

UNIVERSIDADE ESTADUAL DE CAMPINAS



RAIMUNDO NONATO PEREIRA COSTA

“ESTUDO DA ATIVIDADE ANTINEOPLÁSICA DAS
CARBOQUINONAS ATRAVÉS DE DESCRITORES
QUÂNTICOS”

Tese apresentada ao Instituto de Física Gleb Wataghin para obtenção do
título de Mestre em Física.

Orientador: Prof. Dr. Douglas Soares Galvão

Campinas

2006

FICHA CATALOGRÁFICA

BANCA EXAMINADORA



MEMBROS DA COMISSÃO JULGADORA DA TESE DE MESTRADO DE **RAIMUNDO NONATO PEREIRA COSTA – RA 002362**, APRESENTADA E APROVADA AO INSTITUTO DE FÍSICA "GLEB WATAGHIN" DA UNIVERSIDADE ESTADUAL DE CAMPINAS, EM 23/02/2006.

COMISSÃO JULGADORA:

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

Prof. Dr. Varlei Rodrigues – DFA/IFGW/UNICAMP

Prof. Dr. Francisco Carlos Lavarda – Depto. Física/UNESP-Bauru

RESUMO

A descoberta de tratamentos eficazes contra o câncer é um dos maiores desafios da ciência. No Brasil o câncer se tornou a segunda maior causa de morte. As estimativas para o ano de 2005 apontam que ocorrerão 467.440 casos novos de câncer. O investimento para o desenvolvimento de uma droga moderna é de U\$850 milhões e o tratamento com estas variam de U\$4000 a U\$17000 por mês. Estudos de relação atividade-estrutura (SAR) quantitativa têm um importante papel no design de drogas. Neste trabalho foi aplicado a Metodologia de Índices Eletrônicos (MIE) para estudar a correlação da atividade antineoplásica de 36 moléculas derivadas das carboquinonas. A MIE é baseada nos conceitos de densidade local de estado, da qual se extrai o índice Δ , e na diferença de energia dos estados de fronteira, índice η , obtidos por cálculo de mecânica quântica ab initio ou semi-empírico. Os cálculos foram efetuados utilizando o método PM3 (Parametric Method 3) disponível no pacote computacional MOPAC. Os resultados demonstram que a MIE consegue diferenciar com uma precisão de 92% os compostos mais ativos biologicamente. A Análise de Principal Componente (PCA), utilizando-se os parâmetros da MIE, resultou em 100% de acerto na diferenciação da atividade biológica, enquanto com o método de análise hierárquica de grupo (HCA) obteve-se padrão de agrupamento com 94% de acerto. Concluimos que o uso dos índices Δ e η pode ser uma eficaz ferramenta para estudos de relação atividade-estrutura qualitativa das carboquinonas e para o desenvolvimento e aprimoramento de novas drogas.

ABSTRACT

The discovery of efficient treatments against cancer is one of the great challenges of science. In Brazil the cancer became the second greatest cause of death. For the year 2005 467,440 new cases of cancer are expected. The investment for the development of a modern drug is about U\$ 850 millions and the treatment with those drugs varies from U\$ 4,000 to U\$ 17,000 per month. Structure-Activity Relationship (SAR) studies play an important role in design of drugs. In this work the Methodology of Electronic Indices was applied (MEI) to the study of anticancer activity of 36 molecules derived from carboquinones. The MEI is based on the concepts of local density of states, from which the index Δ is extracted, and in the difference of energy of the frontier orbitals, which defines the other MEI parameter η . The geometrical and electronic aspects for the carboquinone set were obtained using the semi-empirical PM3 (Parametric Method 3) available in computational package MOPAC. Our results show that the MEI is able to classify (with an accuracy of 92%) active and inactive compounds. More standards statistical methods such Principal Component Analysis (PCA) and Hierarchic Cluster Analysis (HCA) were also used also producing results of identifying active/inactive molecules with high accuracy (100 and 94%, respectively).

We conclude that the use of the MEI indices Δ and η can be an efficient tool for studies of qualitative SAR studies of carboquinones, as well as, to be used for the development and improvement of new drugs.

SUMÁRIO

Resumo.....	IV
Abstract.....	V
Introdução.....	01
Objetivos.....	09
Métodos.....	11
Resolução da Estrutura Molecular.....	14
<i>Ab Initio</i>	14
Método Semi-empírico: Parametric Method 3.....	19
Propriedades Físico-Químicas para o SAR.....	19
Parâmetros Hidrofóbicos.....	20
Parâmetros Estéricos.....	21
Parâmetros Eletrônicos.....	21
Metodologia dos Índices Eletrônicos.....	25
Análise Quimiométrica.....	29
Análise Componente Principais.....	29
Análise Hierárquica de Grupo.....	31
Resultados.....	33
Resolução das Estruturas Geométricas e Análise Conformacional.....	34
Cálculo das Propriedades Moleculares para SAR.....	41
Cálculo da Energia e LDOS dos Orbitais de Fronteira.....	41
Cálculo dos Parâmetros Críticos Δ_c e η_c	45
Resultados da Análise Quimiométrica: PCA e HCA.....	48

Discussão e Conclusões.....	52
Apêndices.....	57
Métodos Semi-empíricos de Estrutura Eletrônica.....	57
CNDO.....	59
INDO.....	62
NDDO.....	63
MINDO.....	63
Modificado NDDO.....	63
MNDO.....	66
AM1.....	67
PM3.....	67
Referências.....	68

Figuras e Tabelas

Tabela 1. Descritores Topoestrutural, Topoquímico, Geométrico e Quântico.

Tabela 2. Atividade Antileucêmica: Dose Efetiva Média das Carboquinonas.

Tabela 3. Classificação Quanto a Atividade Antineoplásica.

Tabela 4. Calor de Formação das Estruturas Geométricas Resolvidas.

Tabela 6. Valores das Propriedades Físico-Químicas.

Tabela 7. Valores das Energias e LDOS dos Orbitais de Fronteira.

Tabela 8. Valores dos Índices Eletrônicos η e Δ .

Tabela 9. Variância das Componentes Principais PC1 e PC2.

Figura 1. Mecanismo de Ação das Drogas Antineoplásicas.

Figura 2. Mecanismo de Ação Citotóxica das DABQ.

Figura 3. Esquemática Representação da Intercalação das DABQ nas seqüências 5'TGC3' e 5'GNC3'.

Figura 4. Gráfico LDOS versus Energia: Índices da MIE η e Δ .

Figura 5. Análise Hierárquica de Grupo: Exemplo de Dendograma.

Figura 6. Estrutura Geométrica das Moléculas Resolvidas pelo método PM3.

Figura 7. Estrutura Geométrica das Moléculas Resolvidas por *Ab Initio*.

Figura 8. Região para Cálculo dos Parâmetros da MIE: Núcleo Benzeno.

Figura 9. Gráfico Parâmetro ΔL , ΔH e Moléculas.

Figura 10. Gráfico Parâmetro ηL , ηH e Moléculas.

Figura 11. Determinação dos Parâmetros Críticos.

Figura 12. Análise Componente Principais: Gráfico dos Loadings.

Figura 13. Análise Componente Principais: Gráfico dos Scores.

Figura 14. Dendograma da Análise Hierárquica de Grupo das 36 Moléculas.

Listas de Abreviaturas

AO	Orbital Atômico
AM1	Astin Method 1
CF	Calor de Formação
CNDO	Complete Neglect of Differential Overlap
CoMFA	Comparative Molecular Field Analysis
DOS	Density of States
DABQ	2,5-di(1-aziridinil)-p-benzoquinona
ENG	Eletronegatividade
EH	Energia de Hidratação
HOMO	Highest Occupied Molecular Orbital
HCA	Hiearquic Cluster Analysis
INDO	Intermediate Neglect of Differential Overlap
IRE	Integral de Repulsão Eletrônica
IT	Índice Topológico
ITe	Índice Topoestrutural
ITq	Índice Topoquímico
KNN	K th Nearest Neighbor
LUMO	Lowest Unoccupied Molecular Orbital
LDOS	Local Density of States
LCAO	Linear Combination Atomic Orbital
MIE	Metodologia Índices Eletrônicos

MINDO	Modified Intermediate Neglect of Differential Overlap
MLR	Multiple Linear Regression
MED	Dose Média Efetiva
MD	Momento de Dipolo Elétrico
MNDO	Modified Neglect of Diatomic Overlap
NDDO	Neglect of Diatomic Differential Overlap
OA	Orbital Atômico
PM3	Parametric Method 3
PC	Principal Component
PCA	Principal Component Analysis
PLS	Partial Least Squares
PI	Potencial de Ionização
PLZ	Polarizabilidade
QSAR	Quantitative Structure Analysis Relation
RHF	Hartree-Fock-Roothann
RFT	Refratividade
SAR	Structure Analysis Relation
SIMCA	Soft Independent Modeling of Class Analysis
ZDO	Zero Differential Overlap

II. INTRODUÇÃO

Metodologia de predição da atividade biológica de drogas antes que sejam sintetizadas são fundamentais no processo de desenvolvimento e aperfeiçoamento de novas drogas. O estudo de Relação Estrutura-Atividade (SAR) é uma bem estabelecida estratégia de descrição da atividade biológica de drogas. O início de SAR como uma ferramenta prática no design de drogas foi iniciada no início dos anos de 1960s com os trabalhos de Hansch, Free e Wilson[1, 2]. Deste então, vários métodos quimiométricos foram desenvolvidos para estudo de SAR, dentre eles estão: i) métodos de análise exploratória; Principal Analysis (PCA)[3-6] e Hierarchical Cluster Analysis (HCA)[3, 6] ii) os métodos de regressão; Multiple Linear Regression (MLR)[4], Principal Component Regression (PCR)[4] e Partial Least Squares (PLS)[7] iii) método de reconhecimento de padrão Kth Nearest Neighbor (KNN)[7] e Soft Independent Modeling of Class Analysis (SIMCA)[4] iv) Comparative Molecular Field Analysis (CoMFA)[8]. No processo de desenvolvimento de um fármaco se produz (real ou virtualmente) uma grande quantidade de moléculas para testes. Para muitas dessas moléculas não se têm dados de suas propriedades físico-químicas, portanto, é desejável o uso de parâmetros que possam ser calculados diretamente da estrutura molecular. Dentre esses parâmetros estão os índices topológicos, os diversos índices de forma e tamanho molecular, e os parâmetros quânticos.

Numerosos estudos de SAR foram publicados com diversas classes de drogas com parcial sucesso na predição da atividade biológica. Diversas propriedades físico-química, estruturais e conformacionais foram propostas para estudo SAR. Dentre esses estão[9]: i) índices topológicos: índice Wiener; índice de conectividade (de Kier e Hall, Randić), índices de informação teórica, definida a partir das matrizes de grafos;

parâmetros derivados da complexidade vizinha dos vértices em grafos moleculares preenchidos com hidrogênio; índices J de Balaban; ii) Índices Geométricos: volume de van der Waals, número 3D Winner; iii) parâmetros químico-quânticos: energia do mais alto orbital ocupado E_{HOMO} , energia do segundo mais alto orbital ocupado $E_{\text{HOMO}-1}$, energia do mais baixo orbital não ocupado E_{LUMO} , energia do segundo mais baixo orbital não ocupado $E_{\text{LUMO}+1}$, calor de formação. A tabela 1 apresenta uma breve definição de alguns índices utilizados nos estudos de relação estrutura-atividade quantitativo (QSAR). Os índices topológicos (IT) estão divididos em índices topoestruturais (ITe) e topoquímicos (ITq). Os índices Topoestruturais são índices que relacionam a distância dos átomos (vértices) na estrutura molecular (grafo) sem levar em conta a natureza química dos átomos envolvidos na ligação, seus estados de hibridização e o número de valência de cada átomo. Índices Topoquímicos são índices que consideram, além da topologia (conectividade dos átomos), as propriedades dos átomos.

O presente estudo pretende contribuir com uma metodologia (Metodologia de Índices Eletrônicos) de SAR que utiliza parâmetros de química quântica computacional obtidos diretamente a partir da estrutura molecular. A metodologia de Índices Eletrônicos (MIE) é baseada no conceito de densidade local de estados (LDOS), índice Δ , e na diferença de energia dos estados de fronteira, índice η . Esses descritores foram inicialmente propostos para estudo SAR qualitativo de um conjunto de hidrocarbonetos policíclicos aromáticos com propriedade carcinogênica[10]. Desde então, a MIE foi aplicada ao estudo de uma variedade classes de drogas obtendo-se sucesso na predição da sua atividade biológica[11]. O conceito da LDOS pode dar informação da contribuição de

uma região geométrica da molécula para a reatividade química e assim sobre sua atividade biológica.

Tab.1: Descritores Topoestruturais, Topoquímicos, Geométricos e Quânticos[9].

Símbolo	Índices Topoestruturais
I_D^W	Índice informação para distância entre todos pares dos vértices do grafo
$\bar{I}_D^W$	Índice de informação média da medida das distâncias
W	Índice de Wiener
I^D	Grau de complexidade
H^V	Complexidade do vértice do grafo
H^D	Complexidade das distâncias do grafo
IC	Conteúdo informação(CI) das distâncias da matriz particionada pela frequência das ocorrências de distância h
O	Ordem do vizinho quando IC_r encontra seu máximo valor para o grafo preenchido com Hidrogênio
M₁	Parâmetro grupo Zagreb= soma dos quadrados do grau de todos os vértices
M₂	Parâmetro grupo Zagreb= soma dos produtos dos grau de todos os vértices vizinhos
$^h\chi$	Índice de conectividade do caminho de ordem h=0-6
$^h\chi_C$	Índice de conectividade de grupo de ordem h=3-6
$^h\chi_{Ch}$	Índice de conectividade de cadeia de ordem h=3-6
$^h\chi_{PC}$	Índice de conectividade grupo-caminho de ordem h=4-6
P_h	Números de caminho de tamanho h=0-10
J	Índice J de Balaban
	Índices Topoquímicos
I_{ORB}	CI ou complexidade do grafo sem hidrogênio ao máximo vizinho do vértice
IC_r	Média CI ou complexidade de um grafo baseado r-ésima (r=0-6) ordem vizinho do vértice em um grafo sem hidrogênio.
CIC_r e SIC_r	CI complementar e estrutural para o r-ésimo (r=0-6) vértice vizinho em um grafo sem hidrogênio
$^h\chi^b, ^h\chi^v$	Índices de conectividade de caminho (ligação e valência) de ordem h=0-6
$^h\chi_C^b, ^h\chi_C^v$	Índices de conectividade de grupo (ligação e valência) de ordem h=3-6
$^h\chi_{Ch}^b, ^h\chi_{Ch}^v$	Índices de conectividade de cadeia (ligação e valência) de ordem h=3-6
$^h\chi_{PC}^b, ^h\chi_{PC}^v$	Índices de conectividade de grupo-caminho (ligação e valência) de ordem h=4-6

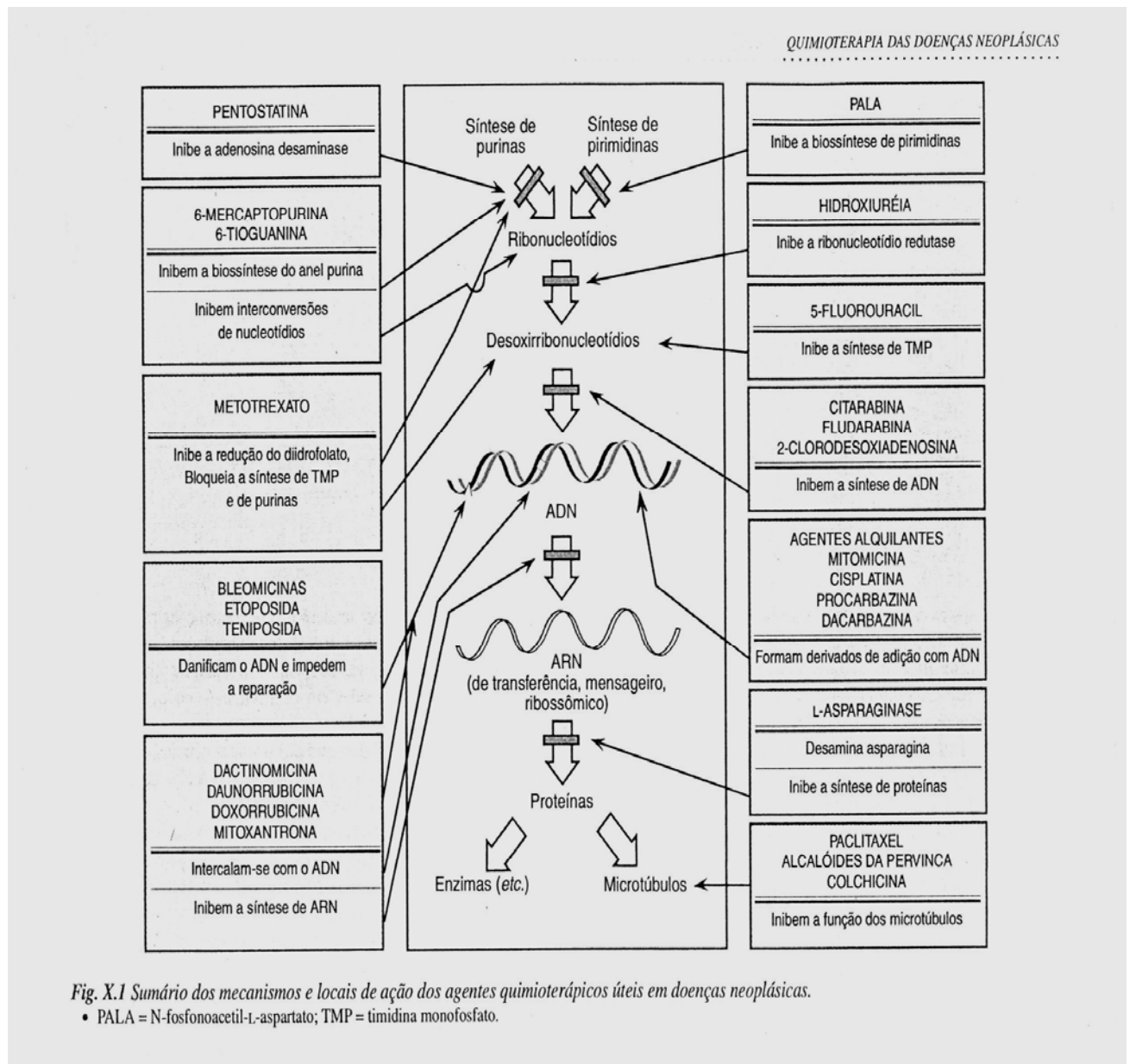
	Índices Geométricos
V_W	Volume de van der Waals
${}^{3D}W, {}^{3D}W_H$	3D número de Wiener para matriz distância geométrica sem e com hidrogênio
	Índices Quânticos
E_{HOMO}	Energia do último orbital molecular ocupado
E_{HOMO-1}	Energia do penúltimo orbital molecular ocupado
E_{LUMO}	Energia do primeiro orbital molecular não ocupado
E_{LUMO+1}	Energia do segundo orbital molecular não ocupado
ΔH_f	Calor de formação
M	Momento de dipolo

Dentre todos os avanços da medicina, o tratamento do câncer tem exercido o maior impacto na transformação da prática médica neste início do século, em razão das descobertas de tratamentos curativos para várias malignidades anteriormente fatais. As abordagens básicas ao tratamento do câncer estão evoluindo constantemente com a exploração de terapias genéticas, manipulações do sistema imune, induções de diferenciação em tecido tumorais e inibição da angiogênese. As drogas empregadas na quimioterapia incluem uma grande variedade de compostos que atuam por vários mecanismos, figura 1. Apesar dos esforços de pesquisadores farmacêuticos no sentido de desenvolvimento de drogas com ação mais seletiva sobre os tecidos neoplásicos, os agentes atualmente disponíveis possuem um índice terapêutico muito estreito e por isso manifestam toxicidade significativa sobre os tecidos normais.

No presente estudo, a atividade antileucêmica de 36 derivados da 2,5-di(1-aziridinil)-p-benzoquinona (DABQ) foi relacionada com os parâmetros da MIE. As carboquinonas foram sintetizadas como modificação da estrutura da Mitomicina C (droga amplamente usada em protocolos de tratamentos antineoplásico) para a procura de um

agente mais potente e menos tóxica. Seus derivados foram avaliados contra vários tumores, e muitos dados biológicos destas drogas foram publicados[11].

Fig.1: Mecanismo de Ação das Drogas Antineoplásicas[12].

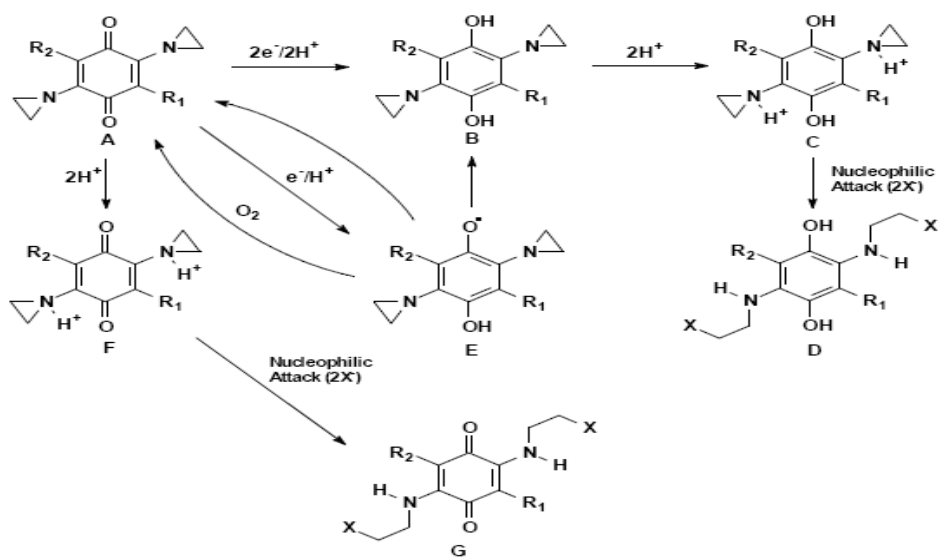


O principal mecanismo pelo quais as drogas antineoplásicas destroem as células cancerígenas envolve a interferência na replicação celular. O modo mais fácil para conseguir este efeito é via alquilação do DNA. A Alquilação invariavelmente ocorre sobre as bases guanina e adenosina. A posição N7 da guanina é o sítio preferido de alquilação, por ser a região mais negativa dentre todas as bases componentes do DNA. A alquilação pode acarretar ao DNA falso pareamento das bases, ligação inter-filamentos, intra-filamentos, inter-hélices ou entre DNA-proteínas.

O mecanismo de ação pelo qual as DABQ lesiona o DNA é mostrado na figura 2 [13]. Os DABQ em solução aquosa podem cruzar o DNA na ausência de redução e este processo é dependente do pH. A eficiência com que as DABQ intercalam no DNA depende da reatividade do segundo grupo aziridinil após o primeiro grupo aziridinil ter alquilado um sítio do DNA.

Fig.2: Mecanismo de Ação Citotóxica das DABQ[13].

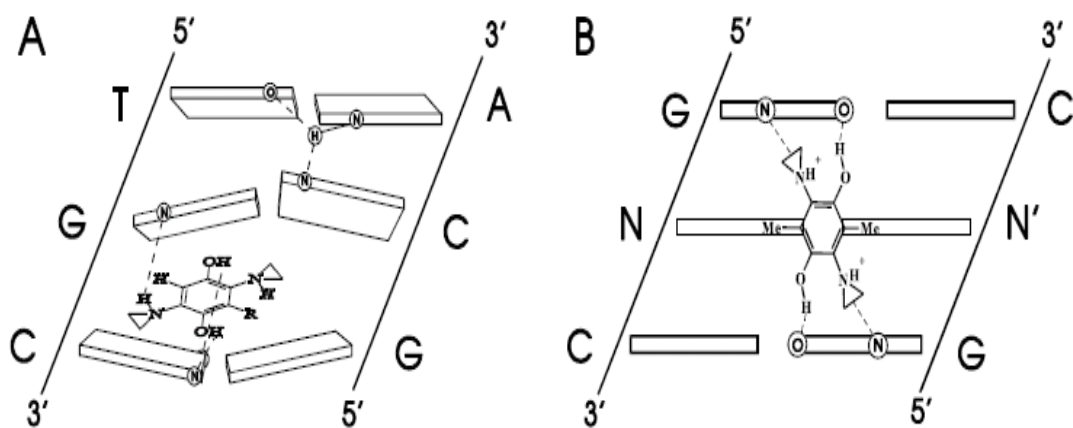
DNA Alkylation of aziridinylquinones



Este é um processo favorável, pois o primeiro grupo aziridinil forma um grupo amina alifático na alquilação. A resultante doação de elétron desta amina para dentro do anel diminui o pK do outro grupo aziridinil. Desta forma o segundo grupo aziridinil terá um menor pK e portanto deve reagir mais facilmente com as bases nitrogenadas. A alquilação pelas DABQ é primariamente no sítio N-7 guanina. Essencialmente a DABQ inicialmente intercala na sequência 5'TGC3' ou 5'GNC3' e o grupo aziridinil alquila a posição N7 na guanina, figura 3 [13].

Fig.3: Esquemática Representação da Intercalação das DABQ e as sequências 5'TGC3' e 5'GNC3'[13].

DNA Alkylation of aziridinylquinones



III. OBJETIVOS

Pretende-se realizar um estudo qualitativo estrutura-atividade com 36 compostos derivados das carboquinonas, através da aplicação da Metodologia dos Índices Eletrônicos. Assim, queremos mostrar que a atividade biológica das drogas pode ser diferenciada através da utilização de apenas dos dois parâmetros, Δ e η , da MIE. A análise estrutural para cálculo dos parâmetros da MIE e dos parâmetros físico-químicos foi realizada utilizando o método semiempírico PM3 (Parametric Method 3)[14-16]. Para validação dos resultados obtidos pela MIE foi realizada a análise estatística com os métodos de análise multivariada de Análise Componente Principal [3, 17] e Análise Hierárquica de Grupo[3, 17] que são dois métodos de quimiometria amplamente empregados para reconhecimento de padrões.

IV. MÉTODOS

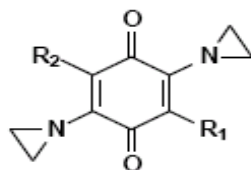
A atividade biológica de 36 derivados da carboquinona foi obtida a partir da avaliação realizada por Yoshimoto et al, contra leucemia linfóide L-210 em camundongos[18]. A dose efetiva mínima (MED) foi utilizada como medida experimental da atividade antineoplásica. A MED é a dose mínima necessária para aumentar em pelo menos em 40% o tempo de vida do animal. Os valores são mostrados na tabela 2, na forma de pMED ($-\log[\text{mol/kg}]$).

Em seguida foram realizados para todas as moléculas os cálculos necessários para a resolução da geometria molecular, cálculo dos índices da MIE e dos descritores das propriedades físico-químicas para estudo de quimiometria, através dos métodos de análise multivariada, PCA e HCA.

Método semi-empírico, Parametric Method 3 (PM3) foi aplicado para resolução da estrutura eletrônica, pois tem se mostrado uma metodologia adequada para tratar compostos orgânicos, além do que, requer menor custo computacional do que métodos *Ab Initio*. A procura conformacional é realizada variando os diedrais para a procura da conformação de menor energia. O método de Monte Carlo, contido no pacote computacional Spartan[19] (www.wavefun.com), foi empregado para varrer o espaço dos conformeros. A partir da determinação da conformação mais estável realizam-se os cálculos das propriedades físico-químicas e dos parâmetros da MIE. Chem2Pac[20] é um software desenvolvido pelo grupo GSONM do Instituto de Física da Unicamp, que permite obter os índices da MIE[20]. Algumas das propriedades físicas químicas foram obtidas pela aplicação QSAR contido no pacote computacional HyperChem[21].

A análise quimiométrica foi realizada com o software EinSight[22] com as propriedades físico químicas calculadas, incluindo-se os parâmetros da MIE.

Tab.2: Atividade Antileucêmica e Dose Efetiva Média das DABQ[18].



Moléculas	R1	R2	pMED
01	C ₆ H ₅	C ₆ H ₅	4.33
02	CH ₃	(CH ₂) ₃ C ₆ H ₅	4.47
03	C ₅ H ₁₁	C ₅ H ₁₁	4.63
04	CH(CH ₃) ₂	CH(CH ₃) ₂	4.77
05	CH ₃	CH ₂ C ₆ H ₅	4.85
06	C ₃ H ₇	C ₃ H ₇	4.92
07	CH ₃	CH ₂ OC ₆ H ₅	5.15
08	CH ₂ OCON(CH ₃) ₂	CH ₂ OCON(CH ₃) ₂	5.16
09	C ₂ H ₅	C ₂ H ₅	5.46
10	CH ₃	(CH ₂) ₂ OCH ₃	5.57
11	CH ₃ O	CH ₃ O	5.59
12	CH ₃	CH(CH ₃) ₂	5.60
13	C ₃ H ₇	CH(OCH ₃)CH ₂ OCONH ₂	5.63
14	CH ₃	CH ₃	5.66
15	H	CH(CH ₃) ₂	5.68
16	CH ₃	CH(OCH ₃)C ₂ H ₅	5.68
17	C ₃ H ₇	CH ₂ CH ₂ OCONH ₂	5.68
18	CH ₃ OC ₂ H ₄	CH ₃ OC ₂ H ₄	5.69
19	C ₂ H ₅	(CH ₃ O)CH ₂ OCONH ₂	5.76
20	CH ₃	(CH ₂) ₂ OCONH ₂	5.78
22	CH ₃	C ₂ H ₅	5.86
23	CH ₃	CH(OC ₂ HSOCOCH ₃)CH ₂ OCONH ₂	6.03
24	CH ₃	CH ₂ CH(CH ₃)OCONH ₂	6.14
25	C ₂ H ₅	CH(CH ₂ OCONH ₂)OCH ₃	6.16
26	CH ₃	CH(CH ₂ CH ₃)CH ₂ OCONH ₂	6.18
27	CH ₃	CH(OC ₂ H ₅)CH ₂ OCONH ₂	6.18
28	CH ₃	(CH ₂) ₃ OCONH ₂	6.18
29	CH ₃	(CH ₂) ₂ OCONH ₂	6.21
30	C ₂ H ₅	(CH ₂) ₂ OCONH ₂	6.25
31	CH ₃	C ₂ H ₅ OH	6.39
32	CH ₃	CH(CH ₃)CH ₂ OCONH ₂	6.41
33	CH ₃	CH(OCH ₃)CH ₂ OCONH ₂	6.41
34	H	N(CH ₂) ₂	6.45
35	(CH ₂) ₂ OH	(CH ₂) ₂ OH	6.54
36	CH ₃	N(CH ₂) ₂	6.77
37	CH ₃	CH(OCH ₃)CH ₂ OH	6.90

1. Resolução da Estrutura Molecular

Para resolução da geometria molecular dos 36 derivados da tabela 2, foi utilizado o método semi-empírico PM3[15, 16] em virtude dos bons resultados com respeito a medidas de distância da ligação, ângulo de ligação, ângulo diedro e calor de formação (CF). Os cálculos de otimização geométrica e obtenção das densidades local e total de estados foram realizados usando o programa Chem2Pac[20] que contém uma série de programas de uso público (MOPAC[23], BABEL[23], RASMOL[23], etc). Este pacote de programas permite realizar a otimização molecular e os cálculos de LDOS. Uma varredura conformacional sistemática variando o ângulo de torção foi realizada usando o software Spartan[®] [19]. A seguir descrevemos de forma sucinta as bases físicas do método de otimização da geometria molecular.

1.1 Método Geometria Molecular: *Ab Initio*

O Hamiltoniano PM3 é um método Hartree-Fock baseado na aproximação de combinação linear de orbitais atômicos (LCAO)[24].

Para um sistema molecular o operador hamiltoniano total H_t é escrito como a soma das energias cinéticas dos núcleos e elétrons[25-28]:

$$H_t = T_e + T_n + V_{ne} + V_{ee} + V_{nn} \quad (\text{Eq.1})$$

Onde: T_e energia cinética dos elétrons.

T_n energia cinética dos núcleos.

V_{ne} energia potencial entre elétrons e núcleos.

V_{ee} energia potencial entre os elétrons.

V_{nn} energia potencial entre os núcleos.

Para a maioria dos sistemas a aproximação de Born-Oppenheimer (BO) introduz um erro pequeno e o problema reduz a equação Schrodinger eletrônica para um conjunto de geometria nuclear[27]. O operador eletrônico, $\mathbf{H}_e$ é dado por:

$$\mathbf{H}_e = \mathbf{T}_e + \mathbf{V}_{ne} + \mathbf{V}_{ee} + \mathbf{V}_{nn} \quad (\text{Eq.2})$$

Onde:
$$\mathbf{T}_e = -\sum_i^N \frac{1}{2} \nabla_i^2$$

$$\mathbf{V}_{ne} = -\sum_i^N \sum_a \frac{Z_a}{|R_a - r_i|}$$

$$\mathbf{V}_{ee} = \sum_i^N \sum_{j>i}^N \frac{1}{|r_i - r_j|}$$

$$\mathbf{V}_{nn} = \sum_a \sum_b \frac{Z_a Z_b}{|R_a - R_b|}$$

Reescrevendo de acordo com o número de elétrons

$$\mathbf{H}_e = \sum_{i=1}^N \mathbf{h}_i + \sum_{i=1}^N \sum_{j>i}^N \mathbf{g}_{ij} + \mathbf{V}_{nn} \quad (\text{Eq.3})$$

$$\mathbf{h}_i = -\frac{1}{2} \nabla_i^2 - \sum_a \frac{Z_a}{|R_a - r_i|} \quad \text{e} \quad \mathbf{g}_{ij} = \frac{1}{|r_i - r_j|}$$

O operador de um elétron $\mathbf{h}_i$ descreve o movimento do elétron i no campo de todos os núcleos, e $\mathbf{g}_{ij}$ é o operador de repulsão entre dois elétrons.

Resolvendo a equação de Schrodinger

$$E = \langle \Phi | \mathbf{H}_e | \Phi \rangle \quad (\text{Eq.4})$$

Obtemos

$$\mathcal{E} = \sum_i^N h_{ii} + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (J_{ij} - K_{ij}) + \mathbf{V}_{nn} \quad (\text{Eq.5})$$

Onde $\mathbf{J}_{ij}$ e $\mathbf{K}_{ij}$ são os operadores de Troca e Coulomb respectivamente:

$$J_{ij} = \iint \psi_i^*(1) \psi_j^*(2) r_{12}^{-1} \psi_i(1) \psi_j(2) dv_1 dv_2 \quad (\text{Eq.6})$$

$$K_{ij} = \iint \psi_i^*(1) \psi_j^*(2) r_{12}^{-1} \psi_i(2) \psi_j(1) dv_1 dv_2 \quad (\text{Eq.7})$$

Reescrevendo a equação em termos dos operadores de $\mathbf{J}$ e $\mathbf{K}$

$$E = \sum_{i=1}^N \langle \phi_i | h_i | \phi_i \rangle + \frac{1}{2} \sum_{ij}^N \left(\langle \phi_j | J_i | \phi_j \rangle - \langle \phi_j | K_i | \phi_j \rangle \right) + V_{nn} \quad (\text{Eq.8})$$

O objetivo agora é determinar um conjunto de orbitais moleculares que minimiza a energia, com a condição de que esses orbitais sejam ortogonais e normalizados.

Aplicando o método dos multiplicadores de Lagrange.

$$L = E - \sum_{ij} \lambda_{ij} \left(\langle \phi_i | \phi_j \rangle - \delta_{ij} \right)$$

$$\delta L = \delta E - \sum_{ij} \lambda_{ij} \left(\langle \delta \phi_i | \phi_j \rangle + \langle \phi_i | \delta \phi_j \rangle \right) = 0 \quad (\text{Eq.9})$$

Resolvendo, encontra-se como solução para um ponto estacionário escrito em termo do operador de Fock, $\mathbf{F}_i$, equação 10.

$$E = \sum_i^N \varepsilon_i - \frac{1}{2} \sum_{ij}^N (J_{ij} - K_{ij}) + V_{nn} \quad (\text{Eq.10})$$

$$\varepsilon_i = \left\langle \phi_i \left| F_i \right| \phi_i \right\rangle = h_i + \sum_j^N (J_{ij} - K_{ij}) \quad (\text{Eq.11})$$

$$F_i = h_i + \sum_j^N (J_j - K_j) \quad (\text{Eq.12})$$

O operador de Fock é um efetivo operador de um elétron descrevendo a energia cinética de um elétron, a atração de todos os núcleos e a repulsão a todos os outros elétrons.

Roothaan em 1951 propôs expandir os orbitais moleculares, ϕ_i , como uma combinação linear de orbital atômico (LCAO) χ_α .

$$\phi_i = \sum_{\alpha}^M c_{\alpha i} \chi_{\alpha} \quad (\text{Eq.13})$$

Onde o χ_α são funções atômicas reais. Para representar exatamente os orbitais moleculares, ϕ_i , as funções de base χ_α , deve formar um conjunto completo, o que requer um número infinito de funções de base. Na prática usamos um número finito, M, de funções de base.

Mantendo a condição de que os orbitais moleculares sejam ortonormal, equação 14.

$$\sum_{\alpha\beta}^M c_{\alpha i}^* c_{\beta j} S_{\alpha\beta} = \delta_{ij} \quad (\text{Eq.14})$$

sendo δ_{ij} o delta de Kronecker e S_{ij} é a matriz de sobreposição para as funções atômicas χ_α e χ_ν , equação 15.

$$S_{\alpha\beta} = \int \chi_{\alpha}^* \chi_{\beta} dr = \langle \chi_{\alpha} | \chi_{\beta} \rangle \quad (\text{Eq.15})$$

E a equação de Hartree-Fock é escrita como:

$$F_i \sum_{\alpha}^M c_{\alpha i} \chi_{\alpha} = \epsilon_i \sum_{\alpha}^M c_{\alpha i} \chi_{\alpha} \quad (\text{Eq.16})$$

Todas as M equações podem ser mais bem escritas na forma matricial como

$$FC = SC\epsilon \quad (\text{Eq.18})$$

$$F_{\alpha\beta} = \langle \chi_{\alpha} | F | \chi_{\beta} \rangle \quad (\text{Eq.19})$$

Resolvendo a equação 19, encontramos:

$$F_{\alpha\beta} = \langle \chi_{\alpha} | h | \chi_{\beta} \rangle + \sum_j^{OM} \langle \chi_{\alpha} | J_j - K_j | \chi_{\beta} \rangle$$

Escrita em termos da matriz densidade $D_{\gamma\delta}$ e das integrais de dois elétrons:

$$F_{\alpha\beta} = \langle \chi_\alpha | h | \chi_\beta \rangle + \sum_{\gamma} \sum_{\delta}^{OA} D_{\gamma\delta} \left(\langle \chi_\alpha \chi_\gamma | g | \chi_\beta \chi_\delta \rangle - \langle \chi_\alpha \chi_\gamma | g | \chi_\delta \chi_\beta \rangle \right) \quad (\text{Eq.20})$$

Onde a matriz densidade é

$$D_{\gamma\delta} = \sum_j^{OM} c_{\gamma j} c_{\delta j} \quad (\text{Eq.21})$$

Escrevendo de forma mais compacta

$$F_{\alpha\beta} = k_{\alpha\beta} + \sum_{\gamma\delta} G_{\alpha\beta\gamma\delta} D_{\gamma\delta} \quad (\text{Eq.22})$$

$$\mathbf{F} = \mathbf{h} + \mathbf{G} \cdot \mathbf{D} \quad (\text{Eq.23})$$

Finalmente, a energia total será dada por:

$$E = \sum_{\alpha\beta}^M D_{\alpha\beta} h_{\alpha\beta} + \frac{1}{2} \sum_{\alpha\beta\gamma\delta}^M (D_{\alpha\beta} D_{\gamma\delta} - D_{\alpha\delta} D_{\beta\gamma}) \langle \chi_\alpha \chi_\gamma | g | \chi_\beta \chi_\delta \rangle + V_{nn} \quad (\text{Eq.24})$$

Para determinar os coeficientes $c_{\alpha i}$, a matriz de Fock deve ser diagonalizada. Contudo, a matriz de Fock é somente conhecida se os coeficientes são conhecidos. Portanto, as equações de Hartree-Fock-Roothaan (RHF) devem ser resolvidas por um processo iterativo. O processo começa com um adequado valor dos coeficientes, forma-se a matriz de Fock, diagonaliza e forma-se nova matriz densidade. O processo continua até que a cada interação não varie os coeficientes ou a energia dentro de determinado limite de precisão. Resolvendo as equações RHF obtém-se M (número de funções de base) orbitais moleculares com N (número de elétrons) orbitais ocupados. Os orbitais virtuais ou não ocupados M-N, não tem interpretação física direta. Para M funções de base existem cerca de M^2 integrais de um elétron e M^4 integrais de dois elétrons.

1.2 Métodos Semiempírico: Parametric Method 3 (PM3)[15, 16]

Métodos semi-empíricos podem reduzir o número de integrais para M^1 . Os principais métodos semi-empíricos envolvem dois tipos de aproximações. Na primeira aproximação os elementos de matriz da hamiltoniana efetiva de 1-elétron são dados por valores empíricos ou semi-empíricos, parametrizados, que tentam correlacionar o resultado dos cálculos aos dados experimentais existentes. Na segunda 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.

O método PM3 foi desenvolvido em 1989 por Stewart[15, 16] como uma reparametrização do Austin Method 1 (AM1)[29]. Os métodos MNDO, AM1 e PM3 são derivados da aproximação Neglect of Diatomic Differential Overlap (NDDO), e diferem somente como a repulsão núcleo-núcleo é tratada e como os parâmetros são obtidos.

O método PM3 envolve 15 parâmetros por átomo: expoentes dos orbitais ζ_s e ζ_p , termos da integral de um elétron, U_s , U_p , β_s e β_p ; termos das integrais de dois elétrons, G_{ss} , G_{sp} , G_{pp} , G_{p2} , H_{sp} ; parâmetros da repulsão núcleo-núcleo α , a , b e c . Detalhes dos métodos semi-empíricos encontram-se no Apêndice A1.

2. Propriedades Físico Química no SAR

A resposta biológica de uma droga ao organismo é relacionada às propriedades lipofílicas, estéricas e eletrônicas. Essas propriedades físico-químicas como descritores moleculares permitem obter informação sobre a relação estrutura-atividade da droga.

2.1 Parâmetros Hidrofóbicos

A afinidade relativa de uma droga entre o meio aquoso e lipídico é uma importante propriedade correlacionada com a atividade da droga devido à direta relação com os fenômenos de transporte, absorção, distribuição e metabolismo. O principal descritor que caracteriza esta propriedade é o coeficiente de partição[4, 30] entre água e 1-octanol.

O Coeficiente de Partição (P) é definido como a relação entre as concentrações do soluto no meio 1-octanol $[C]_{1-octanol}$ e no meio aquoso $[C]_{água}$.

$$P = \frac{[C]_{1-octanol}}{[C]_{água}} \quad (\text{Eq.25})$$

Compostos com $P > 1$ são considerados lipofílicos e com $P < 1$ são hidrofílicos. Medida experimental de Log P é realizado através da técnica de cromatografia líquida (HPLC)[30]. Há diversos métodos de relação estrutura atividade quantitativa (QSAR) para cálculo do Log P a partir de índices topológicos. O método de aditividade de fragmento desenvolvido por Leo e Hansch é amplamente usado para calcular o Log P com a seguinte expressão[30].

$$\text{Log}P = \sum_i a_i f_i + \sum_j b_j F_j \quad (\text{Eq.26})$$

Onde f_i são constantes dos fragmentos e F_j são os fatores de correção. A molécula é quebrada em uma série de fragmentos respeitando um conjunto de regras. Este método faz parte de diversos pacotes de software de química molecular. Neste estudo foi usado o pacote computacional HyperChem®[21].

2.2. Parâmetros Estéricos

A forma ou a configuração estérica de uma molécula tem um profundo efeito sobre as propriedades físico-químicas e da atividade biológica. Efeitos estéricos são influenciados por dois fenômenos químicos. O primeiro, chamado de efeito substituinte, advém de efeitos de grupos de átomos sobre a reatividade de uma parte da molécula. Este efeito geralmente se manifesta como uma influência repulsiva ao ataque do reagente ao sítio reativo. O segundo são as interações moleculares, onde certo grau de complementaridade deva ser necessário para que a reação ocorra[4, 31]. O reconhecimento entre moléculas é governado principalmente pela forma de cada uma dela.

Muitos parâmetros estéricos usam o fato de que os átomos podem ser representados como objetos geométricos tendo uma superfície e volume. O raio de van der Waals é considerado o limite mais externo de um átomo. A partir deste, a área superficial e o volume molecular podem ser estimado e usado para inferir efeitos estéricos[31, 32].

2.3 Parâmetros Eletrônicos

Muitos descritores eletrônicos podem descrever os efeitos ou as forças das interações moleculares. As forças intermoleculares geralmente originam-se das seguintes interações: íon-íon, íon-dipolo, dipolo-dipolo, dipolo-dipolo induzido, pontes de hidrogênio, dispersão[4].

A interação iônica, representada pela constante de ionização (K_i), é usada nos estudos de determinação de potência de drogas. Como descritor a K_i oferece informação

sobre o grau com que uma molécula ioniza e isso relaciona-se fortemente com a absorção e distribuição da droga no organismo[33].

O momento de dipolo elétrico (MD) representa a força de interação do tipo polar e estudos mostram o uso deste como descritor em QSAR[34]. Cálculo a partir da mecânica quântica estima o MD com boa precisão[15].

2.3.1 Interação Eletrostática

A atração inicial entre a droga e seu sítio ativo é quase totalmente do tipo eletrostática. A força atrativa entre as cargas opostas forma três tipos de ligações não covalente: interação carga-carga (ou iônica), carga-dipolo e dipolo-dipolo[4, 17].

A energia potencial de qualquer interação eletrostática é dada por

$$E \propto q_i q_j / D \cdot r_{ij} \quad (\text{Eq.27})$$

Onde q_i e q_j são as cargas separadas pela distância r_{ij} , e D é a constante dielétrica do meio. Para a interação carga-dipolo a equação é [4, 17]

$$E \propto z \mu \cos \varphi / D r^2 \quad (\text{Eq.28})$$

onde z é valência do íon, μ é o momento de dipolo, r é a distância entre o íon e o centro do momento de dipolo e φ é o ângulo entre o dipolo e a linha que une o íon ao centro do dipolo. Para a interação dipolo-dipolo a equação torna [4, 17]:

$$E = A \mu_A \mu_B / D r^3 \quad (\text{Eq.29})$$

onde A é um fator geométrico dependente da posição dos dipolos.

A interação iônica pode ser descrita através do constante de ionização (K_i). A constante K_i , prover informação sobre a extensão com que a droga se ioniza, que, por sua vez, está fortemente relacionada com a absorção e distribuição da droga.

2.3.2 Interação Indutiva

A formação do complexo droga-receptor é geralmente acompanhada por alguma redistribuição de carga. Polarização induzida ocorre quando um dipolo temporário é induzido pelo campo de um íon ou dipolo. A contribuição da polarização para a energia total é dada pela equação 30 e 31 sendo α a polarizabilidade [4, 17].

$$E = \alpha z^2 / D^2 r^4 \quad \rightarrow \text{interação íon-dipolo induzido} \quad (\text{Eq.30})$$

$$E = A \alpha \mu / D^2 r^6 \quad \rightarrow \text{interação dipolo induzido} \quad (\text{Eq.31})$$

Polarizabilidade e Refratividade molecular são propriedades que medem a susceptibilidade da molécula a tornar-se polarizada. Esses descritores são importantes quando interações do tipo dipolo-dipolo induzido estão envolvidas na ligação. Eles são calculados a partir do índice de refração e do volume molar, empiricamente são estimados por técnicas de relação estrutura-propriedade quantitativas (QSPR)[35-37].

2.3.3. Pontes de Hidrogênio

É uma interação de médio alcance entre um átomo de hidrogênio deficiente em elétron e uma região vizinha com alta densidade eletrônica [38]. Quando fortes pontes de hidrogênio são formadas, a distância entre o átomo doador e o receptor é menor do que a soma de seus raios de van der Waals. As pontes de hidrogênio têm geralmente 1/10 da força da ligação covalente. Devido à especificidade atômica, a curta extensão e ser direcionada, as pontes de hidrogênio é um importante fator determinante da especificidade da interação intermolecular, e por isso é reconhecidamente um importante fator para as propriedades físicas e atividades biológicas das moléculas[38]. No meio celular as mais importantes pontes de hidrogênio são aquelas envolvendo o oxigênio e

nitrogênio dos grupos carboxil, hidroxil, carbonil, amino, imino e amida, que são responsáveis por manter a estrutura terciária das proteínas e ácido nucléicos. A interação física responsável pela força da ponte de hidrogênio advém das componentes de interação eletrostática atrativa, transferência de carga, polarização e dispersão e de troca de elétrons repulsivos.

2.3.4. Densidade Eletrônica de Fronteira

Índices de reatividade são classificados de acordo com o tipo de ataque como eletrofílico ou nucleofílico. Os descritores mais simples dessas propriedades são as energias do último orbital molecular ocupado (E_{Homo}) e a energia do primeiro orbital molecular não ocupado (E_{Lumo})[10]. A E_{HOMO} está relacionada ao potencial de ionização da molécula, enquanto a E_{LUMO} está relacionada à afinidade eletrônica. Esses descritores assumem que o HOMO e LUMO são os orbitais moleculares determinantes da probabilidade de um ataque eletrofílico ou nucleofílico.

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.

A dureza e a polarizabilidade estão relacionada com a diferença de energia HOMO-LUMO[39]. Um pequeno gap entre orbitais ocupados e não ocupados indica uma molécula “macia”, e dará uma grande contribuição para a polarizabilidade, isto significa

que maciez é uma medida de quanto a densidade eletrônica pode ser distorcida por um campo externo, como por exemplo o gerado por outra molécula. Estes conceitos têm um importante papel, pois a teoria de dureza e maciez de ácido e base que estabelece que ácidos duros preferam reagir com bases duras e vice-versa[39]. A interação dura-dura é primariamente controlada pelas cargas, enquanto a interação macia-macia pelos orbitais.

3. Metodologia de Índices Eletrônicos (MIE)

A MIE baseia-se em dois descritores quânticos η e Δ , que por sua vez estão fundamentados nos conceitos de densidade de estados[10].

A densidade de estados (DOS) é definida como o número de estados eletrônicos por unidade de energia. A densidade de estados local (LDOS) é a DOS calculada sobre uma região molecular específica. A LDOS descreve a distribuição espacial de estados sobre determinada região molecular.

A DOS representa a distribuição de estados no espaço de energia do sistema. Ela é obtida pelo número de estados eletrônicos por unidade de energia na vizinhança de um valor de energia, (Eq 32), onde ε_i é o autovalor de energia do sistema e o somatório é sobre todos os estados ocupados [11, 28].

$$DOS(E) = \sum_i \delta(E - \varepsilon_i) \quad (\text{Eq.32})$$

Para hamiltoniano baseado na aproximação LCAO os cálculos da LDOS envolvem a contribuição de cada orbital para um nível eletrônico específico e medida pelo coeficiente do orbital molecular elevado ao quadrado, ou seja, pela densidade de probabilidade correspondente ao nível eletrônico de sítio da molécula, (Eq 33). O

somatório é sobre a região molecular selecionada, que é sobre todos os orbitais atômicos selecionados (n_i a n_f)[10, 11].

$$LDOS(E_i) = 2 \sum_{m=n_i}^{n_f} |c_{mi}|^2 \quad (\text{Eq.33})$$

E_i : energia do orbital i ocupado

C_{mi} : coeficiente de expansão dos orbital molecular i em termos dos orbitais atômicos m .

A região molecular definida deve, preferencialmente, ser comum a todas as moléculas do estudo e ter, preferencialmente, algum sítio para ataque nucleofílico ou eletrofílico.

O descritor η está relacionado com as diferenças entre as densidades de estados local de dois orbitais de fronteira, podendo ser entre o último e penúltimo dos orbitais ocupados, HOMO e HOMO-1, respectivamente, (Eq.34), ou entre o primeiro e segundo orbital não ocupado, LUMO e LUMO-1, respectivamente, (Eq.35).

$$\eta H = LDOS(E_{HOMO}) - LDOS(E_{HOMO-1}) \quad (\text{Eq.34})$$

$$\eta L = LDOS(E_{LUMO+1}) - LDOS(E_{LUMO}) \quad (\text{Eq.35})$$

O índice Δ refere-se à diferença de energia dos índices de fronteira (HOMO ou LUMO) de uma determinada região da molécula.

Para os dois últimos orbitais moleculares ocupados definimos o índice ΔH como:

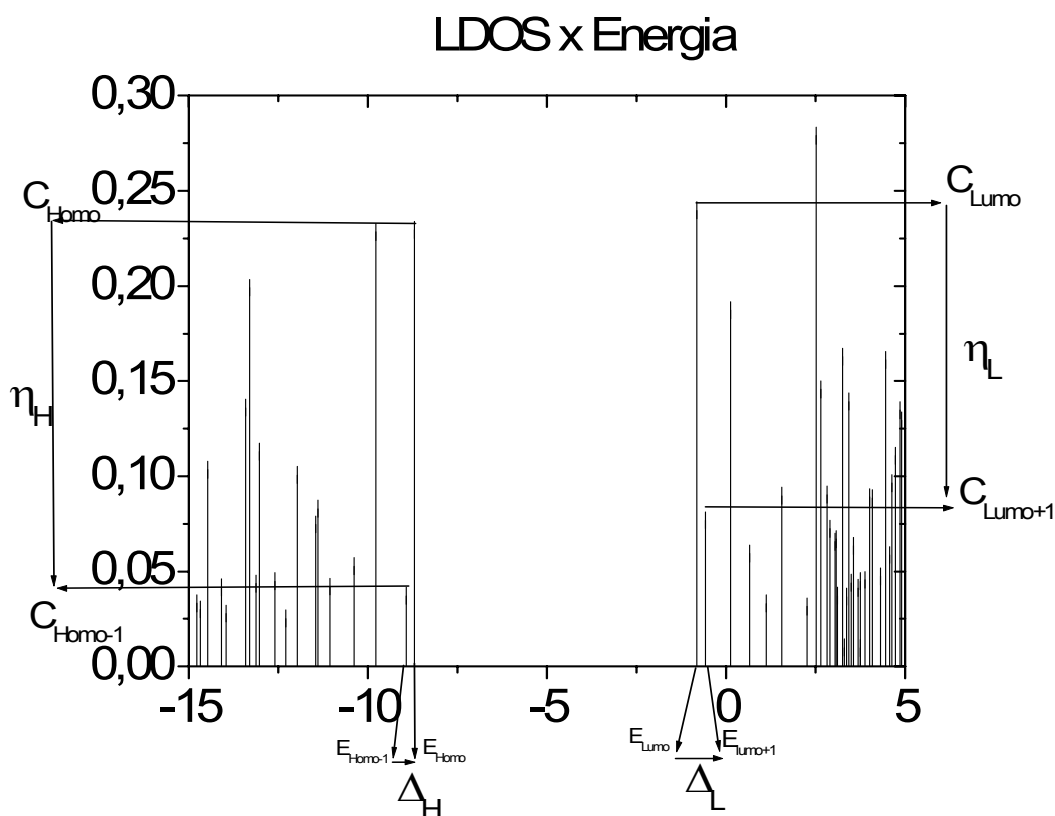
$$\Delta H = E_{HOMO} - E_{HOMO-1} \quad (\text{Eq.36})$$

Para os dois primeiros orbitais moleculares não ocupados definimos o índice ΔL como:

$$\Delta L = E_{LUMO+1} - E_{LUMO} \quad (\text{Eq.37})$$

O gráfico LDOS versus Energia, figura 2, sumariza os conceitos da MIE. O parâmetro Δ contém uma informação global (autovalores), enquanto que o parâmetro η contém uma informação molecular local (LDOS).

Fig.2: Gráfico LDOS versus Energia: Índices da MIE η e Δ .



A partir desses dois índices é possível discriminar qualitativamente as atividades biológicas entre ativas e inativas, através da seguinte regra:

Se $\eta > \eta_c$ ou $\Delta > \Delta_c$ então a molécula será ativa, senão será inativa.

Onde η_c e Δ_c são valores críticos dos índices eletrônicos η e Δ .

Os valores críticos são definidos de modo a obter a maior discriminação (ativas das inativas) das moléculas. O cálculo dos valores críticos η_c e Δ_c , realiza-se através da comparação dos valores da dose efetiva mínima (MED) com os índices da MIE, η e Δ . Para fins de classificação, no presente estudo, foi feita a separação das moléculas em dois grupos, tabela 3:

Grupo 1: Moléculas Ativas $\rightarrow$ MED $<10^{-6}$

Grupo 2: Moléculas Inativas $\rightarrow$ MED $>10^{-6}$

Tab. 3: Classificação quanto à Atividade Antineoplásica[18].

Molécula	MED(mg/Kg)	pMed*	ATV**	Molécula	MED	pMed	ATV
Mol01	4.67735E-05	4.33	0	mol19	1.7378E-06	5.76	1
mol02	3.38844E-05	4.47	0	mol20	1.65959E-06	5.78	1
mol03	2.34423E-05	4.63	0	mol22	1.38038E-06	5.86	1
mol04	1.69824E-05	4.77	0	mol23	9.33254E-07	6.03	1
mol05	1.41254E-05	4.85	0	mol24	7.24436E-07	6.14	1
mol06	1.20226E-05	4.92	0	mol25	6.91831E-07	6.16	1
mol07	7.07946E-06	5.15	1	mol26	6.60693E-07	6.18	1
mol08	6.91831E-06	5.16	1	mol27	6.60693E-07	6.18	1
mol09	3.46737E-06	5.46	1	mol28	6.60693E-07	6.18	1
mol10	2.69153E-06	5.57	1	mol29	6.16595E-07	6.21	1
mol11	2.5704E-06	5.59	1	mol30	5.62341E-07	6.25	1
mol12	2.51189E-06	5.60	1	mol31	4.0738E-07	6.39	1
mol13	2.34423E-06	5.63	1	mol32	3.89045E-07	6.41	1
mol14	2.18776E-06	5.66	1	mol33	3.89045E-07	6.41	1
mol15	2.0893E-06	5.68	1	mol34	3.54813E-07	6.45	1
mol16	2.0893E-06	5.68	1	mol35	2.88403E-07	6.54	1
mol17	2.0893E-06	5.68	1	mol36	1.69824E-07	6.77	1
mol18	2.04174E-06	5.69	1	mol37	1.25893E-07	6.90	1

*pMED=Log-MED. **ATV=Atividade: 0= inativa ; 1= ativa

Como todas as moléculas em estudo possuem algum grau de atividade antineoplásica, esta classificação possibilita separar as moléculas pelo seu grau de atividade biológica. Assim, as moléculas com baixo MED (abaixo de 10^{-6}) são mais ativas do que as com maior MED (acima de 10^{-6}). A atividade biológica é expressa em unidade logarítmica, pMED. Geralmente moléculas com pMED de 5-8 são baixas a levemente ativas e as acima de 8 são altamente ativas.

4. Análise Quimiométrica

Após obtenção de todos os descritores, procuramos saber quais desses estão relacionados com a atividade biológica. Para tal tarefa, usamos análise multivariada baseada em componentes principal e análise de grupo. Para os cálculos da PCA e HCA é utilizado o programa Einsight[22].

4.1 Análise das Componentes Principais (PCA)

É um método de redução de dados cujo objetivo é transformar o conjunto de variáveis em uma combinação linear que comporte a maior variância dos dados originais. A PCA propõe determinar fatores (componentes principais) que mantenha o máximo da variação dos dados com alguns desses fatores quanto possível[3, 6].

A componente principal é extraída de modo que a primeira componente principal, PC_1 , tenha a maior variância dos dados. A PC_1 é uma combinação linear das variáveis X_j , $j=1,2,...,n$, e os coeficientes w_n são escolhidos de tal modo que mantenha a máxima razão da variância da PC_1 e a variância total [3, 6].

$$PC_1 = w_1X_1 + w_2X_2 + ...w_nX_n \quad (\text{Eq.38})$$

Obedecendo a seguinte relação de vínculo[3, 6]:

$$\sum_{j=1}^n w_j^2 = 1 \quad (\text{Eq.39})$$

Assim, a PC_1 é formada de modo que tenha a maior parte da variância. Da variância restante, a PC_2 é extraída. O procedimento se repete até que todas as PC_s sejam geradas. A PC_2 é uma combinação linear das variáveis observadas não correlacionada com a primeira combinação e que mantenha o máximo do resto da variação total. De modo geral, a $PC_{m\text{-ésima}}$ é uma combinação linear ponderada das variáveis que tem maior variação de todas as combinações lineares e está não correlacionada com todas as outras PC_s obtidas[3, 6].

$$PC_{(m)} = w_{(m)1}X_1 + w_{(m)2}X_2 + \dots w_{(m)p}X_p \quad (\text{Eq.40})$$

A posterior PC é ortogonal à anterior e responde pela máxima variância restante. Após remoção de variáveis redundantes, apenas poucas PC_s são necessárias para descrever toda a informação contida nos dados originais.

A matriz de dados $X_{I \times J}$ correspondendo a I moléculas e J descritores, é decomposta em duas matrizes, T e L, tal que [3, 6]

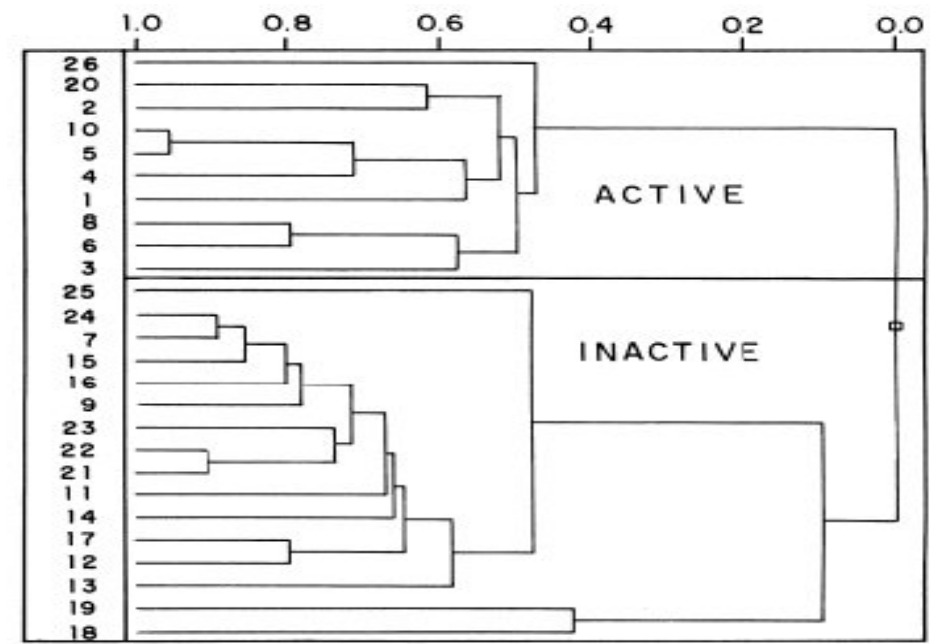
$$\mathbf{X} = \mathbf{T} \cdot \mathbf{L}^T \quad (\text{Eq.41})$$

A matriz T (matriz “score”) representa a posição dos compostos no novo sistema de coordenadas onde as PC_s são os eixos. A matriz L (matriz “loadings”) descreve como os novos eixos, a PC, é gerada do eixo antigo.

4.2 Análise Hierárquica de Grupo (HCA)

HCA[3, 6] é um método estatístico de análise multivariada que se propõe identificar o menor número de grupos, cujos elementos pertencentes ao particular grupo, sejam mais semelhantes um ao outro do que aos elementos de outros grupos. O resultado qualitativo é representado na forma de um dendograma que permite visualizar as amostras no espaço bidimensional, figura 3[4].

Fig. 3. Análise Hierárquica de Grupos: Exemplo de Dendograma [4].



A partir dos gráficos obtidos da PCA (loadings e scores) a HCA verifica a semelhança entre as amostras ou variáveis pela distância entre eles no espaço das componentes principais. As distâncias entre as amostras são calculadas e transformadas,

em uma matriz S , cujos elementos são índices de similaridade dado pela seguinte expressão[3, 6, 7] :

$$S_{kl} = 1 - \frac{d_{kl}}{d_{\max}} \quad (\text{Eq.42})$$

Onde S_{kl} é um elemento da matriz de similaridade S , $d_{\max}$ é a maior distância para qualquer par de amostras e d_{kl} é a distância euclidiana entre as amostras k e l calculada como [6]:

$$d_{kl} = \sqrt{(x_{k1} - x_{l1})^2 + (x_{k2} - x_{l2})^2 + \dots + (x_{kj} - x_{lj})^2} \quad (\text{Eq.43})$$

O índice de similaridade estende-se de 0 a 1. Quanto mais próximo de zero menor a distância entre as amostras k e l , portanto, mais semelhantes são as amostras.

V. RESULTADOS

1. Resolução das Estruturas Geométricas e Análise Conformacional

Inicialmente todas as estruturas foram resolvidas (otimização de geometria) pelo método de mecânica molecular pelo campo CHARMM (Chemistry at HARvard Molecular Mechanics)[40] utilizando o HypeChem[21] e em seguida aplicamos o método semi-empírico PM3[15, 16], utilizando o software Spartan®[19]. A escolha de um método para varredura de todo o espaço conformacional torna-se numa grande tarefa computacional para moléculas muito flexíveis, pois haverá um enorme número de conformações. Para contornar este problema o método de Monte Carlo foi aplicado para varrer todo espaço conformacional de maneira satisfatória. A busca conformacional para todos os diedros relevantes foi realizada aplicando o método de Monte Carlo contido no software Spartan®[19]. O resultado final da resolução geométrica das moléculas é mostrado na figura 4 e o calor de formação na tabela 4. Cálculos *Ab Initio* com a base 6-31G sem busca conformacional foi realizado pelo Spartan®[19] em algumas moléculas (13, 17, 24, 26, 28 e 32) escolhidas aleatoriamente para fins de comparação das estruturas obtidas com o método semi-empírico PM3, figura 7.

2. Cálculos das Propriedades Moleculares para SAR

Os parâmetros para a análise SAR das moléculas otimizadas foram calculados com o software Hypechem[21] e Spartan®[19]. Foram obtidas as seguintes propriedades: Calor de Formação, Volume molecular, Log P, Momento de Dipolo, Polarizabilidade, Refratividade, Carga, Eletronegatividade, Potencial de Ionização, Calor de Hidratação e Dureza. A tabela 5 mostra os valores das energias e LDOS dos orbitais de fronteira.

3. Cálculo da Energia e LDOS dos Orbitais de Fronteira

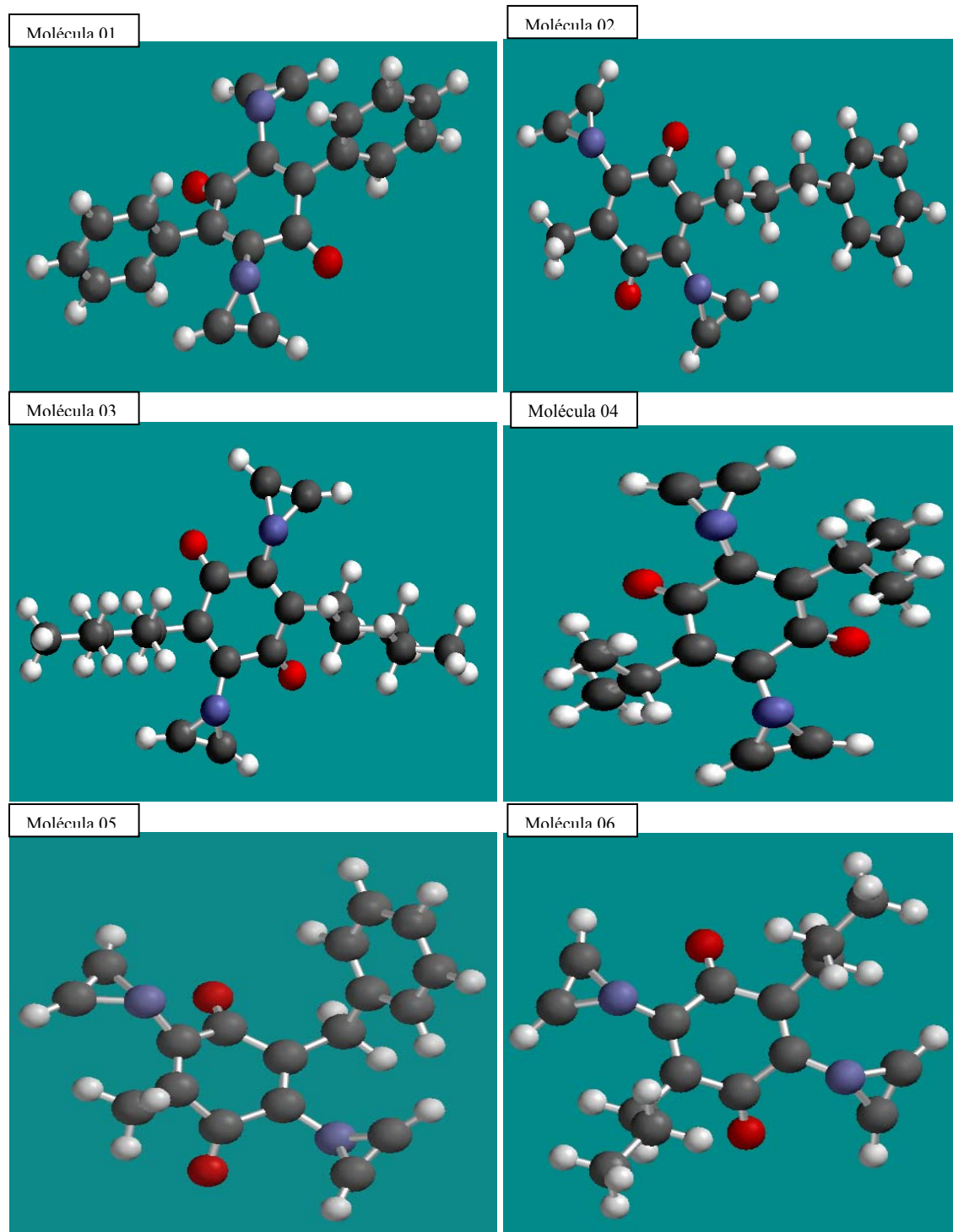
Para os cálculos dos parâmetros de energias e LDOS dos orbitais de fronteira é necessário primeiro a definição de uma região da molécula. A escolha da região pode ser baseada em critério como maior densidade de carga, maior contribuição do orbital atômico para a composição do orbital molecular, região de maior ordem de ligação, região comum a todo o conjunto de moléculas em estudo, etc. Em virtude das moléculas possuírem um núcleo benzoquinona em comum, foi então esta região a escolhida, figura 8. A tabela 6 mostra os valores obtidos de energias e LDOS dos orbitais de fronteira para a região especificada pela figura 8, enquanto na tabela 7 apresentamos os valores dos índices eletrônicos.

Tab.4: Calor de Formação das Estruturas Geométricas Resolvidas.

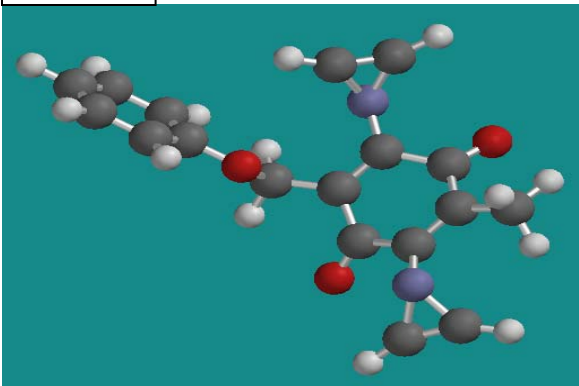
Molécula	Calor de Formação (kcal/mol)	Molécula	Calor de Formação (kcal/mol)
Mol01	197.823842	mol19	4.801396
mol02	145.20355	mol20	49.778377
mol03	86.572696	mol22	123.171529
mol04	112.720896	mol23	-23.806005
mol05	156.5216	mol24	44.285168
mol06	108.214633	mol25	14.470588
mol07	129.036444	mol26	47.085965
mol08	-21.632365	mol27	12.810051
mol09	119.108264	mol28	46.115084
mol10	88.563207	mol29	51.429212
mol11	73.457635	mol30	43.259721
mol12	120.033133	mol31	85.207305
mol13	8.587236	mol32	47.785519
mol14	127.223844	mol33	15.667741
mol15	128.349536	mol34	191.425163
mol16	86.177018	mol35	39.272826
mol17	38.315464	mol36	171.430968
mol18	47.459286	mol37	52.314588

Fig. 6: Estruturas Geométrica das Moléculas Resolvidas com o Método PM3

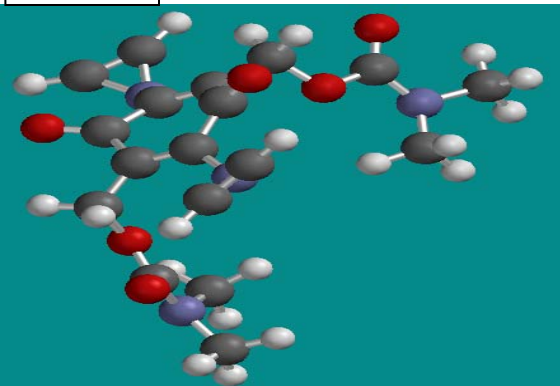
(carbono=preto, nitrogênio=azul, oxigênio=vermelho, hidrogênio=cinza)



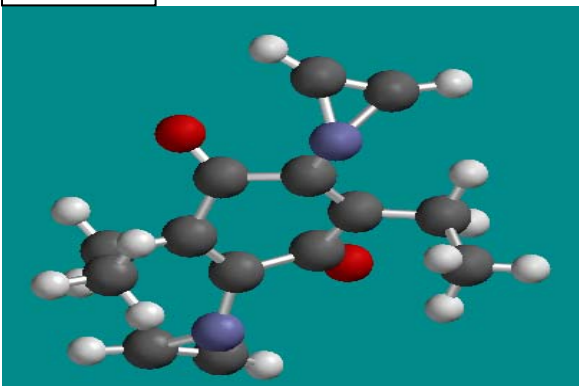
Molécula 07



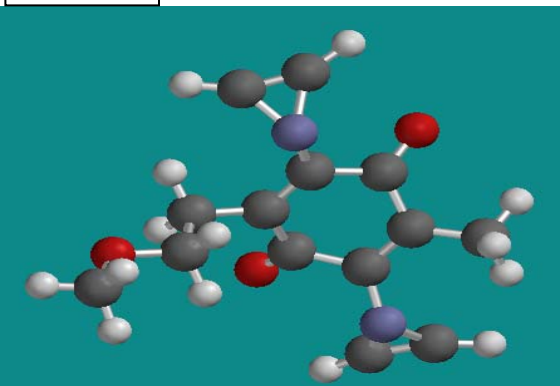
Molécula 08



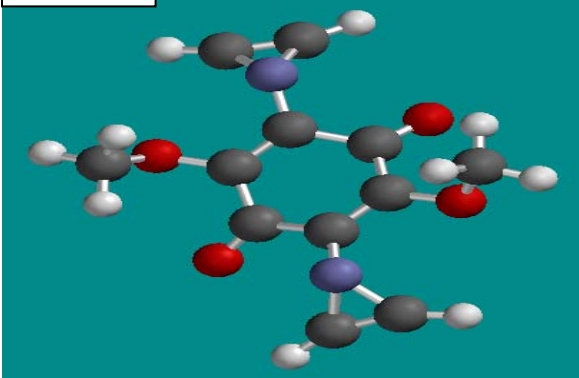
Molécula 09



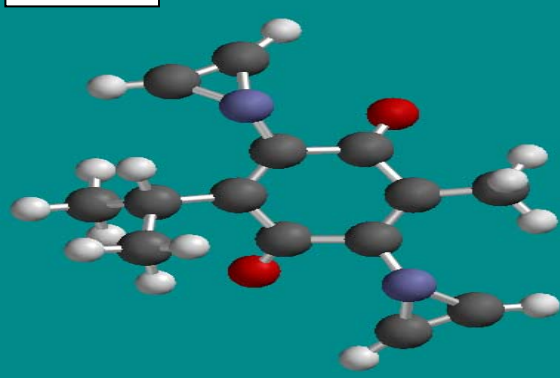
Molécula 10



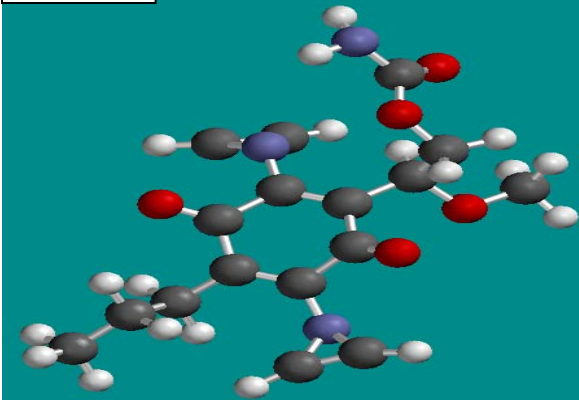
Molécula 11



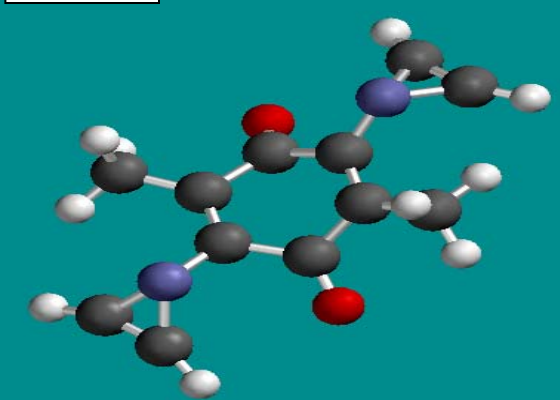
Molécula 12



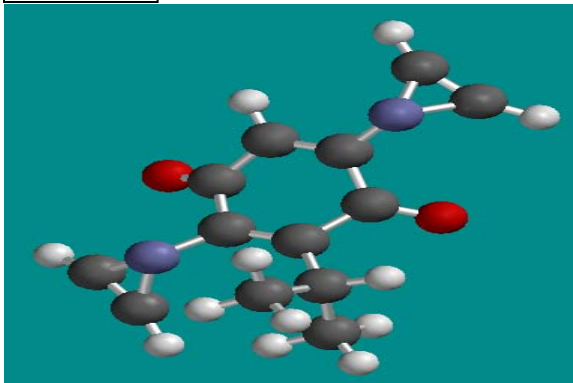
Molécula 13



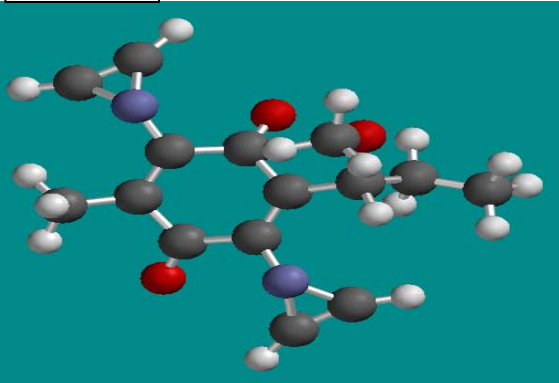
Molécula 14



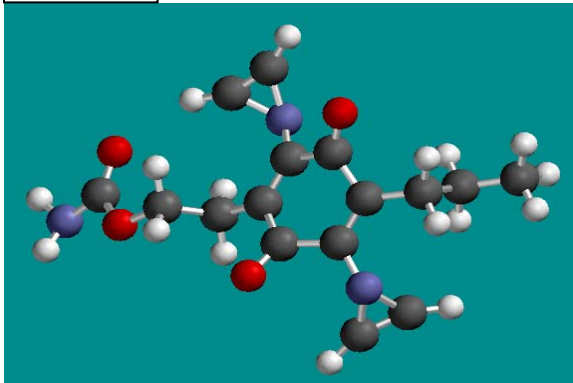
Molécula 15



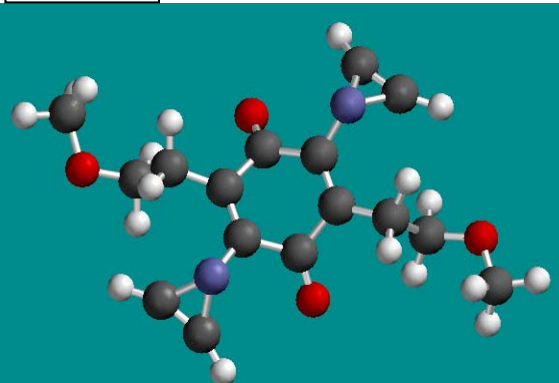
Molécula 16



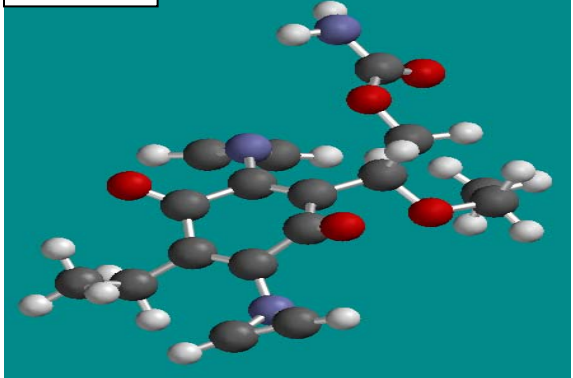
Molécula 17



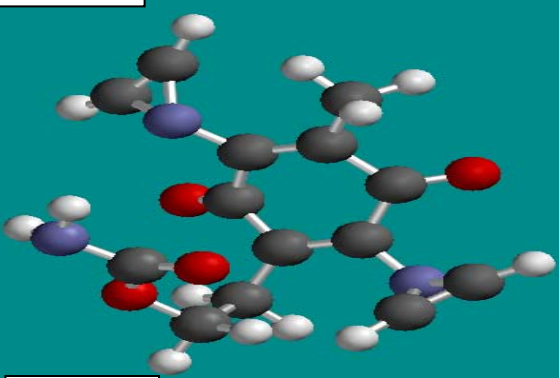
Molécula 18



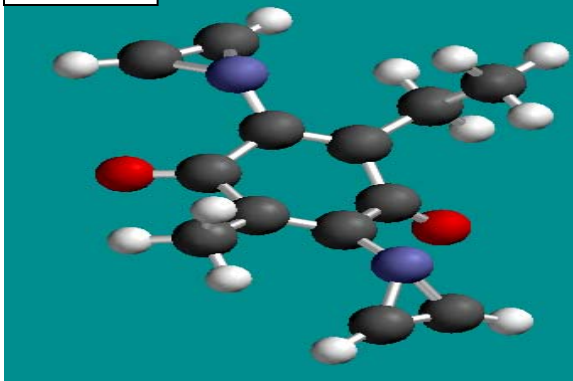
Molécula 19



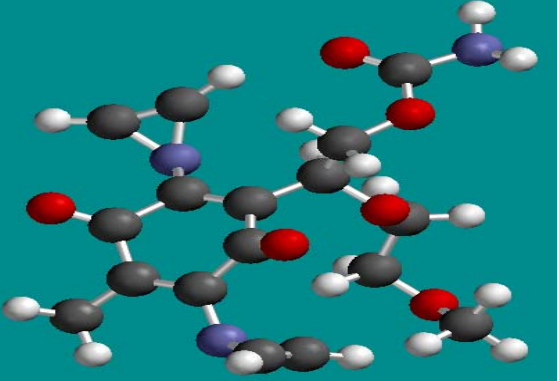
Molécula 20



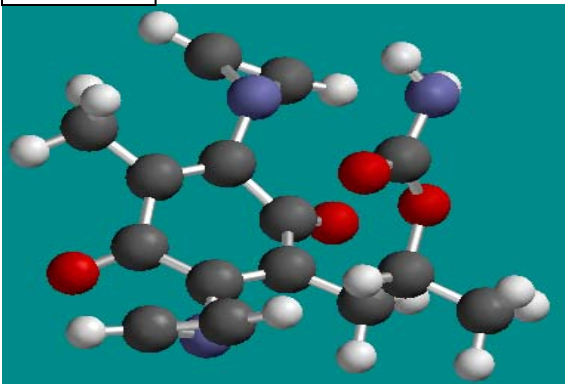
Molécula 22



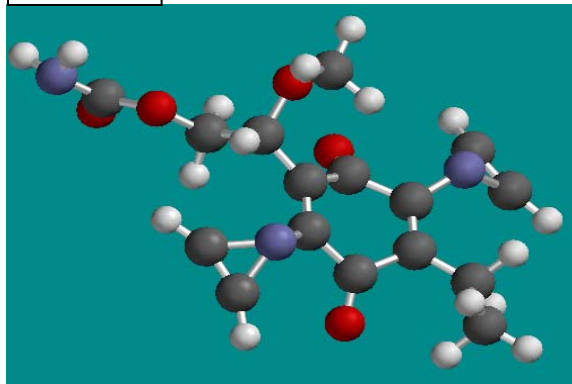
Molécula 23



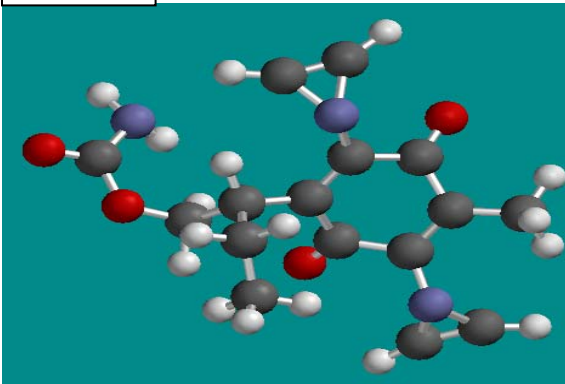
Molécula 24



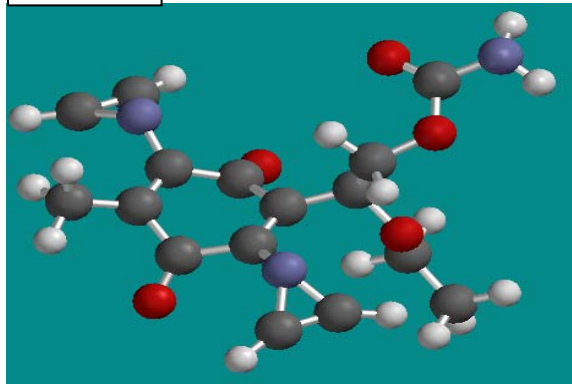
Molécula 25



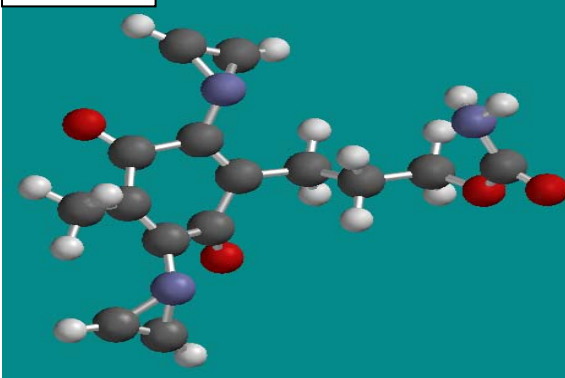
Molécula 26



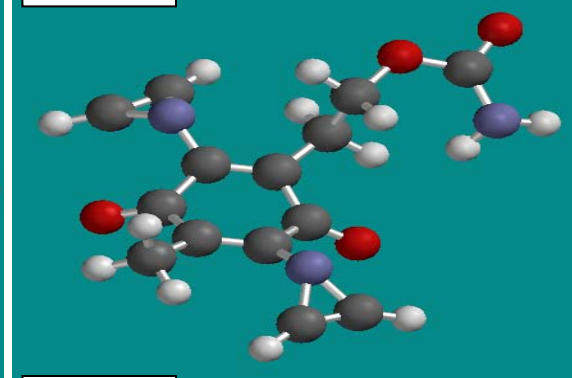
Molécula 27



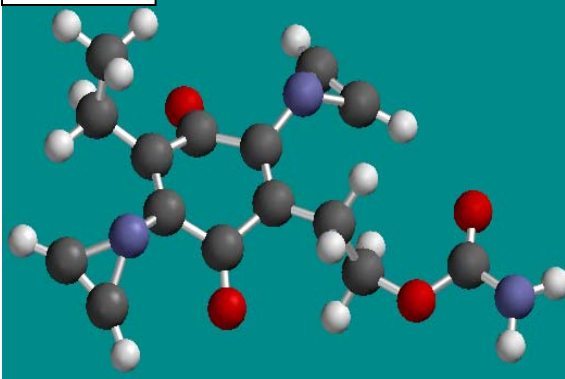
Molécula 28



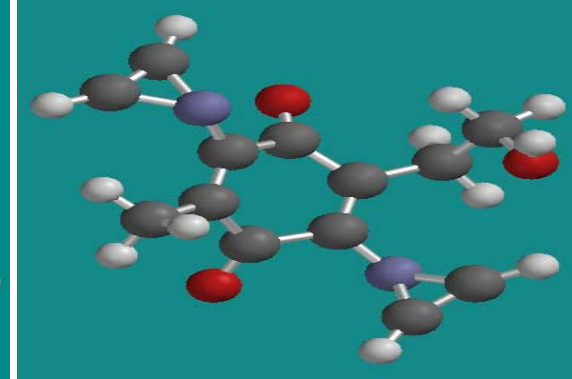
Molécula 29



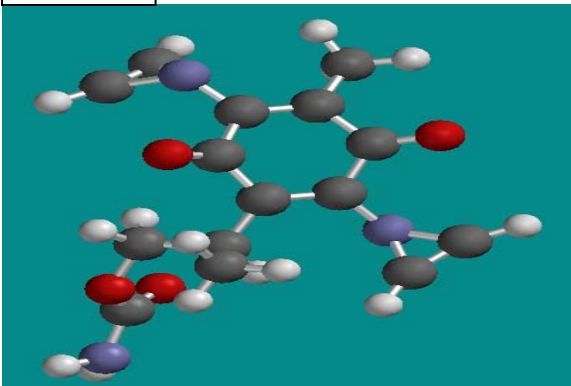
Molécula 30



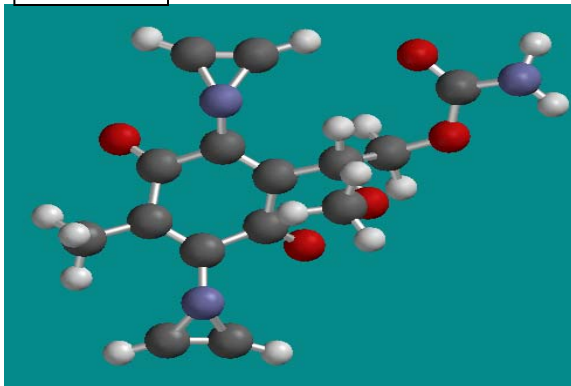
Molécula 31



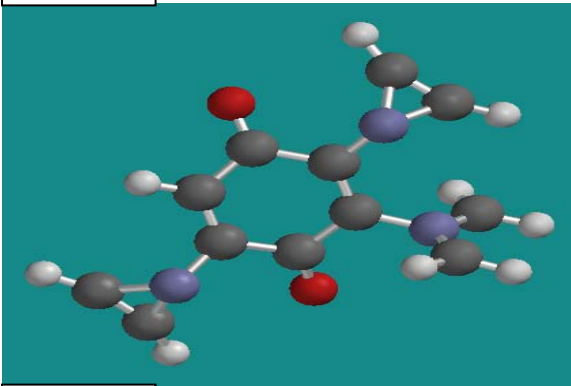
Molécula 32



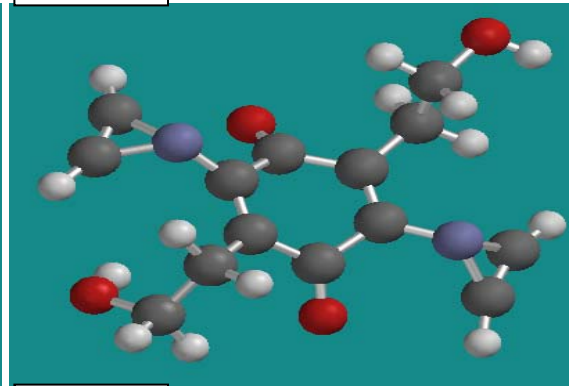
Molécula 33



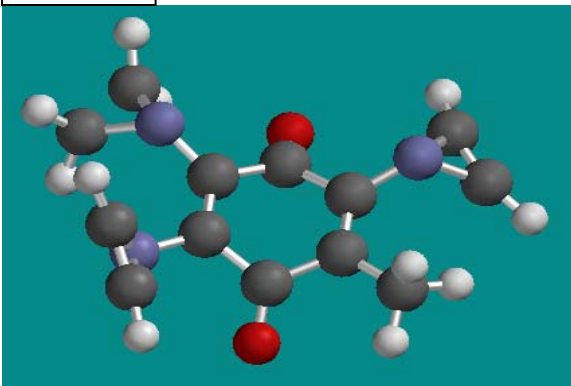
Molécula 34



Molécula 35



Molécula 36



Molécula 37

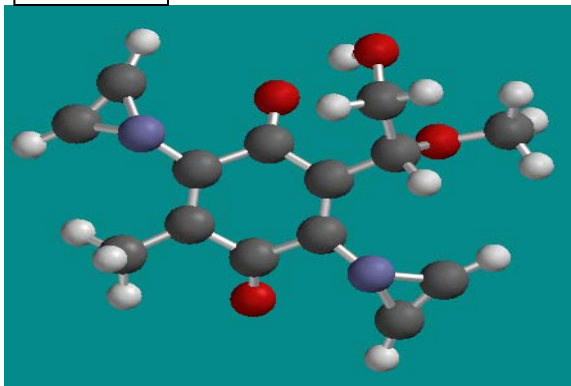
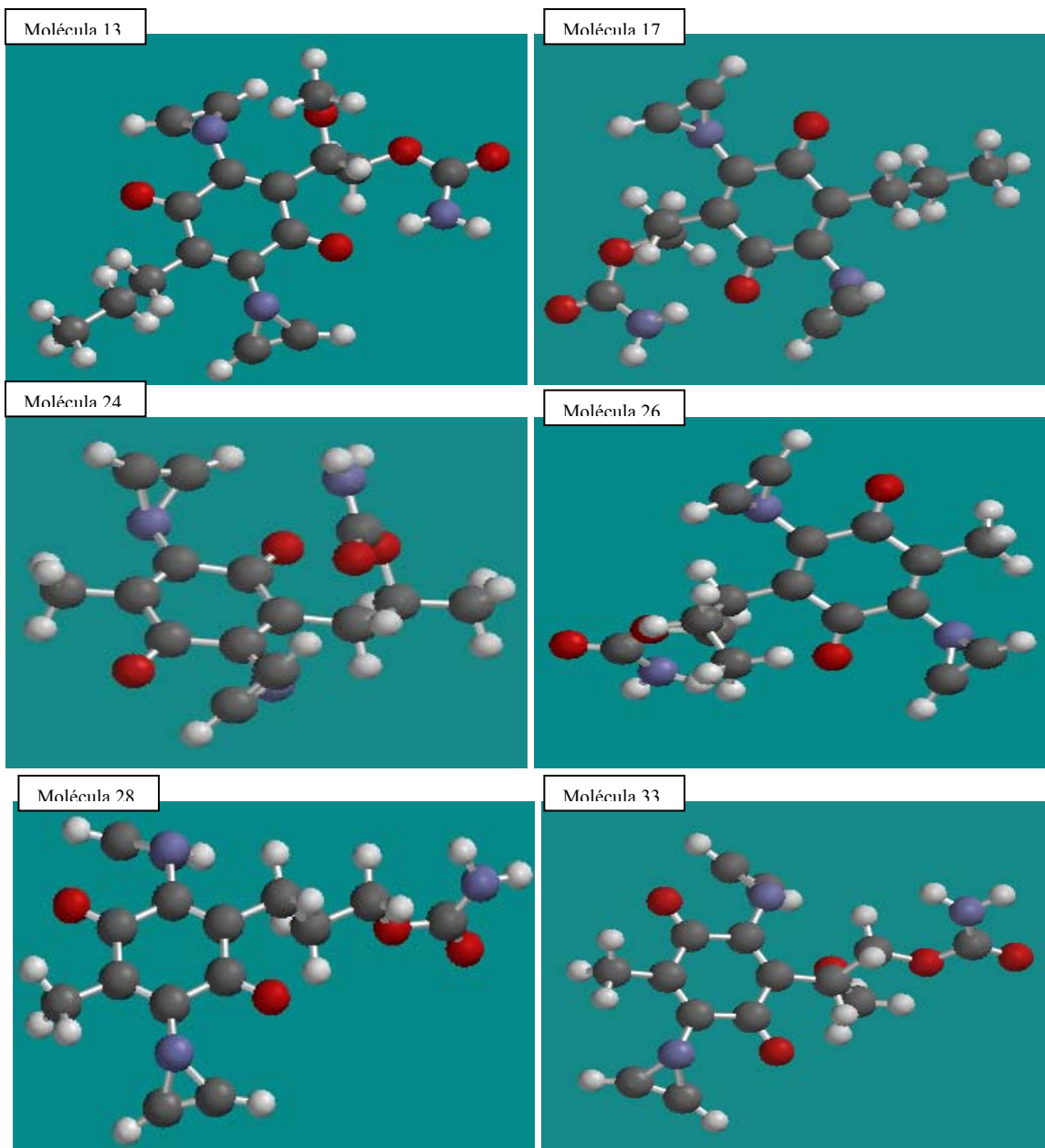


Fig. 7: Estruturas Geométrica das Moléculas Resolvidas com o Método *Ab*

Initio (Base 6-31G)[19].



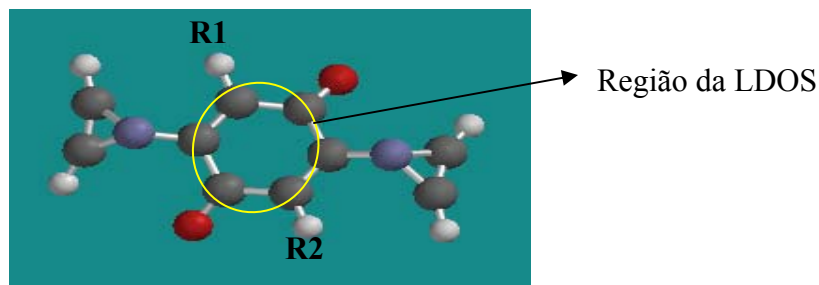
2. Cálculos das Propriedades Moleculares para SAR

Os parâmetros para a análise SAR das moléculas otimizadas foram calculados com o software Hypechem[21] e Spartan[®][19]. Foram obtidas as seguintes propriedades: Calor de Formação, Volume molecular, Log P, Momento de Dipolo, Polarizabilidade, Refratividade, Carga, Eletronegatividade, Potencial de Ionização, Calor de Hidratação, Dureza. A tabela 5 mostra os valores das energias e LDOS dos orbitais de fronteira.

3. Cálculo da Energia e LDOS dos Orbitais de Fronteira

Para os cálculos dos parâmetros de energias e LDOS dos orbitais de fronteira é necessário primeiro a definição de uma região da molécula. A escolha da região pode ser baseada em critério como maior densidade de carga, maior contribuição do orbital atômico para a composição do orbital molecular, região de maior ordem de ligação, região comum a todo o conjunto de moléculas em estudo, etc. Em virtude das moléculas possuírem um núcleo benzoquinona em comum, foi então esta região a escolhida, figura 8. A tabela 6 mostra os valores obtidos de energias e LDOS dos orbitais de fronteira para a região especificada pela figura 8, enquanto na tabela 7 apresentamos os valores dos índices eletrônicos.

Fig.8: Região para Cálculo dos Parâmetros da MIE. Núcleo Benzeno C1 a C6



Tab.5: Valores das Propriedades Físico-Químicas.

Molécula	EH	Log P	RFT	PLZ	MD	ENG	Dureza	P. I	Volume (Å)
Mol01	-5.88	1.11	102.26	38.23	0.8007	-5.169735	3.942815	9.113614	382.92
mol02	-4.2	1.09	95.97	35.91	1.14246	-5.174965	3.935905	9.110768	374.08
mol03	0.39	1.85	98.88	37.26	0.69122	-5.14701	3.96338	9.110641	409.99
mol04	-1.6	0.13	80.37	29.92	0.65525	-5.09564	4.02615	9.122696	326.19
mol05	-4.54	0.29	86.77	32.24	2.64595	-5.146795	3.929385	9.077371	332.93
mol06	-1.31	0.26	80.48	29.92	0.72877	-5.137785	3.966585	9.105936	327.97
mol07	-5.86	-0.26	88.11	32.87	1.5759	-5.2215	3.89005	9.101805	345.82
mol08	-3.54	-2.62	100.87	37.74	1.68857	-5.34674	3.92137	9.26648	431.87
mol09	-2.29	-0.53	71.28	26.25	0.78864	-5.136795	3.970475	9.106484	287.00
mol10	-4.44	-1.79	73.36	26.88	0.95158	-5.174125	3.953795	9.130212	298.77
mol11	-6.39	-3.61	66.26	23.85	0.42489	-5.2516	3.82914	9.080376	270.69
mol12	-2.41	-0.59	71.22	26.25	0.82304	-5.108185	3.979545	9.092134	286.10
mol13	-9.33	-1.77	91.59	34.46	2.53674	-5.181435	3.950275	9.088456	389.71
mol14	-3.19	-1.32	62.07	22.58	0.97904	-5.137645	3.946465	9.085314	246.05
mol15	-3.17	-0.87	66.87	24.41	1.13802	-5.17112	3.92515	9.096127	265.74
mol16	-3.14	-1.16	77.54	28.72	0.91486	-5.134165	3.958515	9.088933	318.82
mol17	-9.75	-1.42	85.63	61.99	3.43526	-5.13122	3.90596	9.037209	358.57
mol18	-8.12	-2.04	72.01	25.64	0.91905	-5.25362	3.92175	9.175244	350.75
mol19	-8.85	-1.82	91.74	34.46	1.74201	-5.119855	3.960615	9.210912	389.04
mol20	-8.99	-2.21	76.43	28.32	3.20443	-5.137695	3.932005	9.069751	317.22
mol22	-2.72	-0.92	66.68	24.41	0.86258	-5.13777	3.95407	9.09201	266.49
mol23	-11.68	-2.72	93.43	35.1	2.92984	-5.19201	3.95657	9.181508	400.28
mol24	-10.09	-1.11	66.46	26.96	3.1928	-5.17711	3.94138	9.118493	337.45
mol25	-9.9	-2.16	86.99	32.63	2.06734	-5.223565	3.957965	9.165254	370.55
mol26	-6.82	-1.41	85.5	31.59	2.87486	-5.29604	3.95772	9.262989	355.94
mol27	-10.23	-2.22	87.14	32.63	2.44901	-5.14667	3.91844	9.02366	370.09
mol28	-8.62	-1.82	81.03	30.16	5.7687	-5.35044	3.90995	9.260273	338.50
mol29	-9.35	-1.19	74.28	26.99	5.76703	-5.417945	3.899155	9.316985	317.13
mol30	-10.11	-1.82	81.03	30.16	2.53498	-5.107825	3.919835	9.022955	337.80
mol31	-6.52	-2.07	68.6	25.05	3.2339	-5.085755	3.898415	8.981548	277.46
mol32	-10.11	-1.81	80.9	30.16	3.43007	-5.115985	3.933555	9.049626	338.24
mol33	-10.7	-2.56	82.39	30.79	1.29929	-5.13023	3.95907	9.085515	349.04
mol34	-6.18	-1.98	66.48	23.86	3.22198	-4.76116	2.92697	7.689845	257.24
mol35	-13.81	-2.82	75.13	27.52	1.14265	-5.32766	3.92824	9.256549	308.35
mol36	-3.11	-2.6	69.44	24.99	1.2049	-5.015245	3.756255	8.771074	274.51
mol37	-8.29	-2.42	74.56	27.52	1.96016	-5.32908	3.88192	9.20874	306.36

EH = Energia de Hidratação; RFT = Refratividade; PLZ = Polarizabilidade; MD = Momento Dipolo; ENG = Eletronegatividade;

P.I = Potencial de Ionização; Log P = logaritmo do coeficiente de Partição.

Tab.6: Valores de Energias e LDOS dos Orbitais de Fronteira.

Molécula	E _{HOMO-1} (eV)	L _{HOMO-1}	E _{HOMO} (eV)	L _{HOMO}	E _{LUMO} (eV)	L _{LUMO}	E _{LUMO+1} (eV)	L _{LUMO+1}
Mol01	-9.12155	0.413651	-9.1126	0.58302	-1.2269	1.4222	-0.01546	0.899796
Mol02	-9.15418	0.504545	-9.1109	0.54747	-1.2391	1.4258	0.13507	1.589642
Mol03	-9.12297	0.445298	-9.1104	0.61452	-1.1836	1.4328	0.16851	1.725334
Mol04	-9.13386	0.44447	-9.1218	0.62668	-1.0695	1.4587	0.20045	1.719743
Mol05	-9.21817	0.461941	-9.0762	0.57916	-1.2174	1.4214	0.10244	1.352324
Mol06	-9.11934	0.458023	-9.1044	0.60429	-1.1712	1.4338	0.17709	1.724174
Mol07	-9.24084	0.005899	-9.1116	0.55331	-1.3315	1.4131	0.02366	1.696902
Mol08	-9.29049	0.512503	-9.2681	0.48202	-1.4254	1.418	-0.1657	1.621441
Mol09	-9.11512	0.44701	-9.1073	0.6103	-1.1663	1.4346	0.18922	1.725678
Mol10	-9.15717	0.479729	-9.1279	0.57969	-1.2203	1.4309	0.12552	1.725216
Mol11	-9.13995	0.424933	-9.0807	0.58129	-1.4225	1.39	-0.00772	1.728773
Mol12	-9.13208	0.50137	-9.0877	0.57756	-1.1286	1.4426	0.1904	1.721232
Mol13	-9.35428	0.489158	-9.1317	0.54529	-1.2312	1.4356	0.05331	1.721989
Mol14	-9.10314	0.450063	-9.0841	0.63368	-1.1912	1.4273	0.17928	1.724563
Mol15	-9.30823	0.475669	-9.0963	0.51502	-1.246	1.4359	0.15379	1.71297
Mol16	-9.23597	0.523341	-9.0927	0.54079	-1.1757	1.4421	0.11599	1.719787
Mol17	-9.17837	0.462275	-9.0372	0.5504	-1.2253	1.4196	0.12017	1.728088
Mol18	-9.20588	0.427646	-9.1754	0.60734	-1.3319	1.426	0.03962	1.720208
Mol19	-9.25578	0.477468	-9.0805	0.56003	-1.1592	1.4362	0.12653	1.722725
Mol20	-9.2533	0.480007	-9.0697	0.54982	-1.2057	1.4256	0.13974	1.731947
Mol22	-9.1141	0.473787	-9.0918	0.59823	-1.1837	1.4297	0.18331	1.726237
Mol23	-9.18656	0.550375	-9.1486	0.41767	-1.2354	1.4392	0.0404	1.708072
Mol24	-9.28047	0.465937	-9.1185	0.56304	-1.2357	1.4271	0.0974	1.731955
Mol25	-9.31901	0.508852	-9.1815	0.52326	-1.2656	1.4436	0.01284	1.71063
Mol26	-9.27741	0.460099	-9.2538	0.59732	-1.3383	1.4444	-0.02422	1.721919
Mol27	-9.15703	0.490095	-9.0651	0.52391	-1.2282	1.4329	0.0938	1.720423
Mol28	-9.43859	0.520416	-9.2604	0.53125	-1.4405	1.4287	-0.08632	1.723609
Mol29	-9.38495	0.485946	-9.3171	0.55161	-1.5188	1.4212	-0.14169	1.722359
Mol30	-9.11208	0.513403	-9.0277	0.51506	-1.188	1.4309	0.16297	1.715019
Mol31	-9.184	0.48952	-8.9842	0.55463	-1.1873	1.4188	0.18016	1.739872
Mol32	-9.19964	0.459126	-9.0495	0.5995	-1.1824	1.4287	0.14139	1.733112
Mol33	-9.13636	0.492346	-9.0893	0.53465	-1.1712	1.444	0.11044	1.714264
Mol34	-9.50909	0.429628	-7.6881	0.00016	-1.8342	1.4048	-0.4083	1.752336
Mol35	-9.31231	0.51113	-9.2559	0.51807	-1.3994	1.4339	-0.06162	1.726763
Mol36	-9.16519	0.493086	-8.7715	0.58703	-1.259	1.4001	0.12577	1.701034
Mol37	-9.32212	0.490003	-9.211	0.53033	-1.4472	1.4172	-0.05501	1.717919

Tab.7: Valores dos Índices Eletrônicos η e Δ .

Molécula	η_H	Δ_H	η_L	Δ_L
mol01	0.1693679	0.009	-0.522417	1.21146
mol02	0.0429243	0.04331	0.163849	1.37413
mol03	0.1692184	0.01258	0.29249	1.35214
mol04	0.1822124	0.01207	0.260997	1.26994
mol05	0.1172181	0.14199	-0.069118	1.31985
mol06	0.1462688	0.01497	0.290363	1.34829
mol07	0.5474067	0.12929	0.28379	1.35511
mol08	-0.030484	0.02238	0.203431	1.25967
mol09	0.1632856	0.00785	0.29103	1.35554
mol10	0.0999584	0.02925	0.294324	1.34585
mol11	0.1563533	0.05921	0.338773	1.41474
mol12	0.0761947	0.04435	0.278654	1.31904
mol13	0.0561292	0.22257	0.28641	1.28447
mol14	0.1836158	0.01903	0.297311	1.37046
mol15	0.039354	0.21196	0.277101	1.39976
mol16	0.017453	0.14329	0.277679	1.29164
mol17	0.0881251	0.14119	0.308502	1.34543
mol18	0.1796987	0.03051	0.294239	1.37149
mol19	0.082566	0.17531	0.286478	1.28577
mol20	0.0698085	0.1836	0.306392	1.34543
mol22	0.1244392	0.02226	0.296546	1.36701
mol23	-0.132705	0.03798	0.268833	1.27584
mol24	0.0970978	0.16198	0.304888	1.33313
mol25	0.014409	0.13748	0.267006	1.27844
mol26	0.1372241	0.02365	0.277506	1.3141
mol27	0.0338131	0.09192	0.287523	1.32203
mol28	0.0108309	0.1782	0.294938	1.35417
mol29	0.0656662	0.06785	0.30115	1.3771
mol30	0.0016552	0.08442	0.284104	1.35096
mol31	0.0651145	0.19983	0.321054	1.3675
mol32	0.1403705	0.1501	0.304381	1.32382
mol33	0.042303	0.04706	0.27029	1.2816
mol34	-0.429467	1.82096	0.347575	1.42589
mol35	0.0069379	0.05641	0.292873	1.3378
mol36	0.093942	0.39369	0.300953	1.38476
mol37	0.0403235	0.11112	0.300705	1.39215

4. Determinação dos Parâmetros Críticos η_c e Δ_c

A comparação dos índices da MIE com os valores de atividade biológica (índice MED) resultou na diferenciação das moléculas entre ativas e inativas com 91,7% de acerto máximo utilizando dois parâmetros da MIE. As figuras 9 e 10 mostram o gráfico para os parâmetro Δ e η em função das moléculas, respectivamente. Nestes gráficos observa-se que não foi possível a discriminação das atividades das moléculas com apenas um parâmetro. Na figura 11 apresentamos os parâmetros críticos, η_L e Δ_L , que permitem obter uma regra simples para separar as moléculas de acordo com suas atividades biológicas. A combinação dos parâmetros Δ_L e η_L apresentou 91,67% de acerto de discriminação entre as moléculas. Os índices Δ_H e η_H tiveram menor grau de acerto da atividade biológica das moléculas.

Regra da MIE

Após comparação dos parâmetros da MIE com os índices de atividade biológica experimental e determinação dos parâmetros críticos, figura 8, pode-se chegar à seguinte conclusão:

$$\Delta_L \geq 1.21146 \text{ e } \eta_L \geq 0.164877 \Rightarrow \text{Moléculas Ativas}$$

Neste caso, somente as moléculas 3, 4 e 6 são erroneamente classificadas.

$$\Delta_L \geq 1.21146 \text{ e } \eta_L \geq 0.260576 \Rightarrow \text{Moléculas Ativas}$$

Neste caso, somente as moléculas 3, 6 e 8 são erroneamente classificadas

$$\Delta_L \leq 1.42589 \text{ e } \eta_L \geq 0.199677 \Rightarrow \text{Moléculas Ativas}$$

Neste caso, somente as moléculas 3, 4 e 6 são erroneamente classificadas.

Fig.9: Gráfico do Parâmetro ΔL e ΔH versus Moléculas.

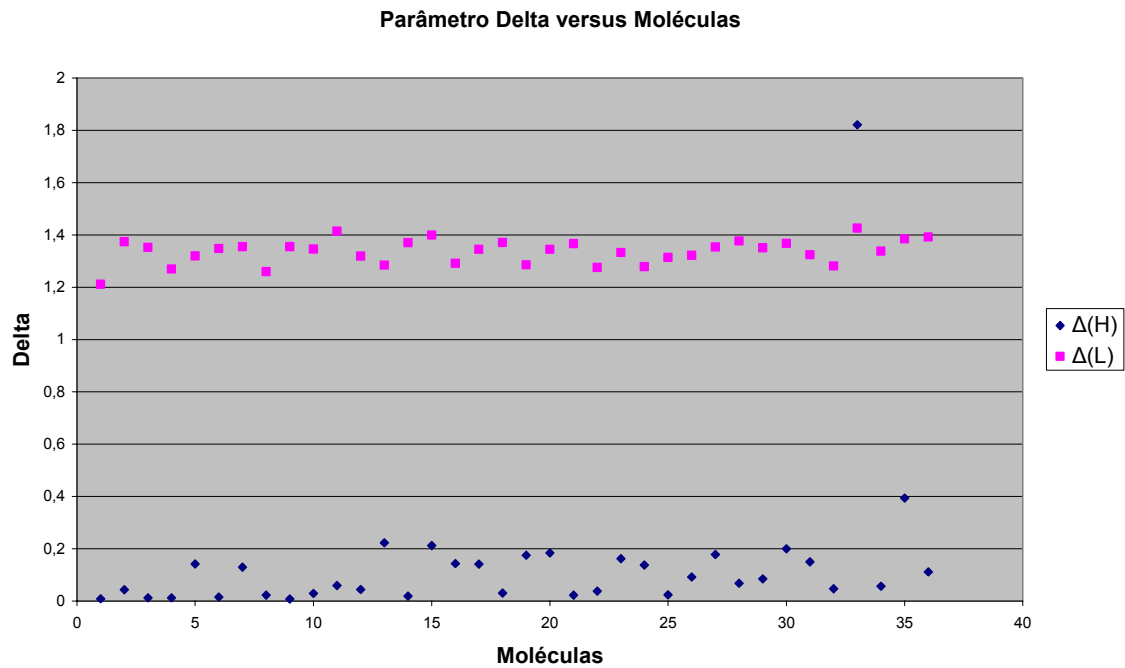


Fig.10: Gráfico do Parâmetro ηL e ηH versus Moléculas.

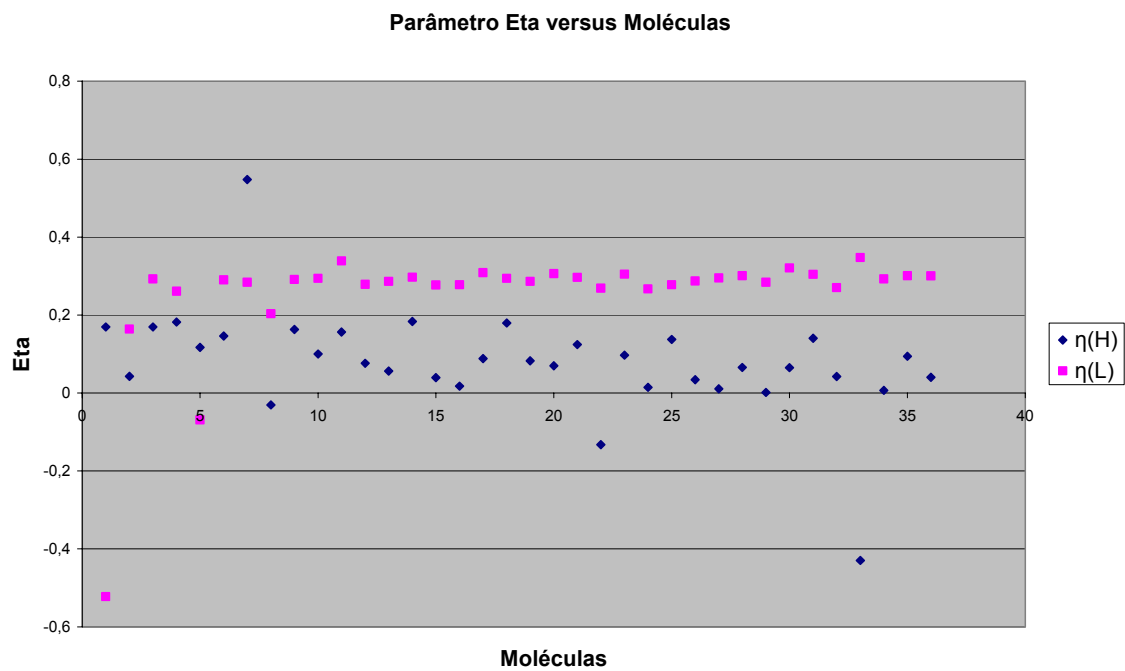


Fig.11: Determinação dos Parâmetros Críticos ΔL_c e ηL_c :

Delta_max= 1.425890 Delta_min= 1.211460

Eta_max= 0.347575 Eta_min= -0.522417

inc_delta= 0.002144 inc_eta= 0.008700

=====

Regra usando combinacao $\Delta L \geq$ e $\eta L \geq$:

$\Delta L = 1.211460$ eV, $\eta L = 0.164877$, Acerto= 91.67 %

Atividade Biológica Teórica (moléculas de 1 a 36)

0 0 1 1 0 1

Atividade Biológica Experimental

0 0 0 0 0 0 1

$\Delta L = 1.275789$ eV, $\eta L = 0.260576$, Acerto= 91.67 %

Atividade Biológica Teórica (moléculas de 1 a 36)

0 0 1 0 0 1 1 0 1

Atividade Biológica Experimental

0 0 0 0 0 0 1

=====

Regra usando combinacao $\Delta L \leq$ e $\eta L \geq$:

$\Delta L = 1.425890$ eV, $\eta L = 0.164877$, Acerto= 91.67 %

Atividade Biológica Teórica (moléculas de 1 a 36)

0 0 1 1 0 1

Atividade Biológica Experimental

0 0 0 0 0 0 1

$\Delta L = 1.425890$ eV, $\eta L = 0.199677$, Acerto= 91.67 %

Atividade Biológica Teórica

0 0 1 1 0 1

Atividade Biológica Experimental

0 0 0 0 0 0 1

5. Análise Quimiométrica: PCA e HCA

Para a análise de PCA foram utilizados 25 descritores os quais representam as características físico-químicas, estereoquímicas e eletrônicas das 36 moléculas em estudo.

- Volume Molecular
- Momento de Dipolo
- Refratividade
- Polarizabilidade
- Dureza: $D=(LUMO-HOMO)/2$ [39]
- Eletronegatividade de Mulligen: $EN=(HOMO+LUMO)/2$ [41]
- Potencial de Ionização
- Índices de Fronteira: energias e LDOS dos orbitais de fronteira HOMO, HOMO-1, LUMO e LUMO+1 da região especificada pela figura 3.
- Índices da MIE: η e Δ
- Energia de Hidratação
- Coeficiente de partição molecular octanol-água ($\log P$),

Utilizando todos esses descritores forma-se uma matriz de dados composta por 36 (moléculas) X 25 (variáveis). A partir desta matriz de dados foi realizado a PCA e HCA. Inicialmente é realizada uma seleção das variáveis que melhor separem as duas classes de moléculas utilizando o menor número de variáveis. Essa seleção de variáveis foi feita por verificação da distribuição das moléculas para determinado conjunto de descritores e observando se existe uma separação entre as moléculas ativas e inativas. A seleção é

realizada até se encontrar um número mínimo de descritores que sejam capazes de distinguir os dois grupos de moléculas.

A separação da classe (ativa e inativa) com 100% de acerto, em relação a atividade experimental, foi obtida utilizando os parâmetros da MIE ΔH , ηH , ΔL , ηL , e Log P. Na figura 12, pode-se observar que usando apenas 3 variáveis Log P, ΔL e ηL ou ΔH ou ηH é suficiente para separar precisamente as moléculas.

A figura 13 apresenta um dos resultados obtidos para a análise PCA. As moléculas são representadas por pontos em um plano de duas componentes principais (PC1 e PC2). As moléculas inativas (mol 1 a mol 6) estão nitidamente separadas das demais moléculas. Outras análises de PCA não revelaram uma discriminação com total acerto.

Fig.12: Análise Componente Principais: Gráfico dos Loadings.

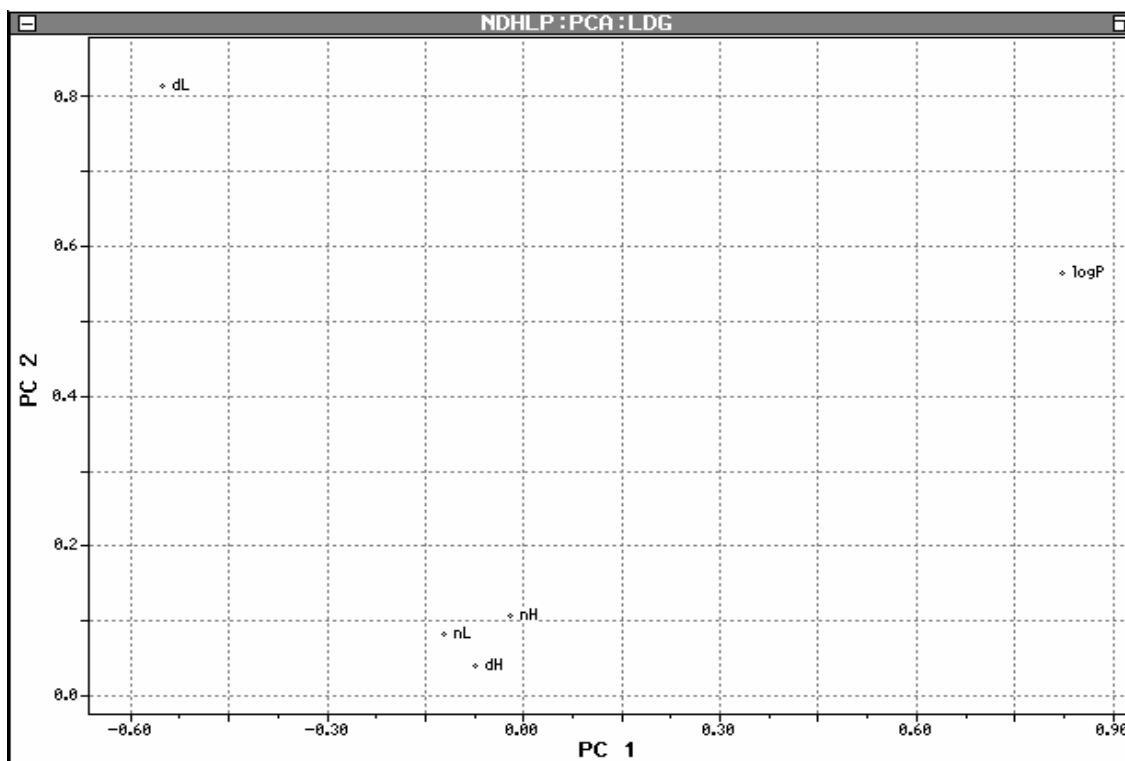
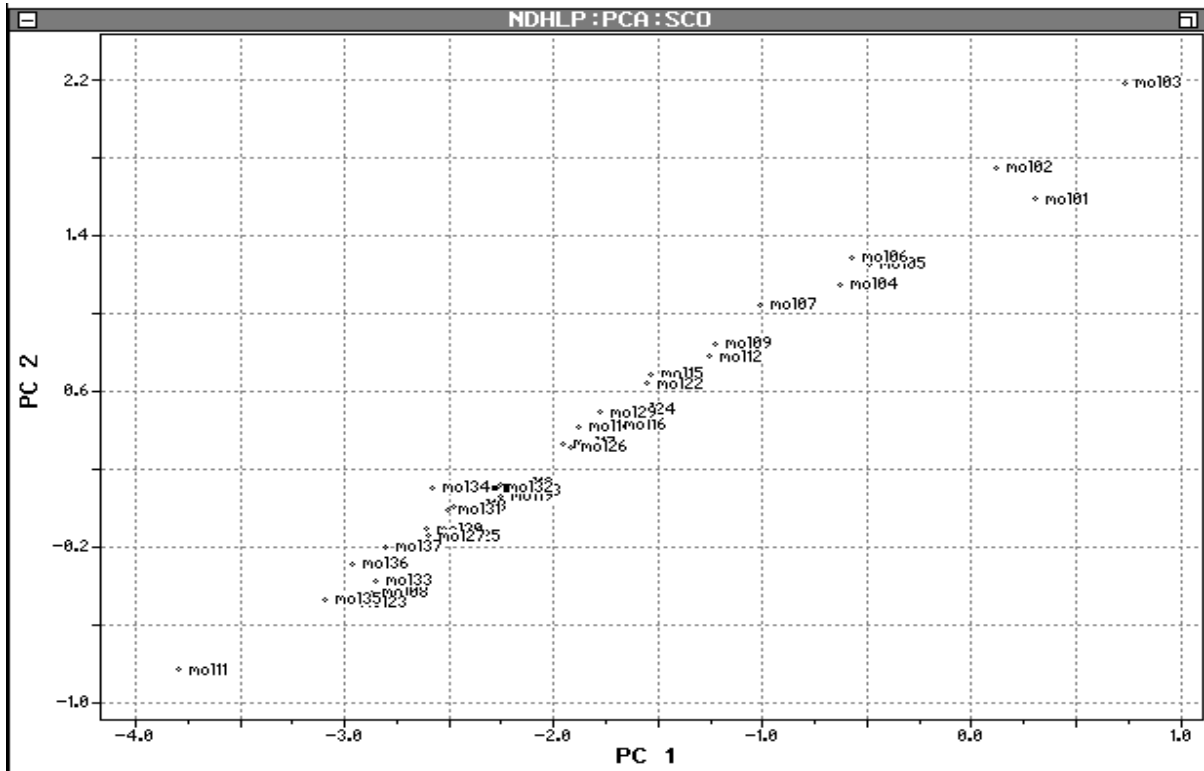


Fig.13: Análise Componente Principais: Gráfico dos Scores.



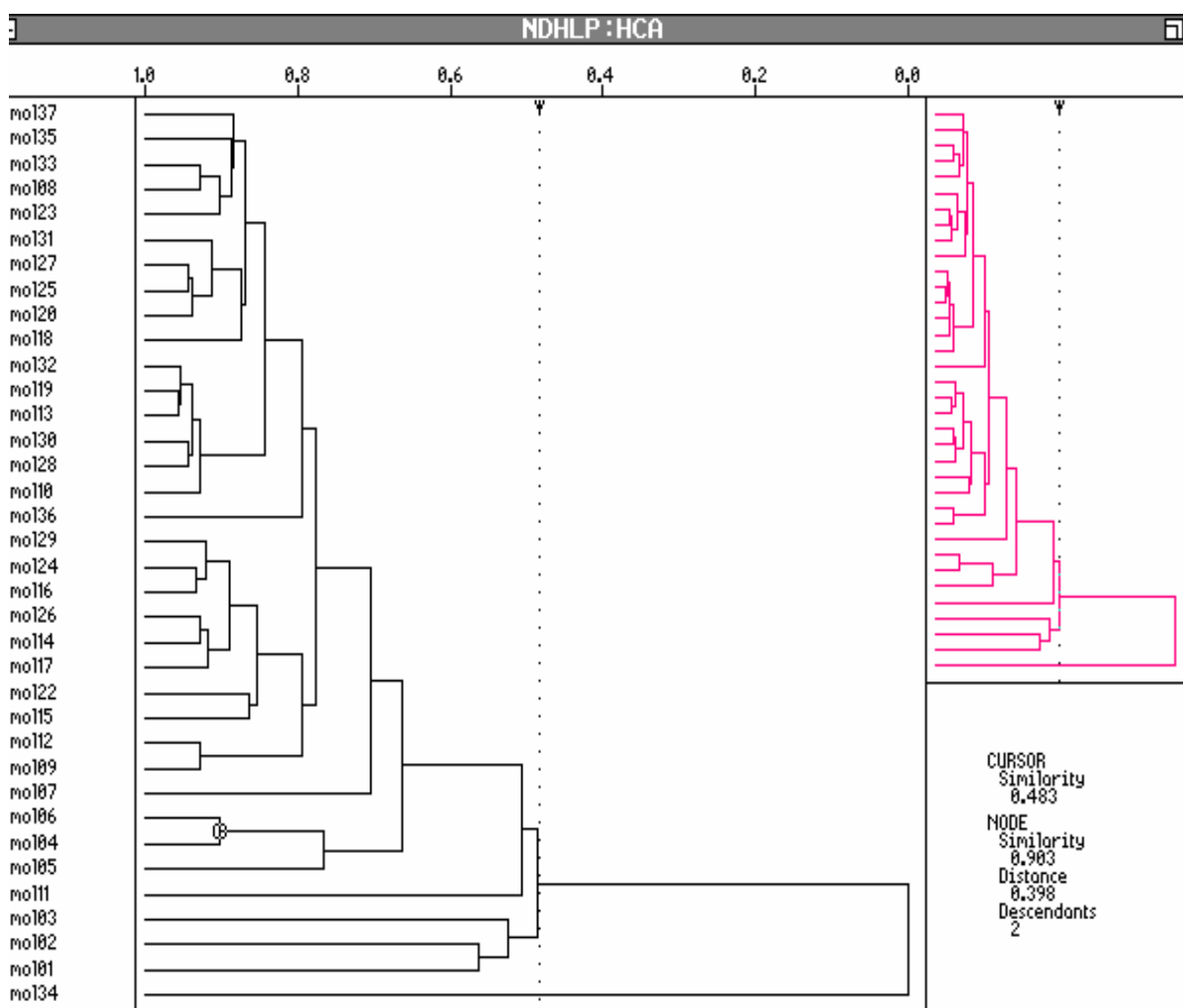
Os eixos PC1 e PC2 correspondem a 86.87% e 10.97% da variância do sistema respectivamente, e o plano PC1 versus PC2 conserva 97.84% da variância total dos dados originais, tabela 8.

Tab. 8: Variância das Componentes Principais PC1 e PC2.

Análise Componente Principais			
	Variância	Percentagem	Acumulativo
PC1	166.02	86.87	86.87
PC2	20.96	10.97	97.84

A figura 14 mostra os resultados obtidos da análise de HCA. Nesta observa-se que as moléculas 11 e 34, ativas, foram agrupadas junto com as moléculas inativas, isso equivale a um acerto de 94,4%. A formação de um único grupo para as moléculas ativas demonstra que os descritores adotados na classificação de atividades foram capazes de avaliar corretamente a similaridade que deve existir entre todos os compostos ativos. Os compostos inativos formaram um único grupo e apresentaram uma similaridade de 0,48 com o grupo ativo.

Fig.14: Dendograma da Análise Hierárquica de Grupo das 36 Moléculas.



VI. Discussão e Conclusões

Devido às dimensões do sistema molecular (graus de liberdade e número de estruturas a serem investigadas) o uso de método semi-empírico foi empregado para análise conformacional. Como os resultados obtidos por métodos semi-empíricos, que determinam as propriedades conformacionais e estereoeletrônicas, são afetados de modo imprevisível pelos parâmetros do método aplicado para resolução da estrutura eletrônica, justifica-se comparar os resultados obtidos por meio do método semi-empírico com os dados experimentais ou resultados obtidos por cálculos *Ab Initio*. A comparação da estruturas resolvidas pelos métodos *Ab Initio* e PM3 associado com busca conformacional, mostra que aparentemente os dois métodos produzem estruturas químicas semelhantes, mas com alguns átomos em posições diferentes, figura 7. Devido ao alto custo computacional o método *Ab Initio* não foi aplicado para uma detalhada pesquisa de conformeros. Isso revela que a busca conformacional é fundamental para se encontrar a estrutura molecular mais estável, mesmo utilizando-se para a resolução métodos *Ab Initio*. A varredura do espaço dos conformeros com o método Monte Carlo foi de grande importância para a obtenção da estrutura molecular mais estável, pois as estruturas moleculares finais encontradas são significativamente mais estáveis do que aquelas não submetidas à análise conformacional.

A região molecular escolhida para a aplicação da MIE foi uma região comum a todas as moléculas. Como o conceito de LDOS fornece informações eletrônicas (reatividade química, resposta óptica, etc.) de uma região molecular específica, é importante que uma região em comum entre as moléculas em estudo, seja usada para fins de comparação.

A análise dos índices eletrônicos η_L , Δ_L , η_H , Δ_H , isoladamente, não discriminam as moléculas ativas das inativas, figuras 9 e 10. Porém, quando dois desses índices são aplicados, resulta uma fácil regra, figura 11, que diferencia as moléculas quanto à sua atividade biológica. Com o uso dos índices Δ_L e η_L é possível discriminar os dois grupos moléculas quanto a sua atividade biológica com um acerto de 91,7%.

A aplicação de métodos quimiométricos PCA e HCA corrobora os resultados obtidos pela MIE. A PCA aplicada às 16 propriedades físico-químicas revela a importância dos descritores Δ e η . Após procura sistemática no espaço das amostras pode-se encontrar que com apenas três parâmetros, η_L , Δ_L e $\log P$, é possível separar as duas classes de moléculas com 100% de acerto, figura 10. Semelhantemente, a análise de HCA resulta num agrupamento das moléculas, porém com apenas 2 moléculas erroneamente agrupadas, figura 11.

Os orbitais não ocupados LUMO e LUMO+1, que compõem os índices críticos η_L e Δ_L , são mais importantes do que os orbitais ocupados HOMO e HOMO-1, para obtenção da regra com maior índice de acerto. Apesar dos orbitais não ocupados não terem um real significado físico, eles forneceram informação a respeito da atividade biológica. Embora os índices eletrônicos da MIE sejam capazes de qualificar as moléculas quanto à sua magnitude de sua atividade biológica, é difícil entender o resultado com uma explicação física sobre estes índices.

A teoria dos orbitais de fronteira procura prever as reatividades químicas a partir das propriedades dos reagentes através do estudo dos orbitais HOMO e LUMO. Sobreposição entre os orbitais HOMO e LUMO é um fator determinante em muitas reações. A razão disso, advém do fato que a forma dos orbitais HOMO e LUMO tem

aspectos semelhantes da densidade eletrônica total. O índice η pode ser interpretado como uma medida indireta da carga local, que se traduz como índice de reatividade química. Já o índice Δ , não está associado a qualquer propriedade física direta. No meio biológico as interações moleculares são estereoespecíficas. A estrutura 3D das biomoléculas grandes e pequenas, suas conformações, são fundamentais para a interação molecular. Isso quer dizer que alguma informação topológica ou estereoquímica deve estar contida nos índices da MIE.

Os índices de reatividades são usualmente classificados como eletrofílico ou nucleofílico, dependendo se a reação envolve um ataque eletrofílico ou nucleofílico. A energia dos orbital HOMO está relacionado ao potencial de ionização de uma molécula, enquanto a energia do LUMO relaciona-se com a afinidade eletrônica. A grandeza dessas quantidades são medidas da susceptibilidade da molécula perder um par de elétrons para um agente eletrofílico ou aceitar um par de elétrons de um agente nucleofílico. As DABQ são agentes alquilantes e age como um agente eletrofílico[13]. Este mecanismo de ação das DABQ corrobora a relevância do LUMO. Este fato está em concordância com o encontrado de que apenas os índices relacionados com o LUMO serem suficientes para discriminação das atividades biológicas das DABQ.

O estudo de estereoquímica biomolecular com métodos físicos, na tentativa de elucidar os mecanismos de interação, é um desafio da pesquisa física atual. O entendimento de como os parâmetros da MIE consegue classificar a atividade biológica das moléculas ainda é uma questão não resolvida, mas que desde já, mostra sua importância para o desenvolvimento e aprimoramento de fármacos com um menor custo computacional.

Em suma, os resultados obtidos com a aplicação dos descritores quânticos da MIE às moléculas derivadas das carboquinonas com atividade antineoplásica, apresentou excelente resultado, sendo eficiente na classificação da atividade biológica destas moléculas. Assim, esta metodologia se mostra promissora como ferramenta para aplicação no design de novas drogas com potencial para serem farmacologicamente ativas. Isso representa uma enorme economia de tempo e custo para que uma droga seja desenvolvida e colocada para uso clínico.

VI. APÊNDICES

Métodos Semi-Empíricos

Introdução

Métodos semi-empíricos foram desenvolvidos devido ao grande custo computacional de métodos *Ab Initio*. Como descrito, a maior dificuldade no cálculo *Ab Initio* são as integrais de dois elétrons que é da ordem de M^4 , sendo M o número de funções da base.

O primeiro passo para reduzir os cálculos é considerar apenas os elétrons de valência, os elétrons do caroço reduz a carga nuclear. Apenas um conjunto mínimo de funções é usado para os elétrons de valência. A eficácia dos métodos semi-empíricos não está no rigor da teoria, mas nos parâmetros ajustáveis que são melhorados para reproduzir as propriedades químicas. Como as integrais são negligenciadas e como as parametrizações são feitas, define o método semi-empírico.

Na teoria semi-empírica, geralmente escreve as funções de base com os símbolos μ , ν , λ e σ . Denotamos um orbital atômico com centro no átomo A como μ_A . Reescrevendo os elementos da matriz de Hartree-Fock com a abreviação da integral de dois elétrons como $\langle \mu\nu | \lambda\sigma \rangle$, resulta:

$$F_{\mu\nu} = h_{\mu\nu} + \sum_{\lambda\sigma}^{AO} D_{\lambda\sigma} [\langle \mu\nu | \lambda\sigma \rangle - \langle \mu\lambda | \nu\sigma \rangle] \quad (\text{Eq.A1})$$

$$h_{\mu\nu} = \langle \mu | h | \nu \rangle \quad (\text{Eq.A2})$$

Sabemos que a maioria das integrais de repulsão entre elétrons tem valores próximos de zero, principalmente aqueles que envolvem o produto $\mu_A \cdot \nu_B \cong 0$. As conseqüências desta aproximação que chamamos de Zero Differential Overlap (ZDO) são:

(1) A matriz de sobreposição S é reduzida a matriz unidade.

$$S_{\mu\nu} = \delta_{\mu\nu} \quad (\text{Eq.A3})$$

(2) As integrais de um elétron envolvendo três centros (dois da função de base e uma do operador) são zero.

(3) Todas as integrais de três e quatro centros, que são as integrais mais numerosas, são desprezadas.

$$\langle \mu\nu | \lambda\sigma \rangle = \delta_{\mu\nu} \delta_{\lambda\sigma} \langle \mu\mu | \lambda\lambda \rangle = \delta_{\mu\nu} \delta_{\lambda\sigma} \gamma_{\mu\lambda} \quad (\text{Eq.A4})$$

Onde $\gamma_{\mu\lambda} = \langle \mu\mu | \lambda\lambda \rangle$. Com a aproximação ZDO a matriz de Fock torna-se:

$$F_{\mu\mu} = H_{\mu\mu}^c - \frac{1}{2} D_{\mu\mu} \langle \mu\mu | \mu\mu \rangle + \sum_{\lambda} D_{\lambda\lambda} \langle \mu\mu | \lambda\lambda \rangle \quad (\text{Eq.A5})$$

$$F_{\mu\nu} = H_{\mu\nu}^c - \frac{1}{2} D_{\mu\nu} \langle \mu\mu | \nu\nu \rangle \quad (\text{Eq.A6})$$

Os métodos de aproximação diferem principalmente no grau com que a ZDO é aplicada para as integrais de repulsão eletrônica.

1. Complete Neglect of Differential Overlap (CNDO)

A aproximação CNDO foi introduzida em 1965 por Pople, Santry e Segal[26, 42].

O método trata apenas dos elétrons de valência. A Hamiltoniana dos elétrons de valência

H_{val} é:

$$H_{\text{val}} = \sum_{i=1}^{nv} \left[-\frac{1}{2} \nabla_i^2 + V(i) \right] + \sum_{i=1}^{nv} \sum_{j>i}^{nv} \frac{1}{r_{ij}} = \sum_{i=1}^{nv} H_{\text{val}}^c(i) + \sum_{i=1}^{nv} \sum_{j>i}^{nv} \frac{1}{r_{ij}} \quad (\text{Eq.A7})$$

$$H_{\text{val}}^c(i) = -\frac{1}{2} \nabla_i^2 + V(i)$$

nv é o número de elétrons de valência, $V(i)$ é a energia potencial do elétron de valência i no campo dos núcleos e dos elétrons do caroço, e $H_{val}^c(i)$ é hamiltoniano de um elétron.

A energia eletrônica molecular é:

$$E = 2 \sum_{i=1}^{nv/2} H_{val,ii}^c + \sum_{i=1}^{nv/2} \sum_{j=1}^{nv/2} (2J_{ij} - K_{ij}) + V_{cc} \quad (\text{Eq.A9})$$

$$V_{cc} = \sum_a \sum_{b>a} \frac{Z'_a Z'_b}{R_{ab}} \quad (\text{Eq.A10})$$

V_{cc} é a energia potencial de repulsão entre dois caroços e Z' é a carga nuclear reduzida do número de elétrons do caroço.

Os elementos da matriz de Fock é calculado pela equação A1 e A2, substituindo respectivamente $F_{\mu\nu}$ e $h_{\mu\nu}$ por $F_{val,\mu\nu}$ e $h_{val,\mu\nu}^c$. O método CNDO usa a aproximação ZDO que despreza todos os produtos dos orbitais atômicos (AO) de elétrons. Contudo, esta aproximação não é invariante para a rotação dos eixos. Para manter a invariância rotacional e hibridizacional quando a aproximação ZDO é aplicada, o método CNDO faz uma adicional aproximação de que as integrais de repulsão eletrônica de dois elétrons $\gamma_{\mu\lambda}$ dependam sobre qual átomo o orbital atômico μ e ν está centrado e não dependa da natureza dos orbitais. Para todos os AO de valência μ_A centrado no átomo A e todos os AO de valência ν_B centrado no átomo B, vale a expressão:

$$\langle \mu_A \mu_A | \nu_B \nu_B \rangle = \gamma_{\mu_A \nu_B} = \gamma_{AB} \quad (\text{Eq.A11})$$

Assim, γ_{AB} é uma média da repulsão eletrostática entre qualquer elétron sobre o átomo A e o elétron no átomo B. Para grandes distâncias interatômicas, R_{AB} , γ_{AB} é aproximadamente igual a R_{AB}^{-1} . O próximo passo é aplicar aproximações para os elementos do operador

$$H_{val}^c(i) = -\frac{1}{2}\nabla_i^2 + V(i) \quad (\text{Eq.A12})$$

Separando as contribuições de cada caroço

$$H^c = -\frac{1}{2}\nabla_i^2 + V_A(i) + \sum_{B \neq A} V_B(i) \quad (\text{Eq.A13})$$

Para os elementos $H_{\mu\nu}^c$

$$H_{\mu_A\nu_B}^c = \frac{1}{2}(\beta_A^0 + \beta_B^0)S_{\mu_A\nu_B} = \beta_{AB}^0 S_{\mu_A\nu_B} \quad \mu \neq \nu \text{ centrado em átomos diferentes} \quad (\text{Eq.A14})$$

$$H_{\mu_A\nu_A}^c = 0 \quad \mu \neq \nu \text{ centrado no mesmo átomo} \quad (\text{Eq.A15})$$

Os parâmetros β_A^0 e β_B^0 são escolhidos de modo que os coeficientes dos orbitais moleculares calculados pelo CNDO de moléculas diatômicas sejam iguais aos dos coeficientes dos orbitais moleculares calculados pelo método *Ab Initio*.

Para os elementos $H_{\mu\mu}^c$

$$H_{\mu\mu}^c(i) = \langle \mu_A | (-1/2)\nabla_i^2 + V_A(i) | \nu_A \rangle + \sum_{B \neq A} \langle \mu_A | V_B(i) | \nu_A \rangle = U_{\mu_A\nu_A} + \sum_{B \neq A} V_{AB}(i) \quad (\text{Eq.A16})$$

Devido a ortogonalidade dos AO, $U_{\mu\nu}$ é zero a menos que sejam os mesmos AO.

E os termos de dois centros $\langle \mu | V_B | \mu \rangle$ e $\langle \mu | V_B | \nu \rangle$ deve ser aproximado de tal modo que seja consistente como as integrais de dois elétrons são tratadas.

$$\langle \mu_A | (-1/2)\nabla^2 + V_A | \nu_A \rangle = \delta_{\mu\nu} \langle \mu_A | (-1/2)\nabla^2 + V_A | \nu_A \rangle = U_{\mu\mu} \quad (\text{Eq.A17})$$

$$U_{\mu\nu} = 0$$

Negligenciar a sobreposição diferencial monoatômico é o mesmo que tomar como zero $\langle \mu | V_B | \nu \rangle$. Para respeitar a condição de invariância requer que todos os elementos da diagonal $\langle \mu | V_B | \mu \rangle$ seja o mesmo para todos os AO sobre o átomo A.

$$\langle \mu_A | V_B | \nu_A \rangle = \delta_{\mu\nu} V_{AB} \quad (\text{Eq.A18})$$

$-V_{AB}$ é a interação de qualquer elétron de valência no átomo A com o caroço do átomo B.

$U_{\mu\mu}$ e V_{AB} são parametrizados da seguinte forma CNDO/2

$$V_{AB} = -Z'_B \gamma_{AB} \quad (\text{Eq.A19})$$

$$U_{\mu\mu} = -\frac{1}{2}(I_\mu + A_\mu) - (Z_A - 1/2)\gamma_{AA} \quad (\text{Eq.A20})$$

Sendo Z'_B a carga do caroço do átomo B, I_μ é o potencial de ionização e A_μ é afinidade eletrônica atômica, Z_A é a carga do átomo A.

2. Intermediate Neglect of Differential Overlap (INDO)

Proposto por Pople, Beveridge e Dobosh em 1967[26, 43], o método INDO é um melhoramento do CNDO. No INDO a sobreposição diferencial entre os AO sobre o mesmo átomo não são negligenciada as integrais de repulsão eletrônica de um centro, mas as integrais de repulsão eletrônica de dois centros são desprezadas.

Integrais de um elétron

$$\langle \mu_A | V_A | \mu_A \rangle = \delta_{\mu\nu} \sum_A \langle \mu_A | V_A | \mu_A \rangle \text{ é independente tipo de orbital (s ou p)} \quad (\text{Eq.A21})$$

Integrais de dois centros

$$\langle \mu_A \nu_B | \lambda_C \sigma_D \rangle = \delta_{\mu_A \lambda_C} \delta_{\nu_B \sigma_D} \langle \mu_A \nu_B | \mu_A \nu_B \rangle \text{ é independente tipo de orbital (s ou p)} \quad (\text{Eq.A22})$$

$$\langle \mu_A \nu_A | \mu_A \nu_A \rangle = \gamma_{AA} \quad (\text{Eq.A23})$$

$$\langle \mu_A \nu_B | \mu_A \nu_B \rangle = \gamma_{AB} \quad (\text{Eq.A24})$$

CNDO e INDO tornaram-se atualmente métodos obsoletos.

3. Neglect of Diatomic Differential Overlap (NDDO).

Desenvolvido por Pople Santry e Segal em 1965[26, 42], o método sugere que a sobreposição diferencial de somente AO centrados em diferentes átomos sejam negligenciados.

$$\langle \mu_A \nu_B | \lambda_C \sigma_D \rangle = \delta_{AC} \delta_{BD} \langle \mu_A \nu_B | \lambda_A \sigma \rangle \quad (\text{Eq.A25})$$

O método foi abandonado em 1977 quando Dewar e Thiel modificaram o método para o MNDO.

4. Modified Intermediate Neglect of Differential Overlap (MINDO/3)

É uma modificação do método INDO, desenvolvido por Dewar et al em 1975[44]. A parametrização de MINDO/3 contém variáveis diatômicas na integral de um elétron de dois centros, o parâmetro β_{AB} é derivado para todos os pares de átomos ligados.

$$\langle \mu_A | h | \nu_B \rangle = \langle \mu_A | (-1/2)\nabla^2 - V_A - V_B | \nu_B \rangle = S_{\mu\nu} \beta_{AB} (I_\mu + I_\nu) \quad (\text{Eq.A26})$$

MINDO/3 foi suplantado por melhores métodos semi-empírico sendo raramente usado na moderna química computacional.

5. Modificado NDDO: MNDO, AM1 e PM3.

MNDO, AM1 e PM3 são parametrizações do modelo NDDO e diferem somente na maneira como a repulsão caroço-caroço é tratada e como os parâmetros são determinados[23, 45]. Cada método considera somente as funções de valência s e p,

tomados do orbital tipo slater (STO) com os correspondentes expoentes ζ_s e ζ_p . O Hamiltoniano e os elementos da matriz de Fock são dados pelas equações A7 e A8.

Os elementos da matriz $H_{\mu_A \nu_B}^c$ (chamado de integral de ressonância do caroço) envolvendo AO centrado em diferentes átomos é proporcional a integral de overlap S.

$$H_{\mu_A \nu_B}^c = \frac{1}{2}(\beta_{\mu_A} + \beta_{\nu_B})S_{\mu_A \nu_B} \quad A \neq B \quad (\text{EqA27})$$

β são parâmetros para cada tipo de AO.

Os elementos de matriz $H_{\mu_A \nu_A}^c$ envolvendo diferentes AO centrado no mesmo átomo é

$$H_{\mu_A \nu_A}^c = \left\langle \mu_A \left| (-1/2)\nabla^2 + V_A \right| \nu_A \right\rangle + \sum_{A \neq B} \left\langle \mu_A \left| V_B \right| \nu_A \right\rangle \quad \text{com } V_B = -Z'_B / r_{1B} \quad (\text{Eq.A28})$$

A primeira integral é zero e a integral $\langle \mu_A | V_B | \nu_A \rangle$ é avaliada como

$$H_{\mu_A \nu_A}^c = - \sum_{A \neq B} Z'_B \left\langle \mu_A \nu_A \left| s_B s_B \right\rangle \quad A=B \text{ e } \mu_A \neq \nu_A \quad (\text{Eq.A29})$$

Onde s_B é um orbital de valência no átomo B e as integrais de repulsão eletrônica (IRE)

$\langle \mu_A \nu_A | s_B s_B \rangle$ é avaliada por um aproximado procedimento.

Para os elementos da matriz $H_{\mu_A \mu_A}^c$ contendo o mesmo AOs duas vezes é

$$H_{\mu_A \mu_A}^c = \left\langle \mu_A \left| (-1/2)\nabla^2 + V_A \right| \mu_A \right\rangle + \sum_{A \neq B} \left\langle \mu_A \left| V_B \right| \mu_A \right\rangle = U_{\mu_A \mu_A} - \sum_{A \neq B} Z'_B \left\langle \mu_A \mu_A \left| s_B s_B \right\rangle \quad (\text{EqA30})$$

O método estabelece a primeira integral como parâmetro $U_{\mu\mu}$ e a segunda integral com a aproximação da integral de repulsão eletrônica.

As IRE são avaliada aplicando a aproximação ZDO que resulta zero para as IRE de três e quatro centros. As IRE de um centro em vez de serem calculadas, são parametrizadas de forma que se ajustem bem aos dados experimentais de energia de

ionização do estado de valência. As IREs de um centro são integrais tipo Coulombiana

$g_{\mu\nu} = \langle \mu_A \mu_A | \nu_A \nu_A \rangle$ ou integrais de troca da forma $h_{\mu\nu} = \langle \mu_A \nu_A | \mu_A \nu_A \rangle$. Para um átomo

com apenas AO de valência s e p existem cinco IRE de um centro.

$$\begin{aligned} g_{ss} &= \langle ss | ss \rangle \\ g_{sp} &= \langle sp | sp \rangle \\ g_{pp} &= \langle pp | pp \rangle \\ g_{p2} &= \langle pp' | pp' \rangle \\ h_{sp} &= \langle ss | pp \rangle \end{aligned} \quad (\text{Eq.A31})$$

O índice(') indica função tipo p que está ao longo de um eixo diferente do primeiro. Os valores das IREs encontrado por ajuste de dados atômicos são menores do que os obtidos por direta integração, devido ao efeito de correlação que atua sobre os elétrons.

Para o sistema de base sp há 22 diferentes IRE de dois centros. As IREs de dois centros não são avaliadas por direta integração, mas são modeladas como interação de multipolo. Elas representam a energia de interação entre a distribuição de carga $\mu_A \nu_A$ no átomo A e $\lambda_B \sigma_B$ no átomo B. Classicamente é igual a soma sobre todas as interações entre o momento de multipolo M_{lm} das duas distribuições de cargas (sendo os sufixos l e m especifica a ordem e orientação do multipolo). Baseado neste clássico conceito, as IREs de dois centros são expandidas em termos da interação de multipolo[46].

$$\langle \mu_A \nu_A | \lambda_B \sigma_B \rangle = \sum_{l1} \sum_{l2} \sum_m [M_{l1m}^A, M_{l2m}^B] \quad (\text{Eq.A32})$$

Cada multipolo é representado por uma apropriada configuração $[M_{lm}]$ de 2^l cargas pontuais de magnitude $e/2^l$, com a separação de carga D_l . A interação entre os multipolos é calculada por aplicar uma apropriada fórmula semi-empírica a cada uma das interações

entre os pontos de carga nas duas configurações e somando sobre todos as interações das cargas pontuais. Expressando a distância entre as cargas pontuais i e j na configuração de interação do átomo A e B por R_{ij} , obtemos:

$$[M_{l1m}^A, M_{l2m}^B] = \frac{e^2}{2^{l1+l2}} \sum_{i=1}^{2^{l1}} \sum_{j=1}^{2^{l2}} f_l(R_{ij}) \quad (\text{Eq.A33})$$

Onde $f_l(R_{ij})$ é uma expressão semi-empírica com a condição de contorno $R_{AB} \rightarrow 0$ e $R_{AB} \rightarrow \infty$ tem que convergir aos clássicos valores de interação. Maiores detalhes do cálculo destas funções veja referências[14, 45, 46].

O termo de repulsão caroço-caroço, expressada como $Z'_A Z'_B / R_{AB}$, devido a aproximação NDDO resulta em uma repulsão entre os átomos mesmo quando suas funções de onda não sobrepõe. Consequentemente, o termo caroço-caroço deve ser modificado e considerado. A específica forma funcional da interação caroço-caroço define o método MNDO ou AM1 ou PM3.

5.1 Modified Neglect of Diatomic Overlap (MNDO)[45]

A repulsão caroço-caroço do modelo MNDO é:

$$V_{nn}^{MNDO}(A, B) = Z'_A Z'_B \langle s_A s_B | s_A s_B \rangle (1 + e^{-\alpha_A R_{AB}} + e^{-\alpha_B R_{AB}}) \quad (\text{Eq.A34})$$

Onde os expoentes α são parâmetros ajustáveis. Interação O-H e N-H são tratados de modo diferente, (Eq.A35).

$$V_{nn}^{MNDO}(A, H) = Z'_A Z'_H \langle s_A s_H | s_A s_H \rangle (1 + (e^{-\alpha_A R_{AH}} / R_{AH}) + e^{-\alpha_H R_{AH}}) \quad (\text{Eq.A35})$$

MNDO também usa a aproximação $\zeta_s = \zeta_p$ para alguns átomos leves. Os parâmetros g_{ss} , g_{sp} , g_{pp} , g_{p2} e h_{sp} são obtidos a partir de dados de espectros atômicos e os outros parâmetros de outros dados moleculares.

5.2 Astin Model 1 (AM1)

Em 1985 Dewar et all[29], desenvolve uma versão melhorada do método MNDO, chamado AM1. MNDO tem tendência de superestimar a repulsão entre os átomos. Para a correção foi adicionada termos gaussiano ao termo caro-caroço. A única diferença entre AM1 e MNDO é que os expoentes dos orbitais de valência, ζ_s e ζ_p , podem ser diferentes para o mesmo átomo. A função de repulsão caroço-caroço é

$$V_{nn}^{AM1}(A, B) = V_{nn}^{MNDO}(A, B) + \frac{Z'_A Z'_B}{R_{AB}} \left(\sum_k a_{kA} e^{-b_{kA}(R_{AB}-c_{kA})^2} + \sum_k a_{kB} e^{-b_{kB}(R_{AB}-c_{kB})^2} \right) \quad (\text{Eq.A36})$$

As constantes a_k , b_k e c_k são ajustadas aos dados moleculares e k varia entre 2 e 4 dependendo do tipo de átomo.

5.3 Parametric Method 3 (PM3)

Stewart em 1989 reparametrizou AM1, dando origem ao método PM3[15]. As IRE de um centro são consideradas como parâmetros em vez de serem obtidas por dados de espectro atômico. Todos os parâmetros são otimizados simultaneamente, incluindo os termos de dois elétrons A expressão da repulsão caroço-caroço (Eq.A36) contem apenas 2 gaussianas por átomo. O método PM3 é essencialmente o AM1 com todos os parâmetros otimizados.

As maiores limitações dos métodos MNDO, AM1 e PM3 é não usar orbitais d, e por isso não resolve bem átomos como enxofre e metais de transição. Thiel e Vityuk em 1996, incluíram orbitais d ao método MNDO derivando o método MNDO/d. Desde que PM3 tem mais parâmetros ajustáveis do que AM1 e MNDO espera-se que os erros diminuam na ordem MNDO>AM1>PM3.

REFERÊNCIAS

1. Hansch, C. and T. Fujita, *Rho-Sigma-Pi Analysis. Method for Correlation of Biological Activity + Chemical Structure*. Journal of the American Chemical Society, 1964. **86**(8): p. 1616-&.
2. Free, S.M. and J.W. Wilson, *Mathematical Contribution to Structure-Activity Studies*. Journal of Medicinal Chemistry, 1964. **7**(4): p. 395-&.
3. Dillon, W.R.G., Mathew, *Multivariate Analysis Methods and Applications*. 1984, New York: John Willey & Sons. 1-52.
4. Beeb, K.R.P., R. J.; Seasholtz, M. B., *Chemometrics: A Pratical Guide*. 1998, New York: Wiley.
5. Sharaf, M.A.I., D. L.; Kowalski, B. R., *Chemometrics*. 1986, New York: Wiley-Interscience.
6. Jackson, J.E., *A User's Guide to Principal Components*. 1991, New York: Willey Interscience. 5-62.
7. Geladi, P. and B.R. Kowalski, *Partial Least-Squares Regression - a Tutorial*. Analytica Chimica Acta, 1986. **185**: p. 1-17.
8. Cramer, R.D., D.E. Patterson, and J.D. Bunce, *Comparative Molecular-Field Analysis (Comfa).1. Effect of Shape on Binding of Steroids to Carrier Proteins*. Journal of the American Chemical Society, 1988. **110**(18): p. 5959-5967.
9. Devillers J.; Balaban, A.T., *Topological Indices and Related Descriptors in QSAR and QSPR*. 1999, Amsterdam: Gordon and Breach Science.
10. Barone, P.M., A. Camilo, Jr., and D.S. Galvao, *Theoretical Approach to Identify Carcinogenic Activity of Polycyclic Aromatic Hydrocarbons*. Physical Review Letters, 1996. **77**(6): p. 1186-1189.

11. Coluci, V.R.B., S.F.; Vendrame, R.; Barone, P.M.V.; Galvao, D.S., *Teoria Eletronica do Cancer e a Origem da Vida: Base para o Desenvolvimento de Parametros "Universais " para Metodos de Estrutura-Atividade.*, ed. E.B.E.E. 2002. 2003, Campinas. 202-223.
12. Gilman, A.G.L., L.E.; Hardman, N., *As Bases Farmacológica da Terapêutica*. 10 ed. 2003, New York: McGraw Hill Interame. 904.
13. Hargreaves, R.H.J., J.A. Hartley, and J. Butler, *Mechanisms of action of quinone-containing alkylating agents: DNA alkylation by aziridinyquinones*. *Frontiers in Bioscience*, 2000. **5**: p. E172-E180.
14. Dewar, M.J.S. and W. Thiel, *Semiempirical Model for 2-Center Repulsion Integrals in Nddo Approximation*. *Theoretica Chimica Acta*, 1977. **46**(2): p. 89-104.
15. Stewart, J.J.P., *Optimization of Parameters for Semiempirical Methods. I. Method*. *Journal of Computational Chemistry*, 1989. **10**(2): p. 209-220.
16. Stewart, J.J.P., *Optimization of Parameters for Semiempirical Methods. 2. Applications*. *Journal of Computational Chemistry*, 1989. **10**(2): p. 221-264.
17. Mannhold, R.K.-L., P.; Timmeman, H., *Chemometric Methods in Molecular Design*. *Methods and Principles in Medicinal Chemistry*, ed. H.v.d. Waterbeemd. Vol. 2. 1994, Weinheim: VCH. 359.
18. Yoshimoto, M.M., H.; Nakao, H.; Shinkai, K; Arakawa, M., *Quantitative Structure-Activity Relationships in 2,5-Bis(1-aziridiny)-p-benzoquinone Derivatives against Leukemia L1210*. *J. Med. Chem*, 1979. **22**: p. 491-496.
19. SPARTAN, Wavefunction Inc ([http://:www.wavefun.com](http://www.wavefun.com)).

20. Cyrillo, M. and D.S. Galvao, *Structure-activity relationship study of some inhibitors of HIV-1 integrase*. Journal of Molecular Structure-Theochem, 1999. **464**(1-3): p. 267-272.
21. *HyperChem Computacional Chemistry*. 1996: Hypercube Inc.
22. *Einsight*. 1991, Infometrix Inc.: Seattle.
23. Stewart, J.J.P., *Special Issue - Mopac - a Semiempirical Molecular-Orbital Program*. Journal of Computer-Aided Molecular Design, 1990. **4**(1): p. 1-45.
24. Roothaan, C.C.J., *New Developments in Molecular Orbital Theory*. Reviews of Modern Physics, 1951. **23**(2): p. 69-89.
25. Levine, I.N., *Quantum Chemistry*. 2000, New Jersey: Prentice Hall.
26. Pople, J.A.B., D.L., *Approximate Molecular Orbital Theory*. 1970, New York: Mcgraw-Hill.
27. Jensen, F., *Introduction to Computacional Chemistry*. 1999, Chichester: John Wiley & Sons.
28. Szabo, A.O., N. S., *Modern Quantum Chemistry*. 1989, New York: Mcgraw-Hill.
29. Dewar, M.J.S.Z., E.G.; Healy, E.F.; Stewart, J.J.P., *AMI: A new General Purpose Quantum Mechanical Molecular Model*. American Chemical Society, 1985. **107**: p. 3902-3909.
30. Leo, A.J., *Calculating Log P(Oct) from Structures*. Chemical Reviews, 1993. **93**(4): p. 1281-1306.
31. Perlman, R.S., *Molecular Surface areas and volumes and their structure-activity relationships*. Physical Chemical Properties of Drugs. 1980, New York: Dekker Publ. 361-385.

32. Bondi, A., *Van Der Waals Volumes + Radii*. Journal of Physical Chemistry, 1964. **68**(3): p. 441-&.
33. Martin, Y.C., *The quantitative Relationships Between pKa, Ionization and Drug Potency: Utility of Model-Based Equations*. Physical Chemical Properties of Drugs, ed. S.H.S. Yalkolvisky, A. A.; Valvani, S. C. 1980, New York: Marcel Dekker.
34. Lien, E.J., et al., *Use of Dipole-Moment as a Parameter in Drug Receptor Interaction and Quantitative Structure Activity Relationship Studies*. Journal of Pharmaceutical Sciences, 1982. **71**(6): p. 641-655.
35. Miller, K.J., *Calculation of the Molecular Polarizability Tensor*. Journal of the American Chemical Society, 1990. **112**(23): p. 8543-8551.
36. Miller, K.J., *Additivity Methods in Molecular Polarizability*. Journal of the American Chemical Society, 1990. **112**(23): p. 8533-8542.
37. Hansch, C., et al., *Aromatic Substituent Constants for Structure-Activity Correlations*. Journal of Medicinal Chemistry, 1973. **16**(11): p. 1207-1216.
38. Fujita, T., T. Nishioka, and M. Nakajima, *Hydrogen-Bonding Parameter and Its Significance in Quantitative Structure-Activity Studies*. Journal of Medicinal Chemistry, 1977. **20**(8): p. 1071-1081.
39. Pearson, R.G., *Hard and Soft Acids and Bases*. Journal of the American Chemical Society, 1963. **85**(22): p. 3533-&.
40. CHARMM - Chemistry at HARvard Molecular Mechanics -
<http://www.charmm.org>.

41. Iczkowski, R. and J.L. Margrave, *Electronegativity*. Journal of the American Chemical Society, 1961. **83**(17): p. 3547-&.
42. Pople, J.A., D.P. Santry, and G.A. Segal, *Approximate Self-Consistent Molecular Orbital Theory.I. Invariant Procedures*. Journal of Chemical Physics, 1965. **43**(10P2): p. S129-&.
43. Pople, J.A., Beveridge, D.L., and P.A. Dobosh, *Approximate Self-Consistent Molecular-Orbital Theory.5. Intermediate Neglect of Differential Overlap*. Journal of Chemical Physics, 1967. **47**(6): p. 2026-&.
44. Bingham, R.C., M.J.S. Dewar, and D.H. Lo, *Ground-States of Molecules.25. Mndo-3 - Improved Version of Mndo Semiempirical Scf-Mo Method*. Journal of the American Chemical Society, 1975. **97**(6): p. 1285-1293.
45. Dewar, M.J.S. and W. Thiel, *Ground-States of Molecules.38. Mndo Method - Approximations and Parameters*. Journal of the American Chemical Society, 1977. **99**(15): p. 4899-4907.
46. Parr, R.G., *3 Remarks on Molecular Orbital Theory of Complex Molecules*. Journal of Chemical Physics, 1960. **33**(4): p. 1184-1199.