



**INSTITUTO DE FÍSICA
GLEB WATAGHIN**

**APLICAÇÃO DE MÉTODOS SEMIEMPÍRICOS AO ESTUDO
DA ESTRUTURA ELETRÔNICA DE COMPOSTOS BIOATIVOS**

por

LARISSA LIMA DO ESPÍRITO SANTO

ORIENTAÇÃO: PROF. DR. DOUGLAS SOARES GALVÃO

**Tese de doutorado apresentada no
Instituto de Física Gleb Wataghin da
Unoversidade Estadual de Campinas.**

Campinas, Agosto de 2001

DEDICATÓRIA

Dedico esse trabalho aos meus pais, Kleber e Isa, que, apesar de não terem a mínima idéia do que seja uma função de onda, há muito são Doutores na 'função' de pais.

AGRADECIMENTOS

Foi muita gente boa que encontrei pelo caminho, todas estarão sempre em meu coração, mesmo que os nomes não estejam nessa listinha feita na correria.

Agradeço

Ao Professor Douglas Galvão pela orientação, pelo carinho, pela amizade e pela dedicação.

Ao amigo Márcio Cyrillo pelas inúmeras discussões proveitosas e pelos “preciosos” programas que desenvolveu e que tanto facilitaram nossa vida.

Aos inúmeros amigos da Pós-graduação : Cabelo, Ana Banana e Super Sérgio, Roma, Lázaro, Martinha, Laura, Pila, Demétrio, Jeferson, Alexandre Urbano, Juliano e Zola. Em especial ao grande amigo de todas as horas, Tersio.

Aos amigos do grupo: Giro, Scheila, Nei, Sérgio e Prof. Bernardo pelas inúmeras vezes que me deram uma mãozinha.

Aos amigos ‘extra’ física: Soraia, Rô, Fê e Eliana.

Aos meus familiares: Meus irmãos, Ivan e Kleber Filho e meus pais, Maria Luiza e Kleber que sempre foram meu porto seguro.

Aos meus ‘novos’ familiares: Sr Ruberval e Dna Nair, Binho e Graciete, Giulia e Geovanna pelo apoio incondicional.

Às amigas da secretaria do DFA: Lúcia e Flávia, sempre sorridentes e prestativas.

Às amigas da BIF: Ângela, Célia e Rita pelos momentos de descontração.

*“Sozinhos e juntos, na dor e no prazer. Nas fases difíceis e nas fáceis de viver...”
Obrigada Cláudio, por tudo. Pelo amor, pelo carinho, pela compreensão e pela paciência.
Sem você este trabalho não se realizaria.*

RESUMO

Na primeira parte desse trabalho apresentamos estudos de estrutura-atividade realizados em uma série de compostos com atividade biológica anticancerígena (derivados do lapachol, da talidomida e das mitomicinas). As caracterizações geométrica e eletrônica desses compostos foram obtidas através de métodos de mecânica molecular e de métodos semiempíricos. Foi possível correlacionar a resposta biológica dos compostos analisados com alguns índices eletrônicos. Em particular, empregamos com êxito uma nova metodologia desenvolvida para tratar o problema da correlação estrutura geométrica com atividade biológica, denominada Metodologia de Índices Eletrônicos. Em uma segunda parte desse trabalho avaliamos a estabilidade genética de seqüências de repetições de nucleotídeos presentes em regiões do DNA, através do método de mecânica molecular AMBER. Concluímos que seqüências formadas por bases ordenadas (microsatélites) possuíam uma estabilidade energética menor do que seqüências formadas por bases distribuídas aleatoriamente sobre a cadeia de DNA.

ABSTRACT

In a first part of this work we present structure-activity studies in a compound series with anti-cancer biological activity (derived from lapachol, thalidomide and mitomycins). The geometric and electronic characterizations of these compounds had been gotten through semiempirical and molecular mechanics methods. We could relate the biological response for analyzed compounds with some electronic index. In particular we successfully used a recently developed methodology to study the correlation between geometric structure and biological activity, called Electronic Index Methodology. In a second part of this work we evaluated the genetic stability of replicated sequences of nucleotides found in DNA regions through the method of molecular mechanics AMBER. We found that sequences formed by repeated base pairs have worse genetic stability when compared against random sequences of DNA.

ÍNDICE

CAPÍTULO 1	4
MÉTODOS DE CÁLCULO	4
1.1 INTRODUÇÃO	4
1.2 MECÂNICA MOLECULAR	6
1.3 MECÂNICA QUÂNTICA.....	12
1.3.1 <i>Aproximação de Born-Oppenheimer</i>	14
1.3.2 <i>Aproximação do Elétron Independente</i>	15
1.3.3 <i>O Procedimento LCAO</i>	19
1.3.4 <i>Métodos Semiempíricos</i>	24
CAPÍTULO 2	27
PROPRIEDADES FÍSICO-QUÍMICAS UTILIZADAS NO ESTUDO	
ESTRUTURA-ATIVIDADE	27
2.1 INTRODUÇÃO	27
2.2 PARÂMETROS HIDROFÓBICOS.....	28
2.2.1 <i>Solubilidade</i>	28
2.2.2 <i>Coefficiente de Partição</i>	28
2.3 PARÂMETROS ESTEROQUÍMICOS	29
2.4 PARÂMETROS ELETRÔNICOS	29
2.4.1 <i>Carga Atômica</i>	29
2.4.2 <i>Momento de dipolo</i>	31

2.4.3	<i>Densidade Eletrônica de Fronteira (DEF)</i>	31
2.4.4	<i>Energias do HOMO (highest occupied molecular orbital) e LUMO (lowest occupied molecular orbital)</i>	32
2.4.5	<i>Superdeslocalizabilidade (S_r)</i>	33
2.4.6	<i>Diferença de Energia HOMO LUMO</i>	34
2.4.7	<i>Metodologia de Índices Eletrônicos</i>	34
CAPÍTULO 3		37
ESTUDO ESTRUTURA ATIVIDADE DE COMPOSTOS DERIVADOS DO		
LAPACHOL		37
3.1	INTRODUÇÃO	37
3.2	DERIVADOS DO LAPACHOL COM ATIVIDADE ANTI-CARCINOGENICA.....	38
3.3	CÁLCULOS E RESULTADOS	40
3.3.1	<i>Composição dos Orbitais Moleculares</i>	45
3.3.2	<i>A Metodologia de Índices Eletrônicos (MIE)</i>	51
3.4	CONCLUSÕES.....	58
CAPÍTULO 4		60
ESTUDO ESTRUTURA-ATIVIDADE DE COMPOSTOS ANÁLOGOS À		
TALIDOMIDA		60
4.1	INTRODUÇÃO	60
4.2	ANÁLOGOS DA TALIDOMIDA	63
4.3	CÁLCULOS E RESULTADOS	66
CAPÍTULO 5		78

ESTRUTURA-ATIVIDADE DE UMA SÉRIE DE INDOLEQUINONAS COM	
ATIVIDADE ANTITUMORAL	78
5.1 INTRODUÇÃO	78
5.2 RESULTADOS E CONCLUSÕES	80
CAPÍTULO 6.....	91
ANÁLISE DA ESTABILIDADE ENERGÉTICA DE MICROSSATÉLITES	
PRESENTES EM CADEIAS DE NUCLEOTÍDEOS.....	91
6.1 INTRODUÇÃO	91
6.2 RESULTADOS E CONCLUSÕES	95
REFERÊNCIAS	106

capítulo 1

MÉTODOS DE CÁLCULO

1.1 INTRODUÇÃO

Por muitos anos a estratégia adotada para a descoberta de novos fármacos consistia em utilizar uma estrutura primária, de atividade conhecida, e tentar encontrar moléculas análogas, que exibiam a atividade desejada, através de um processo exaustivo de tentativa e erro. Além de ser completamente empírico, esse processo era demorado e dispendioso. Da década de 80 para cá, com as diversas conquistas obtidas nas áreas computacional, de química teórica, biotecnologia, cristalografia de proteínas, etc., o desenvolvimento de novos fármacos tem seguido um processo mais racional onde a modelagem molecular tem estado cada vez mais presente.

A modelagem molecular¹ consiste na visualização gráfica e na representação da geometria de uma molécula através de um conjunto de técnicas computacionais que aliam métodos de química teórica a dados experimentais. Com esses métodos teóricos pode-se avaliar e prever certas características moleculares (conformação de mínima energia e propriedades fisicoquímicas) que

são de fundamental importância no entendimento das correlações estrutura-atividade, isso permite o desenvolvimento racional (baseado na estrutura) de novas drogas. Os três principais métodos teóricos computacionais utilizados para se calcular as propriedades moleculares classificam-se em empíricos (mecânica molecular), semiempíricos e *ab initio*. Posteriormente discutiremos cada um deles com mais detalhes.

No desenvolvimento racional de drogas existem dois tipos de formulação. Na primeira delas, chamada direta, são conhecidos o receptor (ou enzima) e suas características tridimensionais e essa informação é utilizada no desenvolvimento do novo fármaco. Na segunda delas, a indireta, a geometria do receptor não é conhecida, e os novos compostos são construídos através de um modelo obtido pela análise de características estereoquímicas e fisicoquímicas de um conjunto conhecido de moléculas ativas/inativas. Nos estudos estrutura-atividade que desenvolvemos sempre utilizamos a formulação indireta pelo fato de estarmos interessados em verificar se a informação a respeito da atividade biológica dos compostos residia em suas próprias estrutura eletrônicas, além disso os mecanismos de ação dos compostos que estudamos ainda não são bem compreendidos, soma-se a isso o fato de não possuímos maquinário computacional necessário para os cálculos que a abordagem direta exige. Na abordagem indireta bons modelos podem ser obtidos através da elaboração de experimentos apropriados para resolver as possíveis contradições observadas.

Seja qual for o método computacional adotado para o cálculo da geometria molecular, devemos fornecer o modelo inicial para a geometria molecular. Esse modelo inicial pode ser derivado de uma informação estrutural experimental

(cristalografia de raios-X ou ressonância magnética nuclear) ou pode ser construído através de programas de modelagem molecular quando dados experimentais não são disponíveis. Existem várias técnicas para se gerar um modelo molecular inicial 3-D. Em todas elas é importante que o programa utilizado mantenha a quiralidade, a valência e a aromaticidade apropriadas para gerar uma estrutura 3D aceitável. Em nossos estudos utilizamos um programa onde geramos um diagrama 2D através da conexão de fragmentos da molécula que eram otimizados através de mecânica molecular para, ao final, obtermos a estrutura 3-D. Uma vez construído o modelo inicial do sistema molecular é necessário explorar as conformações moleculares permitidas. Na maioria das vezes estamos interessados nas estruturas moleculares de mínima energia, embora nem sempre a conformação bioativa seja a de mínima energia.

Teoricamente, os cálculos de mecânica quântica devem fornecer uma descrição completa da energia de uma conformação particular da molécula, porém devido ao grande número de átomos (e de funções de onda) envolvidos, a mecânica molecular é a principal técnica utilizada para avaliar a energia das diferentes conformações moleculares. Uma vez obtida a conformação de mínima energia ela é empregada em cálculos de química quântica.

1.2 MECÂNICA MOLECULAR

A mecânica molecular² é, essencialmente, um tipo de aproximação que se faz para calcular a energia do sistema molecular utilizando-se uma função de energia empírica que tenta descrever as ligações covalentes através de

potenciais tipo 'mola' entre átomos ligados. Já os átomos não ligados são descritos através de potenciais tipo van der Waals e eletrostático. A idéia fundamental da mecânica molecular é de que as ligações têm comprimentos e ângulos naturais e que o sistema molecular tende a adotar uma geometria que convirja para esses valores.

Comumente as contribuições à energia potencial do sistema molecular incluem termos ligantes associados ao estiramento das ligações (E_{lig}), as deformações angulares (E_{ang}) e as rotações internas em torno das ligações simples (E_{die}); e termos não-ligantes, associados às interações de van der Waals (E_{vdW}) e eletrostáticas (E_{ele}). A soma de todos esses termos constitui a energia estereoquímica do sistema que deve ser avaliada de maneira a fornecer uma estrutura molecular de mínima energia:

$$E_{est} = E_{lig} + E_{ang} + E_{die} + E_{vdW} + E_{ele} \quad (1-1)$$

A função energia potencial que descreve as interações entre um conjunto de átomos especificados por suas coordenadas cartesianas (campo de força) varia de acordo com o método de MM escolhido. Cada método possui um campo de força parametrizado especificamente para determinado tipo de molécula³, o que implica em diferentes funções potenciais analíticas, diferentes formas de representar os átomos e diferentes conjuntos de parâmetros ajustáveis. Em geral, as expressões analíticas que aparecem em (1-1) são do tipo:

Estiramento da Ligação (1-2): Representa a deformação da ligação em relação ao seu comprimento de equilíbrio r_0 , sendo K_r a constante de força de estiramento.

$$E_{lig} = \sum_{\text{ligações}} K_r (r - r_0)^2 \quad (1-2)$$

Deformação do Ângulo (1-3): Representa a energia potencial de deformação angular da ligação em relação a um ângulo natural da ligação, θ_0 , sendo K_θ a constante de força de deformação.

$$E_{ang} = \sum_{\text{ângulos}} K_\theta (\theta - \theta_0)^2 \quad (1-3)$$

Rotações Internas em Torno das Ligações Simples (1-4): O potencial de torção (E_{die}) está associado ao fato dos ângulos diedrais possuírem uma simetria preferencial e terem um mínimo de energia para conformações do tipo *cis*, *trans*, *gauche*, etc.

$$E_{die} = \sum_{\text{diedral}} \frac{V_n}{2} [1 + \cos(n\phi - \phi_0)], \quad (1-4)$$

Para $n=1$ e $\phi_0=0$, tem-se a situação onde a energia é mínima para a conformação *trans* e tem uma barreira de rotação V_N para a maior energia da conformação *cis*.

Interações de van der Waals (1-5): Representa a energia potencial associada aos átomos não ligados:

$$E_{vdW} = \sum_{i,j \in vdW} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} \right] \quad (1-5)$$

A_{ij} e B_{ij} são parâmetros que controlam a profundidade e a posição do poço de energia potencial para o par de átomos não ligados e podem ser obtidos através de medidas de polarizabilidade atômica ou através de dados cristalográficos.

As interações não ligantes usualmente incluem uma repulsão de troca quando os átomos estão muito próximos e uma atração de longo alcance vinda de forças e dispersão. O primeiro termo que descreve esses dois tipos de interação é chamado atração de van der Waals. A atração de van der Waals é de curto alcance e tende rapidamente a zero quando os átomos não ligantes são afastados. A repulsão ocorre quando a distância entre os átomos torna-se ligeiramente menor que os seus raios de contato. O termo que descreve a interação atrativa/repulsiva deve promover uma transição suave entre esse dois regimes. Esse termo é geralmente descrito por um potencial do tipo Lenard-Jones (1-5).

Interações eletrostáticas (1-6): Energia potencial associada às interações eletrostáticas:

$$E_{ele} = \sum_{i,j,i \neq j} \frac{q_i q_j}{\epsilon r_{ij}}, \quad (1-6)$$

q_i e q_j são as cargas pontuais de átomos não ligantes separados por uma distância r_{ij} , em um meio com constante dielétrica ϵ . Esse potencial descreve a interação monopolo-monopolo das cargas atômicas q_i e q_j separadas por uma distância r_{ij} . Outros tipos de interações eletrostáticas como as dipolo-dipolo e íon-dipolo também são empregadas em alguns campos de força. A constante dielétrica ϵ é usualmente tomada como um valor constante entre 1.0 e 5.0 e em alguns tipos de cálculo ele pode assumir valores que variem com a distância.

Em adição aos potenciais descritos anteriormente, freqüentemente tem-se termos cruzados que combinam diferentes interações, como, por exemplo, o

estiramento da ligação com a distorção angular. Também alguns campos de força possuem um termo específico para as ligações de hidrogênio (pontes de hidrogênio). Geralmente esse termo é dado por um potencial da forma 12-10 (1-7):

$$E_H = \sum_{ij \in \text{ponte-H}} \frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}} \quad (1-7)$$

Essa função potencial é particularmente importante para moléculas onde as pontes de hidrogênio são fundamentais na estabilização do sistema, como no caso das moléculas de DNA.

Todos os parâmetros que aparecem nas componentes da energia estérea do sistema são ajustados de forma a reproduzir as geometrias de moléculas relativamente simples das quais se tem alguma informação experimental (por exemplo dados de Raio-X), e uma vez ajustados assume-se que parâmetros correspondentes e constantes de força possam ser transferidos para moléculas similares.

O problema de otimização de energia constitui, na prática, em determinar as regiões de mínima energia de uma superfície de potencial multidimensional descrita pelas variáveis de (1-1). Isso é feito minimizando a expressão para a energia estérea do sistema molecular. A maioria das superfícies de energia potencial é extremamente complexa, apresentando vários mínimos locais, e a complexidade aumenta com o aumento do número de graus de liberdade do sistema molecular. Por esse motivo a convergência dos cálculos de otimização de energia não significa necessariamente que um mínimo de energia global foi

encontrado. Para a obtenção de uma geometria de mínima energia muitas vezes é necessário se fazer uma análise extensiva das ligações flexíveis presentes na molécula. Isso é feito através de uma busca conformacional onde varia-se a conformação molecular enquanto se monitora a energia.

Uma outra técnica que pode ser utilizada para explorar o espaço conformacional de uma molécula é a chamada dinâmica molecular^{2,3}. Como o próprio nome sugere, a característica básica da dinâmica molecular é o cálculo da trajetória da molécula. As forças que atuam sobre os átomos são calculadas através da primeira derivada da energia potencial com respeito às posições atômicas. Aplicando-se as equações de movimento de Newton, essas forças podem ser usadas para calcular como a evolução das posições atômicas com o tempo, o que dá origem a uma trajetória dinâmica que representa uma série de conformações moleculares que evoluem com o tempo.

Com os métodos de MM pode-se estudar moléculas na fase condensada, macromoléculas, morfologia cristalina e de interfaces orgânicas e inorgânicas e outros sistemas em que as propriedades de interesse não sejam extremamente dependentes dos efeitos quânticos. A grande popularidade dos métodos de MM vem do fato deles poderem ser empregados no estudo de sistemas moleculares relativamente grandes (centenas de átomos) com um custo computacional razoável. Além disso, esses métodos permitem a análise da composição da energia estérea dos sistema (ligação, angular, van der Waals, etc) e a 'engenharia' dos campos de força pode ser feita com relativa facilidade.

Os programas mais populares de mecânica molecular são baseados nos campos de força de MM2 e MM₃, de Allinger⁴. Em nossos cálculos utilizamos os

métodos MM⁺ e AMBER como implementados no programa HyperChem® versão 5.0. O campo de força do MM⁺ é uma extensão do MM2 e, diferentemente de outros campos de força, os potenciais de estiramento e de deformação angular possuem termos de ordem superior à quadrática. Outra característica do MM⁺ é descrever as interações de átomos não ligantes através de termos 1-3 (termos de Urey-Bradley), enquanto que os outros métodos avaliam essas interações para átomos separados por três ou mais ligações (1-4 e interações maiores). O método AMBER⁵ (Assisted Model Building and Energy Refinement) é baseado num campo de força desenvolvido para proteínas e ácidos nucleicos e inclui termos relativos às ligações de hidrogênio.

Pelo fato de não tratar os elétrons explicitamente, os métodos de MM não são apropriados para descrever situações em que os efeitos eletrônicos são essenciais, tais como a formação de ligações ou as interações entre orbitais. Usualmente as geometrias moleculares obtidas com a MM são utilizadas como geometrias iniciais em cálculos mais precisos onde os elétrons são tratados explicitamente.

1.3 MECÂNICA QUÂNTICA

Diferentemente do tratamento clássico adotado pelos métodos de MM, pode-se tratar o sistema molecular dentro do formalismo da mecânica quântica^{6,7}, e assim obter-se dos cálculos propriedades associadas a estrutura eletrônica do composto. Logicamente os métodos que empregam o formalismo quântico são

muito mais custosos computacionalmente que os métodos de MM, e por isso são limitados a sistemas moleculares não muito extensos.

O primeiro passo para se tratar o sistema molecular quanticamente é estabelecer uma Hamiltoniana para seu conjunto de núcleos e elétrons, em que senão todas, a maioria das interações existentes seja considerada. Um exemplo comum de Hamiltoniana para o sistema molecular é dado pela equação (1-8).

O primeiro termo da Hamiltoniana refere-se à energia cinética de cada núcleo, o segundo termo representa alguma força externa atuando sobre o núcleo (comumente tomada como nula) e o terceiro termo refere-se a repulsão eletrostática entre os núcleos. O próximo termo representa a atração eletrostática entre o elétron i e o núcleo A . Os dois últimos termos são a energia cinética e de repulsão eletrostática dos elétrons, respectivamente. Nesse caso as interações entre todos os momentos magnéticos de spin e o movimento orbital dos núcleos e elétrons e aos efeitos relativísticos de massa foram desprezadas.

$$H = -\frac{1}{2} \sum_A [\nabla_A^2 + V(R_A)] + \sum_{A<B} \frac{Z_A Z_B}{R_{AB}} - \frac{1}{2} \sum_{A,i} \frac{Z_A}{R_{Ai}} + \sum_{i<j} \frac{1}{r_{ij}} - \frac{1}{2} \sum_i \nabla_i^2 \quad (1-8)$$

onde

A, B referem-se aos núcleos.

R_{AB} designa a distância entre os núcleos A e B .

R_{Ai} designa a distância entre os núcleos A e o elétron i .

Z_A refere-se a carga do núcleo A .

i, j referem-se aos elétrons.

r_{ij} designa a distância entre os elétrons i e j .

∇_i^2 é o operador gradiente relativo as coordenadas eletrônicas.

∇_A^2 é o operador gradiente relativo as coordenadas nucleares.

A aplicação desse operador Hamiltoniano ao sistema molecular possui solução exata somente para o caso diatômico, sendo que para o restante dos casos este é um problema matematicamente intratável.

1.3.1 Aproximação de Born-Oppenheimer

Uma das formas de simplificar (1-8) é procurar separar o movimento eletrônico do movimento nuclear, isso é feito mantendo-se os núcleos fixos durante cada ciclo do movimento eletrônico, o que é conhecido como **aproximação de Born-Oppenheimer**. Como a massa nuclear é muito maior que a eletrônica, é uma aproximação razoável considerar que os núcleos mudem de posição tão lentamente que permitam aos elétrons se ajustarem muito rapidamente à nova configuração nuclear. Desse modo os elétrons são tratados como se movessem em um campo de núcleos estacionários, o que permite separar o problema de calcular a energia de um sistema molecular em dois problemas separados, um eletrônico e outro nuclear.

Podemos reescrever o Hamiltoniano total do sistema molecular como sendo a soma de uma parte nuclear (1-9) e outra eletrônica(1-10) e a função de onda total do sistema, que é autofunção da equação de Schödinger, como sendo um produto das funções de onda nuclear e eletrônica (1-11). Assim a, equação de Schödinger para o sistema teria a forma (1-12).

$$H_N = -\frac{1}{2} \sum_A \nabla_A^2 + V(R_A) \quad (1-9)$$

$$H_e = \sum_{A<B} \frac{Z_A Z_B}{R_{AB}} - \frac{1}{2} \sum_{A,i} \frac{Z_A}{R_{Ai}} - \frac{1}{2} \sum_i \nabla_i^2 + \sum_{i<j} \frac{1}{r_{ij}} \quad (1-10)$$

$$\Psi(r_i, R_A) = \Psi_N(R_A) \Psi_e(R_A, r_i) \quad (1-11)$$

$$H\Psi = E\Psi \rightarrow (H_N + H_e)\Psi_N \Psi_e = E\Psi_N \Psi_e \quad (1-12)$$

Como H_e é função somente de r_i , uma vez que os R_A 's são tratados como constantes, podemos escrever uma equação de Schrödinger eletrônica para uma configuração nuclear específica, R_A , da forma:

$$H_e \Psi_e = \varepsilon_e(R_A) \Psi_e \quad (1-13)$$

Supondo a força externa sobre o núcleo nula ($V(R_A)=0$) e admitindo que a função de onda eletrônica varie lentamente com a distância internuclear de forma que possamos omitir a derivada de $\Psi(r, R_A)$ com respeito a R_A , a equação de Schrödinger nuclear assume a forma:

$$H_N \Psi_N \Psi_e \cong \sum_A \Psi_e \nabla_A^2 \Psi_N \rightarrow H_N \Psi_N \cong \varepsilon_N \Psi_N, \quad (1-14)$$

onde ε_N refere-se à energia cinética do núcleo.

Dessa forma a aproximação de Born-Oppenheimer desacopla o movimento do núcleo do movimento eletrônico.

1.3.2 Aproximação do Elétron Independente

Consideremos agora a resolução do problema eletrônico. A equação de Schrödinger eletrônica (1-10) não permite soluções exatas devido às interações intereletrônicas. Uma das aproximações que se faz (conhecida como aproximação

do elétron independente) é considerar que a função de onda que descreve o sistema molecular possa ser escrita como produto das funções de 1-elétron. Neste caso assumimos que as energias cinética e potencial de cada elétron sejam independentes do movimento e da posição dos outros elétrons, de forma que a Hamiltoniana eletrônica total possa ser aproximada pela soma da Hamiltonianas efetivas de 1-elétron (1-15). Podemos resolver a equação de Schrödinger para o elétron independente para uma função de onda ψ que se estende por toda molécula e que descreva o elétron individualmente (1-16). Essa equação possui múltiplas soluções ψ_i e ε_i . As funções ψ_i são chamadas de orbitais moleculares (OM) e a energia ε_i de um elétron no orbital ψ_i é chamada de energia orbital.

$$H_\varepsilon = \sum_i H^{1-ele}(i) \quad (1-15)$$

$$F(r_1)\psi_i(r_1) = \varepsilon_i\psi_i(r_1) \quad (1-16)$$

Como os elétrons possuem spin, além das coordenadas espaciais a função de onda também deve possuir coordenadas de spin, que podem assumir duas configurações, *up* ou *down*. As funções assim construídas são chamadas Spin-Orbitais Moleculares (SOM). Como dois elétrons não podem ocupar funções SOM idênticas (princípio da exclusão de Pauli) uma das formas de se garantir isso é escrever a função de onda multieletrônica como sendo um produto antissimétrico das spin-orbitais moleculares, o que nos leva, para um sistema de N partículas, à expressão:

$$\Psi_e(r_1 \dots r_N) = \sum_p (\mp)^p P_{op} [\psi_1(r_1), \psi_2(r_2), \dots, \psi_N(r_N)], \quad (1-17)$$

onde P_{op} é o operador de permutação, a soma é sobre todas as permutações dos estados $\psi_i(r_i)$ e $(\mp)^p$ indica -1 para uma permutação ímpar e $+1$ para uma permutação par. Essa expressão também pode ser escrita na forma de um determinante, conhecido como determinante de Slater.

A avaliação da energia eletrônica total do sistema molecular ($\varepsilon = \langle \psi | H | \psi \rangle$) quando a função de onda é escrita na forma (1-17) nos leva à:

$$\varepsilon = 2 \sum_i h_{ii} + \sum_i \sum_j (2J_{ij} - K_{ij}), \text{ onde} \quad (1-18)$$

$$h_{ii} = \langle \psi_i | -\frac{1}{2} \nabla_i^2 - \sum_A \frac{Z_A}{|r_i - R_A|} | \psi_i \rangle \quad (1-19)$$

$$J_{ij} = \iint \psi_i^*(r_1) \psi_j^*(r_2) \frac{1}{|r_1 - r_2|} \psi_i(r_1) \psi_j(r_2) dr_1 dr_2 \quad (1-20)$$

$$K_{ij} = \iint \psi_i^*(r_1) \psi_j^*(r_2) \frac{1}{|r_1 - r_2|} \psi_j(r_1) \psi_i(r_2) dr_1 dr_2 \quad (1-21)$$

O termo h_{ii} representa a energia de 1-elétron no orbital molecular ψ_i sujeito a um campo produzido por núcleos fixos. As integrais do tipo J_{ij} são chamadas integrais de Coulomb, e podemos dizer que fisicamente elas representam a interação entre duas densidades de carga ($\psi_i \psi_i^*$ e $\psi_j \psi_j^*$). As integrais K_{ij} são chamadas integrais de troca por possuírem dois produtos de funções de onda que diferem pela troca dos índices eletrônicos r_1 e r_2 . Matematicamente o que temos nesse caso é a interação de uma distribuição eletrônica (sem sentido físico) com outro elétron na mesma distribuição. Fisicamente é difícil atribuir um significado a essas integrais, mas elas são resultado do princípio de antissimetria das funções

de onda moleculares e refletem a energia de estabilização devida à correlação de elétrons com spins paralelos.

A determinação dos orbitais ψ_i é feita aplicando-se o princípio variacional à equação de energia (1-18), sujeita à condição de ortonormalidade da função de onda. A melhor função de onda Ψ é aquela que minimiza a energia do estado fundamental e o método variacional nos mostra que as OM ótimas devem ser autofunções da seguinte equação de autovalores:

$$f(r_1)\psi_i(r_1) = \varepsilon_i\psi_i(r_1) \quad (1-22)$$

onde $f(r_1)$ é o denominado operador de Fock cuja expressão é dada por:

$$f(r_1) = h(r_1) + \sum_j (2J_j(r_1) - K_j(r_1)), \text{ sendo} \quad (1-23)$$

$$h(r_i) = -\frac{1}{2}\nabla_i^2 - \sum_A \frac{Z_A}{|r_i - R_A|} \quad (1-24)$$

$$J_j(r_1) = \int \psi_j^*(r_2) \frac{1}{|r_1 - r_2|} \psi_j(r_2) \quad (1-25)$$

$$K_j(r_1)\psi_i(r_1) = \left[\int \psi_j^*(r_2) \frac{1}{|r_1 - r_2|} \psi_i(r_2) dr_2 \right] \psi_j(r_1) \quad (1-26)$$

$h(r_i)$ é o operador Hamiltoniano de 1-elétron. $J_j(r_1)$ é o chamado operador de Coulomb, pois origina os termos de energia correspondentes as repulsões entre as densidades de cargas (1-20) e $K_j(r_1)$ é o chamado operador de troca, que leva, após aplicado a função spin orbital molecular, às integrais de troca (1-21). Note que se os spins associados aos orbitais ψ_i e ψ_j forem diferentes, os termos

referentes ao operador $K_j(r_1)$ serão nulos. Isso acontece porque a integração sobre o espaço das coordenadas de spin do elétron 1 (ou 2) na equação (1-21) leva à integração sobre duas funções de spins ortogonais.

O problema original, que era encontrar as soluções da equação de Schrödinger independente do tempo (1-8), se reduziu a encontrar as soluções para as equações de autovalores (1-23), conhecidas com equações de Fock. A maior vantagem com relação ao problema anterior é o fato do operador de Fock ser um operador de 1-elétron, porém a equação de autovalores continua tendo um número infinito de soluções.

1.3.3 O Procedimento LCAO

O próximo passo é expandir as funções OM em um conjunto finito de funções de base conhecidas (1-27). Essas funções de base são freqüentemente escolhidas como sendo as funções orbitais atômicas pelo fato dessa escolha facilitar a interpretação dos resultados, pois as propriedades moleculares estão diretamente relacionadas aos átomos constituintes da molécula. Esse procedimento é conhecido como combinação linear dos orbitais atômicos (LCAO - linear combination of atomic orbitals):

$$\psi_i(r_i) = \sum_{\mu} c_{i\mu} \phi_{\mu}, \quad (1-27)$$

Deve-se agora encontrar os valores ótimos para os coeficientes $c_{i\mu}$ que levem ao conjunto de funções auto-consistentes que minimizem a energia eletrônica do sistema, que agora é dada por (1-28) devido a substituição dos OM pela LCAO. Novamente isso é feito minimizando a energia eletrônica total com

respeito aos parâmetros variacionais $c_{i\mu}$, sujeitos à condição de ortonormalidade da funções Ψ :

$$\varepsilon = \sum_{\mu\nu} P_{\mu\nu} H_{\mu\nu} + \frac{1}{2} \sum_{\mu\nu\lambda\sigma} P_{\mu\nu} P_{\lambda\sigma} \left[\langle \phi_\mu \phi_\nu | \frac{1}{r_{12}} | \phi_\lambda \phi_\sigma \rangle - \frac{1}{2} \langle \phi_\mu \phi_\lambda | \frac{1}{r_{12}} | \phi_\nu \phi_\sigma \rangle \right] \text{ onde} \quad (1-28)$$

$$H_{\mu\nu} = \langle \phi_\mu | h | \phi_\nu \rangle \quad (1-29)$$

$$P_{\lambda\sigma} = 2 \sum_i C_{i\lambda} C_{i\sigma} \quad (1-30)$$

$$S_{\mu\nu} = \langle \phi_\mu(r_i) | \phi_\nu(r_i) \rangle \quad (1-31)$$

$H_{\mu\nu}$ (Eq 29) é a matriz densidade da Hamiltoniana de 1-elétron com respeito aos orbitais atômicos ϕ_μ e ϕ_ν . As matrizes $P_{\lambda\sigma}$ e $S_{\mu\nu}$ são chamadas matriz densidade e matriz de superposição, respectivamente.

A expressão da densidade de carga em uma região R da molécula é dada por:

$$\rho(R) = \sum_{\mu\nu} P_{\mu\nu} \phi_\mu(R) \phi_\nu(R) = \sum_{\mu\nu} P_{\mu\nu} S_{\mu\nu} \quad (1-32)$$

O somatório $\sum_{\mu\nu} P_{\mu\nu} S_{\mu\nu}$ sobre todos os orbitais atômicos nos dá o número total de elétrons do sistema. O produto $P_{\mu\nu} S_{\mu\nu}$ está associado à distribuição da população eletrônica na região de superposição entre os orbitais atômicos ϕ_μ e ϕ_ν , e os termos diagonais desse produto, como por exemplo $P_{\mu\mu} S_{\mu\mu}$, estão associados à carga líquida residente no orbital atômico ϕ_μ .

Como resultado desse procedimento temos um conjunto de equações algébricas (Eq. 32) conhecidas como equações de Hartree-Fock-Roothaan (HFR),

$$\sum_{\nu} (F_{\mu\nu} - \varepsilon_i S_{\mu\nu}) c_{\nu} = 0 \quad (1-33)$$

onde a matriz de Fock é dada por:

$$F_{\mu\nu} = H_{\mu\nu} + \sum_{\lambda\sigma} P_{\lambda\sigma} \left[\langle \phi_{\mu} \phi_{\nu} | \frac{1}{r_{12}} | \phi_{\lambda} \phi_{\sigma} \rangle - \frac{1}{2} \langle \phi_{\mu} \phi_{\lambda} | \frac{1}{r_{12}} | \phi_{\nu} \phi_{\sigma} \rangle \right] \quad (1-34)$$

Em termos matriciais as equações de HFR podem ser escritas como:

$$\mathbf{FC} = \mathbf{SCE} \quad (1-35)$$

O problema agora passa a ser resolver esse conjunto de equações algébricas não-lineares. Essa não-linearidade vem do fato do operador de Fock depender dos coeficientes $C_{\mu i}$ ($F=F(C)$), que não são conhecidos *a priori*. Uma maneira de resolver o sistema de equações HFR é através do método iterativo conhecido como método do campo auto-consistente (SCF). Esquemáticamente o procedimento adotado é o seguinte:

- 1- Gera-se aleatoriamente (ou através de um método aproximado) uma matriz P inicial.
- 2- Calcula-se $F_{\mu\nu}$ e $S_{\mu\nu}$.
- 3- Resolve-se as equações HFR através da diagonalização de \mathbf{F} e obtém-se \mathbf{C} e ε .
- 4- Usa-se os valores de C para construir uma nova matriz densidade e retorna-se ao passo 2.

Esse processo iterativo prossegue até que a energia ou a densidade de carga não sofra mudança dentro de uma precisão especificada entre duas iterações sucessivas. Ao atingir essa convergência a solução das equações HFR fornecerá o conjunto de coeficientes de cada orbital molecular na base dos orbitais atômicos juntamente com a energia associada com cada um dos orbitais moleculares.

Podemos sumarizar o procedimento para se realizar um cálculo Hartree-Fock de um sistema molecular como:

- 1- Escreve-se a equação de Schrödinger para o sistema.
- 2- Usa-se um determinante de Slater contendo orbitais moleculares como função de onda.
- 3- Obtém-se as equações não lineares de Hartree-Fock pelo uso do princípio variacional.
- 4- Introduce-se um conjunto finito de bases (orbitais atômicas) para se obter equações algébricas (equações de Hartree-Fock-Roothaan).
- 5- Tenta-se resolver essas equações utilizando um procedimento auto-consistente (SCF).

O maior obstáculo à resolução das equações de Hartree-Fock-Roothaan está no grande número de integrais de dois elétrons a serem calculadas. Essas integrais podem possuir um ($\mu^A \nu^A | \lambda^A \sigma^A$), dois ($\mu^A \nu^A | \lambda^A \sigma^B$), três ($\mu^A \nu^A | \lambda^B \sigma^C$) ou quatro ($\mu^A \nu^B | \lambda^C \sigma^D$) centros, dependendo do número de centros atômicos (A, B, C ou D) nos quais estão localizadas as funções orbitais moleculares (μ , ν , λ ou

σ). Os métodos *ab initio* (primeiros princípios) caracterizam-se pelo cálculo explícito de todas essas integrais, que são originadas da utilização de um conjunto finito de funções de base arbitrárias. São cálculos extremamente custosos computacionalmente, o que restringe sua aplicação a sistemas moleculares pequenos. Outros métodos, que também adotam uma formulação rigorosa da mecânica quântica, utilizam-se de algumas aproximações para tentar resolver essas integrais de dois elétrons. Eles são os conhecidos na literatura como métodos semiempíricos. As aproximações adotadas tornam esses métodos extremamente rápidos computacionalmente em comparação aos métodos *ab initio*, o que permite que sistemas maiores sejam investigados.

O conjunto das funções atômicas que descrevem os orbitais moleculares é chamado conjunto de base. Tanto nos métodos *ab initio* quanto nos métodos semiempíricos é comum a utilização de um conjunto de base mínima. O conjunto de base mínima caracteriza-se pela utilização da combinação linear do menor número possível de funções atômicas que possibilite descrever os orbitais atômicos ocupados do átomo. A formas dessas funções de base pode variar, mas as funções de onda tipo Gaussianas (1-36) são as mais utilizadas nos cálculos *ab initio* enquanto as funções de onda tipo Slater (1-37) são as mais empregadas nos cálculos semiempíricos:

$$\psi(r) = \exp(-\alpha r^2), \quad (1-36)$$

$$\psi(r) = r^{n-1} \exp^{-\zeta r} Y_{lm}(\theta, \phi) \quad (1-37)$$

1.3.4 Métodos Semiempíricos

Os principais métodos semiempíricos envolvem dois tipos de aproximação. Na primeira delas os elementos de matriz da Hamiltoniana efetiva de 1-elétron (1-29) são dados por valores empíricos ou semiempíricos, parametrizados, que tentam correlacionar o resultado dos cálculos aos dados experimentais existentes. Na segunda delas é feita uma aproximação matemática explícita, onde alguns termos relativos às integrais de superposição são desprezados. Essas aproximações são compensadas pela introdução de parâmetros empíricos ajustáveis aos dados experimentais.

As principais diferenças entre os diversos métodos semiempíricos se encontram nas integrais que são desprezadas na resolução das equações de Hartree-Fock-Roothaan, no conjunto de funções de base empregado e nos parâmetros utilizados. Uma característica comum dos métodos que descreveremos a seguir é que em todos eles as integrais de superposição que envolvem dois elétrons em átomos diferentes são desprezadas (ZDO- *zero differential overlap*), o que faz com que todas as integrais de três e quatro centros sejam nulas.

O primeiro método a surgir, o Método de Hückel, despreza completamente as integrais de 2-elétrons. A seguir veio o Método de Hückel Estendido, onde alguns efeitos das interações elétron-elétron foram incluídos através de parametrizações, posteriormente foram desenvolvidos os métodos CNDO (*complete neglect of differential overlap*), INDO (*intermediate neglect of differential overlap*), MINDO (*modified intermediate neglect of differential overlap*), NDDO (*Neglect of diatomic differential overlap*) e MNDO (*Modified Neglect of Differential*

Overlap). Uma descrição detalhada de cada um desses métodos pode ser encontrada na monografia de Pople *et al*⁷. Na Tabela 1-1 apresentamos, resumidamente, as principais diferenças entre alguns desses métodos semiempíricos.

Os métodos AM1⁸ (Austin Model 1) e PM3⁹ (Parametric Method 3) são os mais utilizados atualmente. O método AM1 resulta de uma modificação no MNDO¹⁰ onde a função que descreve a repulsão entre caroços atômicos foi alterada e novos parâmetros foram introduzidos. Essas mudanças produziram previsões mais precisas de barreiras de ativação para muitas reações, bem como uma melhoria de 40% nas previsões do calor de formação de moléculas, se comparadas ao MNDO. O método PM3 é uma reparametrização do AM1, em que os parâmetros utilizados foram obtidos de um número muito maior e representativo de dados experimentais. É vasta a literatura comparando a eficiência desses dois métodos¹¹⁻¹⁶, porém não há consenso a respeito de qual deles reproduz melhor as geometrias e barreiras moleculares.

Deve-se tomar cuidado ao se escolher o método de cálculo utilizado no estudo de determinada série de composto. Cada método possui suas particularidades e devemos analisar qual método é mais indicado ao conjunto de moléculas em questão e às propriedades moleculares nas quais estamos interessados.

Neste trabalho de tese utilizamos de forma extensiva os métodos AM1, PM3 e ZINDO/S¹⁷. Em particular empregamos o método ZINDO para avaliar o GAP de polímeros nitrogenados¹⁸. Por ser este um trabalho que foge aos objetivos

desta tese não o apresentamos. Maiores detalhes podem ser obtidas no artigo acima citado.

MÉTODO	CARACTERÍSTICAS
CNDO (CNDO/1, CNDO/2, CNDO/S)	<ul style="list-style-type: none"> • Despreza as integrais de troca. • Retém as integrais de 2-elétrons do tipo $(\mu\mu \nu\nu)$. • considera uma simetria esférica para os orbitais atômicos.
INDO (INDO/1, INDO/2, INDO/S)	<ul style="list-style-type: none"> • inclui integrais de repulsão de 1-centro entre orbitais atômicos de um mesmo átomo $(\mu\nu^A \lambda\sigma^A)$
MINDO – modificação do INDO (MINDO/2, MINDO/2', MINDO/3,)	<ul style="list-style-type: none"> • usa um conjunto de parâmetros para avaliar as integrais de 1-centro.
NDDO	<ul style="list-style-type: none"> • mantém as integrais de 1-centro do INDO. • adiciona algumas integrais de 2-centros. • inclui a direcionabilidade dos orbitais em um mesmo átomo para as integrais de repulsão.
MNDO – modificação do NDDO (AM1 e PM3)	<ul style="list-style-type: none"> • ignora somente as integrais de superposição quando os orbitais atômicos estão centrados em átomos diferentes.
ZINDO (ZINDO/1, ZINDO/2, ZINDO/S)	<ul style="list-style-type: none"> • versão do INDO. • Inclui interação de configuração (CI) simples e duplas • parametrizado para reproduzir transições espectoscópicas no UV visível (ZINDO/S)

Tabela 1-1 Métodos Semiempíricos

Capítulo 2

PROPRIEDADES FÍSICO-QUÍMICAS UTILIZADAS NO ESTUDO ESTRUTURA-ATIVIDADE

2.1 INTRODUÇÃO

A resposta biológica que determinado composto promove em um organismo vivo depende de diferentes fatores que estão diretamente correlacionados às suas propriedades físico-químicas^{1,19}. Podemos agrupar esses fatores em três classes principais de propriedades físico-químicas: Hidrofóbica, estérea e eletrônica²⁰. A hidrofobicidade está relacionada ao mecanismo de transporte do composto no meio biológico, que é governado principalmente pela partição entre fases lipídicas e aquosas. As propriedades estéreas estão relacionadas aos fatores tais como o tamanho e a forma do composto, o que pode ser determinante para o acoplamento do composto com seu receptor ou enzima e, por fim, temos as propriedades eletrônicas que estão relacionadas à susceptibilidade de determinada ligação ao ataque metabólico eletrofílico ou nucleofílico.

Nos estudos de estrutura-atividade procura-se correlacionar parâmetros originados dessas propriedades à determinada resposta biológica de interesse.

2.2 PARÂMETROS HIDROFÓBICOS

São os parâmetros relacionados à solubilidade e ao transporte do composto no meio biológico.

2.2.1 Solubilidade

A solubilidade de um composto está relacionada à facilidade com que ele é solúvel em diferentes meios, que vão desde meios polares, como a água, até meios apolares, como os lipídios. A importância da solubilidade vem do fato dela estar diretamente ligada ao mecanismo de absorção do composto, uma vez que, inserido no organismo, ele deve atravessar constantemente meios lipídicos e aquosos até chegar a seu receptor.

2.2.2 Coeficiente de Partição

Para tentar descrever o comportamento de um composto no meio celular criou-se um parâmetro chamado coeficiente de partição, P , que descreve a razão das concentrações do composto em dois meios líquidos imiscíveis em equilíbrio. Os meios imiscíveis mais comumente utilizados são a água e o octanol. O octanol tem uma cabeça polar e uma cadeia alquila hidrofóbica que faz com que ele seja um bom modelo para a cadeia lipídica da célula. Assim a mistura octanol/água constitui um bom modelo para a barreira celular. Geralmente se utiliza o logaritmo do coeficiente de partição ($\log P$) como parâmetro hidrofóbico. Um ótimo valor de $\log P$ deve corresponder à máxima probabilidade de um composto encontrar seu receptor em um “caminhar randômico” entre vários compartimentos hidrofílicos e lipofílicos.

2.3 PARÂMETROS ESTEROQUÍMICOS

O volume e a forma das moléculas e dos grupos substituintes são características importantes a serem observadas no controle da atividade biológica. Muitas vezes droga e receptor têm um formato complementar (conceito chave-fechadura), e qualquer modificação que altere a conformação da droga impedindo que haja um encaixe 'ótimo' com o receptor pode levar a um enfraquecimento das interações e uma conseqüente perda ou diminuição de atividade. Podemos citar como parâmetros relacionados às propriedades estéreas do composto o volume molecular, a área superficial e o raio de van der Waals.

2.4 PARÂMETROS ELETRÔNICOS

A distribuição eletrônica sobre uma molécula e a facilidade com que essa distribuição pode ser modificada está diretamente relacionada às forças de interação intermoleculares. Essas interações exercem uma influência direta na resposta biológica que um composto é capaz de induzir. Com o objetivo de modelar essa resposta biológica, vários parâmetros eletrônicos foram e vêm sendo propostos nos diversos estudos de estrutura-atividade. Existem os parâmetros clássicos, tais como o pKa, o momento de dipolo, a ligação de hidrogênio, a refratividade molar, etc., e os parâmetros obtidos dos cálculos de Química Quântica. Nos deteremos aos parâmetros químico-quânticos.

2.4.1 Carga Atômica

Os métodos semiempíricos descritos no capítulo 1 são capazes de calcular as cargas atômicas dos átomos que formam a molécula. Entretanto, os valores

obtidos são diferentes para diferentes métodos, pois os valores das cargas são muito sensíveis ao conjunto de funções de base adotado e aos métodos de cálculo.

Com as funções de onda moleculares escritas como a combinação linear de orbitais atômicos dentro da aproximação ZDO (zero-differential overlap), os elementos diagonais da matriz densidade, $P_{\mu\mu}^{\alpha}$ e $P_{\mu\mu}^{\beta}$, nos dão a população eletrônica no orbital atômico sobre questão. A soma das populações eletrônicas de todos os orbitais atômicos centrados em um dado átomo A nos dá a população eletrônica total do átomo A, chamada P_{AA} .

$$P_{AA} = \sum_{\mu} (P_{\mu\mu}^{\alpha} + P_{\mu\mu}^{\beta}) \quad (2-1)$$

onde o somatório é sobre todos os orbitais atômicos ϕ_{μ} , e α e β distinguem os spins eletrônicos.

A carga atômica líquida do átomo é encontrada subtraindo sua carga nuclear da sua população eletrônica.

$$q_A = Z_A - P_{AA} \quad (2-2)$$

A dificuldade em se obter valores precisos para a carga atômica faz com que sua utilização como parâmetro eletrônico seja recomendável somente em estudos comparativos entre compostos de uma mesma série. Pode-se assim, por exemplo, escolher como parâmetro a carga em um determinado átomo que seja comum a todo o conjunto de moléculas examinadas e verificar se esse parâmetro está correlacionado de alguma forma à resposta biológica desse conjunto de

moléculas. Pode-se também utilizar como parâmetro a diferença de carga em uma determinada ligação.

2.4.2 Momento de dipolo

O momento de dipolo, dentro da aproximação ZDO, é calculado pelo somatório dos produtos da carga líquida localizada em cada núcleo (2-2) pelo vetor posição nuclear. A esse termo são adicionados termos de hibridização referentes às integrais de um centro que medem a contribuição devida ao deslocamento da carga do centro da posição nuclear (2-3). Esse efeito é proporcional aos elementos da matriz densidade fora da diagonal:

$$\mu = \sum_A q_A R_A + \sum_A \text{hibridização} , \quad (2-3)$$

sendo R_A a distância entre a origem e o núcleo A.

As interações tipo íon-dipolo, dipolo-dipolo, dipolo induzido por íon, etc, muitas vezes estão presentes na formação do complexo droga-receptor. Assim o momento de dipolo é um parâmetro que pode apresentar uma forte correlação com a resposta biológica do sistema.

2.4.3 Densidade Eletrônica de Fronteira (DEF)

Esse é um parâmetro que geralmente tem uma forte correlação com a atividade biológica quando o tipo de interação presente é muito localizada. Nesses casos a DEF é preferível à carga atômica total.

2.4.4 Energias do HOMO (highest occupied molecular orbital) e LUMO (lowest occupied molecular orbital)

A energia do HOMO (E_{HOMO}) representa a energia necessária para ‘arrancar’ um elétron da última camada de valência da molécula (potencial de ionização), portanto está relacionada à facilidade com que o elétron pode ser ‘dado’ pela molécula. Já a energia do LUMO (E_{LUMO}) está relacionada à facilidade com que a molécula pode ‘aceitar’ um elétron. Esses parâmetros geralmente estão correlacionados com atividade biológica quando a interação dominante é a transferência de carga entre as moléculas.

Quando ocorre uma ‘aproximação’ dos orbitais moleculares de dois compostos, a interação responsável pelo maior ganho de energia na formação do novo composto é a dos orbitais ocupados de um dos compostos com os orbitais desocupados do outro. Nesse aspecto, as interações entre o HOMO e o LUMO são, comumente, as mais efetivas; isso porque quanto mais próximos em energia estiverem os orbitais interagentes maior será a separação dos níveis de energia do complexo formado pelos dois compostos. Em uma expressão deduzida por Klopman e Salem²¹ pode-se estimar o ganho ou a perda de energia gerada pela interação entre esses orbitais (ocupados e desocupados):

$$\Delta E = -\sum_{ab} (q_a + q_b) \beta_{ab} S_{ab} + \sum_{k < l} \frac{Q_k Q_l}{\epsilon R_{kl}} + \sum_r^{\text{ocup}} \sum_s^{\text{desoc}} - \sum_r^{\text{ocup}} \sum_s^{\text{desoc}} \frac{2 \left(\sum_{ab} c_{ra} c_{sb} \beta_{ab} \right)^2}{E_r - E_s}, \text{ onde (2-4)}$$

q_a e q_b são as populações eletrônicas nos orbitais atômicos a e b .

β e S são as integrais de ressonância e de superposição.

Q_k e Q_l são as cargas totais nos átomos k e l .

ε é a constante dielétrica do meio.

R_{kl} é a distância entre os átomos k e l .

c_{ra} é o coeficiente do orbital atômico a no orbital molecular r , sendo que r se refere aos orbitais moleculares de uma molécula e s se refere aos orbitais moleculares da outra.

E_r é a energia do orbital molecular r .

A interação entre os orbitais moleculares é descrita pelo terceiro termo dessa expressão, que é de segunda ordem e somente é válida quando $E_r \neq E_s$. Podemos observar então que, quanto menor for a diferença de energia entre os orbitais moleculares interagentes, maior será a contribuição à variação de energia do sistema. A interação entre os orbitais moleculares é particularmente significativa quando estamos lidando com sistemas carregados. Nesse caso a maior contribuição para a formação do complexo ânion-cátion vem da interação carga-carga (segundo termo da expressão (2-4)), seguida pela contribuição da interação do HOMO do ânion com o LUMO do cátion.

Devemos ressaltar que se outros fatores intervêm, dificultando a interação HOMO/LUMO, pode-se esperar que orbitais subjacentes (mais baixos que o HOMO e mais altos que o LUMO) tornem-se mais influentes no ganho de energia do sistema.

2.4.5 Superdeslocalizabilidade (S_r)

A S_r é definida como sendo a soma dos quadrados dos coeficientes C_{rj} do r -ésimo orbital atômico do j -ésimo orbital molecular divididos pelas respectivas

energia orbitais^{1,20}. No caso da superdeslocalizabilidade para ataques eletrofílicos, S_r^E , a soma é sobre todos os orbitais ocupados, e no caso de ataques nucleofílicos S_r^N é dado pela soma sobre todos os orbitais desocupados.

2.4.6 Diferença de Energia HOMO LUMO

A diferença de energia entre o HOMO e o LUMO nos dá a energia necessária para promover o elétron do estado fundamental (HOMO) para o estado excitado (LUMO), o que geralmente pode ser correlacionado às absorções óticas do composto.

2.4.7 Metodologia de Índices Eletrônicos

A metodologia de índices eletrônicos (MIE) é uma proposta de abordar o problema farmacológico da discriminação entre compostos ativos/inativos (em relação à determinada resposta biológica), através da utilização de descritores quânticos-moleculares universais. Com a MIE é possível construir regras Booleanas simples, baseadas no conceito de densidade local de estados (LDOS) e nos valores críticos para a diferença de energia entre determinados os orbitais moleculares. Ela tem sido aplicada com sucesso para diferentes classes de compostos biológicos: carcinogênicos²²⁻²⁴, antibióticos²⁵, hormônios esteróides²⁶ e inibidores de protease²⁷. A MIE produz resultados preditivos da mesma ordem de precisão de métodos tradicionais de análise SAR (structure-activity relationship), tais como PCA (principal component analysis) e NN (neural networks)^{26,28,29}, apesar desses métodos utilizarem um número bem maior variáveis e exigirem um maior esforço computacional que a MIE.

As bases dessa metodologia foram elaboradas em um estudo realizado no grupo onde se buscava identificar a atividade carcinogênica de hidrocarbonetos aromáticos policíclicos (PAHs) através de índices eletrônicos e topológicos²². Descobriu-se que a análise da densidade local de estados sobre o anel de maior ordem de ligação associada à diferença de energia entre os orbitais HOMO e HOMO-1 estava diretamente ligada à atividade carcinogênica desses compostos. Com essa metodologia foi possível avaliar a resposta biológica dos PAHs com um acerto maior que 89%.

A densidade de estados eletrônicos (DOS) representa a distribuição de estados no espaço de energia do sistema⁶. Ela é dada pelo número de estados eletrônicos por unidade de energia na vizinhança de um valor particular de energia (2-5), onde ε_i é um autovalor de energia do sistema, e o somatório é sobre todos os estados ocupados.

$$DOS(E) = \sum_i \delta(E - \varepsilon_i) \quad (2-5)$$

Em um sistema finito, como é o caso do sistema molecular, os autovalores de energia ε_i são discretos, assim o espectro do número de níveis de energia versus a energia nos dá uma coleção de picos de larguras nulas (deltas de Dirac). Muitas vezes é útil obter a densidade de estados vinda de um átomo, de um grupo de átomos ou de um tipo particular de orbital; a esse tipo de densidade chamamos densidade local de estados (LDOS). Com as funções de onda moleculares escritas como a combinação linear de orbitais atômicos (LCAO) e dentro da aproximação ZDO (zero-differential overlap) para camada fechada, a expressão para a LDOS é dada pela 2-6, onde E_i é a energia do orbital molecular i e c_{mi} são

os coeficientes de expansão dos orbital molecular i em termos dos orbitais atômicos m . O somatório é sobre os orbitais ou átomos de interesse:

$$LDOS(E_i) = 2 \sum_{m=n_i}^{n_f} |c_{mi}|^2, \quad (2-6)$$

A distribuição de estados no espaço de energia da molécula pode carregar alguma informação importante a respeito da atividade biológica da molécula.

A EIM é baseada em dois maiores descritores, Δ e η . Δ está associado ao valor crítico de separação de energia entre orbitais subjacentes enquanto η está associado à contribuição relativa desses orbitais à densidade local de estados eletrônicos sobre uma região ou átomo da molécula previamente escolhido.

Uma das motivações do presente estudo foi verificar se a metodologia de índices eletrônicos, baseada nos parâmetros Δ e η , poderia ser estendida a compostos orgânicos com estruturas não planares, mais ricas conformacionalmente do que os hidrocarbonetos aromáticos.

Capítulo 3

ESTUDO ESTRUTURA ATIVIDADE DE COMPOSTOS DERIVADOS DO LAPACHOL

3.1 INTRODUÇÃO

O lapachol (2-hidroxi-3-(3-metil-2-butenil) - 1,4-naftoquinona) é uma molécula (Figura 3-1) extraída da casca de uma espécie de árvore (*Tabebuia avellanedae*) encontrada na América do Sul e na Ásia, conhecida por possuir um amplo leque de atividades biológicas. Dentre eles destacam-se sua ação anticancerígena, antiviral, antimalarial e anticoagulante³⁰⁻³⁴. Acredita-se que o mecanismo de ação biológica do lapachol seja a intercalação e a reação com o DNA³⁵.

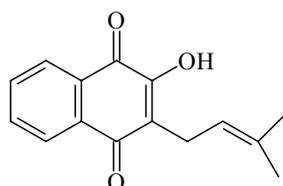
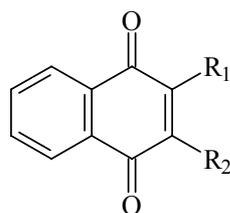


Figura 3-1 – Lapachol

Os primeiros indícios de que esse composto possuía propriedades anticancerígenas foram observados no início dos anos 60. Em 1974 o Instituto Nacional do Câncer dos EUA aprovou o uso do lapachol para testes em seres humanos. Descobriu-se então que o composto apresentava uma potente ação anticoagulante, o que dificultava sua administração oral em altas doses. Esse efeito anticoagulante do lapachol está relacionado ao fato dele interferir no metabolismo da vitamina K, que é a responsável por promover a coagulação e a prevenção de hemorragias. No mesmo ano de 1974 o lapachol foi considerado impróprio para uso em seres humanos devido a seus efeitos colaterais. Apesar disso, continuou-se o estudo da atividade biológica desse composto, principalmente com relação às suas propriedades anticancerígenas. As novas pesquisas foram direcionadas na busca de derivados do lapachol modificados estruturalmente que possuíssem menores efeitos colaterais e que preservassem a atividade antineoplásica.

3.2 DERIVADOS DO LAPACHOL COM ATIVIDADE ANTI-CARCINOGENICA

Alguns derivados do lapachol que apresentavam atividade anti-carcinogênica contra o carcinosarcoma de Walker 256 (Tabela 3-1) foram reunidos em um estudo de estrutura-atividade realizado por Subramaniam *et al*³⁶. Com esse mesmo conjunto de compostos realizamos um estudo de correlação estrutura-atividade, baseado na análise da composição dos orbitais moleculares em termos dos orbitais atômicos e na metodologia de índices eletrônicos (MIE).



Composto	R ₁	R ₂	Atividade (TWI, %)
1	OH	CH ₂ CH=C(Br) ₂	90
2	OH	CH ₂ CH=C(Cl) ₂	85
3	OH	CH ₂ CH=C(CH ₃) ₂ (lapachol)	79
4	OH	CH ₂ CH=C(CH ₃)CH ₂ CH ₂ CH=C(CH ₃) ₂	68
5	OAc	CH ₂ CH=C(CH ₃) ₂	64
6	OH	CH ₂ CH=C(CH ₂ OH)CH ₃	58
7	CH ₃	CH ₂ CH=C(CH ₃) ₂	49
8	OH	CH ₂ CH=C(CH ₂ CH ₃) ₂	42
9	H	CH ₂ CH=C(CH ₃) ₂	41
10	OH	CH ₂ CH=CICH ₃	41
11	OCH ₃	CH ₂ CH=C(CH ₃) ₂	39
12	OH	CH=CBrCH(CH ₃) ₂	34
13	OH	CH=C(CH ₃) ₂	25
14	OH	CH=CH CH ₂ CH ₃	21
15	OH	CH ₂ CH=CH ₂	20
16	OH	CH=CHCH ₂ CH ₂ CH ₂ CH ₂ CH ₃	20
17	OH	CH ₂ CH(OH)C(OH) (CH ₃) ₂	20
18	NH ₂	CH ₂ CH=C(CH ₃) ₂	19
19	OH	CH ₂ CH ₂ CH(CH ₃) ₂	17
20	OH	CH ₂ CH ₂ COH(CH ₃) ₂	10
21	OH	H (lawsona)	0
22	OH	CH ₃	0
23	OH	CH ₂ CH=C(COOH) CH ₃	0
24	OH	CH=CHCH(CH ₃) ₂	0
25	OH	C(CH ₃) ₂ CH=CH ₂	0
26	OH	CH=CHCH ₃	0
27	OH	CH ₂ CH ₂ CCl(CH ₃) ₂	0

Tabela 3-1 Derivados do Lapachol

A atividade biológica desses compostos foi avaliada em termos da inibição do peso tumoral (TWI), dada em porcentagem. Compostos com TWI abaixo de 40% podem ser considerados inativos.

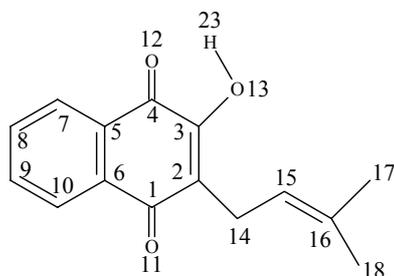
3.3 CÁLCULOS E RESULTADOS

As geometrias iniciais dos compostos, que devem ser fornecidas ao programa que realiza os cálculos semiempíricos, foram obtidas através de métodos de mecânica molecular (MM⁺), uma vez que há na literatura poucos dados experimentais a cerca da geometria desses compostos. Um desses dados se refere ao composto 03³⁷ (lapachol), que foi utilizado para verificarmos o grau de precisão dos parâmetros geométricos obtidos teoricamente.

As geometrias iniciais obtidas da mecânica molecular foram otimizadas através dos métodos semiempíricos AM1 e PM3 (capítulo 1) e no decorrer do trabalho analisaremos, para efeitos comparativos de validade estatística e de congruência, os resultados obtidos com os dois métodos.

As distâncias e os ângulos interatômicos do lapachol (composto 03) obtidos com o AM1 e com o PM3 foram comparados aos dados experimentais de difração de raio-X disponíveis (Tabela 3-2). Com relação aos comprimentos de ligação, o erro médio e o desvio padrão foram de 0.01 para ambos os métodos; em termos percentuais temos os valores 1.12 (0.88) para o erro médio e 1.14 (1.57) para o desvio padrão relativo aos valores teóricos obtidos com o AM1 (PM3). Para os ângulos de ligação (Tabela 3-3) o erro médio e o desvio padrão em termos percentuais foram de 0.91 (1.12) e 0.67 (0.85) para o AM1 (PM3). Esses números

nos mostram uma concordância excelente entre os dados experimentais e teóricos, o que torna confiáveis as geometrias obtidas para os outros compostos.



Ligação	Comprimento de Ligação (Å)				
	Raio-X	AM1	Δ AM1	PM3	Δ PM3
12-4	1.22	1.24	0.02	1.22	0.00
13-3	1.34	1.37	0.03	1.36	0.02
23-13	1.02	0.97	-0.05	0.95	-0.07
11-1	1.22	1.24	0.02	1.22	0.00
4-3	1.48	1.49	0.01	1.50	0.02
4-5	1.47	1.47	0.00	1.48	0.01
3-2	1.35	1.36	0.01	1.36	0.01
1-2	1.47	1.48	0.01	1.49	0.02
2-14	1.51	1.49	-0.02	1.50	-0.01
1-6	1.49	1.48	-0.01	1.49	0.00
10-9	1.38	1.40	0.02	1.39	0.01
10-6	1.38	1.40	0.02	1.39	0.01
9-8	1.38	1.39	0.01	1.39	0.01
8-7	1.40	1.40	0.00	1.39	-0.01
7-5	1.39	1.40	0.01	1.39	0.00
5-6	1.40	1.41	0.01	1.40	0.00
14-15	1.49	1.49	0.00	1.49	0.00
15-16	1.34	1.34	0.00	1.34	0.00
16-17	1.50	1.48	-0.02	1.49	-0.01
16-18	1.50	1.48	-0.02	1.49	-0.01
Média:			0.01	0.01	
Desvio Padrão:			0.01	0.01	
Média (%):			1.12	0.88	
Desvio Padrão(%):			1.14	1.57	

Tabela 3-2 Valores experimentais e teóricos para alguns comprimentos de ligação do lapachol. Δ : diferença entre os valores teóricos e experimentais.

Ângulo	Ângulo de Ligação (graus)				
	Raio-X	AM1	Δ AM1	PM3	Δ PM3
3-13-23	106.5	108.1	1.60	108.3	1.80
12-4-3	119.6	119.7	0.10	120.9	1.30
12-4-5	122.6	123.7	1.10	123.4	0.80
3-4-5	117.8	116.5	-1.30	115.8	-2.00
13-3-4	115.9	117.2	1.30	119.3	3.40
13-3-2	120.9	119.4	-1.50	118.9	-2.00
4-3-2	123.2	123.5	0.30	122.0	-1.20
3-2-1	119.4	120.5	1.10	119.2	-0.20
3-2-14	121.3	122.3	1.00	122.8	1.50
1-2-14	119.3	117.2	-2.10	118.0	-1.30
11-1-2	120.2	121.7	1.50	122.5	2.30
11-1-6	120.4	120.9	0.50	121.1	0.70
2-1-6	119.4	117.4	-2.00	116.4	-3.00
9-10-6	121.0	120.3	-0.70	120.1	-0.90
10-9-8	120.4	120.1	-0.30	120.2	-0.20
9-8-7	119.8	120.0	0.20	120.1	0.30
8-7-5	119.2	120.1	0.90	120.0	0.80
4-5-7	119.2	119.5	0.30	120.5	1.30
4-5-6	119.9	120.6	0.70	119.5	-0.40
7-5-6	120.8	120.0	-0.80	120.0	-0.80
1-6-10	121.0	119.2	-1.80	120.4	-0.60
1-6-5	120.2	121.4	1.20	119.9	-0.30
10-6-5	118.7	119.4	0.70	119.6	0.90
2-14-15	112.0	115.5	3.50	111.7	-0.30
14-15-16	127.6	125.1	-2.50	123.2	-4.40
15-16-17	121.7	121.2	-0.50	122.7	1.00
15-16-18	123.6	123.4	-0.20	121.2	-2.40
17-16-18	114.8	115.4	0.60	116.2	1.40
		Média:	1.08		1.34
		Desvio Padrão:	0.79		1.03
		Média (%):	0.91		1.12
		Desvio Padrão(%):	0.67		0.85

Tabela 3-3 Valores experimentais e teóricos para alguns ângulos de ligação do lapachol. Δ : diferença entre os valores teóricos e experimentais.

Apesar dos comprimentos e ângulos de ligação variarem pouco entre esses dois métodos, observamos que para alguns ângulos de torção a variação era significativa. Enquanto que com o AM1 a planaridade do sistema quinônico era preservada, com o PM3 o anel quinônico sofria uma distorção e os oxigênios carbonílicos tendiam a sair do plano do anel. Fizemos um estudo mais aprofundado desse aspecto e observamos que a quebra de planaridade não estava relacionada à interação dos átomos de oxigênio carbonílicos com átomos de hidrogênio do anel benzênico, uma vez que a introdução de nitrogênios com hibridização *sp* no anel benzênico não resultou em ganho de planaridade para o anel quinônico. O que constatamos foi que o PM3 conservava a planaridade do sistema somente quando o grupo substituinte nas posições 2 ou 3 do anel quinônico era do tipo hidroxila, como é o caso da lawsona (composto 21). Quando um grupo alquila estava presente na monossustituição a planaridade era quebrada. Como o PM3 é parametrizado para descrever melhor as ligações de hidrogênio, a perda de planaridade pode decorrer da tentativa do método de maximizar o ganho energético com esse tipo de interação.

As propriedades eletrônicas obtidas para as estruturas de mínima energia (dipolo, cargas, energias, etc.) foram utilizadas no estudo estrutura-atividade. Alguns dos parâmetros eletrônicos apresentados no capítulo 2, tais como o momento de dipolo, a energia do HOMO (E_{HOMO}), a energia do LUMO (E_{LUMO}) e a diferença de energia entre HOMO e LUMO ($E_{\text{HOMO-LUMO}}$) foram avaliados na tentativa de serem correlacionados à atividade biológica (Tabela 3-4). Esse conjunto de variáveis se mostrou incapaz de separar os compostos em classes ativa e inativa com relação à atividade contra o WM256. Por exemplo, no caso dos

compostos 01 e 17, a energia do orbital de fronteira E_{HOMO} é praticamente a mesma (≈ 10 eV), embora o primeiro composto possua uma atividade alta e o segundo seja inativo. O mesmo acontece com outras variáveis.

composto	Dipolo (Debye)		E_{HOMO} (eV)		E_{LUMO} (eV)		$E_{HOMO-LUMO}$ (eV)	
	PM3	AM1	PM3	AM1	PM3	AM1	PM3	AM1
1	3.28	4.68	-10.04	9.75	-1.6	-1.83	8.44	7.92
2	3.2	4.22	-9.39	-9.7	-1.59	-1.78	7.8	7.92
3	2.26	2.48	-9.35	-9.26	-1.46	-1.61	7.89	7.65
4	2.38	2.52	-9.32	-9.15	-1.46	-1.63	7.86	7.52
5	0.94	0.69	-9.62	-9.35	-1.4	-1.54	8.22	7.81
6	3.65	3.71	-9.58	-9.36	-1.48	-1.64	8.1	7.72
7	2.46	0.73	-9.69	-9.47	-1.24	-1.42	8.45	8.05
8	2.36	2.54	-9.52	-9.32	-1.46	-1.62	8.06	7.7
9	2.01	0.91	-9.7	-9.48	-1.36	-1.47	8.34	8.01
10	3.31	3.62	-9.37	-9.49	-1.53	-1.69	7.84	7.8
11	1.18	2.26	-9.56	-9.38	-1.33	-1.53	8.23	7.85
12	1.76	2.25	-9.92	-9.65	-1.51	-1.65	8.41	8
13	2.14	2.22	-9.13	-8.98	-1.48	-1.59	7.65	7.39
14	2.6	2.5	-9.71	-9.18	-1.48	-1.62	8.23	7.56
15	2.51	2.73	-9.88	-9.83	-1.49	-1.64	8.39	8.19
16	2.29	2.08	-9.07	-8.97	-1.51	-1.64	7.56	7.33
17	1.66	2.38	-10	9.92	-1.52	-1.67	8.48	8.25
18	1.57	1.42	-9.07	-9.02	-1.22	-1.31	7.85	7.71
19	2.4	2.55	-10.26	-9.75	-1.38	-1.61	8.88	8.14
20	3.04	3.58	-9.94	-9.88	-1.53	-1.69	8.41	8.19
21	2.68	3.04	-10.16	-10.19	-1.61	-1.67	8.55	8.51
22	2.42	2.58	-9.78	-9.77	-1.49	-1.62	8.29	8.14
23	3.35	4.58	-10.01	-10	-1.59	-1.78	8.42	8.22
24	2.63	2.6	-9.82	-9.75	-1.47	-1.61	8.35	8.14
25	2.63	2.64	-9.81	-9.76	-1.31	-1.61	8.5	8.15
26	2.6	2.49	-9.74	-9.17	-1.48	-1.62	8.26	7.55
27	2.51	3.72	-9.96	-9.92	-1.56	-1.71	8.4	8.21

Tabela 3-4 Sumário dos resultados obtidos com o AM1 e com o PM3. E_{HOMO} : energia do HOMO, E_{LUMO} : energia do LUMO, $E_{HOMO-LUMO}$: diferença de energia entre HOMO e LUMO.

3.3.1 Composição dos Orbitais Moleculares

O procedimento LCAO descrito no *capítulo 1* - seção 1.2.3 nos permite analisar a composição dos orbitais moleculares (OM) em função dos orbitais atômicos utilizando os coeficientes $C_{i\mu}$ (1-17). Os orbitais moleculares de fronteira, HOMO e LUMO, e os orbitais adjacentes, HOMO-1 e LUMO+1, são os mais indicados para esse tipo de estudo por estarem mais diretamente relacionados à reatividade molecular (*capítulo 2*, seção 1.3). Fizemos uma análise detalhada com respeito à composição dos orbitais moleculares dos derivados dessa série de naftoquinonas e a seguir exporemos os resultados obtidos utilizando os compostos 03 (ativo) e 21 (inativo) como exemplo.

A contribuição de cada orbital atômico para a formação dos orbitais moleculares desocupados LUMO e LUMO+1 pode ser vista na Figura 3-2. Para ambos compostos as maiores contribuições para formação do LUMO vêm dos orbitais π dos átomos C-4 e O-12, enquanto os orbitais π dos átomos C-7 e C-10 são os que mais contribuem para a formação do LUMO+1. Como a única diferença entre esses compostos é o grupo substituinte em R₂, e verificamos que esse grupo não afeta a composição desses orbitais; concluímos que reações que envolvam o LUMO não são as que devem ocorrer no mecanismo de ação biológica desses compostos.

A seguir (Figura 3-3) temos a composição dos OM ocupados HOMO e HOMO-1. Podemos verificar que tanto para o AM1 quanto para o PM3 o composto ativo tem o orbital molecular HOMO-1 formado essencialmente pelas contribuições dos orbitais $p_z(\pi)$ dos átomos C-2, C-3 e O-13 da molécula. Já para o composto inativo esse orbital é formado basicamente pelos orbitais p (ou p_x ou p_y ou p_z) dos

átomos C-6, C-7 e C-10 da molécula. As contribuições para formação do HOMO do composto ativo vêm essencialmente dos orbitais p dos átomos C-15 e C-16 da cadeia lateral enquanto que para os compostos inativos ela se origina dos orbitais p_z dos átomos C-2, C-3 e O-13.

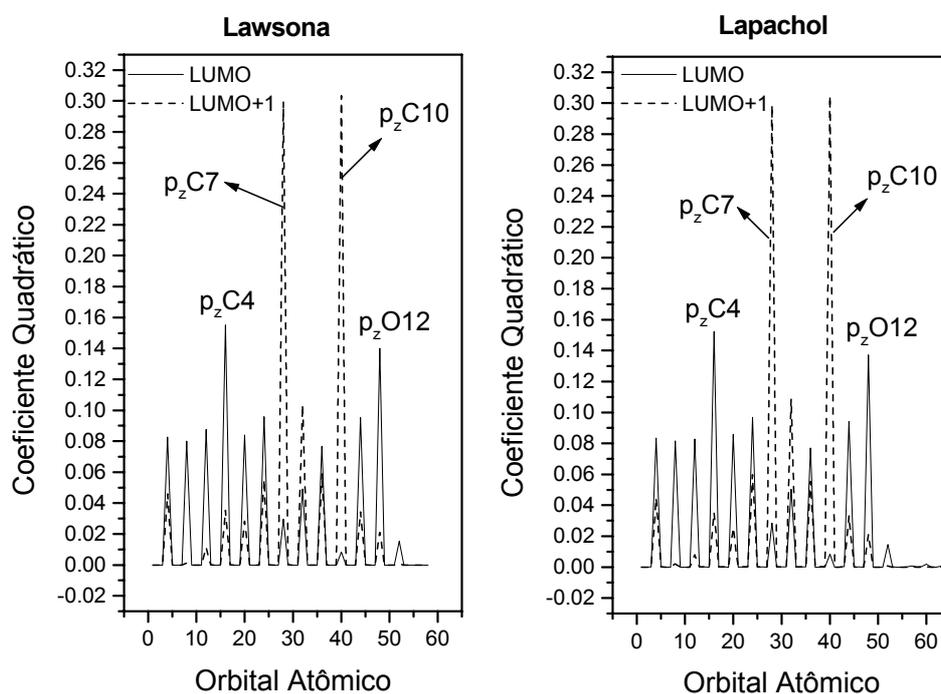


Figura 3-2 Contribuição dos orbitais atômicos para a formação dos orbitais moleculares LUMO e LUMO+1.

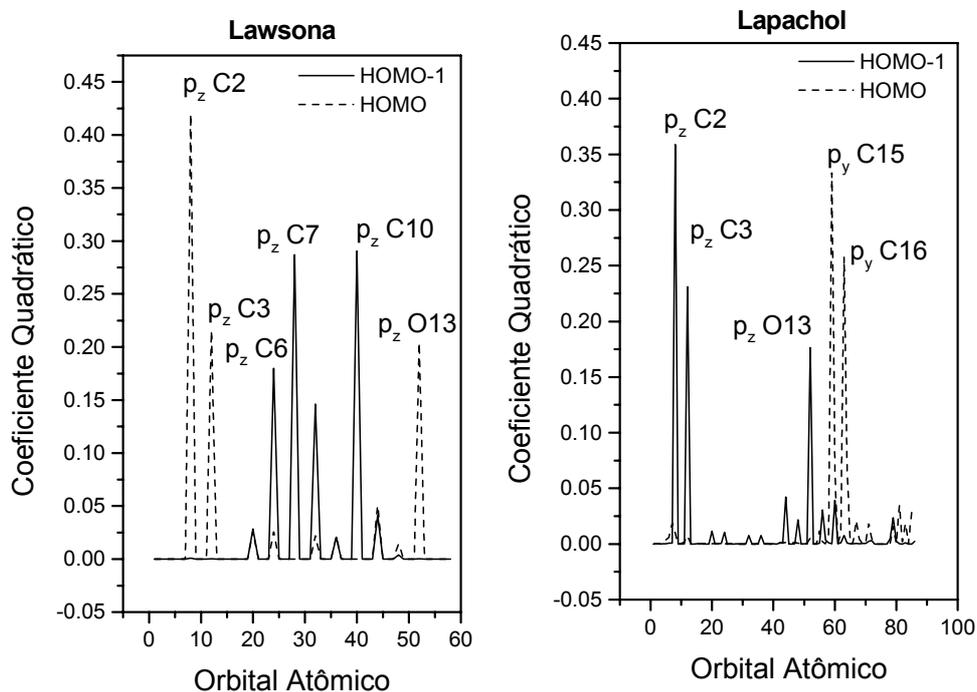


Figura 3-3 Contribuição dos orbitais atômicos para a formação dos orbitais moleculares HOMO-1 e HOMO.

Nos chamou a atenção a inversão das contribuições dos orbitais atômicos dos átomos nas posições 2, 3 e 13 aos orbitais moleculares HOMO e HOMO-1. Ligado ao carbono da posição 2 da lawsona temos somente o hidrogênio, enquanto que no lapachol temos um grupo alquil. Essa alteração faz com que os átomos que contribuíam para a formação do HOMO para o composto inativo passem a contribuir para a formação do HOMO-1 para o composto ativo.

Comparemos agora dois compostos cuja diferença está no grupo que ocupa a posição R₂ (Figura 3-4). No caso do composto 03 (ativo) temos a presença de uma ligação insaturada no grupo substituinte em R₂, enquanto para o

composto 19 (inativo) todas as ligações da cadeia lateral são saturadas. Observamos que a formação dos orbitais moleculares HOMO-1 e HOMO do composto 19 é exatamente a mesma da lawsona, o que indica que, na prática, todo o grupo substituinte presente em R₂ comporta-se como se naquela posição houvesse apenas um átomo de hidrogênio.

Método AM1

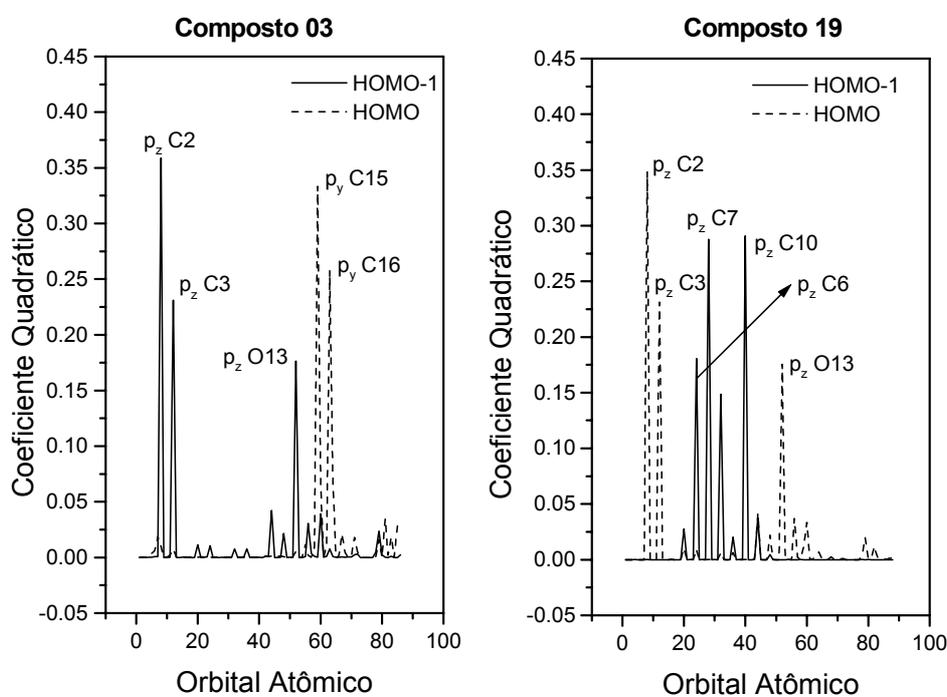


Figura 3-4 Contribuição dos orbitais atômicos para a formação dos orbitais moleculares HOMO-1 e HOMO.

Comparando o composto 03 ao 18 podemos determinar a influência de R_1 na formação dos orbitais moleculares (Figura 3-5). Nessa posição tem-se um grupo hidroxila para o caso do composto 03 e um grupo nitro para o composto 18.

Método AM1

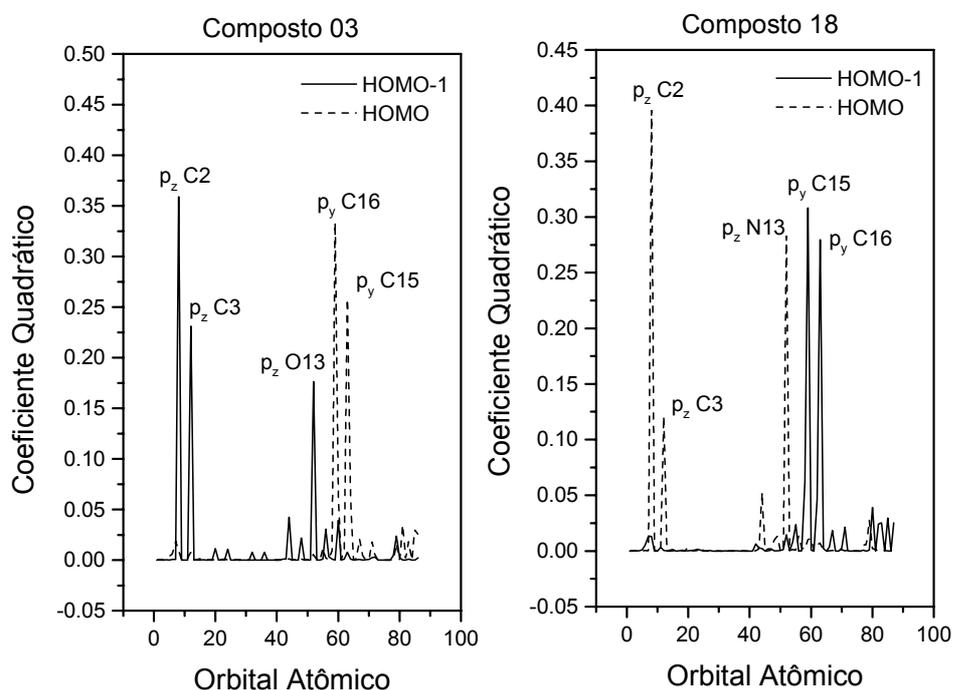


Figura 3-5 Contribuição dos orbitais atômicos para a formação dos orbitais moleculares HOMO-1 e HOMO.

Ambos os grupos substituintes possuem um efeito indutivo negativo e um efeito ressonante positivo, ou seja, eles atraem fortemente os elétrons e aumentam a densidade eletrônica no sistema conjugado. Mas, apesar de terem

propriedades similares, a presença do grupo nitro é suficiente para mudar o comportamento dos orbitais HOMO-1 e HOMO, fazendo com que sua composição se assemelhe àquela dos compostos 19 e 21. Observamos que no presente caso há uma inversão completa das contribuições atômicas aos orbitais moleculares, ou seja, os átomos 2, 3 e 13 participam com as maiores contribuições para a formação do HOMO-1 do composto ativo (03) e do HOMO do composto inativo (18), enquanto que os átomos 15 e 16 são os maiores responsáveis pela formação do HOMO do composto ativo e do HOMO-1 do composto inativo.

Essa análise dos orbitais moleculares nos mostrou a existência de um efeito competitivo entre os substituintes nas posições C-2 e C-3, onde a presença ou não de ligação saturada em R₂ e os aspectos eletrônicos dos grupos em R₁ são fundamentais. Comparando os composto 03 e 18 podemos claramente verificar essa “competição eletrônica”. Apesar dos dados de atividade nos indicarem que a presença da dupla ligação na cadeia lateral seja essencial para a atividade contra o WM256, só a presença dela não é suficiente, como exemplificado no caso composto 18. Para entendermos melhor o que está acontecendo é necessário recorrermos às regras das reações de substituição eletrofilicas. Sabemos que existem grupos que são fortes ativadores desse tipo de reação³⁸, como o NH₂, -NHR, -NR₂ e o OH; como também existem grupos que possuem uma capacidade de ativação moderada (NHCOCH₃, -NHCOR, OCH₃, OR) e fraca (CH₃, -C₂H₅ e C₆H₅). Pelo exposto anteriormente podemos avaliar que os grupos que possuem uma capacidade de ativação maior do que a do OH (NH₂, -NHR e -NR₂) conseguem deslocar a contribuição do HOMO dos átomos da cadeia lateral para os átomos do anel quinônico, enquanto que a contribuição ao HOMO-1 do anel

quinônico é deslocada para a cadeia lateral, tornando o composto inativo. Já os grupos com poder de ativação moderada como o OAc (composto 5) e o OCH₃ (composto 11), podem deslocar a contribuição do HOMO-1 para os átomos do anel benzênico, porém os orbitais que formam o HOMO continuam sendo aqueles dos carbonos da cadeia lateral que participam da dupla ligação, por isso a atividade é mantida. Por essa hipótese podemos sugerir que compostos derivados do lapachol com substituintes em R₂ do tipo NHCOCH₃, -NHCOR ou OR podem apresentar atividade significativa contra o WM256. Já os grupos com atividade fraca na ativação de reações eletrofílicas fazem com que a composição dos orbitais HOMO-1 e HOMO sejam do tipo característico dos compostos inativos, como é o caso do composto 22.

3.3.2 A Metodologia de Índices Eletrônicos (MIE)

A análise da composição dos orbitais de fronteira é um aspecto relevante para se entender a atividade desses compostos, contudo essa análise se torna mais trabalhosa quando os grupos adicionados em R₁ e R₂ se tornam mais complexos do ponto de vista estrutural, podendo as contribuições dos orbitais atômicos para a formação dos orbitais moleculares serem mais difusas sobre a molécula. A metodologia de índices eletrônicos (MIE – *capítulo 2*, seção 2.4.6) nos fornece uma análise mais precisa da composição dos orbitais moleculares em termos espaciais através dos parâmetros Δ e η (*capítulo 2*, seção 2.4.6.1). Como é usual, inicialmente analisamos a densidade local de estados (LDOS – *capítulo 2*-seção 2.4.6) nas regiões comuns a todos os compostos, que no presente caso são

os anéis benzênicos e quinônicos. Desta análise temos a distribuição espacial da contribuição dessas regiões para os níveis de energia da molécula.

Exemplificaremos o uso da metodologia de índices eletrônicos novamente com os compostos 03 (lapachol) e 21 (lawsona). A densidade de estados calculada sobre o anel quinônico desses compostos nos forneceu o padrão espacial apresentado na Figura 3-6. Por conveniência apresentamos somente os valores entre -14 e 4 eV.

Observamos que a distribuição espacial dos orbitais desocupados é muito semelhante para ambos os compostos, o que significa que esses estados não são adequados na descrição dos parâmetros Δ e η da MIE. Já com relação aos níveis ocupados, vemos uma clara inversão nas contribuições relativas dos orbitais HOMO e HOMO-1, o que sugere que os parâmetros Δ e η associados a esses níveis de energia possam estar correlacionados à atividade biológica.

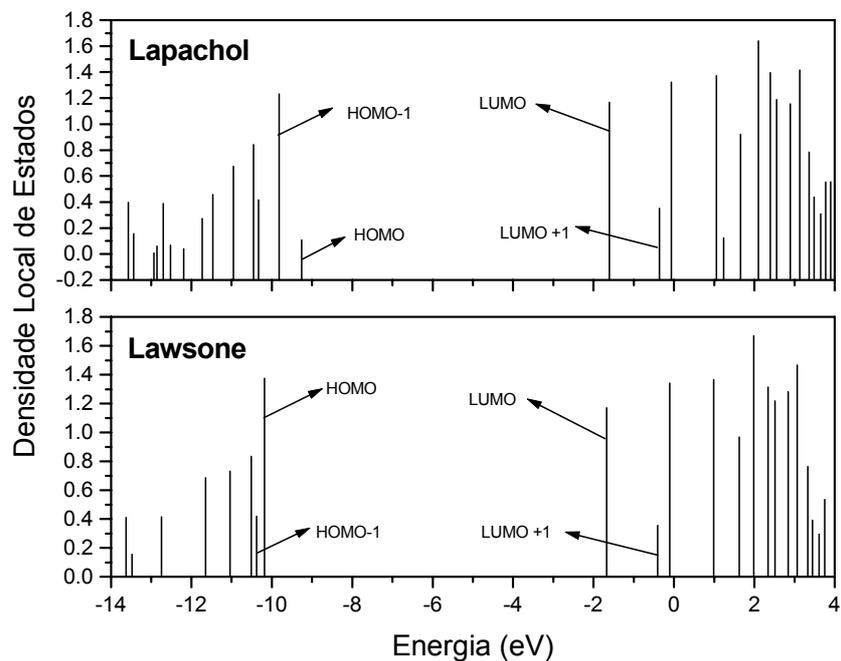


Figura 3-6 Densidade local de estados calculada sobre o anel quinônico.

A densidade local de estados sobre o anel quinônico foi então calculada para cada composto, de onde obtivemos gráficos semelhantes ao da Figura 3-6. Analisamos os dois parâmetros: Δ (associado à diferença de energia entre HOMO e HOMO-1) e η (associado à diferença de intensidade entre esses orbitais). Os valores calculados para esses estão indicados na Tabela 3-5.

Para o presente conjunto de compostos Δ não se mostrou um parâmetro adequado mesmo quando conjugado a η , o que pode ser justificado pelo fato de compostos muito similares apresentarem atividades muito diferentes, como é o caso dos compostos 3 e 19.

composto	Δ (eV)		η	
	<i>PM3</i>	<i>AM1</i>	<i>PM3</i>	<i>AM1</i>
1	0.08	0.40	-1.09	-1.22
2	0.64	0.35	-1.28	-1.24
3	0.53	0.56	-1.32	-1.12
4	0.41	0.15	0.00	-0.03
5	0.63	0.88	-0.31	-0.36
6	0.29	0.50	-1.26	-1.09
7	0.53	0.70	-1.14	-0.34
8	0.31	0.49	-1.26	-1.19
9	0.53	0.72	-0.29	-0.31
10	0.56	0.42	-1.26	-1.23
11	0.70	0.67	-0.33	-1.09
12	0.12	0.39	0.65	-0.74
13	0.69	1.34	0.43	0.18
14	0.35	1.15	0.12	0.27
15	0.27	0.19	1.25	1.15
16	1.27	1.35	0.32	0.29
17	0.35	0.45	0.82	0.59
18	0.63	0.49	0.83	0.97
19	0.50	0.58	0.87	0.78
20	0.45	0.50	0.88	0.82
21	0.21	0.19	0.91	0.95
22	0.56	0.58	0.88	0.84
23	0.13	0.06	1.28	0.28
24	0.22	0.19	1.14	0.62
25	0.41	0.17	1.20	0.50
26	0.28	1.16	0.09	0.28
27	0.41	0.49	1.26	0.82

Tabela 3-5 Parâmetros Δ e η da metodologia de índices eletrônicos obtidos com os métodos AM1 e PM3.

Já o parâmetro η se mostrou um bom descritor da atividade biológica através da relação: Valores de η negativos caracterizam compostos ativos

enquanto que valores de η positivos caracterizam compostos inativos, como pode ser observado pela Figura 3-7.

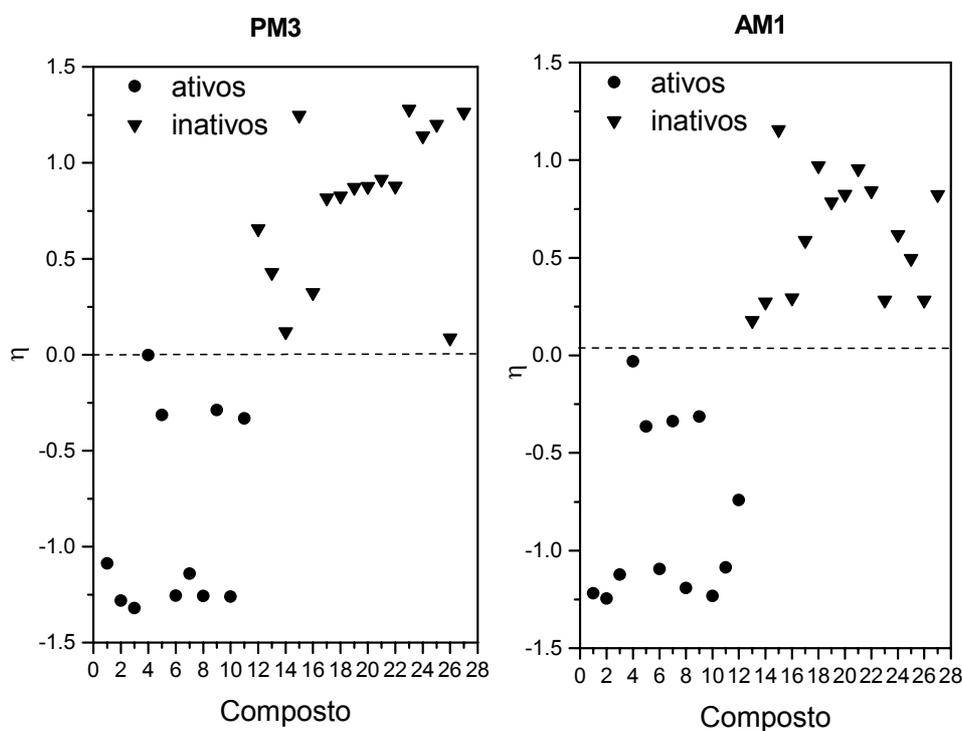


Figura 3-7 Representação gráfica dos parâmetro η onde podemos observar a separação entre compostos ativos e inativos.

É importante observar que o parâmetro η foi um bom descritor da atividade biológica independentemente do método semiempírico (AM1 ou PM3) empregado no cálculo, a despeito das discordâncias entre alguns valores de ângulos diedrais entre esses métodos. Observamos apenas uma discrepância com relação ao composto 12, que pelo PM3 é classificado como inativo e pelo AM1 como ativo (Tabela 3-6). Como as medidas experimentais são passíveis de erros e o número

de compostos com os quais trabalhamos foi relativamente pequeno, não poderíamos, somente com esses dados, estabelecer a superioridade de um método com relação ao outro na previsão da atividade biológica dessa série de compostos.

composto	Atividade Experimental	Atividade prevista por η	
		AM1	PM3
1	A	A	A
2	A	A	A
3	A	A	A
4	A	A	A
5	A	A	A
6	A	A	A
7	A	A	A
8	A	A	A
9	A	A	A
10	A	A	A
11	A	A	A
12	I	A	I
13	I	I	I
14	I	I	I
15	I	I	I
16	I	I	I
17	I	I	I
18	I	I	I
19	I	I	I
20	I	I	I
21	I	I	I
22	I	I	I
23	I	I	I
24	I	I	I
25	I	I	I
26	I	I	I
27	I	I	I

Tabela 3-6 Atividade experimental e atividade prevista pela MIE. A: compostos ativos, I: compostos inativos.

Com o intuito de comparar as correlações estrutura-atividade obtidas por Subramanian e colaboradores as correlações por nós obtidas através da MIE, analisamos algumas estruturas teóricas que ainda não tiveram suas atividades avaliadas (Tabela 3-7). Na Figura 3-8 podemos ver os resultados obtidos pela MIE para o parâmetro η . Todas as previsões foram coincidentes excetuando uma (composto 36), o que mostra a consistência dessa metodologia de análise estrutura-atividade.

Composto	R ₁	R ₂	Previsão Subramanian et al ³⁶	Previsão EIM	
				PM3	AM1
28	OH	CH ₂ C (CH ₃)=CH ₂	A	A	A
29	OH	CH ₂ C (CH ₃)=C(CH ₃) ₂	A	A	A
30	NH ₂	CH ₂ CH=C(Br) ₂	I	I	I
31	SH	CH ₂ CH=C(CH ₃) ₂	I	I	I
32	SH	CH ₂ CH=C(Br) ₂	I	I	I
33	SH	CH ₂ CH=C(Cl) ₂	I	I	I
34	Cl	CH ₂ CH=C(CH ₃) ₂	A	A	A
35	Cl	CH ₂ CH=C(Br) ₂	I	I	I
36	Cl	CH ₂ CH=C(Cl) ₂	A	I	A
37	Cl	CH ₂ CH=C(CH ₃)CH ₂ CH ₂ CH=C(CH ₃) ₂	A	A	A

Tabela 3-7 Comparação da previsão da atividade biológica contra o WM256 obtida pela EIM e pelo estudo de Subramanian et al.

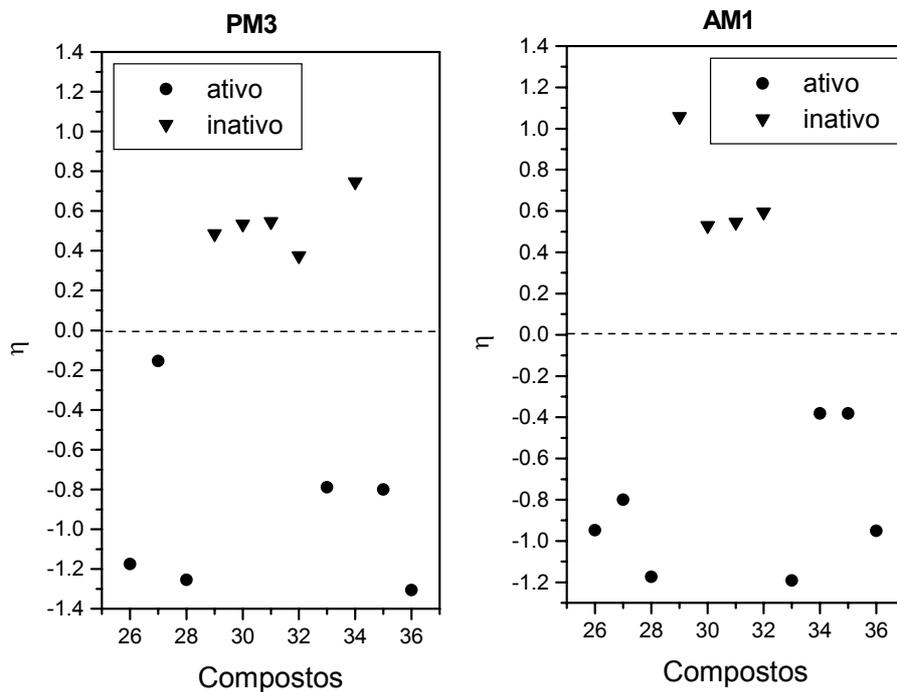


Figura 3-8 Representação gráfica do parâmetro η para os compostos hipotéticos da tabela 3-7.

3.4 CONCLUSÕES

Para o conjunto de moléculas derivadas do lapachol, duas metodologias se mostraram apropriadas na correlação da estrutura com a atividade anti-carcinogênica: A análise da composição dos orbitais moleculares HOMO e HOMO-1 e a metodologia dos índices eletrônicos. Embora essas duas formas de abordar o problema estejam interligadas, a MIE é um método mais eficiente por correlacionar a atividade biológica da molécula com uma região (anel quinônico)

que não pertence aos grupos substituintes. Isso nos permite analisar estruturas onde esses grupos sejam pouco similares.

A análise que fizemos a respeito da eletrofilicidade dos grupos substituintes e o fato dos orbitais desocupados não estarem envolvidos nos parâmetros associados à atividade biológica (tanto na análise da composição dos orbitais moleculares quanto na MIE) indicam que o mecanismo de ação dos derivados do lapachol contra o carcinossarcoma de Walker não envolve a redução da quinona (o que é mais comum), e que eles atuam preferencialmente como agentes oxidantes nos processos de reação envolvidos.

Capítulo 4

ESTUDO ESTRUTURA-ATIVIDADE DE COMPOSTOS ANÁLOGOS À TALIDOMIDA

4.1 INTRODUÇÃO

O composto N-fitalamida-glutaramídeo, mais conhecido como talidomida, é uma droga com propriedades imunomoduladoras que ficou conhecida em todo mundo no início dos anos 60³⁹ como responsável pelo o nascimento de milhares de crianças com deformações congênitas.

Desde sua descoberta, em 1953, até os dias de hoje, vários trabalhos apareceram na literatura mencionando sua potencialidade no tratamento de inúmeros distúrbios fisiológicos, o que fez da talidomida uma droga com características singulares do ponto de vista farmacológico.

Introduzida no mercado mundial em 1956 como sedativo e tranqüilizante, logo levantou suspeitas sobre sua atividade teratogênica devido ao aumento do

número de nascimento de crianças com defeitos físicos nos países onde era comercializada. No início dos anos 60 foi comprovada sua teratogenicidade e sua comercialização foi proibida. Apesar dessa trágica história, a utilização desse composto nunca foi totalmente abandonada e, em 1965, uma descoberta acidental mostrou que a talidomida era eficaz no tratamento de pacientes com hanseníase que apresentavam um quadro epidérmico inflamatório conhecido como eritema nodoso leprosal (ENL). A partir de então ressurgiu o interesse de laboratórios em pesquisar os mecanismos de ação da talidomida contra o ENL e em desenvolver compostos análogos que não apresentassem propriedades teratogênicas ou que tivessem essa propriedade atenuada⁴⁰⁻⁴². Hoje em dia já se conhece os efeitos terapêuticos dessa droga em vários tipos de doenças tais como artrite reumatóide, fotodermatites, tuberculose, síndrome da imunodeficiência adquirida (AIDS), hanseníase, esclerose múltipla e, mais recentemente, mostrou-se sua propriedade como potente inibidor de metástase tumoral⁴³. Apesar desse vasto campo de aplicações farmacológicas, os mecanismos pelos quais a talidomida atua no organismo humano continuam desconhecidos. Recentemente mostrou-se que a talidomida tem a propriedade de inibir a produção do chamado fator de necrose tumoral- α ⁴⁴ (TNF- α). O TNF- α é uma citocina produzida por monócitos ativados, responsável pela resposta do sistema imunológico a agentes inflamatórios⁴⁵; níveis altos de TNF- α são altamente prejudiciais ao organismo e estão associados a diversas doenças de caráter autoimune e inflamatório⁴⁰. Portanto, pelo seu leque de aplicações, é de se esperar o interesse de diversos laboratórios farmacêuticos em descobrir

análogos à talidomida que sejam mais potentes e que não possuam seus indesejáveis efeitos colaterais, tais como a sedação, a indução de neuropatias e a teratogenicidade.

A talidomida é formada por um anel fitaloil unido a um anel glutaramídeo, e possui um carbono quiral na posição α amino do anel glutaramídeo que possibilita a existência das formas opticamente ativas S e R⁴² (Figura 4-1).

Estudos de atividade feitos com os enantiômeros R e S e com suas misturas racêmicas mostraram que não há nenhuma diferença significativa entre as formas opticamente ativa e opticamente inativa em termos da atividade inibitória do TNF- α . Na realidade sabe-se que a interconversão de um isômero a outro é efetuada rapidamente⁴⁶⁻⁴⁸, tanto em plasma (*in vitro*) quanto *in vivo*, o que dificulta um estudo de atividade biológica restrito apenas a um dos isômeros e faz com que a maioria dos dados de atividade seja relativo às misturas racêmicas da talidomida ou de seus análogos nas duas formas enantioméricas.

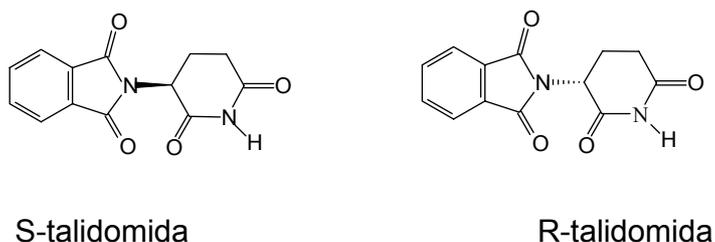
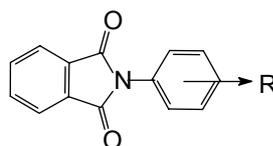


Figura 4-1– Estruturas S e R da talidomida

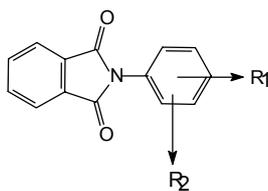
4.2 ANÁLOGOS DA TALIDOMIDA

Em artigo recente⁴⁹, uma série de fitalimidias N-substituídas (Tabelas 4-1 a 4-5) foram preparadas e tiveram suas atividades medidas com relação à produção e a inibição do fator de necrose tumoral- α . Essa série foi construída partindo-se de um composto padrão, o N-fenilfitalimida, e realizando-se modificações do tipo substituição orto no grupo N-fenil (Tabela 4-1), substituição dialquil no grupo N-fenil (Tabela 4-2), variação do anel succinimídeo (Tabela 4-3), substituição no anel benzênico condensado (Tabela 4-4) e halogenação do anel benzênico condensado (Tabela 4-5).



Composto	R	Quantidade de TNF- α (%)
01	-	100
02	2-Me	100
04	2-iPr	158
06	2-MeO	95
07	2-NO ₂	89
08	3-iPr	158
09	4-iPr	139

Tabela 4-1 – Análogos a fenilfitalimida com substituição ortofenil.

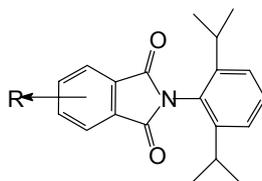


Composto	R ₁	R ₂	Quantidade de TNF- α (%)
01			100
11	2-Me	6-Me	161
12	3-Me	5-iPr	136
14	2-Et	6-Et	453
15	2-iPr	6-Me	234
17	2-iPr	2-iPr	602

Tabela 4-2 Análogos a fenilfitalimida com substituição dialquil.

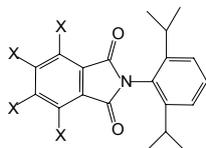
Composto	R ₁	Quantidade de TNF- α (%)
18		1252
19		214
21		134

Tabela 4-3 Análogos a fenilfitalimida com anel succinimídeo modificado.



Composto	R	Quantidade de TNF- α (%)
23	4-NO ₂	792
25	4-NH ₂	248
27	4-OH	239

Tabela 4-4 Análogos a fenilfitalimida com substituição no anel benzênico.



Composto	X	Quantidade de TNF- α (%)
30	F	3750
32	Br	235

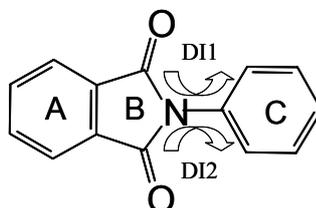
Tabela 4-5 Análogos a fenilfitalimida com halogenação do anel benzênico.

Esses compostos têm uma ação bidirecional com relação à modulação do TNF- α , ora aumentando ora diminuindo sua produção, dependendo de quais células estimuladas e quais os indutores utilizados. Trataremos aqui apenas do

sistema TPA/HL-60, que refere-se as células HL-60 (células de leucemia humana) estimuladas com 12-O-tetradecanoil-forbol 13 acetato (TPA). A atividade biológica do composto nesse sistema é representado pela porcentagem de TNF- α produzida pelas células HL-60 na presença do TPA em comparação a um composto tomado como padrão⁴⁹.

4.3 CÁLCULOS E RESULTADOS

A flexibilidade do ângulo de torção formado entre anéis condensados (benzeno (A) e succinimídeo(B)) e o anel fenil (C), denominado DI1(DI2) (Figura 4-2), exigiu uma busca conformacional cuidadosa a fim de encontrarmos as estruturas de mínima energia.



**Figura 4-2 – N-fenilfitalimida . DI1
(DI2) –ângulo de torção entre os
anéis B e C.**

Essa busca conformacional foi realizada variando-se o ângulo de torção DI1(DI2) de 0 a 360 graus, com passos de 10, e fazendo-se a otimização total de geometria para cada conformação. Esses cálculos foram realizados tanto com o

método AM1 quanto com o método PM3 para que pudéssemos verificar o grau de concordância entre eles. Resultados típicos para compostos ativos e inativos expressos através das barreiras de rotação são apresentados nas Figuras 4-3, A e B, respectivamente.

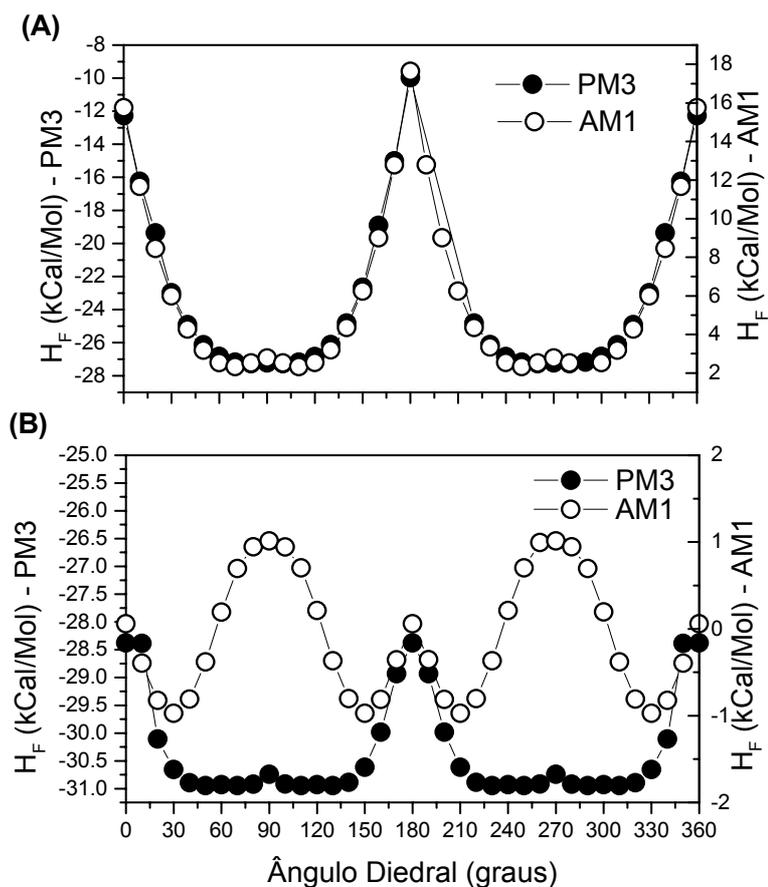


Figura 4-3 Barreiras rotacionais típicas obtidas com os métodos AM1 e PM3 para compostos ativos (A) e inativos (B).

Observem que há uma concordância relativa entre esses métodos, especialmente em se tratando de compostos ativos (Figura 4-3A). O padrão de

comportamento da barreira de rotação para essa classe de compostos é caracterizada pela presença de um pico central em torno de 180° separando duas regiões de mínimo (60-120, 240-300). Para os compostos inativos têm-se a presença de um pico adicional em torno de 90° , que é especialmente ressaltado com o método AM1 (Figura 4-3B). Essa diferença de comportamento pode estar relacionada às interações que ocorrem entre o átomo de hidrogênio em C-2 e o oxigênio succinimídeo, que é negativamente carregado. Como interações com essas características são altamente direcionais, a ligação se torna mais fraca quando o ângulo de torção se aproxima de 90° , o que provoca a elevação da barreira de rotação.

Comparando os compostos 11 e 12 podemos associar essas diferenças de comportamento em termos das interações envolvendo os átomos de hidrogênio. Os resultados obtidos sugerem que para os análogos às N-fenilfitalimidas a presença do átomo de hidrogênio nas posições C-2 e C-6 pode causar a perda de atividade, um vez que, em todos os compostos onde isto ocorre a atividade medida é fraca ou inexistente.

Uma vez obtidas as estruturas de menor calor de formação iniciamos o processo de investigação dos parâmetros eletrônicos que poderiam estar relacionados à atividade biológica dos compostos. Na Tabela 4-6 A e B apresentamos um sumário dos resultados obtidos com os métodos AM1 e PM3, respectivamente. Parâmetros tais como as energias dos orbitais de fronteira, a diferença de energia entre esses orbitais e o momento dipolar foram analisados e não se mostraram correlacionados diretamente com a atividade. Observamos, por

(A)

Molécula	H _F (kCal/Mol)	E _H (eV)	E _{H-L} (eV)	Dip (Debye)	DI1 (degrees)	DI2 (degrees)
01	-12.14	9.39	8.24	2.55	48.17	67.08
02	-19.93	9.52	8.40	2.46	98.37	73.05
04	-28.39	9.60	8.49	2.42	102.68	77.33
06	-58.82	9.74	8.69	1.06	69.19	90.11
07	-29.39	9.67	8.53	2.54	57.76	81.18
08	-30.94	9.32	8.18	2.51	47.90	66.34
09	-31.67	9.25	8.11	2.30	48.75	68.12
11	-27.27	9.43	8.31	2.60	79.88	99.86
12	-30.94	9.29	8.16	2.22	68.53	49.01
14	-35.04	9.45	8.32	2.52	76.84	96.32
15	-35.60	9.47	8.33	2.56	79.81	100.30
17	-43.82	9.49	8.38	2.45	78.59	96.42
18	26.47	9.05	7.16	2.43	90.41	89.55
19	-14.20	9.42	8.98	3.45	67.57	106.03
21	5.09	9.33	8.50	1.47	74.66	112.29
23	-61.34	9.49	8.27	3.39	96.39	78.61
25	-46.67	9.17	8.14	3.44	97.13	77.72
27	-89.13	9.48	8.37	3.14	78.24	96.69
30	-207.61	9.73	7.70	1.14	96.44	78.45
32	-2.15	9.61	7.99	1.30	99.44	80.54

(B)

Molécula	H _F (kCal/Mol)	E _H (eV)	E _{H-L} (eV)	Dip (Debye)	DI1 (graus)	DI2 (graus)
01	13.95	-9.05	7.89	3.21	29.27	29.26
02	8.81	-9.22	8.13	2.96	63.93	55.78
04	0.60	-9.28	8.21	2.87	65.53	60.16
06	-35.98	-9.43	8.41	1.63	61.24	59.96
07	-10.39	-9.40	8.22	4.03	67.97	76.06
08	-3.24	-8.99	7.85	1.63	29.69	29.63
09	-6.38	-8.86	7.72	2.91	29.17	29.16
11	2.34	-9.29	8.21	3.01	69.64	69.63
12	-0.97	-8.94	7.81	2.87	29.78	29.72
14	-7.66	-9.29	8.23	2.92	74.58	76.82
15	-5.82	-9.28	8.22	2.92	71.21	72.59
17	-13.88	-9.32	8.27	2.84	90.0	76.25
18	43.05	-8.87	7.37	2.33	97.90	97.64
19	12.86	-9.31	9.00	3.60	79.82	104.21
21	37.78	-9.39	8.67	1.53	89.36	90.89
23	-38.86	-9.37	8.13	2.60	75.45	75.81
25	-16.45	-9.12	8.25	4.45	75.17	74.9
27	-58.00	-9.32	8.24	3.32	104.66	104.69
30	-178.51	-9.56	7.59	0.62	75.93	75.91
32	18.11	-9.48	7.70	0.62	60	63.05

Tabela 4-6 Sumário dos cálculos com AM1 (A) e PM3 (B). H_F: calor de formação, E_H: energia do HOMO, E_{H-L}: diferença de energia HOMO-LUMO, Dip: dipolo e ângulos de torção (DI1 e DI2).

exemplo, que os compostos 1 e 19 possuem aproximadamente o mesmo valor de energia para o HOMO (9.39 e 9.42, respectivamente), embora possuam atividades opostas. O mesmo foi observado para outras magnitudes.

Prosseguimos então o estudo SAR através da abordagem da MIE. Realizamos os cálculos de densidade local de estados em algumas regiões e átomos dos compostos e observamos que a informação a respeito da atividade biológica estava fortemente localizada sobre o átomo de nitrogênio do anel succinimídeo.

Da análise da densidade local de estados sobre o nitrogênio (padrões típicos para os compostos inativos e ativos são apresentados nas Figura 4-4 A e B) observamos que, em geral, os compostos inativos têm uma alta contribuição ao HOMO, enquanto que para os compostos ativos isto não acontece.

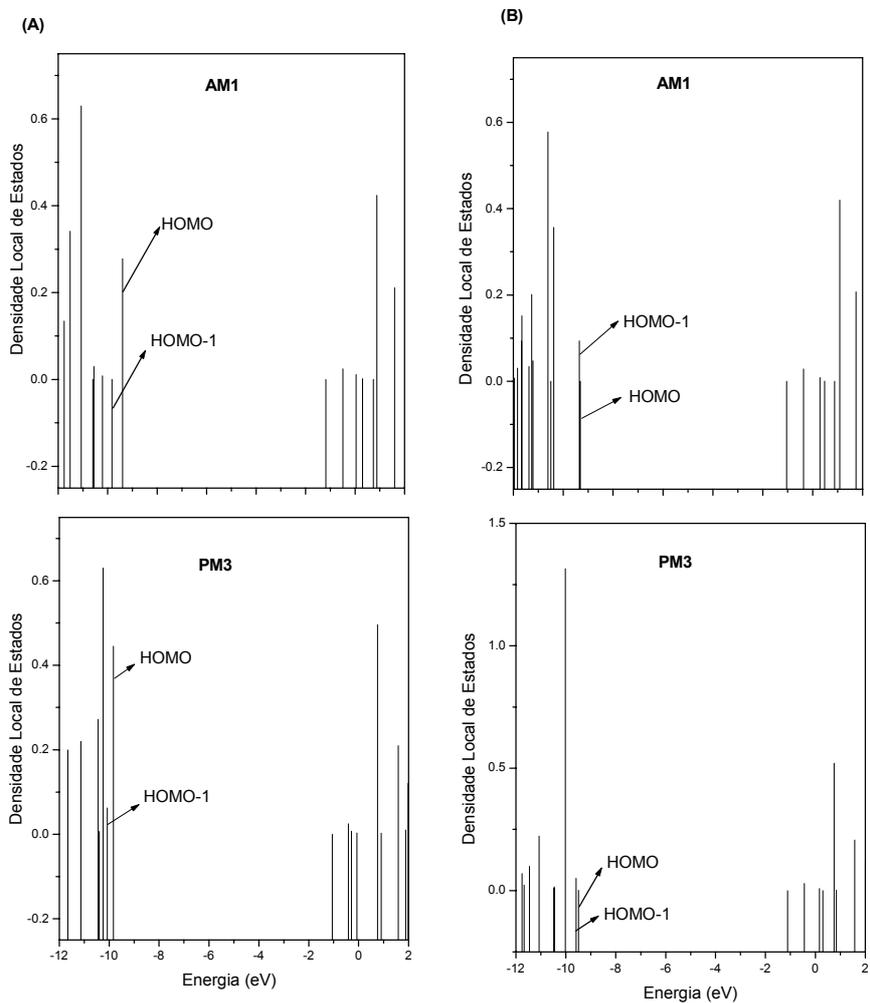


Figura 4-4 Densidade local de estados calculada sobre o átomo de nitrogênio (A) para um composto inativo e (B) para um composto ativo.

Uma das vantagens de se usar se usar a representação de densidade de estados é o fato dela proporcionar um modo rápido e prático de determinar a distribuição espacial dos orbitais moleculares LDOS. Este procedimento produz o mesmo nível de informação obtido com representações gráficas 3-D dos orbitais, como a exemplificada na Figura 4-5 para o composto ativo da Figura 4-4 B.

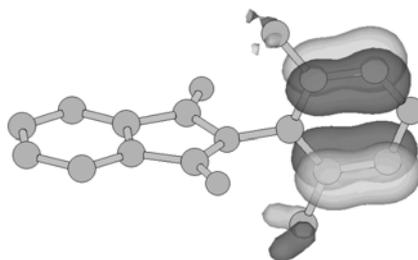


Figura 4-5 Representação 3-D do orbital atômico HOMO para um composto ativo.

Prosseguindo com a análise MIE, calculamos os valores para os índices Δ e η associados ao nitrogênio succinimídeo e aos orbitais HOMO e HOMO-1, que estão apresentados na Tabela 4-8.

molécula	AM1		PM3	
	Δ (eV)	η	Δ (eV)	η
01	0.65	0.33	0.47	0.49
02	0.33	0.11	0.27	-0.02
04	0.28	0.15	0.17	0.00
06	0.38	0.24	0.21	0.33
07	0.43	0.29	0.25	0.38
08	0.49	0.28	0.35	0.40
09	0.81	0.28	0.58	0.39
11	0.01	0.13	0.14	0.03
12	0.29	0.29	0.12	0.39
14	0.07	-0.03	0.16	-0.05
15	0.06	-0.03	0.17	-0.02
17	0.03	-0.09	0.10	-0.05
18	0.52	0.07	0.39	-0.61
19	0.05	-0.07	0.08	0.89
21	0.04	-0.02	0.28	0.60
23	0.04	-0.09	0.11	-0.05
25	0.13	0.05	0.28	0.03
27	0.03	-0.10	0.10	-0.05
30	0.10	-0.07	0.15	-0.03
32	0.09	-0.07	0.14	-0.02

Tabela 4-7 Índices Δ e η da MIE calculados para o nitrogênio succinimídeo.

O poder discriminatório desses parâmetros pode ser melhor visualizado através das Figuras 4-6 e 4-7 onde é clara uma separação entre esses dois grupos de compostos.

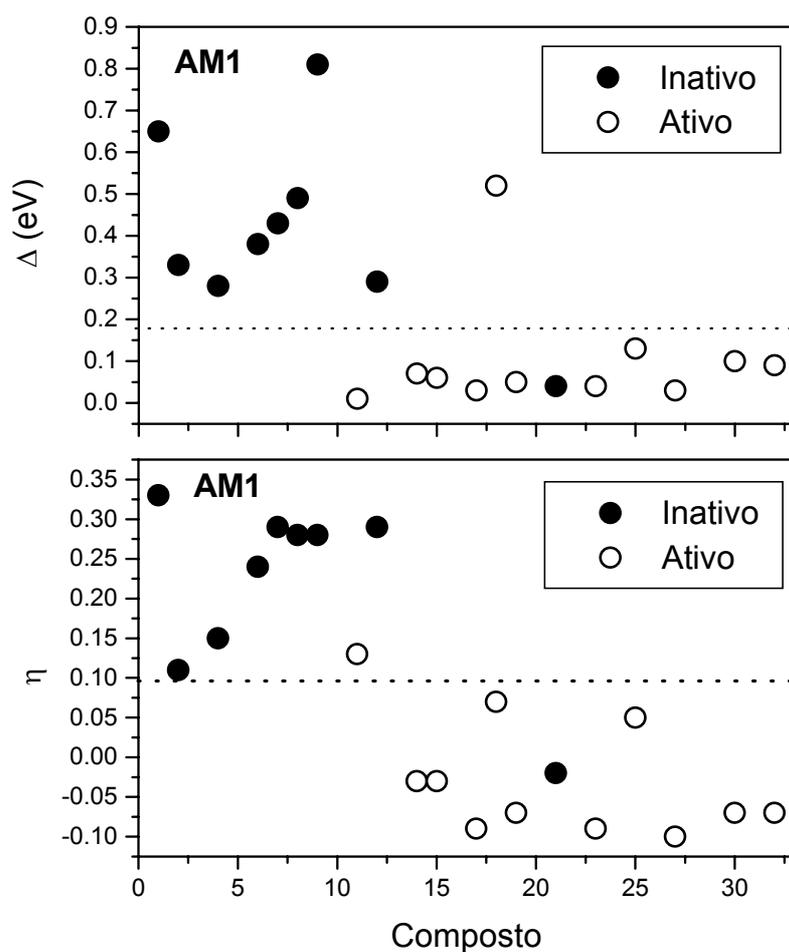


Figura 4-6 Representação gráfica dos valores de Δ (A) e η (B) da MIE para compostos ativos (○) e inativos (●) obtidos com o método AM1.

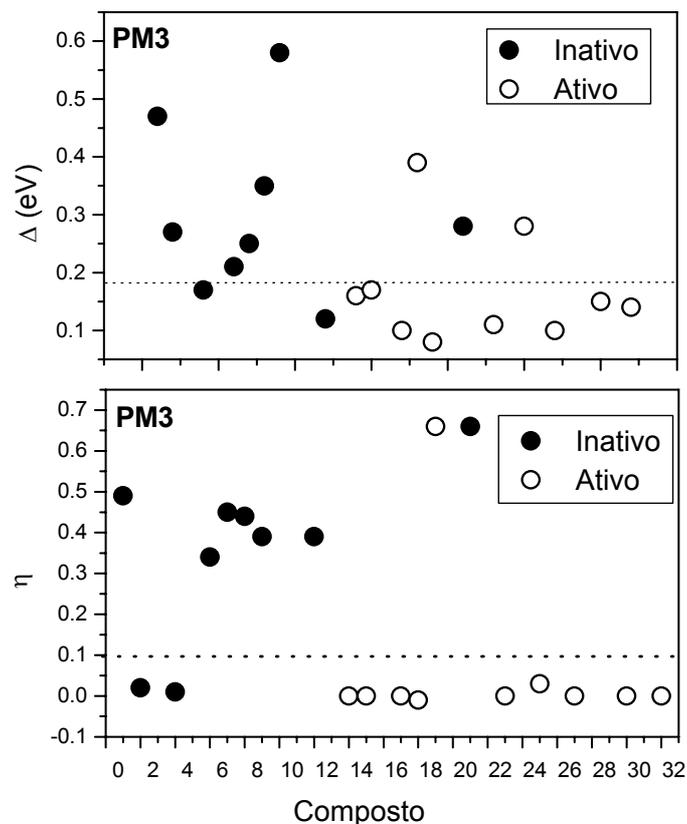


Figura 4-7 Representação gráfica dos valores de Δ (A) e η (B) da MIE para compostos ativos (○) e inativos (●) obtidos com o método PM3.

Embora o PM3 e o AM1 produzam resultados similares, o AM1 é superior em termos estatísticos. Aqui, em contraste com outras classes de compostos estudados, cada variável independentemente (especialmente para o AM1) pode ser usada para discriminar os compostos ativos e inativos com alta precisão. No presente caso o parâmetro η representa a melhor escolha:

- Δ , 90% e 75% de acerto para AM1 e PM3, respectivamente;
- η , 90% e 85% de acerto para AM1 e PM3, respectivamente.

Para testar a consistência da proposição original da MIE construímos regras Booleanas com esses parâmetros e verificamos que o poder de predição não é afetado. As regras Booleanas foram construídas da Tabela 4-8 e das Figuras 4-6 e 4-7, e podem ser assim enunciadas:

Valores Críticos :

- $\Delta \leq 0.15$ (+); $\Delta > 0.15$ (-)

- $\eta \leq 0.10$ (+); $\eta > 0.10$ (-)

Regras:

$\Delta(+)$ e $\eta (+) \Rightarrow (+)$ Ativo

$\Delta (-)$ e $\eta (-) \Rightarrow (-)$ Inativo

$\Delta (-)$ e $\eta (+) \Rightarrow (+)$ Ativo

$\Delta (+)$ e $\eta (-) \Rightarrow (-)$ Inativo

Na Tabela 4-10 apresentamos um sumário desses resultados. Os mesmos valores críticos podem ser aplicados tanto ao AM1 quanto ao PM3, embora o AM1 produza melhores resultados em termos de acerto de previsão (90%) em comparação com o PM3 (85%). Também aplicando o método AM1 encontramos um número menor de termos cruzados (+ - ou - +).

Molécula	AM1				PM3			
	Δ (eV)	η	MIE	D.E.	Δ (eV)	η	MIE	D.E.
01	-	-	I	I	-	-	I	I
02	-	-	I	I	-	+	A	I
04	-	-	I	I	-	+	A	I
06	-	-	I	I	-	-	I	I
07	-	-	I	I	-	-	I	I
08	-	-	I	I	-	-	I	I
09	-	-	I	I	-	-	I	I
11	+	-	I	A	+	+	A	A
12	-	-	I	I	+	-	I	I
14	+	+	A	A	-	+	A	A
15	+	+	A	A	-	+	A	A
17	+	+	A	A	+	+	A	A
18	-	+	A	A	-	+	A	A
19	+	+	A	A	+	-	I	A
21	+	+	A	I	-	-	I	I
23	+	+	A	A	+	+	A	A
25	+	+	A	A	-	+	A	A
27	+	+	A	A	+	+	A	A
30	+	+	A	A	+	+	A	A
32	+	+	A	A	+	+	A	A

Tabela 4-8 Esquema para discriminação entre compostos ativos e inativos. A classificação obtida com a MIE (A-ativo, I-inativo) é contrastada com os dados experimentais (D.E.). A linha sombreada indica os casos onde não houve concordância entre a EIM e os D.E..

É importante salientar que essas regras foram construídas utilizando-se um pequeno número de compostos e que seria desejável trabalharmos com um banco de dados mais abrangente, apesar disso os resultados são úteis e contribuem para a validação da MIE como uma ferramenta nos estudos SAR.

Explorando as similaridades estruturais construímos um grupo de novos potenciais compostos e os analisamos dentro das regras Booleanas construídas. As estruturas propostas foram elaboradas tendo em mente grupos substituintes presentes em drogas similares.

Na tabela abaixo (Tabela 4-11) apresentamos os resultados obtidos. Pelas regras observadas com a MIE seriam ativos os primeiros 4 compostos e inativos os dois últimos compostos. Vale ressaltar que nenhum desses compostos foi investigado experimentalmente e isto constituiria um bom teste para a MIE.

composto	R ₁ =R ₂	AM1			PM3		
		Δ(eV)	η	Prev. MIE	Δ(eV)	η	Prev. MIE
1	CHCHCOOH	0.07(+)	-0.10(+)	A	0.13(+)	0.06(+)	A
2	OCOCH ₃	0.04(+)	-0.13(+)	A	0.15(+)	0.05(+)	A
3	CH ₂ (NOH) ₂	0.01(+)	-0.03(+)	A	0.01(+)	0.04(+)	A
4	CHCHBr	0.01(+)	-0.15(+)	A	0.18(-)	0.03(+)	A
5	NO ₂	0.22(-)	0.21(-)	I	0.31(-)	0.51(-)	I
6	CH ₂ NO ₂ OH	0.17(+)	0.15(-)	I	0.12(+)	0.30(-)	I

Tabela 4-9 Parâmetros MIE para compostos hipotéticos. A previsão para a atividade desses compostos é indicada (Prev. MIE).

Capítulo 5

ESTRUTURA-ATIVIDADE DE UMA SÉRIE DE INDOLEQUINONAS COM ATIVIDADE ANTITUMORAL

5.1 INTRODUÇÃO

As mitomicinas⁵⁰⁻⁵⁷ formam um conjunto de potentes antibióticos naturais descobertos no início da década de 50. A mitomicina C⁵⁸⁵⁹ (Figura 5-1), produzida pela bactéria *Streptomyces lavadulae*, é o representante mais popular dessa família de antibióticos e vem sendo utilizada, desde a década de 60, como agente quimioterápico no tratamento de vários tipos de tumores⁶⁰ (câncer de mama, gastrointestinal e do tecido craniano).

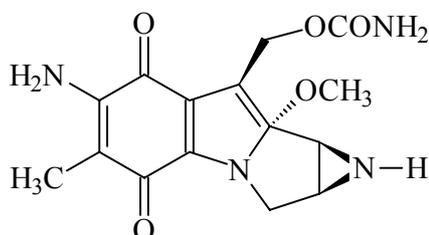


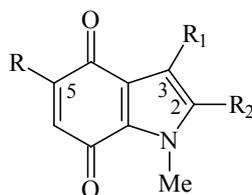
Figura 5-1 – Representação esquemática da Mitomicina C

5.2 RESULTADOS E CONCLUSÕES

Neste trabalho realizamos um estudo sistemático de uma série de indolequinonas construídas através da dimetilação do sistema indol (Tabela 1), preparadas e avaliadas por Naylor e colaboradores⁷⁶.

A atividade biológica desses compostos foi avaliada *in vitro* através dos índices $C_{50}(\text{ar})$ e $C_{50}(\text{N}_2)$, que representam a concentração necessária para matar 50% das células tumorais em um meio aeróbico e hipóxico (pobre em oxigênio), respectivamente. Em nossa análise nos concentramos na atividade avaliada no meio aeróbico, embora, como verificamos *a posteriori*, as conclusões obtidas também fossem aplicáveis aos índices de atividade biológica medida em meio hipóxico. Consideramos inativos todos compostos com valores de $C_{50}(\text{ar})$ superiores a $100\mu\text{M}$.

A presença do grupo Aziridinil (Az) na posição C-5 confere uma alta atividade ao composto, enquanto que a presença do grupo MeO nessa mesma posição está, geralmente, associada a uma baixa atividade. Estudos QSAR (quantitative structure activity relationship) anteriores tentaram correlacionar a atividade antitumoral desses compostos com parâmetros tais como o coeficiente de partição octanol-água, o potencial de redução da quinona e o tamanho dos substituintes, porém os resultados obtidos foram relativamente insatisfatórios. Sabe-se, que as presenças do anel aziridino, do substituinte carbamoiloxil ($\text{CH}_2\text{OCONH}_2$) e do esqueleto mitoseno são de suma importância para que se tenha uma boa atividade, porém não há explicação em termos de estrutura eletrônica do porquê da importância desses grupos.



COMPOSTOS	R	R ₁	R ₂	C ₅₀ (AR) ^a μM	C ₅₀ (N ₂) ^b μM
18	MeO	CO ₂ Me	c-Pr	103.5	59.8
19	Az	CO ₂ Me	c-Pr	1.72	0.93
20	MeO	CH ₂ OH	c-Pr	540	820
21	Az	CH ₂ OH	c-Pr	0.603	0.0058
24	MeO	CH ₂ OCONH ₂	c-Pr	29	0.44
25	Az	CH ₂ OCONH ₂	c-Pr	3.33	0.074
35	MeO	CO ₂ Me	CH(Me) ₂	150	150
36	MeO	CH ₂ OH	CH(Me) ₂	800	200
37	Az	CH ₂ OH	CH(Me) ₂	25.7	0.79
40	MeO	CH ₂ OCONH ₂	CH(Me) ₂	20.9	0.111
41	Az	CH ₂ OCONH ₂	CH(Me) ₂	47.5	5.46
43	MeO	Me	CH(Me) ₂	1.236	0.657
44	Az	Me	CH(Me) ₂	0.33	0.188
45	MeO	CH ₂ OH	c-Hex	122.4	20.5
46	Az	CH ₂ OH	c-Hex	34.5	0.93
52	MeO	CH ₂ OH	Me	1077	284.8
53	Az	CH ₂ OH	Me	0.149	0.0116
62	MeO	CH ₂ OH	H	220	240
63	Az	CH ₂ OH	H	0.153	0.0086
65	MeO	CH ₂ OCONH ₂	H	3.1	0.037
66	Az	CH ₂ OCONH ₂	H	0.00019	0.00013

Tabela 5-1 Análogos à indolequinona. Me: Metil, c-Pr: Ciclopropano, c-Hex: ciclohexano.

a e b extraídos da referência.

O estudo dessa série de indolequinonas com atividade antitumoral foi um dos primeiros trabalhos que realizamos, e nosso objetivo imediato era verificar se poderíamos estender a metodologia de índices eletrônicos (MIE), que anteriormente havia sido aplicada aos hidrocarbonetos aromáticos, para esse conjunto de compostos. Por esse motivo apresentaremos os resultados obtidos pelos os métodos semiempíricos de forma sucinta, sem entrarmos nos detalhes a

respeito dos cálculos realizados para a obtenção das estruturas de mínima energia, embora a busca conformacional feita para a obtenção dessas estruturas tenha sido um trabalho exaustivo devido à grande riqueza conformacional desses compostos.

Os cálculos de estrutura eletrônica foram realizados com o método semiempírico PM3 dentro do programa MOPAC6 e, posteriormente, repetimos alguns cálculos também com o AM1 e verificamos uma perfeita concordância entre os métodos no que diz respeito às conclusões obtidas pela MIE. Na Tabela 5-2 apresentamos alguns parâmetros obtidos com o PM3. Uma vez não sendo verificada nenhuma correlação direta da atividade biológica com esses parâmetros eletrônicos, prosseguimos com a análise da densidade local de estados.

A primeira região a ser analisada foi o anel indol. Um resultado típico para a densidade de estados calculada sobre essa região pode ser visto na Figura 5-3, onde comparamos os compostos 36 e 37, cuja única diferença estrutural está no grupo substituinte que ocupa a posição C-5.

Observamos que o contraste essencial na distribuição espacial de estados no espaço de energia desses dois compostos está nas diferenças de energia e nas intensidades relativas dos orbitais HOMO, HOMO-1 e HOMO-2.

Composto	ΔH_F (kCal/Mol)	Dipolo (Debye)	E_{HOMO} (eV)	E_{LUMO} (eV)	$E_{HOMO-LUMO}$ (eV)
18	-123.13	3.60	-9.58	-1.31	8.27
19	-50.61	2.72	-9.51	-1.39	8.12
20	-92.25	3.08	-9.36	-1.27	8.09
21	-62.06	3.31	-9.33	-1.36	7.97
24	-125.80	4.99	-9.51	-1.42	8.09
25	-53.89	2.80	-9.39	-1.33	8.06
35	-161.67	3.65	-9.61	-1.30	8.30
36	-131.50	3.22	-9.38	-1.26	8.11
37	-58.82	3.47	-9.36	-1.36	8.00
40	-164.88	4.94	-9.54	-1.41	8.12
41	-93.78	0.85	-9.30	-1.21	8.09
43	-91.81	2.64	-9.15	-1.10	8.04
44	-19.23	2.65	-9.14	-1.19	7.95
45	-138.81	3.21	-9.36	-1.26	8.10
46	-66.17	3.20	-9.35	-1.36	7.99
52	-121.32	3.16	-9.33	-1.28	8.05
53	-44.68	1.82	-9.06	-1.24	7.82
62	-109.31	1.30	-9.28	-1.17	8.10
63	-36.67	1.65	-9.27	-1.26	8.01
65	-145.43	5.07	-9.76	-1.50	8.26
66	-72.73	5.21	-9.70	-1.60	8.10

Tabela 5-2 Sumário dos cálculos obtidos com o PM3.

Para o composto inativo (36) a intensidade desses esses orbitais segue o padrão: HOMO>HOMO-1>HOMO-2. Para o composto ativo (37) o padrão seguido é: HOMO>>HOMO-1<<HOMO-2. Com relação às diferenças de energia, a substituição aziridinil aproxima os níveis de energia HOMO/HOMO-1 em 39% e os níveis HOMO-1/HOMO-2 em 58% em relação à substituição metoxil. Esses dois comportamentos ilustram o efeito eletrônico que a substituição aziridinil provoca nas indolequinonas. Através dos parâmetros Δ e η da metodologia de índices eletrônicos podemos avaliar mais precisamente esse comportamento e verificar se

esse mesmo padrão é compartilhado pelo restante dos compostos ativos e inativos dessa série.

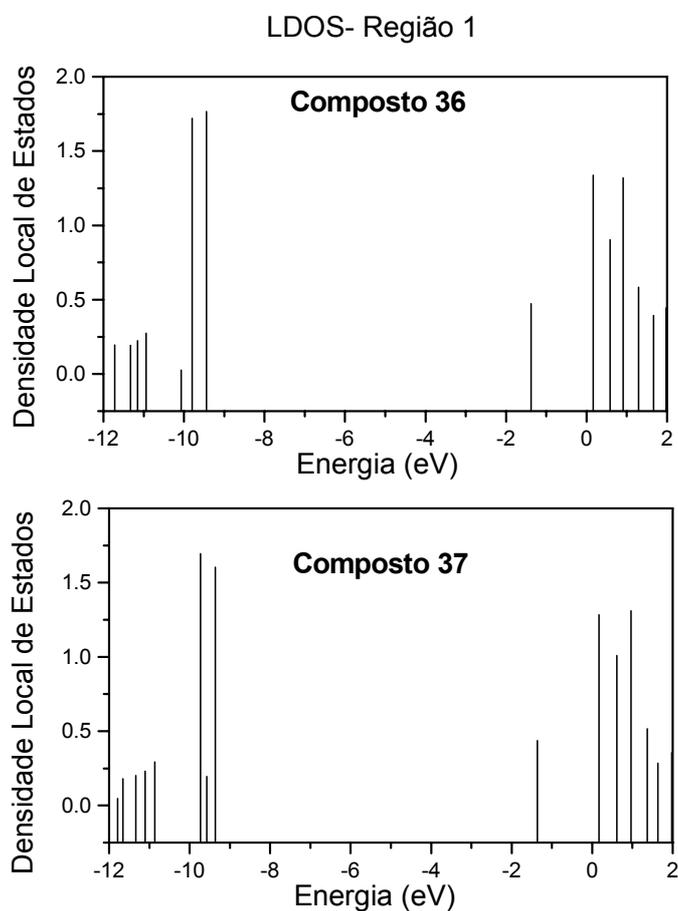


Figura 5-3 LDOS calculada sobre o anel indol. Para os compostos 36 (inativo) e 37 (ativo).

- Anel indol - Orbitais moleculares HOMO e HOMO-1

Os parâmetros Δ e η da metodologia de índices eletrônicos associados aos orbitais HOMO e HOMO-1 para a LDOS calculada sobre o anel indol podem ser vistos na figura abaixo (Figura 5-3).

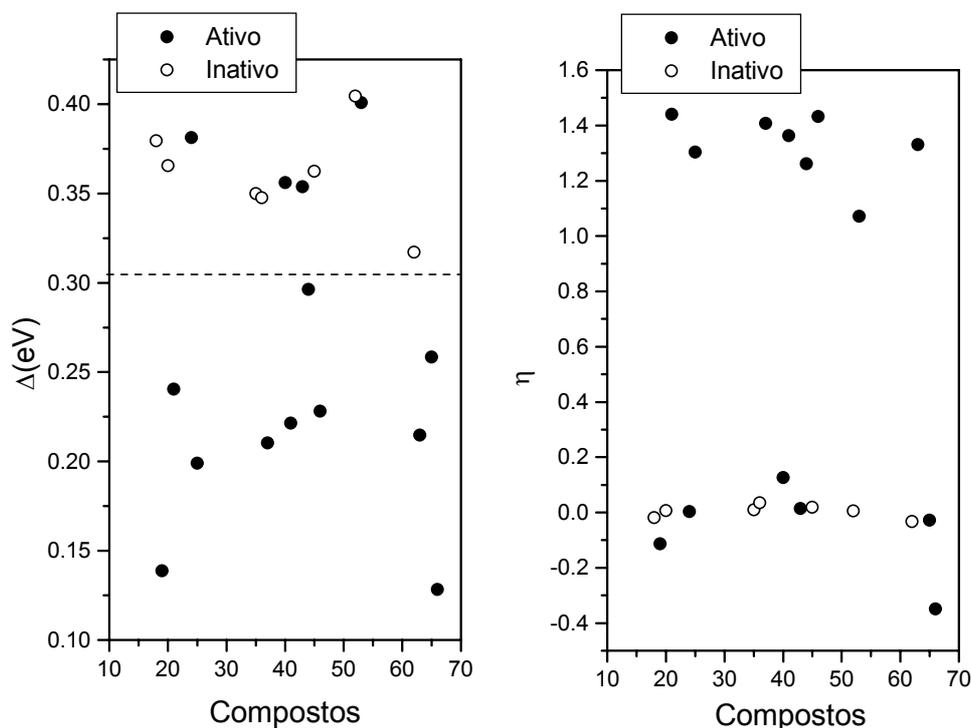


Figura 5-4 Parâmetros Δ e η calculados sobre o anel indol.

Alguns aspectos devem ser observados. Para esse caso o parâmetro Δ é um bom descritor dos compostos inativos, pois permite que eles sejam agrupados espacialmente, sem exceções. Podemos estabelecer um valor limite para Δ acima do qual o composto é classificado com inativo (como, por exemplo, 0.3 eV), porém alguns compostos ativos não se encaixam nessa classificação. São eles os compostos 24, 40, 43 e 53. Excetuando o composto 53, os outros compostos possuem uma atividade alta apesar de apresentarem um grupo metoxil (MeO) ocupando a posição C-5. Isso nos mostra que o parâmetro Δ associado aos níveis de energia HOMO e HOMO-1 dessa região reflete essencialmente a presença ou

não do grupo aziridinil. Com relação ao parâmetro η , a presença do grupo Az aumenta a diferença de intensidade relativa entre os orbitais HOMO e HOMO-1 e, novamente, poderíamos sugerir um valor para η abaixo do qual os compostos seriam inativos. Podemos concluir então que com esses dois parâmetros podemos discriminar com relativo sucesso os compostos ativos e inativos, porém eles não são capazes de traduzir o efeito da presença do grupo carbamoiloxil na posição 3-indol dos compostos 5-metoxindolequinonas (como no caso dos compostos 24, 40 e 65).

Continuando a análise LDOS sobre o anel indol, pesquisamos os índices Δ e η associados aos orbitais HOMO-1 e HOMO-2, porém não obtivemos resultados substanciais. Partimos então para o estudo da LDOS calculada sobre o anel quinônico, que denominamos região 2.

- Anel Quinônico - Orbitais moleculares HOMO-2 e HOMO-1

Como dissemos anteriormente, antes de atuar como agentes alquilantes esses compostos sofrem uma redução enzimática no anel quinônico (a quiniona é transformada em hidroquinona), o que sugere que esse anel possa guardar alguma correlação com a atividade antitumoral, embora a facilidade de redução não esteja diretamente relacionada à maior atividade antitumoral.

Novamente ilustraremos com os compostos 36 e 37 um resultado típico para a densidade de estados sobre a região 2 (Figura 5-5). Observamos que, para essa região, as diferenças de intensidade entre os orbitais HOMO e o HOMO-1 não são tão expressivas quanto no caso da região 1, e realmente

verificamos que os parâmetros Δ e η relativos a esses orbitais não se mostraram bons descritores da atividade biológica.

A intensidade HOMO-2 é a que mais se diferencia entre esses dois compostos. Enquanto que para o composto inativo o padrão de intensidade seguido é HOMO-2>>HOMO-1, para o composto ativo essa diferença é reduzida pela metade.

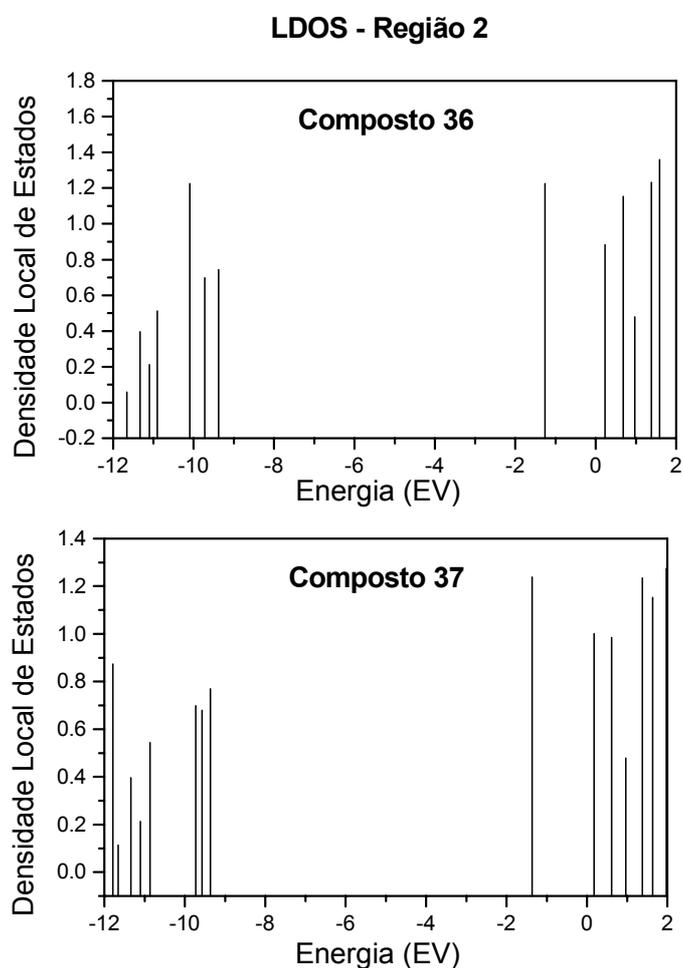


Figura 5-5 LDOS calculada sobre o anel quinônico.

Na figura abaixo podem ser vistos os parâmetros Δ e η da metodologia de índices eletrônicos associados aos orbitais HOMO-2 e HOMO-1 calculados sobre o anel quinônico.

Observamos que as menores diferenças de energia entre esses orbitais estão associadas aos compostos ativos (valor máximo de 0.31 eV). Há dois compostos, 18 e 35, que apesar de serem inativos estão agrupados com o conjunto ativo. Ambos compostos apresentam o grupo éster CO_2Me na posição R_1 , o que sugere que os efeitos desse grupo não sejam sentidos pelo orbital HOMO-2. Um outro composto, 43, é classificado erroneamente como inativo. Os resultados anteriores indicam que o índice C_{50} para esse composto poderia estar incorreto.

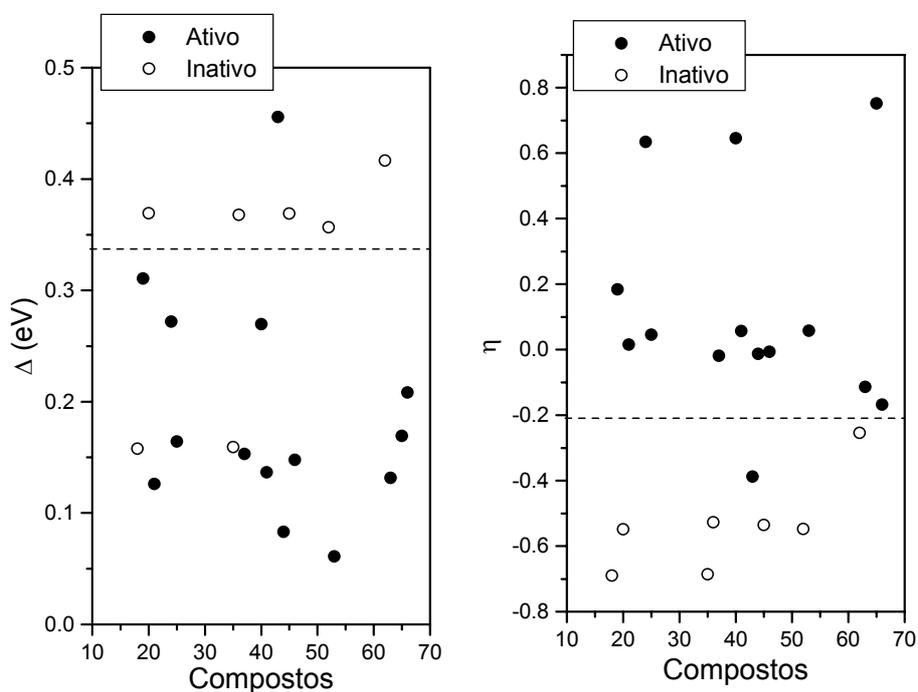


Figura 5-6 Parâmetros Δ e η calculados sobre o anel quinônico.

O índice η relativo aos orbitais HOMO-1 e HOMO-2 é o que melhor distingue as duas classes de compostos, inclusive registrando o efeito do grupo carbamato em compostos com substituição *MeO* na posição C-5. Poderíamos estabelecer um valor de corte para η em torno de -0.2, abaixo do qual o composto seria inativo.

Esses resultados nos mostram a relevância do orbital molecular HOMO-2 e do anel quinônico na análise estrutura-atividade das indolequinonas, o que é visto claramente ao compararmos a localização desse orbital sobre as indolequinonas 36 e 40 (Figura 5-7). Ao contrário dos outros substituintes que ocupam a posição 3-indole, o grupo carbamoiloxil é capaz de deslocar para si a contribuição ao HOMO-2, que nos outros compostos com substituição 5-Metoxindolequinonas está localizada no anel quinônico. Apesar do grupo carbamoiloxil estar diretamente ligado ao anel indol, o efeito eletrônico que ele ocasiona na estrutura é melhor descrito pelo anel quinônico.

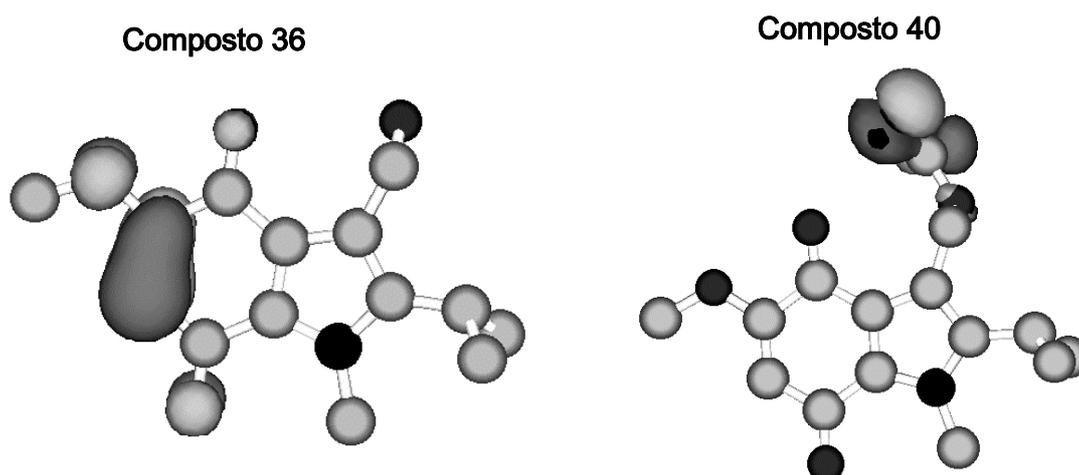


Figura 5-7 Representação da isosuperfície do orbital HOMO-2 para o composto 36 (inativo) e 40 (ativo).

Como era o objetivo inicial desse estudo, os resultados obtidos nos mostraram a eficácia da metodologia de índice eletrônicos no estudo sistemático da correlação estrutura-atividade para compostos mais complexos do ponto de vista estrutural do que os hidrocarbonetos aromáticos policíclicos.

Particularmente, a atividade anticancerígena desses agentes biorredutivos está associada aos índices Δ e η da região 2 e aos índices MIE relativos à região 1 da molécula.

Há na literatura alguns trabalhos que avaliam a eficiência de análogos às indolequinonas como substratos a enzima DT-diaforase, presentes em muitos tumores sólidos. Uma das perspectivas futuras para esse trabalho é aplicar a MIE ao estudo desse tipo de atividade biológica correlacionando-a também com a atividade anticancerígena.

capítulo 6

ANÁLISE DA ESTABILIDADE ENERGÉTICA DE MICROSSATÉLITES

PRESENTES EM CADEIAS DE NUCLEOTÍDEOS

6.1 INTRODUÇÃO

A informação genética que cada indivíduo carrega é armazenada em seus cromossomos celulares (Figura 6-1), os quais são formados por moléculas de DNA (deoxyribonucleic acid).

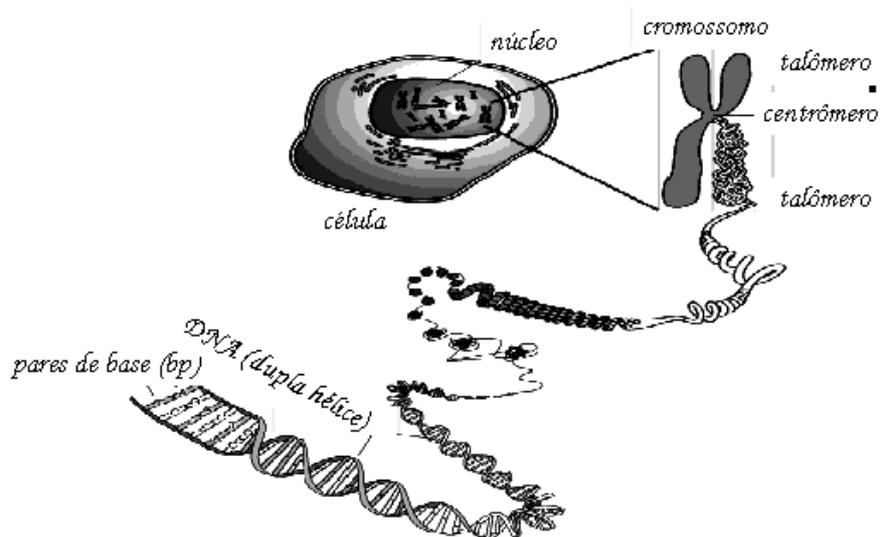


Figura 6-1 Representação do cromossomo.

O DNA é uma longa cadeia polimérica, constituída por duas fitas antiparalelas de forma helicoidal (Watson e Crick), mantidas unidas através de

pontes de hidrogênio (Figura 6-2). Cada fita de DNA é formada por uma seqüência linear de blocos unitários, chamados nucleotídeos, que por sua vez são constituídos por três grupos funcionais: uma base nitrogenada (Adenina(A), Citosina (C), Guanina (G) ou Timina (T)), uma molécula de açúcar e um grupo fosfato. A disposição das bases nitrogenadas nas fitas de DNA é de tal forma que, para cada tipo de base nitrogenada em uma das fitas, exista somente um outro tipo de base correspondente na outra fita (A-T e C-G). Essa propriedade ímpar permite que a informação genética sempre possa ser recuperada se a seqüência de bases em uma das fitas é conhecida.

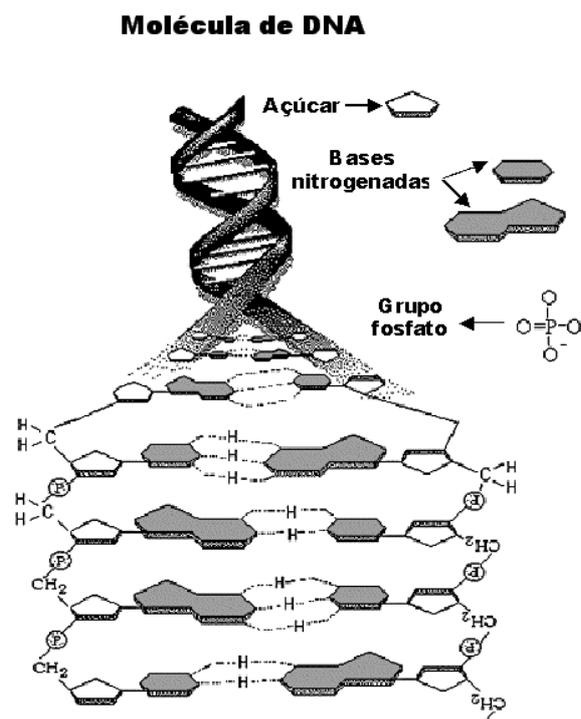


Figura 6-2 Representação esquemática da molécula de DNA e de seus constituintes.

Apesar de ser formado por milhares de nucleotídeos, somente algumas regiões do DNA, chamadas genes, são capazes de fornecer à célula a informação

de que ela necessita para exercer a importante função de codificar proteínas. Por ser vital à sobrevivência do indivíduo, o gene deve ser copiado e transmitido precisamente, de tal forma que todas as células “filhas” carreguem a mesma informação contida nas células “mães”. Esse mecanismo de cópia e transmissão de dados exige uma grande estabilidade genética, que é mantida graças a um mecanismo extremamente preciso de replicação, aliado a um mecanismo eficiente de reparação que corrige eventuais erros envolvidos nesse processo. Apesar da excelência do maquinário genético de cópia, as vezes ocorrem falhas (um nucleotídeo errado a cada 10^9 copiados) e se produz um erro na transmissão de dados que pode adquirir um caráter permanente, fenômeno a que se chama mutação. Os tipos mais comuns de mutação que ocorrem durante o processo de duplicação são : exclusão, inserção, expansão ou translocação de nucleotídeos (Figura 6-3).

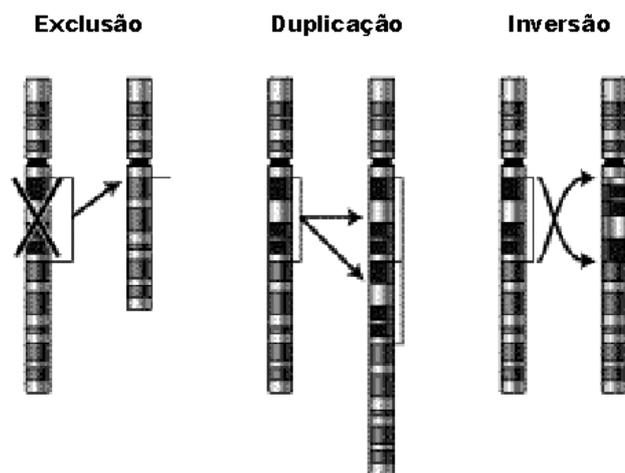


Figura 6-3 Representação de algumas mutações que ocorrem durante o processo de replicação do DNA.

Algumas regiões do DNA, tanto de seres procariotes quanto eucariotes, apresentam extensas seqüências de repetições de oligômeros (mono, di, tri ou tetranucleotídeos). Estudos recentes mostram que estas repetições possuem uma instabilidade genética intrínseca⁷⁷⁻⁸⁰, manifestada com freqüentes mudanças de comprimento devido à expansão ou exclusão das unidades que se repetem. A origem, a evolução, e a regra biológica para essas repetições de oligômeros que ocorrem no DNA, chamadas microssatélites (MS) ou repetições de seqüências simples (SSR- simple sequence repetition), são alvo de grande interesse dentro da biologia molecular, principalmente devido ao papel que desempenham na organização do genoma e nos processos evolutivos.

Pesquisas recentes associam a expansão das SSR a diversos tipos de distúrbios hereditários⁸¹⁻⁸⁵. A expansão descontrolada do dímero $(CA)_k$, por exemplo, está associada ao desenvolvimento do câncer de ovário⁸⁶, enquanto que expansões do trímero $(CAG)_k$ são causadoras de diversos distúrbios neuromusculares e neurodegenerativos dominantes⁸⁷. A maioria dessas doenças apresenta um comportamento clínico conhecido como antecipação, que é o aumento da severidade e/ou a diminuição da idade na qual a doença se manifesta a cada nova geração. Esse fenômeno está diretamente correlacionado ao tamanho da repetição oligomérica.

A análise de seqüências de DNA disponíveis em bancos de dados genéticos (como o GenBank) revelou que, embora as SSR sejam abundantes nas regiões do DNA que não são responsáveis pela codificação de proteínas (introns e regiões intragênicas), elas são relativamente raras em regiões codificadoras (éxons)⁸⁸⁻⁹¹. Esse fato sugere que a pressão evolutiva sofrida pelas regiões do

DNA que contêm o código genético para a elaboração de proteínas tenha sido maior do que a pressão sofrida pelas regiões não-codificadoras, o que fez com que as regiões codificadoras fossem mais conservadas durante o processo evolutivo.

A abordagem mais comum a respeito da presença das SSR no genoma humano tem sido a da análise estatística^{88-91,92} Na maioria dos trabalhos busca-se modelar a ocorrência de determinados oligômeros no genoma através de funções de distribuição de probabilidades. Particularmente tem-se dado destaque às repetições diméricas⁹², onde estudos indicam uma correlação de curto alcance para regiões codificadoras de proteínas (distribuição exponencial), e de longo alcance para regiões não-codificadoras (lei de potências). Nossa proposta foi abordar esse problema do ponto de vista da estabilidade energética, o que parece ser um caminho promissor para o entendimento da organização do genoma humano, podendo trazer informações relevantes a respeito do por que da instabilidade genética intrínseca das SSR.

6.2 RESULTADOS E CONCLUSÕES

As bases moleculares de algumas desordens genéticas vêm gradativamente sendo estabelecidas, e um grande número delas está associada à expansão das repetições triméricas. Em particular as repetições do tipo $(CAG)_k$, $(CGG)_k$ e $(AAG)_k$ já foram associadas a pelo menos dezessete tipos de doenças hereditárias⁹³⁻⁹⁵, sendo somente a expansão $(CAG)_k$ responsável por pelo menos oito desordens neurodegenerativas dominantes. Essas características nos fizeram

iniciar nossos estudos de estabilidade energética utilizando esse tipo de repetição trimérica.

Construímos as seqüências de nucleotídeos utilizando o banco de dados do programa HyperChem 5.0. Trabalhamos com uma conformação de dupla hélice na forma B, e com o grupo ribose na forma 2' endo. A limitação computacional nos levou a realizar somente cálculos de mecânica molecular do tipo *single point*, que foram feitos com o campo de força AMBER, desenvolvido para descrever esse tipo de estrutura (ácidos nucléicos e aminoácidos).

Trabalhamos com seqüências típicas de 600 nucleotídeos (600 pares de base), o que representa, aproximadamente, 76.800 átomos (incluindo os hidrogênios). Primeiramente geramos uma seqüência formada completamente por repetições triplas e, posteriormente, mantendo o tamanho da seqüência e a quantidade de bases, fomos diminuindo gradativamente o número de repetições presentes, ao mesmo tempo em que monitorávamos o comportamento da energia total e de suas componentes (energia de estiramento, de ligação, de torção, de van der Waals, de ponte de hidrogênio e eletrostática).

As repetições do tipo $(CAG)_k$ foram as primeiras a serem analisadas, onde k variou de zero a duzentos. Na Figura 6-4 apresentamos os resultados obtidos para a energia total. Há uma clara tendência de aumento da energia à medida em que k aumenta.

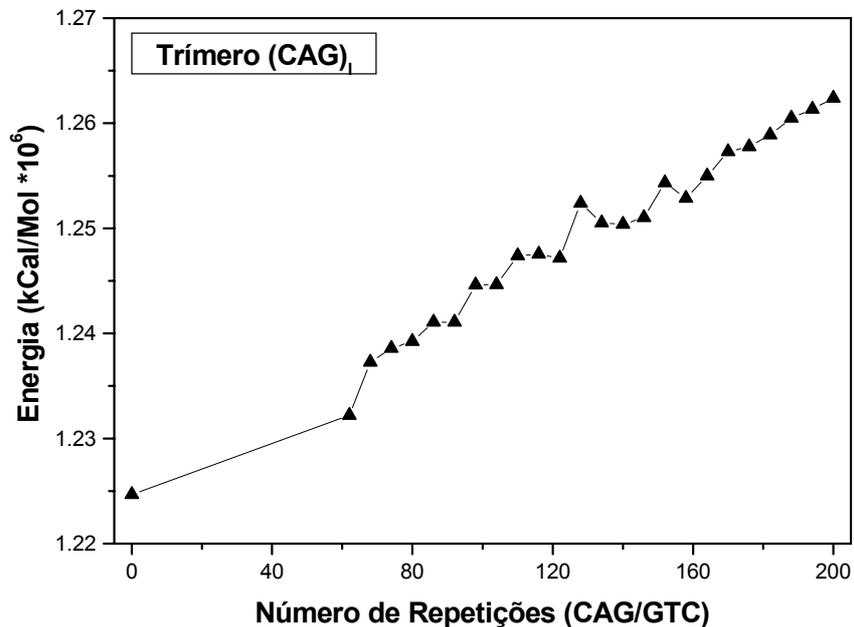


Figura 6-4 Energia total para uma seqüência trimérica CAG.

O fato de a energia total aumentar à medida em que o sistema se torna mais organizado (maior número de repetições) sugere que a instabilidade genética que essas seqüências apresentam pode estar relacionada à sua instabilidade energética.

A análise das componentes da energia nos permite avaliar como cada uma delas contribui para a evolução da energia total da cadeia polimérica à medida em que k varia. Na Figura 6-5 seguinte temos a representação normalizada dessas energias. Na Figura 6-5A as energias são normalizadas pela média e na Figura 6-5B elas são normalizadas pelas energias da seqüência onde $k = 0$ (seqüência completamente aleatória).

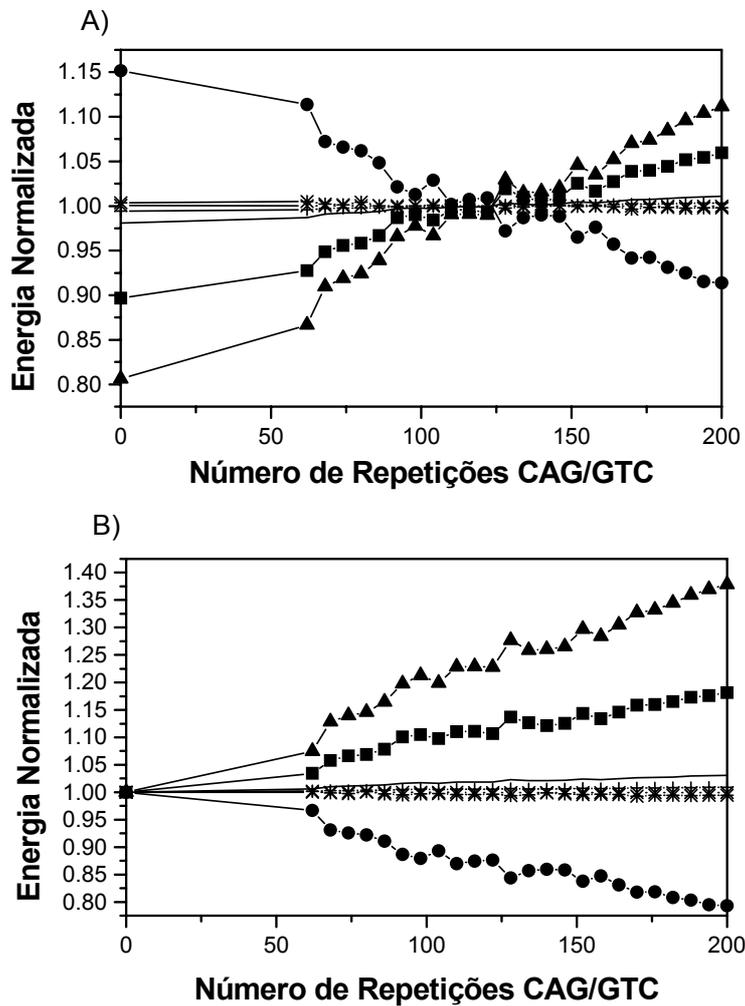


Figura 6-5

Figura 6-5 – A) Gráfico das componentes da energia total normalizadas pela média. Energia de estiramento (•); energia angular (■); energia de torção (+); energia de van der Waals (▲); energia de ligação de hidrogênio (*); energia eletrostática (x) e (-) energia total. B) Gráfico das componentes da energia total normalizadas pela seqüência $k=0$.

Observe que a energia de estiramento é a que mais diminui à medida em que k aumenta, enquanto que a energia de van der Waals seguida da energia angular são as que mais aumentam à medida que o número de repetições CAG/GTC aumenta. As energias de torção, eletrostática e de ligação de hidrogênio sofrem pouca variação. Isso nos mostra que, enquanto para longas repetições as ligações entre os átomos das seqüências são menos tensionadas, o acoplamento entre os átomos das bases (representando pela energia de van der Waals) é enfraquecido.

Repetimos o mesmo tipo de construção para a seqüência $(CTA)_k$ e novamente observamos o mesmo tipo de comportamento, embora nesse caso a energia de ligação de hidrogênio tenha um comportamento similar à energia de ligação, colaborando para a diminuição da energia total a medida que o número de repetições $(CTA)_k$ aumenta.

Os resultados para a repetição (AAG) também podem ser vistos abaixo (Figura 6-8). Para esse caso o aumento da energia total é mais suave, o que já esperávamos pelo fato do grande número de bases adenina em relação à guanina possibilitar o surgimento de seqüências monoméricas do tipo $(A)_k$ onde as bases foram distribuídas aleatoriamente. Esse tipo de repetição também pode desestabilizar a seqüência.

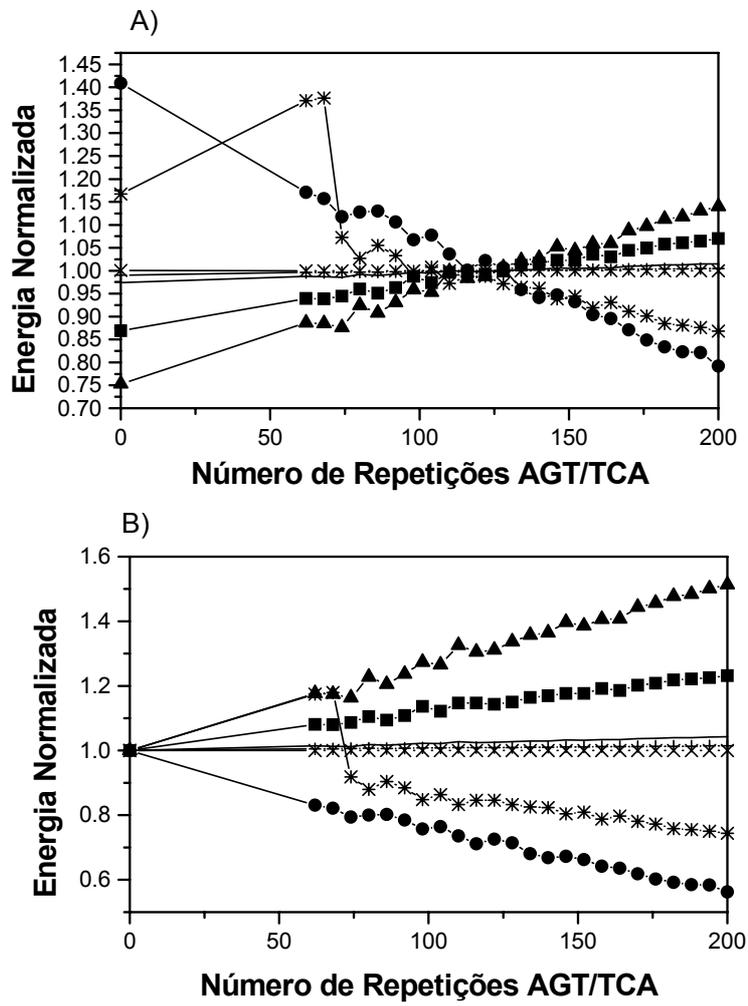


Figura 6-6

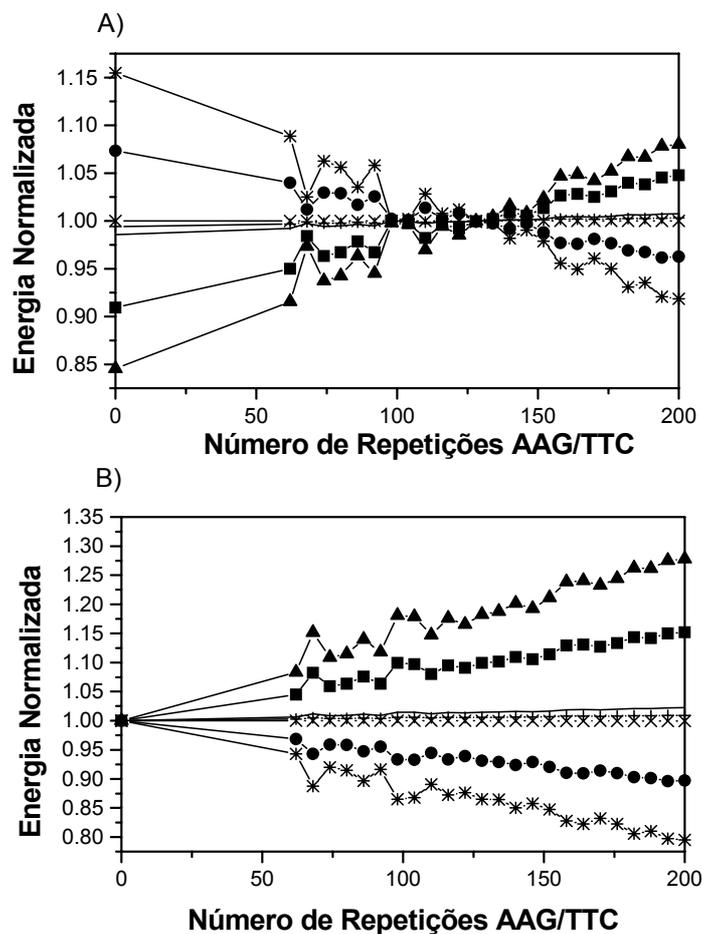


Figura 6-7 A) Gráfico das componentes da energia total normalizadas pela média e B) pela seqüência k=0

A imensa quantidade de informação genética disponível atualmente em bancos de dados genéticos nos permitiu analisar várias seqüências de repetições associadas a diferentes tipos de desordens hereditárias. As seqüências com as quais trabalhamos foram retiradas do banco de dados GenBank (www.ncbi.nlm.nih.gov), e posteriormente foram convertidas em um formato que pudesse ser lido pelos programas computacionais disponíveis. Nesse processo

encontramos algumas dificuldades, tais como: localizar as regiões gênicas ou intragênicas onde as repetições associadas às doenças estavam presentes, encontrar a seqüência de repetições dentro da seqüência global fornecida e descartar as seqüências redundantes.

Uma vez localizada a seqüência de repetições (cromossomo e gene), extraíamos aproximadamente 300 nucleotídeos da seqüência total fornecida (quando possível) que englobasse a região das SSR. Poderíamos construir esse tipo de seqüência distribuindo aleatoriamente as bases A, T, C e G sobre um determinado comprimento de nucleotídeos, e, posteriormente, inserindo microsatélites na seqüência formada; entretanto a informação genética embutida em uma seqüência real não seria avaliada.

Localizamos seqüências SSR do DNA humano associadas à algumas desordens causadas pelas repetições triméricas CAG, AAG e CGG. Podemos citar: Distrofia Miotônica⁹⁶, Doença de Huntington⁹⁷, Ataxia Espinocerebelar Tipo 1, Atrofia Muscular Espinal e Bulbar, Atrofia Dentatorubral-Palidolusiana, Síndrome do Cromossomo X Frágil⁹⁸ e Doença de Kennedy. Calculamos a energia da seqüência da maneira como ela havia sido extraída do GenBank. Posteriormente distribuímos aleatoriamente a região da seqüência onde as repetições estavam presentes. Em todos os casos citados anteriormente a energia total das seqüências onde as SSR foram gradativamente suprimidas foi inferior à respectiva seqüência com as SSR intactas.

Um fato interessante de se ressaltar é que muitas dessas desordens genéticas associadas às SSR estão localizadas em regiões do DNA consideradas 'lixo' (junk DNA) Essas regiões (íntrons e regiões intragênicas) constituem

aproximadamente 97% do genoma humano e, embora tenham sido tratadas por muito tempo como lixo genético, trabalhos recentes evidenciam sua importância em muitas funções regulatórias desempenhadas pelo DNA, e também mostram sua influência no comportamento das regiões codificadoras de proteína. A própria associação das SSR presentes nessas regiões a vários distúrbios genéticos evidencia sua importância no processo evolutivo.

Um outro estudo que fizemos do qual obtivemos resultados interessantes foi a avaliação da energia total para seqüências de repetições onde um nucleotídeo (ou mais) era introduzido em posições estratégicas, quebrando a continuidade da seqüência que se repetia. A tendência geral que observamos foi a da diminuição da energia para esses casos, e, avaliando várias posições de interrupção, verificamos que para cada tipo de repetição havia uma posição preferencial onde ocorria uma maior estabilização da cadeia (menor energia). Esse resultado é similar ao de um trabalho de 1994 de Eichler⁹⁹ e colaboradores em que se verificou que a indivíduos normais apresentavam em seu genótipo seqüências GGC interrompidas por uma base A (GGC...GGAGGC...) enquanto que aqueles indivíduos que perdiam o códon AGG adquiriam uma grande tendência em expandir essas repetições triméricas, desenvolvendo a síndrome do cromossomo X frágil.

O debate na literatura a respeito da randomicidade ou não do genoma humano está longe de chegar a um consenso. Enquanto alguns tentam identificar padrões de repetição no genoma⁹¹, estudos estatísticos indicam uma forte pressão evolutiva contra a presença das SSR nos éxons. Em um trabalho de 1997, Li,

propõe que, para (dobrar) atingir o estado nativo, as proteínas devem ter propriedades estatísticas próximas aquelas de uma seqüência randômica¹⁰⁰.

Uma forma interessante de abordar a questão da informação contida nas seqüências de DNA é através de uma teoria elaborada por Shannon conhecida como teoria da informação¹⁰¹. Shannon relacionou a quantidade de informação carregada por um evento com a probabilidade do mesmo ocorrer. Esse conceito de incerteza ou entropia da informação foi formalizado através da equação 6-1.

$$H = -\sum_{i=1}^n p_i \log_2 p_i \quad (6-1)$$

sendo p_i a probabilidade do evento i ocorrer.

É interessante notar que a quantidade de informação (medida em bits) é inversamente proporcional à probabilidade de ocorrência do evento, ou seja, a ocorrência de eventos raros traz muito mais informação ao sistema do que a ocorrência de eventos comuns.

De acordo com a equação 6-1, se, por exemplo, uma fonte de mensagens puder fornecer a um observador (receptor) quatro elementos diferentes com igual probabilidade, digamos as letras A, T, C ou G, a incerteza na informação será de 2 bits por símbolo transmitido. Se a distribuição de probabilidades não for homogênea, ou seja, se cada símbolo for transmitido com uma probabilidade diferente, haverá uma diminuição na entropia da informação, pois o fato das probabilidades serem desiguais já é uma informação dada ao sistema, o que conseqüentemente reduz a incerteza. Para o caso de um dispositivo que produza apenas um símbolo, a incerteza na informação é nula ($\log_2 1=0$).

Como o DNA é, essencialmente, um banco de dados que carrega informação genética, acreditamos ser um caminho promissor analisar a estrutura do código genético (padrões e informação) utilizando a formulação da teoria da informação, ou seja, utilizando a relação entre ordem/desordem e a informação transmitida por seqüências de DNA.

Referências

- ¹ Cohen N. C., *Guidebook on Molecular Modeling in Drug Design*. Academic Press, San Diego, California, 1996.
- ² Clark, T. *A Handbook of Computational Chemistry- A Practical Guide to Chemical Structure and Energy Calculations*. John Wiley & Sons, EUA, 1985.
- ³ HyperChem©. *Computational Chemistry*, Hypercube, 1996
- ⁴ Allinger, N. L. *Journal of American Chemical Society* **1977**, 99, 8127.
- ⁵ Weiner, S. J. et al *Journal of American Chemical Society* **1984**, 106, 765.
- ⁶ Lowe, J. P. *Quantum Chemistry*. Academic Press, San Diego, California, 1978.
- ⁷ Pople, J.A., Beveridge, D. L., *Approximate Molecular Orbital Theory*. McGraw-Hill, 1970.
- ⁸ Dewar, M.J.S.; Zoebish, E.G.; Healy, E.F.; Stewart, J.J.P. *J. Am. Chem. Soc.* **1985**, 107, 3902.
- ⁹ Stewart, J.J.P. *J. Comp. Chem.* **1991**, 209, 10, 221; MOPAC program, version 6.0, Quantum Chemistry Exchange Program No. 455.
- ¹⁰ Dewar, M.J.S.; McKee, M.L. *J. Am. Chem. Soc.* **1977**, 99, 5231.
- ¹¹ Zerner, M.C. *in Reviews in Computational Chemistry II*, Lipkowitz, KB, and Boyd, DB, Eds., VCH Publishers, 1991, 313.
- ¹² Scano, P.; Thompson, C. *J. Comp. Chem.* **1991**, 12, 172.
- ¹³ Zoos, Z.G.; Galvão, D.S. *J. Phys. Chem.*, **1994**, 98, 1029 and references therein.
- ¹⁴ Koll, A.; Rospenk, M.; Jagodzinska, E.; Dziembowska, T. *J. of Molecular Structure* **2000**, 552, 193-204.
- ¹⁵ Gorb, L.; Korkin, A.; Leszczynski, J.; Varnek, A.; Mark, F.; Schaffner, K. *J. Mol. Struct. (THEOCHEM)* **1998**, 425, 137-145.

-
- ¹⁶ Palafox, M.A.; Melendez, F.J. *J. Mol. Struct. (THEOCHEM)* **1999**, 459, 239-271.
- ¹⁷ Ridley, J.; Zerner, M.C. *Theo. Chim. Acta* **1976**, 42, 223.
- ¹⁸ Santo, L. L. E.; Durrant, S. F.; Rangel, E. C. *et al J VAC SCI TECHNOL A* **2000**, 18: (5), 2466-2471.
- ¹⁹ Korolkovas A., *Essential of Molecular Pharmacology*. John Wiley & Sons, NY, 1970.
- ²⁰ Karcher W., Devillers, J. *Practical Applications of Quantitative Structure-Activity Relationship (QSAR) in Environmental Chemistry and Toxicology*. Kluwer Academic Publishers, Holland, 1990.
- ²¹ Fleming, I. *Frontier Orbitals and Organic Chemical Reactions*. John Wiley & Sons, Gran Bretanha (Great Britain) 1976.
- ²² Barone, P.M.V.B.; Camilo Jr, A.; Galvão, D.S. *Phys. Rev. Lett.* **1996**, 77, 1186.
- ²³ Braga, R.S.; Barone, P.M.V.B.; Galvão, D.S. *J. Mol. Struct. (THEOCHEM)* **1999**, 464, 257.
- ²⁴ Barone, P.M.V.B.; Braga, R.S.; Camilo Jr. A.; Galvão, D.S. *J. Mol. Struct. (THEOCHEM)* **2000**, 505, 55-66.
- ²⁵ Santo, L.L.E.; Galvao, D.S. *J. Mol. Struct. (THEOCHEM)* **1999**, 464, 273.
- ²⁶ Braga, R.S.; Vendrame, R.; Galvão, D.S. *J. Chem. Inf. Comput. Sci.* **2000**, 40, 1377.
- ²⁷ Cyrillo, M.; Galvao, D.S. *J. Mol. Struct. (THEOCHEM)* **1999**, 464, 267.
- ²⁸ Vendrame, R.; Braga, R.S.; Takahata, Y.; Galvão, D.S. *J. Chem. Inf. Comput. Sci.* **1999**, 39, 1094.
- ²⁹ Vendrame, R.; Braga, R.S.; Takahata, Y.; Galvão, D.S. *J. Mol. Struct. (THEOCHEM)*, in press.
- ³⁰ Morrison, R. K.; Brown, D. E.; Oleson, J.J., Cooney, D. A. *Toxicol. Appl. Pharmacol* **1970**, 17, 1.
- ³¹ da Silveira, J.C. *Pytochem.* **1975**, 14, 1829.

-
- ³² Jacobsen, N. *Acta Chem. Scand.* **1973**,27, 3211.
- ³³ *Trans R Soc Trop Med Hyg.* **1991** May-Jun;85(3):372-4.
- ³⁴ *Planta Med.* **1994** Aug;60(4):373-4.
- ³⁵ Lagrota, M.H.C.; Wigg, M.D.; Pereira, L.O.B.; Fonseca, M.E.F.; Pereira, M.A.; Guimarães, J.C. *Rev. Microbiologia* **1983**, 14, 21.
- ³⁶ Subramaniam, S.; Ferreira, M.M.C.; Trsic,M. *Structural Chemistry* **1998**, 9, 1,47.
- ³⁷ S. Subramaniam . Tese. Instituto de Química da UNICAMP, Universidade de Campinas, Campinas, São Paulo, Brasil, **1998**.
- ³⁸ T. W. Graham Solomons, *Organic Chemistry*, John Wiley & Sons, New York, **1988**.
- ³⁹ Hashimoto, Y. *Current Medicinal Chemistry.* **1998**, 5, 163-178.
- ⁴⁰ Shannon, E.D.; Morales, M.J.; Sandoval, F. *Immunopharmacology* **1997**, 35, 203-212.
- ⁴¹ Corral, L.G.; Muller, G.W.; Moreira, A.L.; Chen, Y.; Wu, M.; Stirling, D.; Kaplan, G. *Molecular Medicine* **1996**, 2, 506-515.
- ⁴² Miyachi, H.; Koiso, Y.; Shirai, R.; Niwayama, S.; Liu, J.O.; Hashimoto, Y. *Chem. Pharm. Bull.* **1998**, 46, 1165-1168.
- ⁴³ Shah, J.H.; Swartz, G.M.; Papathanassiou, A.E.; Treston, A.M.; Fogler, W.E.; Madsen, J.W.; Green, S.J. *J. Med. Chem.* **1999**, 42, 3014-3017.
- ⁴⁴ Sampaio, E.P.; Sarno, E.N.; Galilly, R.; Cohn, Z.A.; Kaplan, G. *J. Exp. Med.* **1991**, 173, 669.
- ⁴⁵ Muller, G.W. *CHEMTECH* **1997**, 21-25.
- ⁴⁶ Eriksson, T. *et al Chirality* 1995, 7, 44-52.
- ⁴⁷ Shannon, E.J.; Morales, M.J.; Sandoval, F. *Immunopharmacology* 35 (1997) 203-212.
- ⁴⁸ Shah, J.H. *et al J. Med. Chem.* **1999**, 42, 3014-3017.
- ⁴⁹ Miyachi, H.; Azuma, A.; Ogasawara, A.; Uchimura, E.; Watanabe, N.; Kobayashi, Y.; Kato, F.; Kato, M.; Hashimoto, Y. *J. Med. Chem.* **1997**, 40, 2858-2865.

-
- ⁵⁰ Tulinski, A.; van de Hende, J. H. *Journal of the American Chemical Society* **1967**, *7*, 2905.
- ⁵¹ Ogawa, K.; Nomura, A.; Fujiwara, T.; Tomita, K. *Bull of Chem. Soc. of Japan* **1979**, *52*, 8, 2344.
- ⁵² Shirahata, K.; Hirayama, N. *J. Am. Chem. Soc.* **1983**, *105*, 7199.
- ⁵³ Hirayama, N.; Shirahata, K. *Acta Cryst.* **1987**, B43, 555.
- ⁵⁴ Hirayama, N.; Shirahata, K. *Acta Cryst.* **1989**, C45, 1780.
- ⁵⁵ Hirayama, N. *Acta Cryst.* **1991**, C47, 604.
- ⁵⁶ Hirayama, N.; Arai, H.; Kasai, M. *Acta Cryst.* **1996**, C52, 1806.
- ⁵⁷ Hirayama, N.; Arai, H.; Kasai, M. *Acta Cryst.* **1996**, C52, 2365.
- ⁵⁸ S.K. Arora, *Life Sciences*, *24* (1979) 1519.
- ⁵⁹ K. Shirahata and N. Hirayama, *J. Am. Chem. Soc.*, *105* (1983) 7199.
- ⁶⁰ Tomasz, M.; Palom, Y. *Pharm. Ther* **1997**, *76*, 73.
- ⁶¹ Rao, S. N.; Singh, U. C.; Kollman, P. *J. Am. Chem. Soc.* **1986**, *108*, 2058.
- ⁶² Toamasz, M.; Lipman, R.; Chowdary, D.; Pawlak, J.; Verdine, G. L.; Nakanishi, K. *Science* **1986**, *235*, 1204.
- ⁶³ Toamasz, M.; Chawla, A. K.; Lipman, R. *Biochemistry* **1988**, *27*, 3182.
- ⁶⁴ Teng, S.; Woodson, S. A.; Crothers, D. M. *Biochemistry* **1989**, *28*, 3901.
- ⁶⁵ Arora, S. K.; Cox, M. B.; Arjunan, P. *J Med Chem* **1990**, *33*, 3000.
- ⁶⁶ Kunz, K. R.; Iyengar, B. S.; Dorr, R. T.; Alberts, D. S.; Remers, W. *J Med Chem* **1991**, *34*, 2281.
- ⁶⁷ Maliepaard et al *J Med Chem* **1993**, *36*, 2091.
- ⁶⁸ Zhou, R.; Skibo, B. *J Med Chem* **1996**, *39*, 4321.
- ⁶⁹ Saunders et al *Biochemical Pharmacology* **2000**, *59*, 993
- ⁷⁰ Hodges, J. C.; Remmers, W. A. *J Med Chem* **1981**, *24*, 1184.

-
- ⁷¹ Orlemans, E. O. M.; *et al J Med Chem* **1989**, 32, 1612.
- ⁷² Cotteril, A. S.; *et al J Med Chem* **1994**, 37, 3834.
- ⁷³ Naylor, M. A.; *et al J Med Chem* **1998**, 41, 2720.
- ⁷⁴ Beall, H. D. *et al Bioorganic & Medicinal Chemistry Letters* **1998**, 8, 545.
- ⁷⁵ Phyllips, R. M.; *et al J Med Chem* **1999**, 42, 4071.
- ⁷⁶ Naylor, M. A. *J Med Chem* **1997**, 40, 2335.
- ⁷⁷ Schlötter, C.; Tautz, D. *Nucleic Acid Research* **1992**, 20, 2, 211.
- ⁷⁸ Wells, R. D. *The Journal of Biological Chemistry* **1996**, 271, 6, 2875.
- ⁷⁹ Baccola A.; Gellibolian, R.; Shimizu, M.; Amirhaeri, S.; Kang, S.; Ohshimat, K.; Larson, J. E.; Harvey, S. C.; Stollar, B. D.; Wells, R. D. *The Journal of Biological Chemistry* **1997**, 272, 27, 16783.
- ⁸⁰ Ratnaparkhe, M. B.; Tekeoglu, M.; Muehlbauer, F., J. *Theor Appl Genet* **1998**, 97, 515.
- ⁸¹ Kleiderlein, J. K.; Nisson, P. E.; Jessee, J.; Li, W.B.; Becker, K. G.; Derby, M. L.; Ross, C. A.; Margolis, R. L. *Human Genetic* **1998**, 103, 666.
- ⁸² Verkerk, A. J. M. H. *et al Cell* **1991**, 65, 905.
- ⁸³ Lin, B.; Nasir, J.; Kalchman, M. A.; McDonald, H.; Zeisler, J.; Goldberg, Y. P.; Hayden, M. R. *Genomics* **1995**, 25, 707.
- ⁸⁴ Rubinsztein, D. C. *et al American Journal of Human Genetic* **1996**, 59, 16.
- ⁸⁵ Virtaneva, K. *et al. Nature Genetics* **1997**, 15, 393.
- ⁸⁶ Inov, Y.; Peinado, M. A.; Malkhosyan, S.; Shibata, D., Perucho, M. *Nature* **1993**, 365, 558.
- ⁸⁷ Kleiderlein, J. K.; Nisson, P. E.; Jessee, J.; Li, W.B.; Becker, K. G.; Derby, M. L.; Ross, C. A.; Margolis, R. L. *Human Genetic* **1998**, 103, 666.
- ⁸⁸ Dokholyan, N. V.; Buldyrev, S. V.; Halvlin, S.; Stanley, H. E. *Physical Review Letters* **1997**, 79, 25, 5182.

-
- ⁸⁹ Mackiewicz, P.; Gierlik, A.; Kowalczyk, M.; Szczepanik, D.; Dudek, M. R.; Cebrat, S. *Physica A* **1999**, 103.
- ⁹⁰ Buldyrev, S. V.; Dokholyan, N. V.; Halvlin, S.; Stanley, H. E.; Stanley, R. H. R. *Physica A* **1999**, 273, 19.
- ⁹¹ Oiwa, N. N. *Physical Review Letters* **2000**, 85, 11, 2396.
- ⁹² Dokholyan, N. V.; Buldyrev, S. V.; Halvlin, S.; Stanley, H. E. *J. Theor. Biol.* **2000**, 202, 273.
- ⁹³ MacDonald, M. E.; et al *Cell* **1993**, 72, 971.
- ⁹⁴ Ambrose, C. M. et al *Somatic Cell and Molecular Genetics* **1994**, 20, 1, 27.
- ⁹⁵ Wells, R. D.; Parniewski, P.; Pluciennik, A.; Bacolla, A.; Gellibolian, R.; Jaworski, A. *The Journal of Biological Chemistry* **1998**, 273, 31, 19532.
- ⁹⁶ Brook, J. D. et al *Cell* 1992, 68, 799.
- ⁹⁷ Butterworth, N., J.; Williams, L.; Bullock, J. Y.; Love, D. R.; Faull, R. L. M.; Dragunow, M. *Neuroscience* **1998**, 87, 1, 49.
- ⁹⁸ Ji, J.; Clegg, N. J.; Peterson, K. R.; Jackson, A. L.; Laird, C. D.; Loeb, L. A. *Nucleic Acid Research* **1996**, 24, 14, 2835.
- ⁹⁹ Eichler, E. E. et al. *Nature Genetics* **1994**, 8, 88.
- ¹⁰⁰ Li, W. *Computational Chemistry* **1997**, 21, 848.
- ¹⁰¹ Ash, R. *Information Theory* **1995**, Dover, New York.