11	122,77	121,07	122,43
C9-O2-C10	117,70	118,00	116,57	C12-C11-C6	125,90	123,14	125,02
C1-N1-C2	118,40	117,78	120,72	C11-C12-C13	129,10	123,25	124,30
O1-N1-C1	123,50	118,11	121,42	C12-C13-C14	117,83	119,10	118,54
N1-C1-S1	114,50	115,67	113,74	C13-C14-C15	120,85	120,01	120,01
C17-O5-C21	116,80	117,48	116,14	C14-C15-C16	119,75	120,54	120,64
C16-O4-C20	116,02	113,60	113,72	C15-C16-C17	119,72	118,78	118,56
C15-O3-C19	118,19	116,46	116,14	C16-C17-C18	120,18	120,27	120,61
C14-C15-O3	123,15	123,38	124,09	C17-C18-C13	120,27	120,23	120,02
C4-C9-C8	118,95	120,14	120,34	C18-C13-C14	119,18	120,17	120,02
C8-C7-C6	122,53	120,50	120,97				

Tabela 3.3: Valores para o ângulos de ligação obtidos experimentalmente e através dos métodos PM3 e AM1.

Deve-se ressaltar que, dentre os parâmetros coletados e sumarizados nas tabelas 3.1, 3.3 e 3.2, alguns correspondem á geometria dos grupos laterais, que, naturalmente sofrem os maiores desvios por causa do empacotamento cristalino. Como poderemos verificar, estes dados contribuem para o aumento do valor do desvio padrão, principalmente, do ângulo de ligação e do ângulo diedral. Para cada um dos parâmetros calculados, encontramos o desvio padrão dos resultados teóricos em relação aos respectivos valores experimentais, ou seja utilizamos a seguinte relação

$$\sigma_n^{A/P} = \sqrt{\frac{\sum_{i=1}^N (x_{ni}^{A/P} - rx_{ni})^2}{N}}, \quad n = 1, 2 \text{ e } 3, \quad (3.1)$$

onde os índices 1, 2 e 3 representam, respectivamente, o comprimento de ligação, ângulo de ligação e ângulo diedral. Por exemplo, σ_1^A significa o desvio padrão para os valores de comprimento de ligação obtidos utilizando o método AM1.

As medidas do desvio padrão estão sumarizadas na tabela 3.4, a seguir

	Comprimento de Ligação (Å)	Ângulo de ligação átomos(graus)	Ângulo diedral(graus)
AM1	0.0047	0.3839	4.3211
PM3	0.0048	0.4796	1.8543

Tabela 3.4: Desvio padrão dos resultados teóricos em relação aos dados experimentais.

Na comparação entre as geometrias experimentais de raios-x e as geometrias obtidas teoricamente, encontramos que

- Os desvios para os comprimentos de ligação obtidos pelos dois métodos são, praticamente, iguais.
- O desvio do AM1 é ligeiramente menor do que o do PM3 para os ângulos de ligação .
- Para os ângulos diedrais, o desvio padrão do AM1 corresponde a mais que o dobro do desvio do PM3.

Assim, podemos concluir que, mesmo com o AM1 apresentando resultados melhores com relação aos comprimentos e ângulos de ligação, o PM3 mostra-se mais apropriado para a otimização geométrica de C_{rx} , visto que os resultados obtidos para os ângulos diedrais, são bastante superiores aos obtidos pelo AM1. Para verificar isto, podemos recorrer à tabela 3.2 e determinarmos a grande diferença dos resultados encontrados pelos dois métodos, principalmente para os átomos que constituem o esqueleto básico. Portanto, baseado nestes resultados, o método PM3 foi escolhido para a investigação da estrutura geométrica das moléculas de Combretastatina e derivados.

3.2 Metodologia de Otimização das moléculas

Como vimos acima, na seção 3.1.1, o método selecionado para a otimização de nosso conjunto de moléculas é o método semi-empírico PM3. O principal objetivo desta seção é descrever a metodologia desenvolvida para se encontrar as estruturas moleculares das quais retiraremos as informações necessárias para a aplicação da MIE.

Adotaremos o seguinte critério:

As geometrias de menor energia, obtidas computacionalmente num ambiente de vácuo, serão consideradas as estruturas que representam as fontes primárias de

informações teóricas para a investigação da correlação com a atividade biológica para a Combretastatina.

Nosso critério está baseado na observação de que a maioria dos sistemas físicos isolados tendem a uma configuração de mais baixa energia.

3.2.1 Procedimentos computacionais

Definido nosso critério de busca, necessitamos elaborar uma espécie de roteiro de procedimentos que garantam, que ao final do processo de otimização, as estruturas, verdadeiramente, possuam mínima energia conformacional. Os procedimentos adotados devem ser idênticos para cada uma das dezoito moléculas definidas na tabela 1.1.

Suponhamos que o mapa de energia de uma determinada molécula, em termos de um parâmetro que defina um grau de liberdade molecular, seja dado pela figura 3.5.

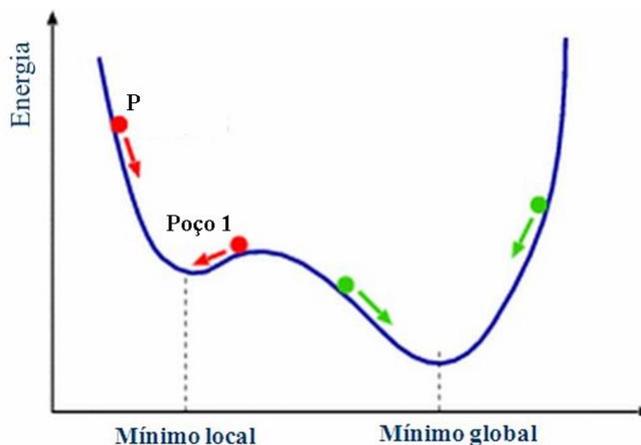


Figura 3.5: Curva de energia associada à mudança na geometria molecular.

Se a estrutura inicial da molécula corresponder, por exemplo, ao ponto (P) no mapa, iniciado o processo de otimização, quando a energia estrutural estiver dentro do Poço 1, o critério de convergência adotado pode ser alcançado e, conseqüentemente, a energia da estrutura obtida ser, na verdade, apenas um mínimo local dentro do mapa. Para garantir que a busca pelas estruturas de menor energia seja efetuada inequivocamente, devemos garantir que o processo de otimização percorra o maior número de configurações possíveis.

O esqueleto básico da Combretastatina é constituído por dois anéis aromáticos **A** e **B** e uma **Ponte**, formada por quatro átomos de carbono. A ligação dupla entre os carbonos da

ponte confere bastante rigidez a essa região molecular, ou seja, alterar a estrutura da ponte requer muita energia. Uma vez que a geometria das ligações dos grupos laterais depende da disposição relativa entre os anéis, devemos assegurar que o nosso processo de otimização seja efetuado para um grande número de configurações da ponte. Se definirmos o ângulo diedral próprio como o ângulo formado pelos planos que contém três átomos consecutivos da ponte, podemos efetuar a otimização para um grande número de possíveis configurações moleculares cobrindo, desta forma, o domínio de variação do diedral próprio. Assim podemos construir um mapa de energia em termos deste ângulo diedral.

O mapeamento da energia foi feito utilizando o programa CACHE, considerando a **Ponte** como uma haste rígida. Realizamos a variação do ângulo diedral, alterando o seu valor a passos de 5 graus, varrendo os 360 graus possíveis. A cada passo efetuado, todas as estruturas moleculares, excetuando a *Ponte*, são otimizadas. Nas figuras 3.6 e 3.7 são mostrados o mapa de energia em termos do diedral e as geometrias moleculares correspondentes a um máximo (figura 3.6) global e a um mínimo (figura 3.7) da energia estrutural

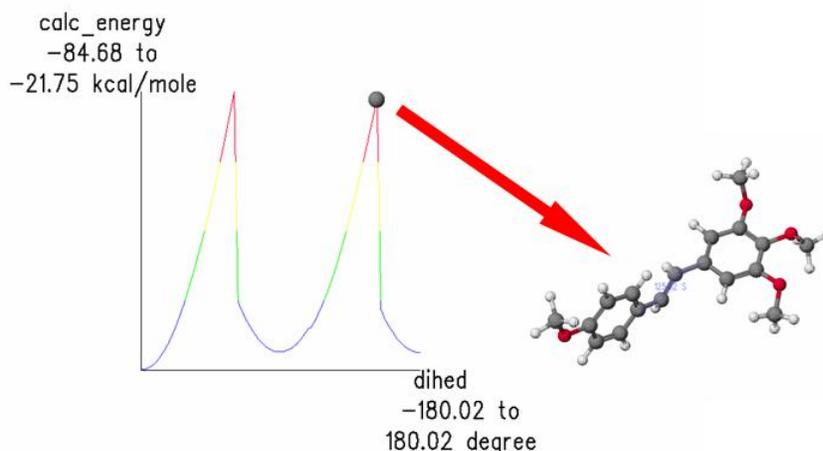


Figura 3.6: Geometria molecular correspondente a um máximo na energia.

Com o mapa da energia de cada uma das dezoito moléculas¹, determinamos as três estruturas de menor energia para cada uma delas. Reotimizamos cada uma das moléculas deste novo conjunto liberando o relaxamento do diedral e utilizando critérios de convergência mais refinados. Destas três moléculas, selecionamos a que possui o menor calor de formação para ser nossa molécula de trabalho. Os resultados comuns da otimização

¹Ver tabela 1.1

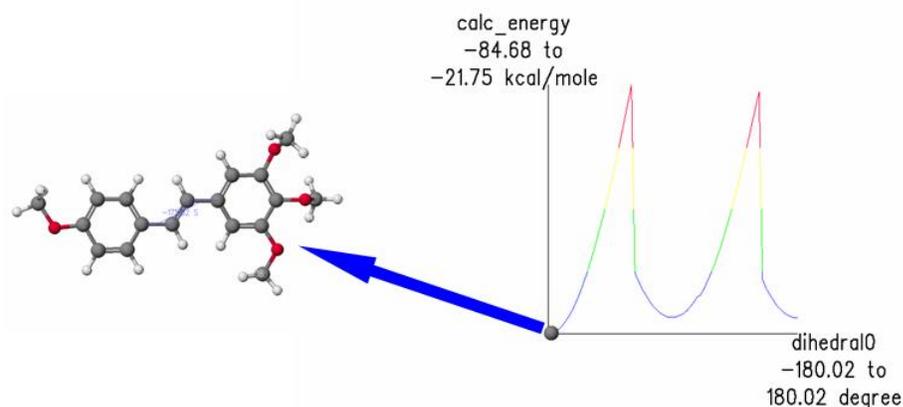


Figura 3.7: Geometria molecular correspondente a um mínimo na energia.

geométrica para conjunto de compostos deste trabalho é:

As geometrias moleculares de menor energia estrutural possuem aproximada co-planaridade dos anéis aromáticos e configuração *trans* para os átomos da ponte.

Os procedimentos de busca geométrica descritos acima, possibilitam, com maior segurança, a obtenção das estruturas moleculares de mais baixa energia, pois força-se a otimização das estruturas fora de possíveis regiões de mínimo local. De posse do conjunto de moléculas devidamente otimizadas, podemos calcular as propriedades que serão utilizadas para o estudo da correlação entre suas propriedades e a atividade da Combretastatina e Derivados.

3.3 Estudo da estrutura-atividade

Nesta seção utilizaremos três metodologias distintas para a investigação da estrutura-atividade do nosso conjunto de moléculas. Primeiramente analisaremos os resultados obtidos pela MIE e, posteriormente, para corroborar estes resultados, utilizaremos as metodologias de análise multivariada: PCA e HCA.

3.3.1 A MIE aplicada a Combretastatina e derivados

Na seção 2.4 destacamos que uma das principais características da MIE é a possibilidade de se estudar os compostos levando em consideração suas propriedades locais. O esqueleto básico da Combretastatina e derivados pode ser dividido em três regiões, Anéis A, B e a ponte, como podemos ver na figura 3.8.

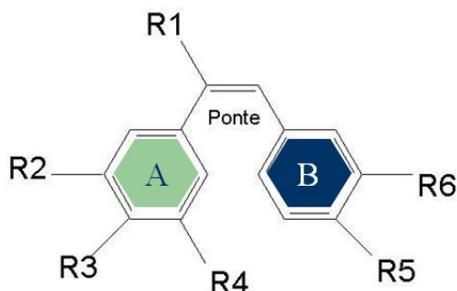


Figura 3.8: Regiões onde foram calculadas a densidade de estados

Calculamos as contribuições individuais de cada uma destas porções moleculares para a densidade total de estados da molécula, ou seja, calculamos a densidade local de estados, restrita a cada uma das regiões da figura 3.8 utilizando a equação (2.86). Isto é feito utilizando-se o programa Chem2Pac. A partir dos valores obtidos da LDOS para os orbitais de fronteira, construímos o conjunto de descritores dados pela expressão (2.87). Por outro lado, os valores de energia para os orbitais de fronteira, obtidas por meio da equação (2.88), nos fornecem o outro parâmetro relevante para MIE. Utilizaremos as seguintes definições:

$$\Delta_H = \text{HOMO} - \text{HOMO}_{-1}$$

$$\Delta_L = \text{LUMO}_{+1} - \text{LUMO}$$

$$\eta_H = \text{LDOS}_{\text{HOMO}} - \text{LDOS}_{\text{HOMO}_{-1}}$$

$$\eta_L = \text{LDOS}_{\text{LUMO}_{+1}} - \text{LDOS}_{\text{LUMO}}$$

As tabelas 3.5, 3.6 e 3.7 mostram os resultados obtidos para os parâmetros Δ e η para os anéis A, B e a ponte, respectivamente.

	η_H	Δ_H	η_L	Δ_L	IC_{50}
C01	0.4803315	0.75069	-0.4033569	0.97172	0.12
C02	0.5936072	0.3916	-0.5806974	0.56475	0.13
C03	-0.3426816	0.36422	1.323302	0.36532	0.13
C04	0.1357518	0.43066	0.5562871	0.79046	0.18
C05	-0.7077786	0.82573	1.033115	0.87969	0.2
C06	-0.0392521	0.65375	0.8468041	0.95564	0.5
C07	-1.2058065	1.01335	-0.6628903	1.15318	0.65
C08	-0.4107413	0.67078	-0.4881468	0.55259	0.7
C09	-0.7474253	0.98008	-0.7194485	0.35156	1.2
C10	-0.6522283	0.75321	-0.6666668	0.79106	1.25
C11	-0.616581	0.67231	0.8265954	1.06018	1.3
C12	-0.6712009	0.77001	1.0449642	0.71056	2
C13	0.3676843	0.69018	-0.4548492	0.99895	2.5
C14	0.4511835	0.74278	-0.3173476	0.98271	3.1
C15	-0.6717248	0.75116	1.0564859	0.89213	3.4
C16	0.0157281	0.69739	0.9412414	0.57813	4.5
C17	-0.3030174	1.06444	0.9860814	0.69587	5.5
C18	-0.4359419	1.00092	-0.5858258	0.73097	6.6

Tabela 3.5: Valores de Δ e η para o anel A.

	η_H	Δ_H	η_L	Δ_L	IC_{50}
C1	0.7548800	0.75069	0.9860764	0.97172	0.12
C2	0.030059	0.3916	-0.3946191	0.56475	0.13
C3	0.290346	0.36422	-0.7984249	0.36532	0.13
C4	0.3072335	0.43066	-0.0427517	0.79046	0.18
C5	0.4815401	0.82573	-0.48899	0.87969	0.2
C6	-0.1903644	0.65375	-0.3312041	0.95564	0.5
C7	0.7220372	1.01335	1.3188674	1.15318	0.65
C8	0.0275415	0.67078	1.0754699	0.55259	0.7
C9	0.4143085	0.98008	1.3286891	0.35156	1.2
C10	0.3757373	0.75321	1.2923866	0.79106	1.25
C11	0.3263045	0.67231	-0.2900719	1.06018	1.3
C12	0.4445876	0.77001	-0.5115222	0.71056	2
C13	-0.8152579	0.69018	1.2239505	0.99895	2.5
C14	-0.7318639	0.74278	0.8965183	0.98271	3.1
C15	0.4410776	0.75116	-0.523745	0.89213	3.4
C16	-0.212753	0.69739	-0.4176575	0.57813	4.5
C17	0.031885	1.06444	-0.4520339	0.69587	5.5
C18	0.3525497	1.00092	1.2876501	0.73097	6.6

Tabela 3.6: Valores de Δ e η para o anel B.

	η_H	Δ_H	η_L	Δ_L	IC_{50}
C1	0.7385883	0.75069	-0.8495485	0.97172	0.12
C2	-0.2925104	0.3916	0.8516712	0.56475	0.13
C3	0.2321421	0.36422	-0.8696665	0.36532	0.13
C4	0.5351487	0.43066	-0.687762	0.79046	0.18
C5	0.6943658	0.82573	-0.903	0.87969	0.2
C6	0.6925167	0.65375	-0.7415106	0.95564	0.5
C7	0.8980378	1.01335	-0.9483412	1.15318	0.65
C8	0.3801848	0.67078	-0.9357861	0.55259	0.7
C9	0.6470071	0.98008	-0.9287376	0.35156	1.2
C10	0.6559358	0.75321	-0.9540081	0.79106	1.25
C11	0.6154597	0.67231	-0.8807117	1.06018	1.3
C12	0.7465122	0.77001	-0.9098358	0.71056	2
C13	0.7575054	0.69018	-0.8784219	0.99895	2.5
C14	0.7529522	0.74278	-0.8150408	0.98271	3.1
C15	0.7703548	0.75116	-0.9104734	0.89213	3.4
C16	0.4891514	0.69739	-0.9231077	0.57813	4.5
C17	0.4380467	1.06444	-0.9220528	0.69587	5.5
C18	0.5232488	1.00092	-0.944615	0.73097	6.6

Tabela 3.7: Valores de Δ e η para a **Ponte**.

Avaliamos separadamente os dados correspondentes à cada uma das três regiões. Os melhores resultados foram obtidos para os orbitais HOMO e HOMO₋₁ sobre o anel B. A regra que nos permite separar as moléculas em 2 grupos (ativas e inativas) é dada por:

Se $\eta_H \geq 0.46$ ou $\Delta_H \leq 0.66$ a molécula é ativa; caso contrário é inativa.

O resultado da regra aplicada ao nosso conjunto de dezoito moléculas está sumarizado na tabela abaixo:

Ativas	C01(A)	C02(A)	C03(A)	C04(A)	C05(A)	C06(A)	C07(A)
Inativas	C08(A)	C09(I)	C10(I)	C11(I)	C12(I)	C13(I)	
Inativas	C14(I)	C15(I)	C16(I)	C17(I)	C18(I)		

Tabela 3.8: Classificação de atividade segundo a MIE.

As letras A (Ativa) e I (Inativa), entre parênteses, a frente de cada uma das moléculas, indicam a classificação, obtida experimentalmente, a partir do índice IC_{50} , dado na tabela 1.2. Se compararmos as previsões teóricas com as experimentais, verificamos que há uma concordância de 94,4% na previsão de atividade da MIE.

Para obtermos uma avaliação estatisticamente independente destes resultados, realizamos

novas análises teóricas da atividade do nosso conjunto de compostos. Foram utilizadas as técnicas de análise multivariada PCA (Análise de Componente Principais) e HCA (Análise Hierárquica de Clusters), muito populares em investigações químicas de fármacos e alimentos.

3.3.2 Investigação da Atividade Biológica da Combretastatina e Derivados através das técnicas: PCA e HCA

No capítulo 2, seção 2.3, fizemos um esboço da teoria por trás do PCA e do HCA. Utilizaremos, tanto o PCA, como o HCA, como metodologias independentes para corroborar os resultados encontrados pela MIE. Para a aplicação do PCA selecionamos, inicialmente, 34 descritores:

- as energias dos orbitais de fronteira: HOMO-1 (H-), HOMO (H), LUMO (L) e LUMO+1 (L+);
- as diferenças entre as energias dos orbitais acima: Δ_H e Δ_L ;
- As LDOS sobre as três regiões (Anéis A, B e a ponte); H-A, H-B, H-P, HA, HB, HP, LA, LB, LP, L+A, L+B e L+P;
- a diferença entre as LDOS acima: η_{HA} , η_{HB} , η_{HP} , η_{LA} , η_{LB} e η_{LP} ;
- energia eletrônica (EE);
- momento de dipolo (Dipolo);
- refratividade (Ref);
- polarizabilidade (Pol);
- massa (m);
- volume (v);
- calor de formação (CF);
- energia de hidratação (EH);
- coeficiente de partição octanol-água (log P);
- potencial de ionização .

Na Tabela 3.9 encontram-se os valores obtidos para cada um dos 34 descritores utilizados.

	H-	H+A	H+B	H+P	H	HA	HB	HP	L	LA	LB	LP	L+	L+A	L+B	L+P	AH
C01	-9.226	0.257	1.375	0.138	-8.475	0.738	0.620	0.877	-0.636	0.570	0.796	0.952	0.336	0.167	1.782	0.103	0.751
C02	-9.229	0.151	0.747	0.844	-8.837	0.745	0.777	0.551	-0.472	0.934	0.824	0.598	0.093	0.353	0.429	1.443	0.392
C03	-9.392	0.547	0.910	0.370	-8.998	0.204	1.200	0.603	-0.344	0.568	0.808	0.875	0.022	1.891	0.010	0.006	0.364
C04	-9.133	0.527	0.721	0.193	-8.702	0.662	1.028	0.728	-0.692	0.809	0.560	0.963	0.099	1.365	0.517	0.275	0.431
C05	-9.287	1.292	0.266	0.165	-8.416	0.584	0.747	0.860	-0.707	0.766	0.580	0.954	0.173	1.799	0.091	0.051	0.826
C06	-9.238	0.803	0.779	0.110	-8.585	0.764	0.589	0.803	-0.804	0.655	0.703	0.956	0.152	1.502	0.371	0.215	0.654
C07	-9.311	1.813	0.001	0.002	-8.298	0.607	0.724	0.900	-0.804	0.668	0.675	0.950	0.349	0.005	1.994	0.002	1.013
C08	-9.336	0.963	0.681	0.504	-8.666	0.553	0.709	0.885	-0.490	0.492	0.895	0.938	0.063	0.004	1.971	0.002	0.671
C09	-9.598	1.298	0.347	0.217	-8.618	0.551	0.761	0.864	-0.382	0.725	0.643	0.929	-0.031	0.005	1.972	0.001	0.980
C10	-9.394	1.296	0.314	0.162	-8.641	0.644	0.690	0.818	-0.869	0.669	0.681	0.955	-0.078	0.003	1.974	0.001	0.753
C11	-9.244	1.148	0.410	0.259	-8.572	0.531	0.737	0.875	-0.899	0.901	0.446	0.964	0.161	1.728	0.156	0.083	0.672
C12	-9.301	1.346	0.239	0.125	-8.531	0.675	0.684	0.872	-0.531	0.763	0.595	0.960	0.179	1.807	0.084	0.049	0.770
C13	-9.307	0.197	1.420	0.143	-8.617	0.565	0.605	0.901	-0.699	0.591	0.563	0.966	0.300	0.137	1.787	0.088	0.690
C14	-9.243	0.252	1.379	0.136	-8.502	0.704	0.648	0.889	-0.660	0.586	0.779	0.954	0.323	0.269	1.676	0.139	0.743
C15	-9.332	1.410	0.177	0.113	-8.581	0.738	0.618	0.883	-0.727	0.752	0.607	0.962	0.165	1.808	0.084	0.051	0.751
C16	-9.546	0.284	1.255	0.245	-8.848	0.299	1.042	0.735	-1.121	1.012	0.418	0.923	-0.543	1.953	0.000	0.000	0.697
C17	-10.033	0.666	0.916	0.399	-8.969	0.363	0.948	0.837	-1.303	0.967	0.452	0.922	-0.607	1.953	0.000	0.000	1.064
C18	-9.481	1.070	0.595	0.360	-8.480	0.634	0.948	0.883	-0.698	0.586	0.688	0.945	0.033	0.001	1.975	0.000	1.001
	AL	η HA	η LA	η HB	η LB	η HP	η LP	m	v	Ref	Pol	Log p	EH	CF	EE	Dipolo	PI
C01	0.972	0.480	-0.403	0.755	0.986	0.739	-0.850	268.4	867.2	32.260	32.190	4.540	-7.890	-46.63	-20657	1.579	8.475
C02	0.565	0.594	-0.581	0.030	-0.395	-0.293	0.852	330.4	996.1	42.410	35.930	3.180	-12.860	-3.24	-18620	1.096	8.837
C03	0.365	-0.343	1.323	0.290	-0.798	0.232	-0.870	344.4	1039.1	47.010	37.770	3.580	-12.720	-138.20	-32055	2.763	8.998
C04	0.790	0.136	0.556	0.307	-0.043	0.535	-0.688	330.4	983.7	42.410	35.930	3.180	-12.930	-134.74	-29568	3.491	8.702
C05	0.880	-0.708	1.033	0.482	-0.489	0.694	-0.903	330.4	932.4	36.410	33.460	3.040	-7.850	-84.41	-25045	1.964	8.416
C06	0.956	-0.039	0.847	-0.190	-0.331	0.693	-0.742	358.4	1101.3	52.040	39.600	3.900	-11.240	-143.97	-33633	1.617	8.585
C07	1.153	-1.206	-0.663	0.722	1.319	0.898	-0.948	252.4	845.0	30.970	31.550	4.710	-2.020	-3.24	-18620	1.096	8.298
C08	0.553	-0.411	-0.488	0.028	1.075	0.380	-0.936	284.4	891.3	35.490	33.290	5.280	-0.590	-20.33	-15622	1.609	8.666
C09	0.352	-0.747	-0.719	0.414	1.329	0.647	-0.929	360.4	1095.0	50.560	38.870	4.320	-5.090	-64.41	-24941	2.195	8.618
C10	0.791	-0.652	-0.667	0.376	1.292	0.656	-0.954	304.3	887.1	31.280	31.530	3.360	-11.170	-133.48	-25132	1.986	8.641
C11	1.060	-0.617	0.827	0.326	-0.290	0.615	-0.881	314.4	966.3	41.020	35.300	3.360	-7.000	-89.84	-27400	2.042	8.572
C12	0.711	-0.671	1.045	0.445	-0.512	0.747	-0.910	298.4	950.8	39.210	34.660	3.910	-4.020	-140.87	-33419	2.836	8.531
C13	0.999	0.368	-0.455	0.815	1.224	0.758	-0.878	226.3	722.8	18.530	26.680	3.600	-10.940	-50.54	-22872	3.034	8.617
C14	0.983	0.451	-0.317	-0.732	0.897	0.753	-0.815	254.3	819.1	27.750	30.350	4.220	-8.760	-38.28	-18821	1.755	8.502
C15	0.892	-0.672	1.056	0.441	-0.524	0.770	-0.910	298.4	958.8	39.200	34.660	4.000	-4.520	-60.34	-24795	2.060	8.581
C16	0.578	0.016	0.941	-0.213	-0.418	0.489	-0.923	280.3	749.4	18.250	26.410	4.340	-10.040	-149.01	-21388	5.135	8.848
C17	0.696	-0.303	0.986	0.032	-0.452	0.438	-0.922	282.2	740.1	16.770	25.680	4.770	-3.550	-147.49	-21425	4.121	8.969
C18	0.731	-0.436	-0.586	0.353	1.288	0.523	-0.945	256.3	807.1	26.270	29.620	4.650	-2.210	-36.86	-18851	2.976	8.480

Tabela 3.9: Valores calculados dos 34 descritores utilizados, inicialmente, pelo PCA.

É importante ressaltar, que a busca no PCA não privilegia qualquer variável. Baseado nos 34 descritores utilizados, devemos encontrar a melhor aglomeração dos compostos em grupos de atividade. Como vimos na seção 2.3.2, o descarte de descritores é feito segundo a sua relevância para a distribuição dos pontos. Assim, a melhor separação encontrada para o nosso conjunto de moléculas pode ser visto na figura 3.9.

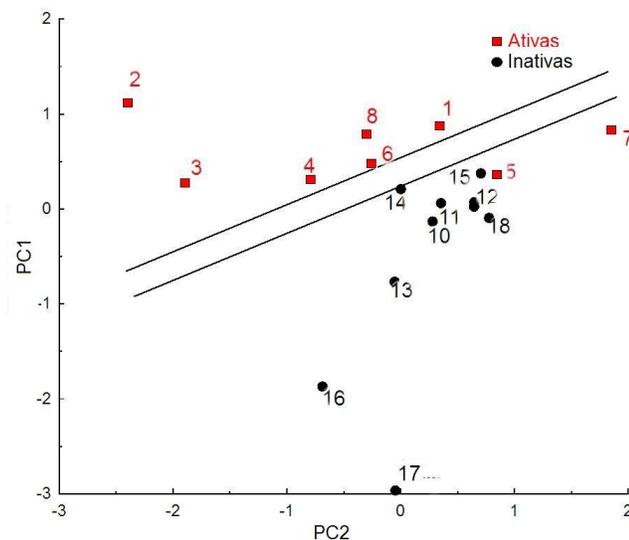


Figura 3.9: Melhor separação utilizando o PCA.

As componentes principais são dadas por

$$\begin{aligned}
 PC1 &= -0.43[H_-] + 0.39[H_-A] - 0.39[H_-P] + 0.43[H] + 0.48[HP] - 0.15[LB] + 0.44[\Delta_H] \\
 &\quad + 0.20[\eta_H B] - 0.12[v] + 0.10[CF] \\
 PC2 &= 0.54[H_-] + 0.13[H_-A] + 0.10[H_-P] + 0.33[H] - 0.49[HP] + 0.46[LB] - 0.25[\Delta_H] \\
 &\quad + 0.22[\eta_H B] + 0.38[v] + 0.35[CF]
 \end{aligned}
 \tag{3.2}$$

Note que as somas dos quadrados dos coeficientes tanto de PC1, quanto de PC2, são a unidade. Isto significa que variáveis com coeficientes maiores, contribuem mais significativamente para a separação dos compostos segundo sua atividade biológica. Fica evidente, portanto, que as maiores contribuições para as componentes principais PC1 e PC2, são de origem eletrônica. Os valores da energia do $HOMO_{-1}$ e da densidade de estados de HOMO sobre a **Ponte**, mostraram-se como as variáveis mais importantes para a separação obtida. Além disso, os parâmetros

utilizados para construir a classificação segundo a MIE estão presentes nas componentes principais, o que evidencia a importância destes parâmetros para a classificação da atividade anticarcinogênica da Combretastatina.

Podemos verificar que os compostos estão claramente separados em dois subgrupos de atividade (ativos e inativos) e que, se compararmos o resultado encontrado pelo PCA com relação à classificação baseada nos dados experimentais em termos do IC₅₀, temos uma correspondência de 89%, como pode ser visto na tabela abaixo.

Ativas	C01(A)	C02(A)	C03(A)	C04(A)	C06(A)	C08(A)
Inativas	C05(A)	C07(A)	C09(I)	C12(I)	C10(I)	C11(I)
Inativas	C13(I)	C14(I)	C15(I)	C16(I)	C17(I)	C18(I)

Tabela 3.10: Classificação de atividade segundo o PCA.

Este resultado, obtido a partir de uma metodologia amplamente utilizada no estudo de reconhecimento de padrões e que faz uso de um grande número de descritores, confirma a superioridade das variáveis eletrônicas selecionadas pela MIE na investigação da atividade da Combretastatina e derivados.

O resultado do HCA, construído com base na distribuição dos compostos no espaço das componentes principais, nos mostra mais claramente como estes grupos se aglutinam e a similaridade existente entre eles.

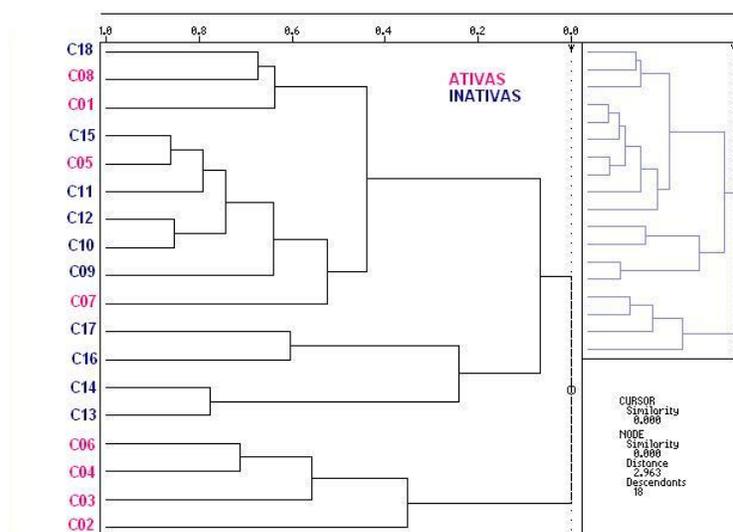


Figura 3.10: Resultado da análise de HCA.

Na figura 3.10, podemos observar a formação de quatro Grupos:

$$G_1 = \{C01(A), C08(A), C18(I)\}$$

$$G_2 = \{C07(A), C09(I), C10(I), C11(I), C05(A), C15(I)\}$$

$$G_3 = \{C13(I), C14(I), C16(I), C17(I)\}$$

$$G_4 = \{C02(A), C03(A), C04(A), C06(A)\}$$

O grupo G_4 , composto, somente, por moléculas ativas possui similaridade nula com qualquer um dos outros compostos estudados, o que inclui todos os classificados como inativos, tanto pelo PCA, como pela Metodologia de Índices Eletrônicos (MIE). Assim, os demais grupos, devem ser classificados como grupos de moléculas inativas. É importante notar que as moléculas C05 e C07, assim como no resultado do PCA, pertencem a um grupo composto, somente, por moléculas inativas(G_2). Além disso, as moléculas C01 e C08, experimentalmente ativas, dentro do HCA, foram classificadas como inativas. Analisando a tabela 3.8, podemos verificar que C08 também foi classificada como inativa pela MIE. A similaridade entre os grupos G_1 e G_2 é de aproximadamente 0.44. Já o grupo G_3 é formado, exclusivamente, por moléculas inativas e possui similaridade muito baixa (em torno de 0.06) com os grupos G_1 e G_2 . A figura 3.11 reproduz o agrupamento descrito na figura 3.10 dentro do plano formado pelas duas componentes principais.

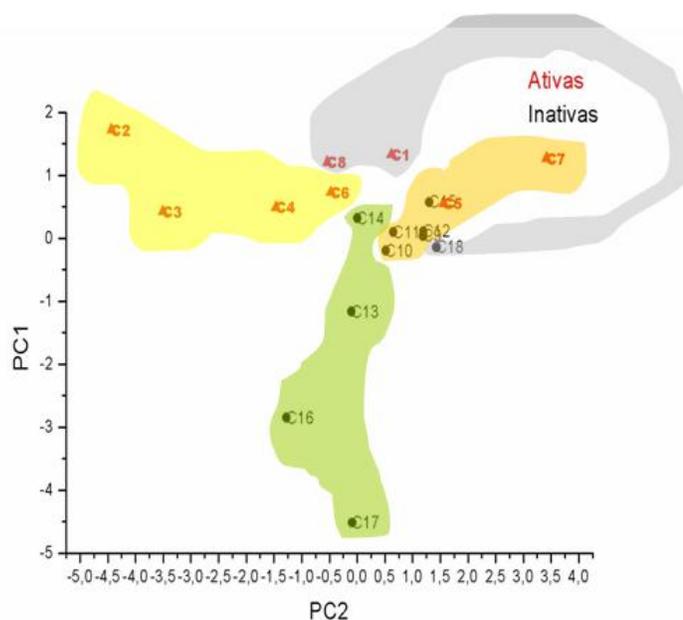


Figura 3.11: Agrupamento do HCA no plano formado pelas duas componentes do PCA.

Os resultados das metodologias de análise multivariada, tão utilizadas para o reconhecimento de padrões, evidenciam a contribuição das variáveis de origem eletrônica, e, principalmente a importância dos parâmetros utilizados pela MIE para a separação da Combretastatina segundo sua atividade anti-carcinogênica.

3.4 Conclusões e perspectivas

Mais uma vez, fica evidente o poder da MIE para classificação da atividade anticarcinogênica da Combretastatina e derivados. A utilização de apenas dois parâmetros de origem eletrônica, associados aos orbitais de fronteira (cuja importância para a interação entre moléculas está bem estabelecida) e que, de forma indireta, levam em consideração a importância das propriedades eletrônicas locais, são capazes de reproduzir a classificação da atividade biológica obtida a partir de dados experimentais com um impressionante percentual de acerto. A tabela 3.11, abaixo, mostra o percentual de acerto de cada uma das três metodologias utilizadas neste trabalho.

Metologia	Acerto
MIE	94,4%
PCA	89,0%
HCA	78,0%

Tabela 3.11: Percentual de acerto, comparado à atividade experimental, das metodologias utilizadas neste trabalho.

Os índices de acerto, com relação à reprodução da atividade biológica experimental, do PCA e do HCA são menores que o da MIE, mesmo utilizando um número maior de parâmetros de comparação. No nosso caso os resultados do PCA e do HCA estão baseados na combinação de 10 descritores, muitos deles eletrônicos, mas mesmo assim, estas metodologias não são capazes de reproduzir a atividade experimental com a mesma eficiência que a MIE. Isso evidencia a importância dos descritores Δ e η , dados pelas equações (2.88) e (2.87), frente aos descritores comumente usados.

Concluimos que a MIE e seus parâmetros eletrônicos podem ser aplicados no estudo e predição da atividade biológica de novos derivados da Combretastatina.

O grande sucesso demonstrado pela aplicação da Metodologia de Índices Eletrônicos às mais

diversas classes de compostos, abre novas perspectivas de investigação da razão pela qual estas simples regras nos conduzem, de algum modo, rumo a critérios de afinidade entre compostos orgânicos candidatos à reação . Utilizando novas vertentes do estudo de sistemas complexos fora do equilíbrio [45, 46, 47], temos caminhado em direção à possibilidade de fundamentação teórica da MIE, o que traria segurança sobre a aplicabilidade da metodologia e conseqüentemente, mudaria as bases das metodologias de busca de novos compostos, não só com finalidade terapêutica, mas sim, de compostos orgânicos que desempenhem qualquer outro tipo de atividade. Existe a esperança de que os parâmetros Δ e η surjam, naturalmente, quando tratarmos o problema dentro de um ambiente no qual as interações persistentes possam ser consideradas, ao invés das interações transitórias comumente tratadas nes teorias clássicas de perturbação. Entretanto, esta nova etapa, está na sua fase inicial, buscando a estrutura teórica capaz de sustentar as bases para o estudo de sistemas complexos fora do equilíbrio. Como sabemos, o não-equilíbrio é uma condição necessária para o estabelecimento de reações entre moléculas e portanto, o regime ideal para o estabelecimento da fundamentação teórica da Metodologia de Índices Eletrônicos.

Referências Bibliográficas

- [1] *Estimativa 2006: Incidência de câncer no Brasil*, Rio de Janeiro, INCA (2005);
- [2] R. W. Ruddon, *Cancer Biology*, Oxford University Press, (1987);
- [3] H. Lodish *et al*, *Molecular Cell Biology*, W. H. Freeman and Co., (2000);
- [4] Figuras retiradas do sítio <http://www.gcarlson.com>;
- [5] S. Ducki, G. Mackenzie, N.J. Lawrence, J.P. Snyder, *J. Med. Chem.*, **48**, 457-465, (2005);
- [6] M. Medarde *et al*, *J. Enz. Inhib. Med. Chem.*, **19**, 521, (2004);
- [7] G. Attard *et al.*, *Path. Biology*, , , (2005);
- [8] C. Borrel *et al.*, *Bioorg. Med. Chem. Lett.*, **13**, 3853, (2005);
- [9] B. Goggiola *et al.* , *Bioorg. Med. Chem. Lett.*, **15**, 3551, (2005);
- [10] L. Wang *et al.*, *J. Med. Chem.*, **45**, 1697, (2002);
- [11] Figuras retiradas do endereço eletrônico <http://www.origene.com>;
- [12] G. R. Pettit *et al*, *Canadian Journal of Chemistry*, **60**, 1374, (1982);
- [13] G. R. Pettit *et al*, *Experientia*, **45**, 209-211, (1989);
- [14] D. A. Ramirez, *et al*, *J. Chem. Crystall.* **35**, 227, (2005);
- [15] J. P. Liou *et al.*, *J. Med. Chem.*, **47**, 4247, (2004);
- [16] G. C. Tron *et al.*, *J. Med. Chem.*, **48**, 3260, (2005);
- [17] G. R. Pettit *et al.*, *Journal of Medicinal Chemistry*, Dezembro, (2002);
- [18] J. A. Hadfield *et al.*, *Euro. J. Med. Chem.*, **40**, 529, (2005);

- [19] N. J. Lawrence *et al.*, *Bioorg. Med. Chem. Lett.*, **11**, 51, (2001);
- [20] C. Rappl *et al.*, *Biochemistry*, **45**, 9210, (2006);
- [21] C. Bailly *et al.*, *J. Med. Chem.*, **46**, 5437, (2003);
- [22] A. B. S. Maya *et al.*, *J. Med. Chem.*, **48**, 556, (2005);
- [23] A. B. S. Maya *et al.*, *Bioorg. Med. Chem. Lett.*, **13**, 2097, (2005);
- [24] N. J. Lawrence *et al.*, *J. Comb. Chem.*, **3**, 421, (2001);
- [25] J. A. Pople, D. Beveridge, *Approximate Molecular Orbital Theory*. 1. ed. New York, McGraw-Hill, (1970).
- [26] A. Szabo, N.S. Ostlund, *Modern Quantum Chemistry*, McGraw-Hill, New York, (1989);
- [27] I.N. Levine, *Quantum Chemistry*, Prentice Hall, New Jersey, (1991);
- [28] J.D.M. Vianna, A. Fazzio, S. Canuto, *Teoria Quântica de Moléculas e Sólidos*, Livraria da Física, São Paulo, (2004);
- [29] K. Hoffman, R. Kunze, *Álgebra Linear*, LTC, Rio de Janeiro, 1979;
- [30] C. C. J. Roothaan, *Rev. Mod. Phys.*, **23**, 69, (1951);
- [31] J. A. Pople, D. P. Santry, G. A. Segal, *J. Chem. Phys.*, **43**, S129, (1965);
- [32] J. A. Pople, D. P. Santry, *J. Chem. Phys.*, **43**, S136, (1965);
- [33] J. A. Pople, D. P. Santry, *J. Chem. Phys.*, **44**, 3289, (1966);
- [34] , D. P. Santry, G. A. Segal, *J. Chem. Phys.*, **47**, 158, (1967);
- [35] J. Pople, D. Beveridge, P. A. Dobosh, *J. Chem. Phys.*, **47**, 2026, (1967);
- [36] M.J.S. Dewar *et al.*, *Science* **187**, 1037 (1975);
- [37] M.J.S. Dewar e W. Thiel, *J. Am. Chem. Soc.*, **99**, 4899-4907 (1977);
- [38] M.J.S. Dewar e H.S. Rzepa, *J. Am. Chem. Soc.*, **100**, 5777-5784 (1978);
- [39] W. Härdle, L. Simar, *Applied Multivariate Statistical Analysis*, Springer-Verlag Berlin Heidelberg, (2003);

- [40] V. R. Coluci, S. F. Braga, R. Vendrame, P. M. V. B. Barone, *Esc. Bras. Estrut. Elet.*, Livraria da Física, 202 , (2003);
- [41] P.M.V.B. Barone, A. Camilo Jr., D.S. Galvão, *Phy. Rev. Let.*, **77**, 1186, (1996);
- [42] R. Hoffmann, *J. Chem. Physics.*, **39**, 1397, (1963);
- [43] J. J. P. Stewart, *J. Comp. Chem.*, **10**, 209, (1989);
- [44] J. J. P. Stewart, *J. Comp. Chem.*, **10**, 221, (1989);
- [45] I. Prigogine, *Non-equilibrium Statistical Mechanics*, Interscience Publishers, (1962);
- [46] I. Prigogine, *From Being to Becoming: Time and Complexity in the Physical Sciences*, W. H. Freeman and Co., (1981);
- [47] T. Pretosky, I. Prigogine, *Chaos, Solitons and Fractals*, **7**, 441, 1996.