

*Este exemplar corresponde à redação
final da tese de mestrado defendi-
da pelo aluno Márcio Cyrillo e
aprovada pela comissão julgadora
2/6/98 B.S.H.*

**ESTUDO ESTRUTURA-ATIVIDADE
DE ALGUNS INIBIDORES DA ENZIMA
INTEGRASE DO VÍRUS HIV-1**

Por: Márcio Cyrillo ✓
Orientador: Douglas S. Galvão

Dissertação apresentada como parte
dos requisitos para a obtenção do
título de Mestre em Física pelo
Instituto de Física "Gleb Wataghin"
da Universidade Estadual de Campinas
UNICAMP.

Campinas, Junho de 1998.

UNIDADE	IFGW
N.º CHAMADA:	
V.	EX.
TOMBO BC/	34995 / 1140
PROD.	395/98
C	<input type="checkbox"/>
D	<input checked="" type="checkbox"/>
PREÇO	R\$ 11,00
DATA	12/06/98
N.º CPD	6400/160.FS-6

**FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO IFGW- UNICAMP**

C989e **Cyrillo, Márcio**
Estudo estrutura-atividade de alguns inibidores de enzima
integrase do vírus HIV-1 / Márcio Cyrillo. - Campinas,
SP: [s.n.], 1998.

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

1. AIDS. 2. HIV. 3. Química quântica. II. Universidade
Estadual de Campinas, Instituto de Física "Gleb Wataghin".
III. Título.

Instituto de Física "Gleb Wataghin"

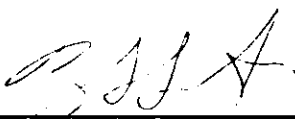
UNIVERSIDADE ESTADUAL DE CAMPINAS

Secretaria de Pós-Graduação

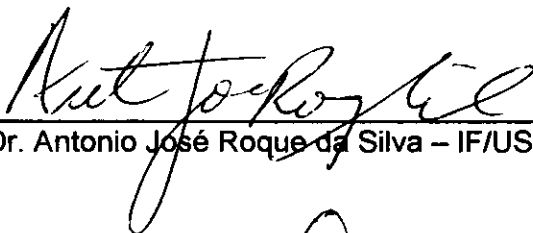
CAIXA POSTAL 6165
CEP 13083-970 – Campinas-SP - Brasil
Tel: (019) 788-5305
Fax: (019) 289-2424

MEMBROS DA COMISSÃO JULGADORA DA TESE DE MESTRADO DE MÁRCIO CYRILLO APRESENTADA E APROVADA AO INSTITUTO DE FÍSICA "GLEB WATAGHIN", DA UNIVERSIDADE ESTADUAL DE CAMPINAS, EM 08/06/98.

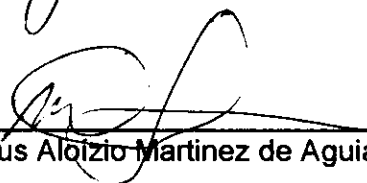
COMISSÃO JULGADORA:



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



Prof. Dr. Antonio José Roque da Silva – IF/USP



Prof. Dr. Marcus Aloizio Martinez de Aguiar – IFGW/UNICAMP

ESTUDO ESTRUTURA-ATIVIDADE DE ALGUNS INIBIDORES DA ENZIMA INTEGRASE DO VÍRUS HIV-1

Por: Márcio Cyrillo

Orientador: Douglas S. Galvão

Campinas, Junho de 1998.

Gostaria de agradecer primeiramente a toda minha família, sem o apoio da qual muito provavelmente eu não teria caminhado para tão longe dos caminhos que me imaginava quando criança. O primeiro computador, comprado com dificuldade, certamente fez valer o investimento, apesar dos 16Kb de memória, que hoje as calculadoras mais simples possuem. Em especial, agradeço meus irmãos Marcelo e Maciel pela força e maturidade mostradas ao longo dos anos e que me servem de referência. A minha irmã Eliana pela ajuda nos momentos difíceis. À Cássia pela paciência ou falta dela e proximidade constante.

A meu orientador Douglas Galvão, não só pelo grande apoio, incentivo e demonstração de competência durante todas as fases deste trabalho, mas também pela oportunidade que me proporcionou de trabalhar nesta área fascinante, onde a interface entre as ciências é tênue e os desafios grandes e motivadores.

A todos os membros do Grupo de Sólidos Orgânicos e Novos Materiais pelo apoio, companheirismo e disposição em testar os meus programas. Em especial a Larissa Lima, pelas críticas construtivas que teceu em relação a eles.

Aos membros do coral Gilberto Mendes, em especial a Ivaldo Pessôa, pela amizade e momentos valiosos que tivemos juntos.

Aos membros das secretarias do DFA e da pós-graduação do IFGW por todo o apoio operacional que proporcionaram.

Ao CNPq pelo apoio financeiro.

Enfim, a todos que de alguma forma contribuíram para a realização deste trabalho, frase utilizada por aqueles que temem o esquecimento de alguém.

Obscurum per obscurius
Ignotum per ignotius[†]

Divisa alquímica.

[†] Do Latim: “Ir para o obscuro e o desconhecido por aquilo que é ainda mais obscuro e desconhecido”.

A meu Pai, minha irmã e minha Cássia,
minhas três melhores referências de vida.

Índice	
Agradecimentos	III
Resumo	IX
<i>Abstract</i>	X
Introdução	1
Capítulo I – AIDS, HIV e a enzima integrase	8
I.1 Introdução	8
I.2 A anatomia do HIV-1	10
I.2.1 Os vírus	10
I.2.2 O HIV	11
I.2.3 A capa protetora	12
I.2.4 O núcleo	12
I.3 O processo de replicação do HIV-1 e sua inibição	13
I.3.1 A invasão	15
I.3.2 A transcrição do RNA	16
I.3.3 A fusão do genoma viral com o da célula	16
I.3.4 A maturação dos novos vírus	18
I.4 Pesquisas com inibidores	19
I.5 Considerações finais	21
Referências	23
Capítulo II – Teoria de Índices Eletrônicos	25
II.1 Introdução	25
II.2 Metodologia e cálculos utilizando o método de Hückel	28
II.3 Cálculos utilizando o método semi-empírico PM3	32
II.4 Conclusões	35
Referências	39

Capítulo III – Problema e metodologia	41
--	-----------

III.1 Introdução	41
III.2 Os compostos inibidores de Integrase	42
III.3 Métodos de Química Computacional	48
III.3.1 Mecânica Molecular	50
III.3.2 Métodos mecânico-quânticos	53
III.4 A adaptação da teoria de índices eletrônicos	56
Referências	61

Capítulo IV – Análise conformacional: software HyperSpin	62
---	-----------

IV.1 Introdução	62
IV.2 HyperChem	63
IV.3 A busca conformacional refinada	65
IV.4 HyperSpin: o programa para análise conformacional	67
IV.5 Versões e componentes do HyperSpin	70
IV.5.1 HyperSpin 1.0 (16-bit)	70
IV.5.2 HyperSpin 2.0 (32-bit)	73
IV.5.3 Script Editor	77
IV.5.4 HyperGraph	86
IV.5.5 HyperRemote	88
IV.5.6 LogViewer	90
IV.6 Análises conformacionais: Mecânica Molecular vs. Semi-empírico	92
IV.7 Usuários e sites para download	95
Referências	98

Capítulo V – Cálculo da densidade de estados: software Chem2Pac	99
--	-----------

V.1 Introdução	99
V.2 O método de Hückel e a Densidade Local de Estados Eletrônicos (LDOS)	100
V.2.1 O método HMO	100
V.2.2 Definições de DOS e LDOS	104
V.3 O cálculo LDOS do ponto de vista computacional	107
V.4 O programa Chem2Pac	112
V.4.1 O preparo de inputs para o MOPAC	114
V.4.2 Cálculos LDOS	120
V.5 Usuários e implementações futuras	126
Referências	128

Capítulo VI – Resultados e Conclusões	129
VI.1 Introdução	129
VI.2 Os trabalhos com os monômeros	129
VI.2.1 O preparo das moléculas	131
VI.2.2 Cálculos e resultados	133
VI.2.3 As regras	139
VI.3 Os trabalhos com os dímeros	140
Conclusões	150
Perspectivas futuras	152
Referências	154
Apêndice A – Métodos de Mecânica Quântica	155
A.1 Aproximação de Hartree-Fock	155
A.2 Hartree-Fock restrito (camada fechada) – Equações de Roothaan	163
A.2.1 A introdução de uma base	164
A.2.2 A densidade de carga	165
A.2.3 Equações de Roothaan	167
A.3 Métodos semi-empíricos	171
A.3.1 O método de Hückel Estendido (EHT)	171
A.3.2 Aproximações NDO	174
A.3.3 Aproximações MNDO, AM1 e PM3	175
Referências	181
Apêndice B – Rotinas DDE para o HyperChem	182
Referências	191

Adendo de primeiro de Julho de 1998:

A versão eletrônica de todo o manuscrito desta tese está disponível, juntamente com os slides apresentados durante a sua defesa no dia 8 de Junho, pelo seguinte endereço:

<http://www.ifi.unicamp.br/~cyrillo/mestrado/index.htm>

Recentemente uma nova e simples metodologia foi proposta com o objetivo de correlacionar índices eletrônicos à atividade biológica de alguns compostos carcinogênicos denominados Hidrocarbonetos Aromáticos Policíclicos (HAPs ou PAHs). Tal metodologia baseia-se na análise da separação em energia entre os chamados orbitais de fronteira e das suas contribuições relativas para a densidade local de estados calculada sobre regiões específicas das moléculas. No presente trabalho, adaptamos essa metodologia ao estudo de alguns novos compostos inibidores da proteína integrase do vírus HIV-1 (*Human Immunodeficiency Virus type 1* ou Vírus da Imunodeficiência Humana tipo 1), uma das enzimas essenciais à replicação do mesmo. Nossos cálculos foram efetuados utilizando o método semi-empírico PM3 (*Parametric Method 3*) disponível no pacote MOPAC, com o auxílio de dois outros programas por nós independentemente desenvolvidos. Nossos resultados mostram que, similarmente ao caso dos PAHs, é possível estabelecer regras simples correlacionando a atividade biológica de inibição da integrase a certos índices eletrônicos.

Recently a new and simple methodology has been proposed to correlate electronic indices to the biological activity of a class of carcinogenic Polycyclic Aromatic Hydrocarbons (PAHs). This methodology is based on the energy separation values between frontier orbitals and in their relative contribution to the local density of electronic states over specific molecular regions. In this work we adapted this methodology to the study of some new HIV-1 (*Human Immunodeficiency Virus type 1*) Integrase inhibitors. Integrase is a HIV enzyme essential for effective viral replication. Our calculations were carried out using the semi-empirical Parametric Method 3 (PM3) available in MOPAC package as well as two other independent programs developed as part of this work. Our results show that similarly to the PAHs it is possible to derive very simple rules correlating inhibitory integrase activity to electronic indices.

A rápida disseminação da Síndrome da Imunodeficiência Adquirida (AIDS ou SIDA)^[1] e as dificuldades encontradas no desenvolvimento de drogas eficientes e baratas e/ou uma vacina para a mesma, têm estimulado uma intensa pesquisa tanto acadêmica como industrial.

Acredita-se que o agente causador da AIDS seja o Vírus da Imunodeficiência Humana (HIV)^[2], um retrovírus. Retrovírus são vírus que carregam seu material genético na forma de RNA, que precisa ser transcrito reversamente em DNA complementar antes que o processo usual de síntese de proteínas possa ocorrer^[3]. Este aspecto tem sido largamente explorado do ponto de vista farmacológico, pois uma grande variedade de drogas anti-HIV baseia-se na inibição das enzimas essenciais ao ciclo de vida do vírus.

Entre tais enzimas, três foram identificadas como sendo essenciais ao processo de replicação virótica, e têm sido os principais alvos dos recentes desenvolvimentos de fármacos anti-AIDS: **transcriptase reversa**, **protease** e **integrase**^[4]. As duas primeiras têm sido amplamente estudadas com inúmeros inibidores aprovados pela FDA (*Federal Drug Administration*) e disponíveis no mercado. Como exemplo temos o **AZT** (*3'-azido-3'-deoxythymidine*)^[4] e **indinavir**^[5].

A terceira e menos estudada enzima é a integrase, que se tornou conhecida por seu papel fundamental no processo de replicação do HIV^[6,7]. Tal característica, em associação ao fato de não ser nativa ao corpo humano, torna-a um alvo muito importante para um ataque seletivo e eficiente ao vírus.

Com o aumento da expectativa de vida dos pacientes de AIDS, o alto poder tóxico dos atuais inibidores comercialmente disponíveis e o desenvolvimento de resistência por parte do vírus tornaram-se fatores de extrema importância^[8]. Uma solução promissora para tais problemas tem sido o uso de coquetéis de drogas capazes de atacar simultaneamente as três diferentes enzimas virais citadas anteriormente.

No final de 1996 foi anunciado na XI conferência internacional sobre AIDS, realizada em Vancouver, casos de pacientes nos quais o nível de infecção pelo HIV tinha caído a índices quase nulos^[9]. Eles haviam recebido alguns meses de terapia com os coquetéis disponíveis na época, que incluíam pelo menos um tipo de inibidor de protease^[10]. Nos primeiros seis meses de 1997, entretanto, constatou-se que aquele 1% de vírus que restara nos organismos dos doentes voltara a se multiplicar. Desta vez em linhagens mais resistentes ao coquetel de drogas^[11].

A busca por novos inibidores continua e aqueles baseados na inibição da enzima integrase têm se mostrado os mais promissores para que novos coquetéis mais efetivos possam ser desenvolvidos.

A integrase é uma enzima que catalisa a integração de uma cópia do DNA viral, sintetizado previamente pela enzima transcriptase reversa, no núcleo das células do sistema imunológico, os alvos do vírus HIV^[12]. Sua atividade farmacológica *in vitro* tem sido investigada nos últimos anos por muitos grupos de pesquisa e classificada através do índice IC_{50} . Este mede a concentração efetiva do composto inibidor, necessária para inibir 50% de uma determinada amostra de integrase^[12].

Dos estudos mais recentes sobre os inibidores de integrase, distingue-se claramente um padrão estrutural específico, onde a maioria dos compostos

apresenta múltiplas unidades aromáticas. Este padrão pode ser descrito de maneira geral como consistindo de duas arilas* separadas por um conector apropriado. Pelo menos uma delas deve conter o padrão 1,2-dihidróxido.

O padrão, arila hidroxilada – conector – arila (hidroxilada), é observado na esmagadora maioria de inibidores de integrase com potência inibitória significativa. Alguns exemplos podem ser vistos na figura 1, a seguir. Vale ressaltar, no entanto, que nem toda molécula que apresentar este padrão será necessariamente um inibidor de integrase.

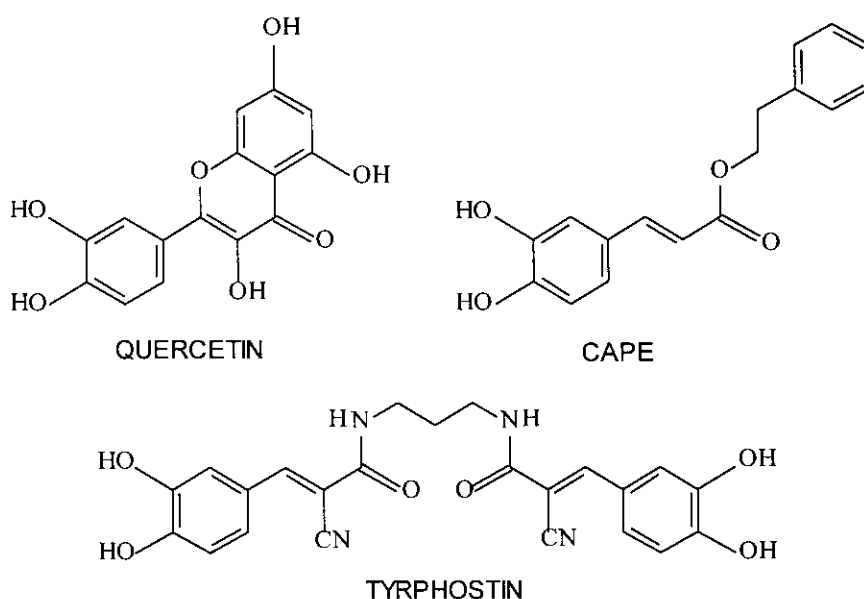


FIGURA 1: Alguns exemplos de inibidores de Integrase do HIV que obedecem ao padrão topológico proposto pelos experimentais^[12].

Fugindo a esta “regra topológica” destacam-se alguns compostos sintetizados recentemente pelo mesmo grupo de pesquisa que propôs o padrão anterior^[12] e que, ao contrário do que se esperava, apresentam atividade inibitória significativa. Alguns

* Tipicamente anéis benzênicos contendo ou não heteroátomos.

deles podem ser vistos na figura 2 e a explicação para tal atividade anti-HIV permanece ainda como uma questão em aberto. Exatamente por isto tais compostos são atraentes do ponto de vista teórico.

No primeiro capítulo desta tese, apresentaremos alguns dados sobre a AIDS, uma descrição sobre a anatomia do vírus HIV e a forma como este infecta as células do sistema imunológico. Voltaremos também a descrever alguns detalhes sobre as pesquisas de fármacos inibidores da integrase e as perspectivas desses estudos.

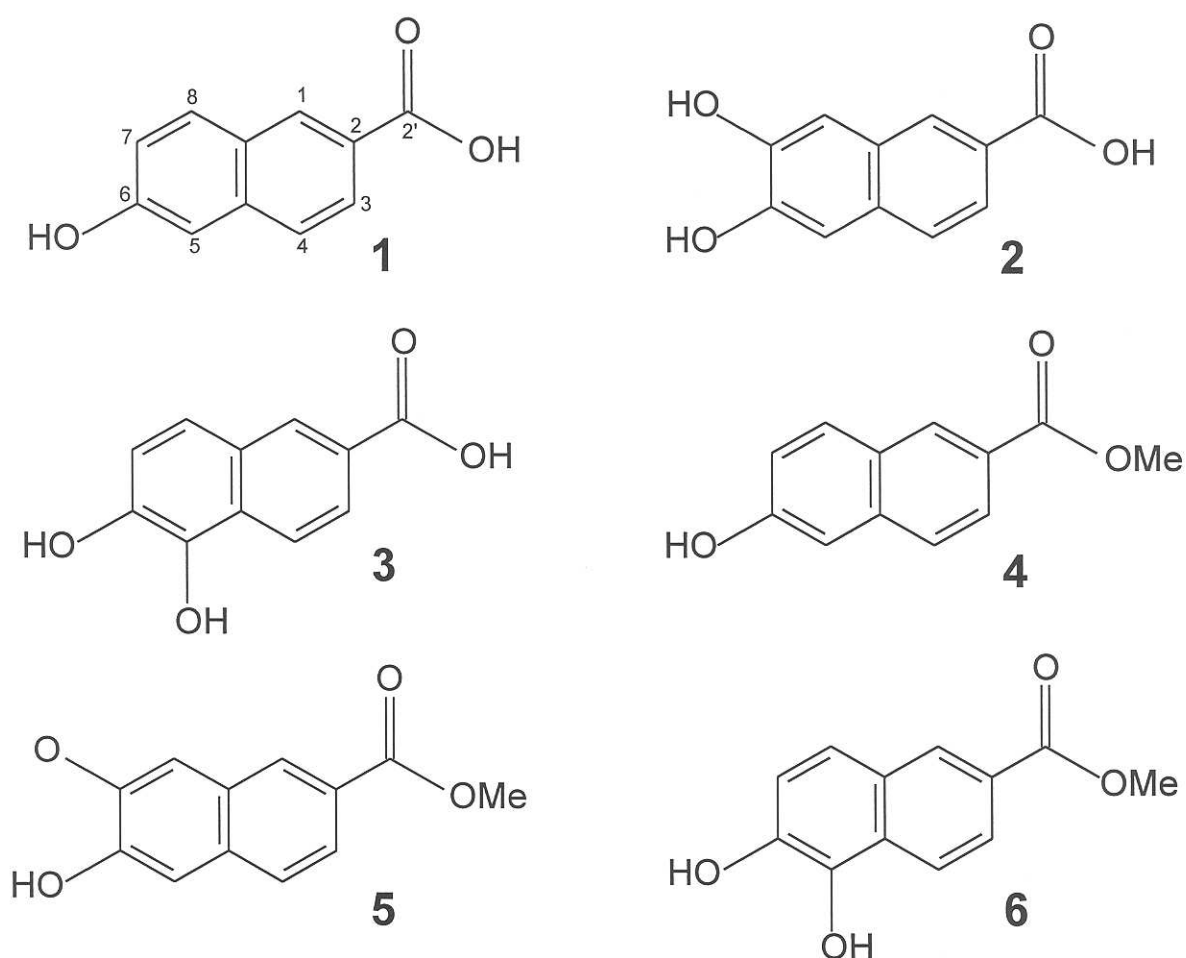


FIGURA 2: Alguns compostos que fogem ao padrão topológico da maioria dos inibidores de integrase, mas que possuem potência inibitória significativa.

No capítulo 2 discutiremos com algum detalhe a motivação para o presente trabalho: uma simples e nova metodologia recentemente proposta com o objetivo de relacionar índices eletrônicos à atividade biológica de uma classe de compostos carcinogênicos denominados Hidrocarbonetos Aromáticos Policíclicos (PAHs)^[13]. Esta metodologia, batizada recentemente de "Teoria de Índices Eletrônicos"^[14], está baseada na separação em energia entre os orbitais de fronteira (HOMO = *Highest Occupied Molecular Orbital* e LUMO = *Lowest Unoccupied Molecular Orbital*), bem como na contribuição relativa destes orbitais para a densidade de estados eletrônicos calculada sobre regiões específicas das moléculas.

O objetivo do presente trabalho é basicamente adaptar a metodologia da Teoria de Índices Eletrônicos ao estudo de alguns novos compostos inibidores da enzima integrase do vírus HIV-1.

Pretendemos investigar a aplicação do método à sistemas mais complexos que os PAHs[†]. Para tal, além das ferramentas comumente utilizadas para cálculos semi-empíricos, como o pacote de programas MOPAC^[15], desenvolvemos outros dois programas adicionais que fazem parte do material desta tese.

Estes dois programas foram muito úteis, facilitando bastante os cálculos envolvidos e a interpretação dos resultados da presente tese e de trabalhos realizados por outros membros de nosso grupo. Como exemplo temos uma tese de mestrado recentemente defendida em nosso instituto^[14], que estendeu a Teoria de Índices Eletrônicos ao estudo dos compostos derivados do Taxol com relativo sucesso e que fez uso desses programas em vários passos de seu

[†] Os PAHs, por serem essencialmente planares, apresentam muitas simplificações no que se refere ao cálculo de densidade de estados e à sua interpretação.

desenvolvimento. Apesar de suas complexidades individuais, os programas desenvolvidos são razoavelmente fáceis de serem utilizados e capazes de fornecer as informações mais relevantes ao tipo de estudo que parte de nosso grupo vem realizando nos últimos anos.

No capítulo 3 descreveremos os procedimentos utilizados na adaptação da metodologia utilizada, enquanto que os capítulos 4 e 5 serão reservados para a apresentação dos dois programas desenvolvidos. Finalmente, no capítulo 6, descreveremos os resultados e as conclusões obtidas para o estudo dos inibidores de integrase^[16], além de apresentarmos nossas perspectivas futuras.

Apresentamos ainda dois apêndices: o apêndice A dedica-se a um aprofundamento nos métodos de mecânica quântica e na descrição de alguns métodos semi-empíricos; o apêndice B descreve algumas rotinas especiais escritas em Visual Basic, que utilizamos neste trabalho como parte dos programas desenvolvidos.

- ¹ Thaisrivongs, S.; Turner, S.R.; Strohbach, J.W.; TenBrink, R.E.; Tarpley, W.G.; McQuade, T.J.; Heinrichson, R.L.; Tomasselli, A.G.; Hui, J.O. e Howe, W.J., *J. Med. Chem.* **1993**, 36, 942.
- ² Miertus, S; Furlan, M; Tossi, A. e Romeo, D., *Chem. Phys.* **1996**, 204, 173.
- ³ Volker, E.J., *J. Chem. Ed.* **1993**, 70, 3.
- ⁴ Hiebl, J.; Zbiral, E.; Balzarini, J. e De Clercq, E., *J. Med. Chem.* **1992**, 35, 3016.
- ⁵ Balter, M., *Science* **1996**, 274, 1988.
- ⁶ LaFemina, R.L.; Schneider, C.L.; Robibins, H.L.; Callahan, P.L.; LeGrow, K; Roth, E.; Schleif, W.A. e Emini, E.A., *J. Virol.* **1992**, 66, 7414.
- ⁷ Sakai, H.; Kawamura, M.; Sakuragi, J.; Sakuragi, S.; Shibata, R.; Ishimoto, A.; Ono, N.; Veda, S. e Adachi, A., *J. Virol.* **1993**, 67, 1169.
- ⁸ De Clercq, E., *J. Med. Chem.* **1995**, 38, 2491.
- ⁹ Wilson, E. K., *C&EN* Julho **1996**.
- ¹⁰ "AIDS, a 1% da cura", *Revista Superinteressante*, Outubro **1996**, número 10.
- ¹¹ Mathé, G., *Biomed & Pharmacother*, **1997**, 51, 413 e referências citadas.
- ¹² Zhao, H.; Neamati, N.; Mazumder, A.; Sunder, S.; Pommier, Y. e Burke Jr., T.R., *J. Med. Chem.* **1997**, 40, 1186 e referências citadas.
- ¹³ Barone, P.M.V.B.; Camilo Jr., A. e Galvão, D.S., *Phys. Rev. Lett.* **1996**, 77, 1186.
- ¹⁴ Paupitz, R., "Estudo estrutura-atividade do Taxol e derivados", *Tese de Mestrado*, **1998**, IFGW-UNICAMP.
- ¹⁵ Stewart, J.J.P., *J. Comp. Chem.* **1991**, 10, 209-221; MOPAC program, version 6.0, Quantum Chemistry Exchange Program No. 455.
- ¹⁶ Cyrillo, M. e Galvão, D.S., *THEOCHEM*, "Structure-activity relationship study of some new inhibitors of HIV-1 integrase", submetido para publicação em Outubro de **1997**.



Capítulo I

AIDS, HIV e a enzima Integrase

I.1 Introdução

AIDS, SIDA ou Síndrome da Imunodeficiência Adquirida tornou-se um dos mais significativos focos das preocupações humanas da atualidade, sendo um grande consumidor de recursos materiais e humanos, bem como fator de desequilíbrio social emergente^[1]. Desde sua identificação em 1980, a AIDS já dizimou cerca de 8 milhões de pessoas em todo o mundo. Estima-se ainda que mais de 20 milhões carreguem atualmente o vírus HIV em seus organismos^[1]. Tais doentes em potencial contribuem cada vez mais para o alastramento da epidemia que já não possui alvos preferenciais, como eram os grupos de risco: homossexuais e usuários de drogas.

Acredita-se que Vírus da Imunodeficiência Humana (HIV) seja o agente causador da AIDS^[2]. As poucas controvérsias que surgiram a este respeito são desacreditadas cada vez mais pela importância que os ataques ao HIV vêm alcançando no tratamento anti-AIDS^[3,4,5,6,7,8].

O HIV foi isolado pela primeira vez por Robert Gallo em 1984^[9]. Desde então, uma verdadeira batalha tem sido travada no sentido de decifrar os detalhes da infecção e multiplicação do vírus e suas mutações, procurando os mais alternativos meios de atacá-los dentro e fora das células humanas^[1,8].

Como seu próprio nome sugere, o HIV infecta as células do sistema imunológico. O alvo mais provável é composto pelas células CD-4, que são as

comandantes das defesas do nosso organismo. Por atacar justamente as células designadas a defender o organismo humano contra invasores, o HIV é um alvo muito difícil do ponto de vista farmacológico.

Além disso, o HIV é um retrovírus, que são vírus que carregam sua informação genética na forma de RNA (ácido ribonucleico) em vez do usual DNA (ácido desoxirribonucleico)^[3]. Esta é uma característica comum aos vírus primitivos, que apesar de terem uma estrutura bastante simples, cometem muitos erros durante seu processo de replicação. Como resultado, tem-se vírus com um altíssimo grau de mutabilidade, o caso do HIV^[10]. Este fato, contraditoriamente, coloca este vírus em vantagem frente às mais sofisticadas terapias que visam eliminá-lo dos pacientes com AIDS, pois uma droga que se mostre eficiente num dado momento pode simplesmente perder seus efeitos quando uma nova linhagem de vírus mais resistentes aparecer. Estima-se que isso possa acontecer em um prazo de 12 semanas a contar da data de início do tratamento com uma determinada droga^[10].

A explicação para essa rapidez no desenvolvimento de tolerância fundamenta-se na teoria de evolução das espécies e na velocidade de replicação do HIV: cerca de 10 bilhões de vírus são produzidos diariamente no organismo de um doente de AIDS^[1], a maior parte deles sendo derrotada pelo sistema imunológico e remédios como o AZT. Porém, devido ao grande número de mutações sofridas pelo HIV, novas linhagens mais resistentes aparecem e se multiplicam. Logo após o curto período em que o paciente parece melhorar, o vírus em suas novas formas, toma conta novamente do organismo.

As mutações podem ocorrer em várias posições do genoma do HIV, o que afeta a estrutura de certas proteínas virais que serão produzidas. Em alguns casos os erros podem afetar cerca de 5 em 10 vírus replicados *in vivo*^[10].

Antes de partirmos para mais detalhes sobre as formas de se atacar o HIV, vamos descrever brevemente sua anatomia e seu processo básico de replicação.

I.2 A anatomia do HIV-1

I.2.1 Os vírus

Os vírus são estruturas razoavelmente simples, despojadas das organelas reprodutivas que praticamente toda célula possui. Por não terem a capacidade de se reproduzir sozinhos, não são considerados, por alguns pesquisadores, como sendo seres vivos* e necessitam infectar uma célula viva para se replicarem. Utilizam para isso o aparato reprodutivo das mesmas.

Logo após a infecção, conseguem integrar seu material genético com o da célula normal, cujo DNA passará a conter todas as informações necessárias para a produção de novos vírus. Como o DNA comanda a síntese de novas proteínas (através do RNA mensageiro), a célula infectada produzirá cópias de novos vírus desenfreadamente, até que se esgotem as fontes de energia e ela morra.

* Não pretendemos colocar este aspecto em discussão, dadas as controvérsias que o cercam.

A figura 1.1 a seguir mostra um desenho esquemático da anatomia do vírus HIV-1^[1]. Em sua forma exterior o HIV é praticamente esférico com diâmetro aproximado de 1100Å; como grande parte dos vírus isolados e estudados até a atualidade, é composto basicamente de duas partes: um envelope e um núcleo^[3].

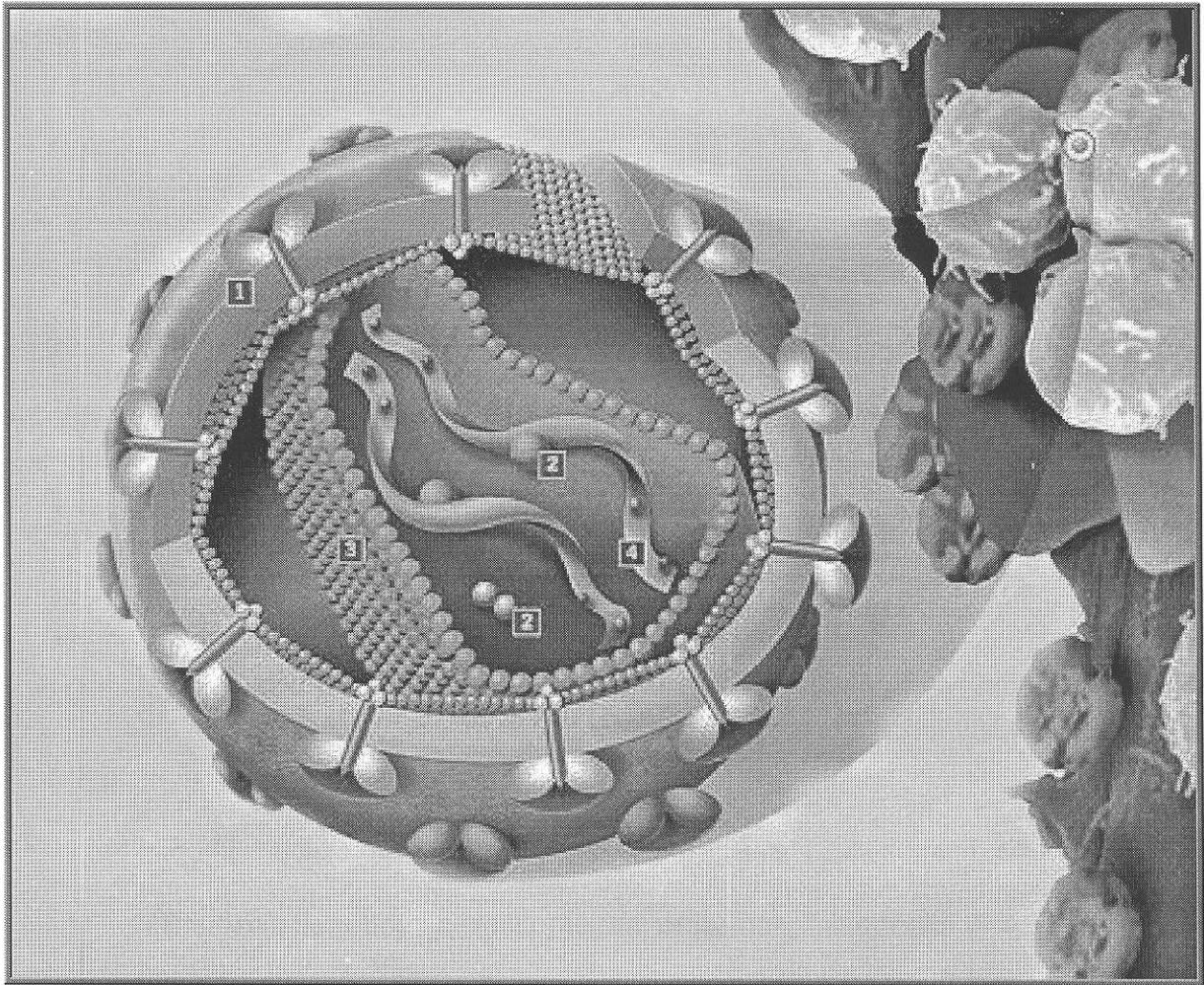


FIGURA 1.1: Anatomia do vírus HIV-1.

Dois tipos de HIV foram identificados até o momento: HIV-1 e HIV-2, sendo que este último, uma variedade mutante do primeiro, é bem menos freqüente nos doentes de AIDS e semelhante ao SIV (*Simian Immunodeficiency Virus*), o vírus da imunodeficiência dos símios, que é também um retrovírus muito parecido estruturalmente com o HIV^[3].

Vamos descrever agora os componentes da estrutura do HIV em referência aos números da figura 1.1 anterior.

I.2.3 A capa protetora

1 → É o envelope formado por uma camada dupla de lípideo similar à membrana plasmática das células humanas. Contém três proteínas:

- A que forma a camada externa, **gp120[†]** (*knob-like surface protein*);
- A que faz a interface entre o interior e o exterior do vírus, **gp41** (*transmembrane protein*);
- A que forma a camada interna, **p17** (*matrix protein*).

I.2.4 O núcleo

→ O núcleo tem uma forma alongada e protege em seu interior o material genético, bem como as proteínas essenciais à replicação do vírus. Possui em sua superfície uma capa protetora formada pela proteína **p24**, chamada de capsídeo^[3]. Tal proteína ficou famosa por

[†] O índice **gp** significa glicoproteína e **p** proteína; já os números que os seguem dão o tamanho aproximado da molécula em kilodaltons, medido em experiências de eletroforese^[3].

servir de rastro: pela quantidade de moléculas **p24** no sangue é possível estimar o número de vírus que circulam pelo corpo^[1].

4 → A informação genética está armazenada na forma de duas fitas de RNA (característica dos retrovírus), cada uma delas coberta pela proteína **p7** (*nucleocapsid protein*) que as protege formando um complexo protéico^[3]. As extremidades dessas fitas recebem nomes especiais: terminais 3' e 5' ^[3].

2 → As proteínas, neste caso também enzimas, importantes ao processo de replicação do HIV e presentes no seu núcleo, são três:

- a transcriptase reversa (**p66**), abreviaremos TR;
- a protease (**p11**), abreviaremos PR;
- a integrase (**p32**), abreviaremos IN.

O HIV possui também uma proteína com função ainda não identificada, chamada de **p6**.

1.3 O processo de replicação do HIV-1 e a sua inibição

Como dito anteriormente, as enzimas TR, PR e IN são essenciais ao processo de replicação do HIV. A maioria das drogas desenvolvidas até o momento, com o objetivo de bloquear a ação do HIV, é baseada na inibição das mesmas. Isto ocorre pelo usual processo de competição: uma enzima é inibida quando algum outro elemento liga-se a ela no lugar de seu substrato “natural”. Esta ligação ocorre

no mesmo sítio ativo em que o substrato se ligaria à enzima, na ausência do inibidor, pelo processo de chave-fechadura^[11]. A concentração de inibidor é um parâmetro muito importante, pois determina a maior ou menor probabilidade de a enzima se ligar a ele.

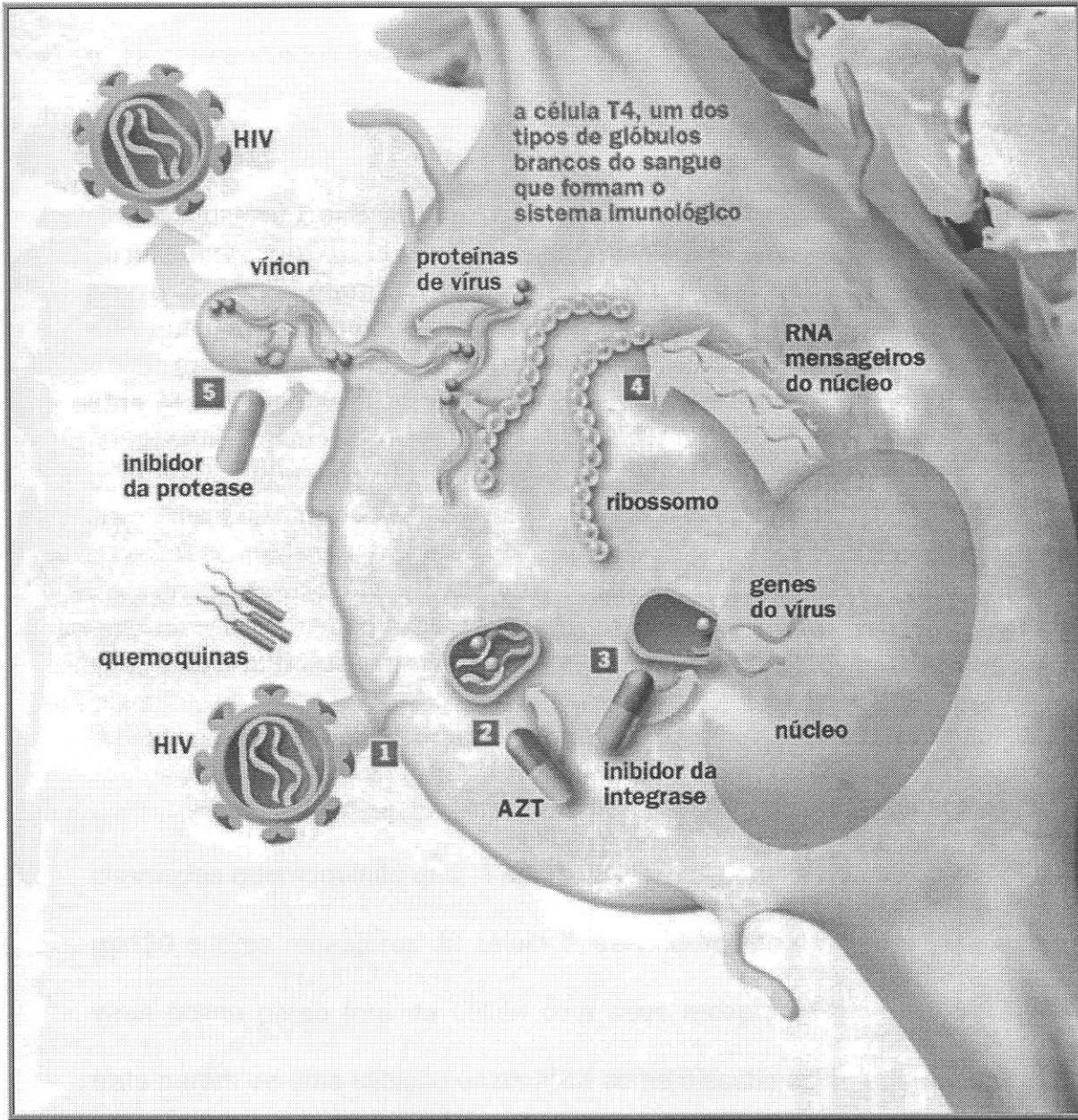


FIGURA 1.2: Representação esquemática do processo de replicação do HIV-1.

Um inibidor eficiente deve, portanto, “imitar” estruturalmente o substrato da enzima que irá inibir. Como veremos nas próximas seções, o substrato da TR é o RNA viral, da IN o DNA viral e o da PR a poliproteína viral **Pr160**, cuja origem ficará clara mais adiante^[3].

Vamos separar o processo de replicação do HIV em seus estágios mais importantes, referenciados aos números da figura 1.2^[1] acima

1.3.1 A invasão

1 → A molécula **gp120** é a chave feita sob medida para abrir as células do sistema imunológico, as chamadas células T. Uma vez que esta proteína se ligue ao receptor de uma célula T (no exemplo da figura acima uma célula T4), o processo de invasão se inicia com a **gp41** fundindo-se à membrana plasmática e o núcleo do vírus conseguindo atingir o citoplasma celular^[1,3]. Ali sua parede é desfeita[‡], liberando seu conteúdo: as fitas de RNA e enzimas; o processo de invasão, então, continuará.

Tem-se investigado recentemente a ação de algumas substâncias chamadas quemoquinas que, a princípio, seriam capazes de se ligar melhor à **gp120** que os receptores da célula T, servindo como um tipo de inibidor^[1]: o vírus ficaria preso fora da célula com seus receptores “ocupados”; não se trata porém de uma inibição enzimática como discutimos anteriormente.

A vantagem no uso das quemoquinas reside no fato de serem substâncias não tóxicas ao organismo humano, pois são secretadas pelo próprio sistema imunológico. Tal fato contrasta-se com o alto poder tóxico da

maioria dos inibidores de TR e PR testados até o momento. A utilização prática das quemoquinas esbarra, porém, na dificuldade de controlar sua produção a concentrações suficientes para uma inibição significativa^[1].

I.3.2 A transcrição do RNA

2 → Como dito anteriormente, o HIV possui sua informação genética na forma de duas fitas de RNA em vez de uma fita dupla de DNA. Para que o genoma do HIV possa ser integrado ao núcleo da célula hospedeira, é necessário primeiramente que o seu RNA seja convertido em DNA complementar. Tal processo é catalisado pela enzima transcriptase reversa (TR) e ocorre no citoplasma da célula hospedeira. Os inibidores de TR atuam neste momento, substituindo o seu substrato, o RNA viral. Um dos mais conhecidos é o AZT (3'-azido-3'-deoxythimidina), pioneiramente utilizado nos tratamentos anti-AIDS com bons resultados iniciais^[4]. Seu uso é atualmente limitado pelo desenvolvimento de resistência e alto poder tóxico. Além do AZT, temos outras substâncias derivadas: ddI, ddC, d4T e 3tC, todas aprovadas pela FDA (*Food and Drug Association*)^[1].

I.3.3 A fusão do genoma viral com o da célula

3 → Uma vez transcrito na forma de DNA, o genoma do HIV precisa ser integrado ao código genético (DNA), da célula hospedeira. Para tal, é transportado ao núcleo juntamente com a enzima integrase (IN). Esta catalisa

† Processo ainda desconhecido^[3].

as duas etapas necessárias para a completa integração dos códigos genéticos:

- Inicialmente dois nucleotídeos são retirados de cada terminal 3' do DNA viral. Esta etapa é chamada de *3'-processing*^[3,12];
- O processo de integração pode, então, continuar com uma série de reações tipo cortar e colar capazes de emendar o DNA viral ao DNA da célula hospedeira. Esta etapa é chamada de *strand transfer* e executada em ambas as extremidades 3' e 5' ^[3,12].

O pedaço de DNA proveniente do vírus, após passar por essas duas etapas e ter se integrado totalmente ao DNA da célula hospedeira, é chamado de provírus e é composto por 9749 pares de bases^[3].

Um inibidor eficiente de integrase deve ser capaz de inibir as duas etapas supracitadas. As pesquisas mais recentes mostram que, justamente por este motivo, os inibidores de integrase podem ser mais eficientes que os demais, baseados na inibição das outras duas enzimas^[8,12], pois o processo inibitório aconteceria em dois momentos, aumentando a probabilidade de uma inibição mais eficiente.

Não há ainda nenhum inibidor de integrase aprovado pela FDA, apesar da imensa quantidade de compostos propostos e testados^[12,13,14,15].

Após o processo de integração descrito anteriormente, tem início uma fase de “trégua”, quando o provírus pode permanecer inativo por período indeterminado e, por fazer parte do próprio DNA celular, não pode ser detectado pelo sistema imunológico como intruso.

A continuidade da infecção pode, no entanto, ser desencadeada repentinamente^[16]. Em grande parte dos casos isto acontece quando outra doença ativa o sistema imunológico.

4 → Após a ativação do HIV, o DNA infectado (através da enzima RNA polimerase) comandará a síntese de RNAs mensageiros (mRNAs) com o tamanho igual ao do genoma do vírus^[3]. Estes deixam o núcleo em direção aos ribossomos, onde servem de modelo para a síntese das proteínas que formarão um novo HIV. Estas proteínas estarão concentradas na forma de grandes poliproteínas (**Pr160**), que possuem em sua estrutura quase todos os componentes para a formação de um novo vírus, ligados uns aos outros^[1,3].

5 → A enzima protease (PR) catalisa a quebra dos pedaços da **Pr160** de maneira a separar aqueles que farão parte do novo vírus^[1,3]. Após se juntarem, o novo vírus fica na forma de um vírion (um vírus imaturo), que precisa passar por um processo de maturação antes de poder deixar a célula.

Neste processo final, a PR exerce um papel fundamental^[3], hidrolizando certas reações.

Como dito anteriormente, a maior parte das proteínas virais vêm da poliproteína **Pr160**. Desta não fazem parte, porém, as moléculas que constituirão a nova capa protetora do vírus que são obtidas através de pedaços de RNA viral presente no citoplasma^[3].

Os inibidores de protease atuariam impedindo o processo de maturação do vírus, sendo muito importantes na situação em que a doença já está ativa no organismo do paciente. Muitos deles vem sendo testados nos últimos anos com excelentes resultados iniciais^[17] e alguns foram inclusive aprovados pela FDA, sendo utilizados em novas terapias baseadas no ataque a múltiplas enzimas^[18].

1.4 Pesquisas com inibidores

Como discutido anteriormente, a principal e mais promissora forma de se atacar o HIV consiste na inibição das três enzimas fundamentais ao seu processo de replicação: TR, PR e IN. As duas primeiras têm sido extensivamente estudadas nos últimos anos: suas estruturas cristalográficas foram determinadas, inclusive quando ligadas aos seus substratos^[19,20,21] e inúmeros testes *in vitro* foram realizados. Tais informações são fundamentais ao desenvolvimento de compostos inibidores mais eficientes e de novas estratégias de ataque^[8,19].

No final de 1996, foi realizada em Vancouver a XI conferência internacional sobre AIDS, onde foi apresentada pelo pesquisador David Ho, o que seria quase a

“cura” para a AIDS^[1,18,22]. Ho, utilizando um coquetel de drogas contendo o AZT, o ddI (parente do AZT) e alguns inibidores de protease, mostrou que a carga viral no sangue de seus pacientes tinha caído a níveis quase indetectáveis^[22]. Aproximadamente 98,9% do total de vírus dos doentes tinham sido eliminados, restando a dúvida de onde estariam os restantes 1,1% e de como poderiam ter escapado ao ataque das drogas^[1].

Alguns meses depois, porém, foi constatado que esses resultados tinham sido superestimados e que o tratamento biterápico[§], como o de Ho, deveria ser expandido^[23].

A integrase, a menos estudada das três enzimas do HIV, aparece como esperança no desenvolvimento de drogas mais eficientes. Sua estrutura cristalográfica já foi determinada^[24], mas nenhuma informação desse tipo sobre a ligação com seu substrato está disponível até o momento, o que dificulta o processo de desenvolvimento de inibidores.

Acredita-se que um coquetel contendo drogas capazes de inibir as três enzimas virais simultaneamente seja muito mais promissor, porque as mutações simultâneas de mais de uma enzima, mesmo para o HIV, ocorreriam mais raramente^[3,25]. Algumas drogas vêm sendo testadas com o objetivo de inibir a integrase e alguns resultados tem sido promissores^[12,13,14,15].

[§] Atacando dois alvos virais simultaneamente.

Com o aumento da expectativa de vida dos pacientes de AIDS, graças ao fato de a doença estar cada vez mais controlável, o poder tóxico das drogas utilizadas nas terapias passou a ser um parâmetro muito importante.

Quando o tempo de vida de um aidético era medido em meses, não havia tempo para que ele sentisse os efeitos colaterais das drogas utilizadas no combate ao HIV. Sabe-se que a maioria dos inibidores de TR e PR, além dos efeitos colaterais mais simples como náuseas constantes e diarreias, podem causar um comprometimento permanente da medula óssea e alguns tipos de câncer^[1].

Muitos dos compostos com grande atividade anti-AIDS são também, em contrapartida, muito tóxicos. Na busca por compostos menos tóxicos, encontram-se aqueles com menor atividade^[12]. Muitas vezes é necessário optar por elementos mais tóxicos e que possuam grande atividade inibitória, para que, ministrados em menores concentrações, dêem melhores resultados^[12].

O desenvolvimento de novas drogas é demorado e caro. Para se ter uma idéia, no ano passado a indústria de inibidores de protease movimentou cerca de 800 milhões de dólares com previsão de aumento significativo em 98. Novos coquetéis, incluindo inibidores de integrase, devem ser aprovados até o final deste ano.

Um dos métodos atualmente mais utilizados para se encontrar novos compostos que possuam atividade anti-HIV é denominado *pharmacofore search*. Consiste basicamente em escolher um composto bastante ativo como padrão e, escolhendo algumas características estruturais do mesmo como referência, procurar por compostos semelhantes que possam ser ativos. As buscas são feitas em bases

de dados bastante completas como a do NCI (*National Cancer Institute*) que contem as estruturas 2D e 3D de grande parte dos compostos já estudados na pesquisa contra o câncer^[26]. Estas pesquisas são rápidas e abrangentes, o que confere a esse método uma importância bastante significativa^[13].

No entanto, os testes *in vitro* devem ser realizados em algum momento para verificar se as moléculas escolhidas possuem realmente atividade biológica significativa. Na maioria dos casos a resposta é negativa, o que causa perda de tempo e dinheiro.

Um método mais eficiente capaz de apontar os compostos potencialmente mais ativos seria de enorme valia. Nosso trabalho pretende trilhar os primeiros passos neste sentido.

- ¹ "AIDS, a 1% da cura", *Revista Superinteressante*, Outubro **1996**, número 10.
- ² Miertus, S; Furlan, M; Tossi, A. e Romeo, D., *Chem. Phys.* **1996**, 204, 173.
- ³ Volker, E.J., *J. Chem. Ed.* **1993**, 70, 3.
- ⁴ Hiebl, J.; Zbiral, E.; Balzarini, J. e De Clercq, E., *J. Med. Chem.* **1992**, 35, 3016.
- ⁵ Balter, M., *Science* **1996**, 274, 1988.
- ⁶ LaFemina, R.L.; Schneider, C.L.; Robibins, H.L.; Callahan, P.L.; LeGrow, K; Roth, E.; Schleif, W.A. e Emini, E.A., *J. Virol.* **1992**, 66, 7414.
- ⁷ Sakai, H.; Kawamura, M.; Sakuragi, J.; Sakuragi, S.; Shibata, R.; Ishimoto, A.; Ono, N.; Veda, S. e Adachi, A., *J. Virol.* **1993**, 67, 1169.
- ⁸ Mitsuya, H.; Yairchoan, R. e Broder, S., "Molecular Targets for AIDS therapy", *Science* **1990**, 249, 1533-1543.
- ⁹ Gallo, R.C.; Sarin, P.S.; Galmann, E.P.; Robert-Guroff, M.; et al., *Science* **1983**, 220, 865-867.
- ¹⁰ Roberts, J.D. et al., *ibid.* **1988**, 231, 850.
- ¹¹ Marzzoco, A. e Torres, B.B. "Bioquímica básica", Editora Guanabara, **1990**.
- ¹² Zhao, H.; Neamati, N.; Mazumder, A.; Sunder, S.; Pommier, Y. e Burke Jr., T.R., *J. Med. Chem.* **1997**, 40, 1186 e referências citadas.
- ¹³ Hong, H.; Neamati, N.; Wang, S. et al., *J. Med. Chem.* **1997**, 40, 930-936.
- ¹⁴ Mazumder, A.; Neamati, N.; Sunder, S. et al., *J. Med. Chem.* **1997**, 40, 3057-3063.
- ¹⁵ Zhao, H.; Neamati, N.; Hong, H. et al., *J. Med. Chem.* **1997**, 40, 242-249.
- ¹⁶ Zagury, D. et al., *Science* **1986**, 231, 850.
- ¹⁷ Jadhav, P.K.; Ala, P.; Woerner, F.J. et al., *J. Med. Chem.* **1997**, 40, 181-191 e referências citadas.
- ¹⁸ Ho, D.D. et al., *Nature* **1995**, 373, 123-126 e referências citadas.
- ¹⁹ Wlodawer, A. et al., *Science* **1989**, 245, 616 e referências citadas.
- ²⁰ Navia, M.A. et al., *Nature* **1989**, 337, 615.
- ²¹ Lapattoet, R. et al., *ibid.* **1989**, 342, 299.

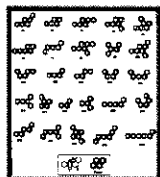
²² Wilson, E. K., *C&EN* Julho **1996**.

²³ Mathé, G., *Biomed & Pharmacother*, **1997**, 51, 413 e referências citadas.

²⁴ Dyda, F.; Hickman, A.B.; Jenkins, T.M. *et al.*, *Science* **1994**, 266, 1981-1984.

²⁵ Wei, X.; Ghosh, S.K. *et al.*, *Nature* **1995**, 373, 117.

²⁶ Milne, G.W.A.; Nicklaus, M.C. *et al.*, "The NCI Drug Information System 3D Database", *J. Chem. Inf. Comput. Sci.* **1994**, 34, 1219-1224.



Capítulo II

Teoria de Índices Eletrônicos

II.1 Introdução

Neste capítulo discutiremos o estudo da teoria desenvolvida recentemente por Barone *et al.*^[1], como parte de uma tese de doutoramento defendida neste instituto, que foi capaz de classificar em termos qualitativos os compostos conhecidos como Hidrocarbonetos Aromáticos Policíclicos (PAHs) quanto à presença ou ausência de atividade carcinogênica.

Esta metodologia, chamada de Teoria de Índices Eletrônicos, pretende relacionar índices eletrônicos de moléculas de interesse biológico às suas atividades.

A figura 2.1 mostra os 26 compostos estudados inicialmente naquele trabalho, colocando-os em ordem decrescente de atividade carcinogênica: aproximadamente até o décimo sexto estão os elementos muito cancerígenos*, os demais apresentam atividade bastante tênue ou nula.

Os PAHs têm sido objeto de intensas pesquisas nos últimos anos, sendo que em várias frentes, muitos grupos trabalharam no intuito de obter parâmetros teóricos confiáveis que permitissem distinguir, ao menos qualitativamente, quais deles são cancerígenos.

* Os primeiros são capazes de causar câncer em 80% dos ratos em que são aplicados via tintura de pele.

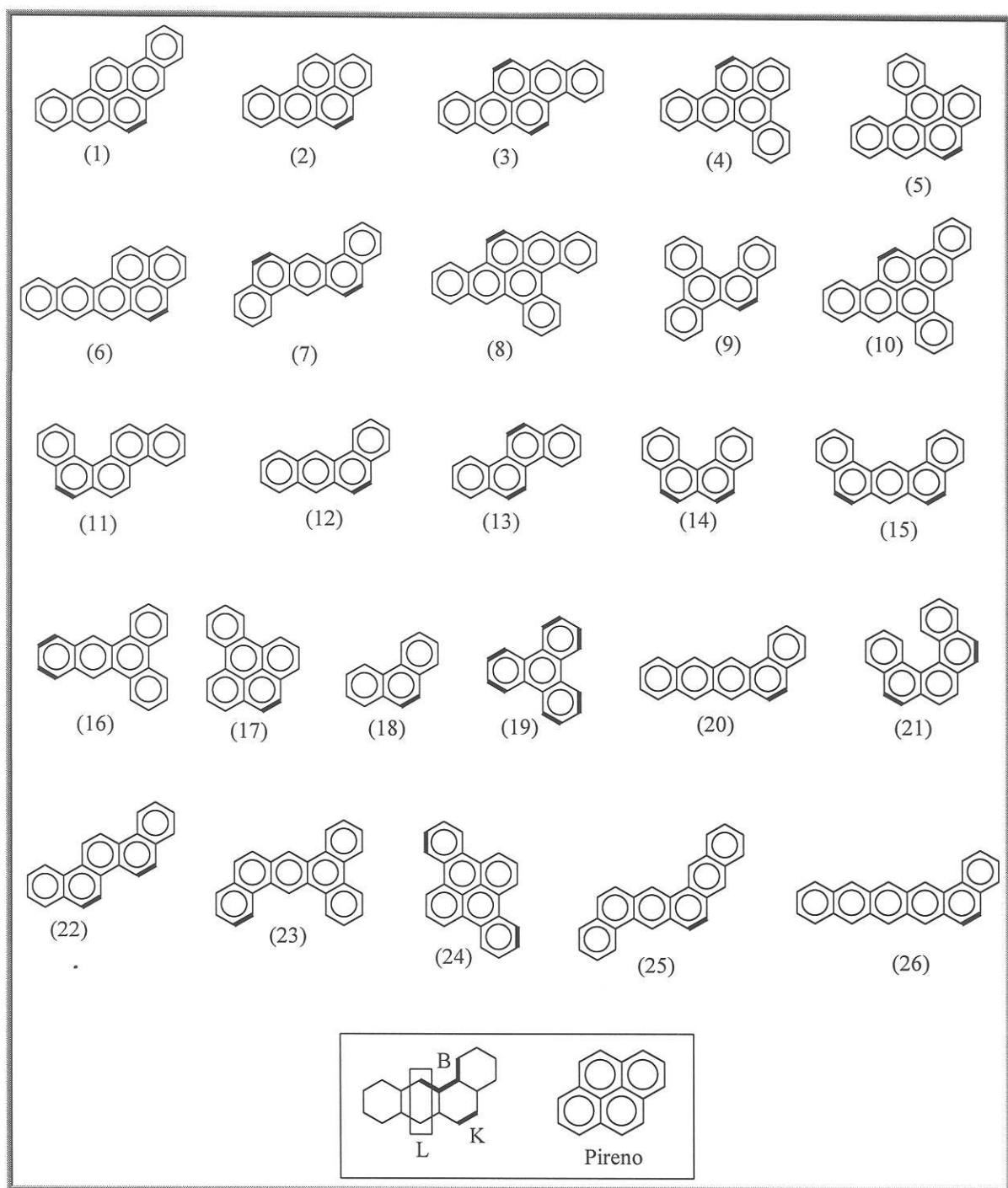


FIGURA 2.1: os 26 Hidrocarbonetos Aromáticos Policíclicos estudados inicialmente com a Teoria de Índices Eletrônicos. No quadro menor temos as regiões K, L e de Baía (B) e a estrutura do Pireno.

Estes compostos têm estruturas bastante semelhantes embora seu poder carcinogênico, medido pelo índice de Iball^[2] varie muito. Este índice é definido como a porcentagem de câncer de pele ou papiloma induzido em ratos submetidos à aplicação de tintura do composto carcinogênico na pele, dividido pelo período de latência e multiplicado por 100. A atividade cancerígena destes compostos foi estudada de maneira sistemática na década de 30 por Cook e colaboradores^[3]. Testes experimentais mostraram a existência de algumas correlações entre a intensidade da atividade biológica (expressa pelo índice de Iball) e certos aspectos topológicos dessas moléculas.

Os primeiros estudos teóricos, realizados por Coulson^[3], Schmidt^[4] e Svarthölm^[5], exploraram tais aspectos na tentativa de correlacionar a atividade química às geometrias e às características eletrônicas das moléculas. Schmidt e Svarthölm mostraram que certas regiões dos PAHs deveriam ter densidades eletrônicas mais altas que outras e, conseqüentemente, seriam mais susceptíveis a reações eletrofílicas. Svarthölm calculou as valências livres dos elétrons π e as associou com a reatividade das moléculas. Naquele trabalho surgiram os conceitos que depois serviriam de base para a chamada teoria das regiões K,L e Baía^[1] (quadro menor da figura 2.1).

O desenvolvimento de uma teoria completa de carcinogênese química, baseada em cálculos quânticos, foi realizado por Pullman e Pullman^[6,7], que demonstraram (utilizando o método de Hückel^[8]) que os PAHs apresentavam alta densidade eletrônica na região K. Índices eletrônicos para os elétrons π das regiões K e L foram utilizados por Pullman na tentativa de correlacioná-los a dados

experimentais de ação cancerígena. Nenhum índice isolado foi suficiente para descrever os dados disponíveis, mas alguns índices compostos foram propostos^[6,7].

II.2 Metodologia e cálculos utilizando o método de Hückel

A teoria das regiões K e L e teorias similares baseadas em índices eletrônicos têm sido testadas há décadas, mas nenhuma delas é totalmente consistente com os dados experimentais de carcinogênese disponíveis na literatura. O mesmo acontece com outras teorias mais recentes, que incluem um grande número de parâmetros e técnicas sofisticadas de análise de dados ou mesmo de redes neurais e inteligência artificial, enquanto outras tentam encontrar correlações por meio de análise estatística, ou reconhecimento de padrões^[9,10,11,12,13,14].

Mesmo com o uso de todas essas técnicas sofisticadas não foi possível encontrar, pelo menos de maneira satisfatória, correlações que pudessem distinguir de maneira completa as moléculas ativas das inativas.

Com esse objetivo foi desenvolvida a Teoria de Índices Eletrônicos. Esta lança mão de recursos computacionais bastante simples, pois utiliza como ferramenta de cálculo o método de Hückel, que é o modelo mais simples para tratar sistemas moleculares utilizando mecânica quântica e uma parametrização experimental. O método de Hückel, apesar da simplicidade de sua forma, justamente por estar baseado nos resultados experimentais com os quais foi calibrado, pode reproduzir ótimos resultados qualitativos^[1,15,16].

A calibração experimental aparece quando as integrais, decorrentes da formulação da equação de Schrödinger e da aproximação LCAO (*Linear Combination of Atomic Orbitals*), são substituídas por valores obtidos com a

comparação de resultados experimentais para um conjunto de moléculas escolhido como padrão. Por este motivo o método de Hückel é conhecido como um método empírico, pois apesar de partir de uma formulação teórica, não resolve as integrais decorrentes.

Quando um método resolve parte dessas integrais e parametriza as outras ele é denominado de "semi-empírico". Como exemplo temos o *MNDO (Modified Neglect of Differential Overlap)*^[17], *AM1 (Austin Method 1)*^[17], *PM3 (Parametric Method 3)*^[18], etc., que diferem entre si pelas simplificações utilizadas na resolução das integrais e pela forma como as parametrizam.

A Teoria de Índices Eletrônicos baseia-se na densidade de estados eletrônicos calculada sobre regiões quimicamente ativas das moléculas, o que resulta em regras simples de identificação da atividade cancerígena dos PAHs. Foram estudados com sucesso os 26 PAHs da figura 2.1, que já haviam sido investigados por Pullman.

Basicamente, a teoria utiliza os conceitos de densidade local e total de estados eletrônicos. A densidade total de estados eletrônicos (DOS) é definida como o número de estados eletrônicos por unidade de energia, enquanto que a local (LDOS), é esta densidade calculada sobre uma certa região da molécula. Essas grandezas têm como vantagem o fato de trazerem informações a respeito do comportamento bioquímico, reatividade química, etc., sem que seja necessário recorrer aos conceitos de regiões K-L ou Baía^[1].

Procurou-se analisar a DOS e LDOS calculadas sobre diversas regiões das moléculas, incluindo as regiões K-L e Baía, que normalmente são usadas para definir índices críticos para a atividade carcinogênica. Tal análise preliminar não

revelou a existência de um padrão que pudesse ser correlacionado à atividade biológica dos compostos, porém foram encontradas outras regiões mais promissoras levando em consideração o conceito de ordem de ligação.

A ordem de ligação entre dois átomos mede a contribuição total dos elétrons π para a ligação entre eles^[1]. A análise da LDOS foi feita sobre os cálculos para o anel que contém a mais alta ordem de ligação nas moléculas (estes anéis podem ser vistos na figura 2.1 pela ligação em negrito). Destes cálculos surgiram padrões bem definidos que englobaram todas as 26 moléculas consideradas e se traduziram em 3 regras simples explicitadas a seguir^[1]:

A primeira regra é baseada na arquitetura molecular e na diferença em energia entre os orbitais HOMO (último orbital ocupado) e seu antecessor em energia, o HOMO-1 (a esta diferença daremos o nome de Δ), sendo aplicada somente às moléculas que contêm estruturas tipo pireno (ver quadro menor da figura 2.1):

- Se a molécula contiver uma estrutura tipo pireno e $\Delta > 0.25\beta$ ($\beta \approx 2.4$ eV é um parâmetro do método de Hückel), será fortemente carcinogênica; do contrário será inativa.

As demais estruturas são analisadas com duas outras regras semelhantes que além da separação em energia, referenciam-se pelas contribuições às densidades de estados:

- Se a contribuição do HOMO para a LDOS for a maior de todas (maior pico), a molécula será completamente inativa.
- Se a contribuição do HOMO para a LDOS for maior que a do HOMO-1 (mas não for o maior pico) e se $\Delta > 0.15\beta$ a molécula terá atividade carcinogênica forte ou moderada; já se a contribuição do HOMO-1 for maior que a do HOMO, a molécula apresentará atividade fraca ou será inativa.

Como dito anteriormente, tais regras podem ser utilizadas para qualquer uma das moléculas da figura 2.1 com 100% de acerto.

Uma análise destas regras mostra que a atividade carcinogênica dos PAHs é determinada basicamente por três fatores: a arquitetura molecular, um valor mínimo para a diferença em energia Δ e as contribuições relativas dos orbitais HOMO e HOMO-1 para a LDOS calculada no anel de maior ordem de ligação. Esta última poderia também ser traduzida entre as contribuições para os orbitais LUMO (primeiro orbital desocupado) e LUMO+1, já que os resultados de densidade de estados com o método de Hückel para o caso dos PAHs (moléculas altamente simétricas) são simétricos em relação ao *gap* HOMO-LUMO.

O resultado que mais se destaca desse trabalho é justamente a existência de um valor mínimo para a diferença em energia Δ para que exista atividade carcinogênica. Esse índice indica que um orbital de fronteira mais isolado, ou seja um HOMO proporcionalmente afastado do HOMO-1, é a primeira condição para a ação cancerígena da substância; no caso das moléculas pirênicas, tal condição é suficiente, enquanto que para as outras é apenas condição necessária onde faz-se

necessária ainda a análise da competição entre as contribuições relativas para a LDOS desses dois orbitais moleculares.

Uma limitação natural do método de Hückel é não poder ser aplicado diretamente a moléculas não planares, pois está baseado na separação σ - π e não considera explicitamente as geometrias dos sistemas, mas apenas suas topologias (os tipos de átomos e com quem eles estão ligados). Levando este fato em consideração, o passo seguinte no estudo da aplicação da presente teoria foi verificar sua extensibilidade. A idéia foi, então, utilizar para os cálculos mecânico-quânticos o método semi-empírico mais sofisticado PM3, que leva em consideração não só a topologia, mas também a geometria das moléculas[†]. Tal método está disponível em várias implementações comerciais e de domínio público.

O objetivo de tal esforço foi desenvolver uma metodologia que pudesse ser aplicada também aos compostos não planares. Uma confirmação importante da teoria viria com a concordância entre os índices eletrônicos derivados pelos dois métodos diferentes.

II.3 Cálculos utilizando o método semi-empírico PM3

As moléculas tiveram sua geometria otimizada através do método semi-empírico PM3 contido no pacote MOPAC^[19] e como resultado dos cálculos tem-se, além de muitas informações como calor de formação, momentos de dipolo, etc., o anel de maior ordem de ligação para cada uma delas. Estes coincidiram com os valores obtidos com o método de Hückel.

[†] Os métodos semiempíricos podem, inclusive, ser utilizados para otimizar geometrias de moléculas.

Além disso, a partir dos elementos da matriz de coeficientes, pode-se calcular as densidades de estados. Maiores detalhes sobre como esses cálculos foram feitos serão apresentados no capítulo 5.

Como resultado para os PAHs, observou-se que os indicadores eletrônicos foram os mesmos que os obtidos com o método de Hückel^[20] e que a diferença entre as regras encontradas anteriormente e as obtidas com o PM3, ficava por conta de uma nova calibração no valor da diferença em energia Δ .

Tal fato é compreensível, pois a parametrização dos dois métodos é diferente: o PM3 descreve melhor a função de onda do sistema, já que despreza menos integrais que o método de Hückel, como por exemplo as integrais de *exchange*.

As novas regras são as seguintes^[20,21]:

1- Se a molécula contiver uma estrutura tipo pireno e $\Delta > 0.5$ eV, será fortemente carcinogênica; do contrário, será inativa.

Para as outras moléculas, a classificação é feita com a análise do Δ e também da competição de contribuição relativa dos orbitais HOMO e HOMO-1 para a LDOS na região do anel com maior ordem de ligação:

2- Se a contribuição do HOMO para a LDOS for maior que a do HOMO-1 e se $\Delta > 0.3$ eV, a molécula apresentará atividade carcinogênica forte ou moderada; se $\Delta < 0.3$ eV, a molécula terá atividade fraca ou será completamente inativa.

3- Se a contribuição do HOMO-1 for maior que a do HOMO, a molécula apresentará atividade fraca ou será inativa.

Como no caso anterior, esse conjunto de regras pode ser aplicado com 100% de eficácia aos HAPs da figura 2.1 e sendo assim, temos:

Regra	Aplica-se às moléculas (ver figura 2.1)
1	1,2,3,4,5,6,7,8,10,17 e 24
2	7,9,13,14,18,19,21,22, e 25
3	11,12,15,16,20,23 e 26

O significado destas regras é, basicamente, o mesmo já determinado no cálculo com o método de Hückel.

Recentemente decidimos investigar se as regras obtidas com o PM3 poderiam também ser alternativamente expressas em termos dos orbitais LUMO e LUMO+1, no lugar de HOMO e HOMO-1. Como dito anteriormente isso era possível com os resultados do método de Hückel para os PAHs, graças à simetria que aparece na matriz de autovetores. Surpreendentemente, também no caso do PM3, as regras podem ser rescritas, apesar da perda de simetria nos resultados com este método semi-empírico.

Sugerimos também mudanças no método de cálculo da densidade de estados, passando de uma situação contínua e com ajuste Lorentziano para uma situação discreta. Isso foi importante nas análises de sistemas mais complexos que os PAHs, onde o número de níveis de energia é muito maior. Essa quantidade muito maior de estados causava uma sobreposição indesejável dos picos de densidade e uma má descrição das funções Lorentzianas. O cálculo discreto foi feito em

trabalhos posteriores^[22,23,24] e revelou-se mais interessante. Os detalhes serão discutidos no capítulo 5.

II.4 Conclusões

A teoria desenvolvida conseguiu de maneira muito simples resolver satisfatoriamente um problema que estava em aberto desde a década de 30 e isso sem a utilização de grandes recursos computacionais[‡].

A característica mais interessante do método é a seguinte: sem considerar os detalhes do processo biológico, o que tornaria o problema complicadíssimo ou mesmo insolúvel do ponto de vista físico, pode-se tirar informação qualitativa sobre a atuação biológica das moléculas, bastando para isso analisar suas características estruturais. Poderíamos então dizer que a atuação biológica, pelo menos no caso dos PAHs, estaria vinculada a uma assinatura eletrônica que dependeria não do meio em que o processo ocorre, mas apenas das moléculas em si.

Com tal possibilidade, na extensibilidade desse método a novos compostos e situações biológicas diferentes, reside uma perspectiva bastante atraente: poder selecionar dentre um grande número de compostos aqueles com maior potencialidade biológica para uma posterior análise química detalhada. Com isso poder-se-ia poupar muito tempo e dinheiro, já que testes *in vitro* e *in vivo* são muito demorados e caros.

Com essa perspectiva, muito trabalho a nível teórico vem sendo realizado no nosso grupo. O primeiro deles foi a expansão da análise, usando o método de

[‡] Um cálculo típico de Hückel para uma molécula de porte médio demora alguns segundos em um PC486.

Hückel, para os PAHs metilados^[22]. Tais compostos, cerca de 90, não são mais planares devido à presença dos grupos metil (CH_3). Na figura 2.2 a seguir, podemos ver alguns dos compostos estudados. Para que seja possível usar o método de Hückel estes grupos são tratados como uma perturbação no parâmetro do átomo em que estão ligados. As mesmas regras levam a 90% de acerto, considerando todos os 90 compostos, um resultado excelente, pois devemos considerar ainda que há muitas incertezas na determinação do índice de Iball para parte dos compostos.

A análise para os compostos metilados usando-se o PM3 também foi feita e, como sugerido, com densidade de estados discreta. Tal estudo resultou em regras mais simples ainda e também com cerca de 90% de acerto^[23].

Estes resultados nos levam a concluir que, apesar de ser difícil explicar como um processo biológico tão complexo possa ser explicado do ponto de vista físico de maneira tão simples, essa metodologia deve ser ainda extensivamente investigada, pois mostra uma potencialidade muito útil e promissora.

Surge, então, a seguinte pergunta: será que podemos expandir a teoria a situações mais complexas e para os mais variados problemas biológicos? Dados os últimos trabalhos realizados por nosso grupo^[24,25] e os resultados da presente tese, podemos dizer que a resposta a essa pergunta é, *a priori*, afirmativa.

Vale ressaltar, no entanto, que a adaptação da metodologia inicial não é simples e direta. Para cada problema biológico diferente, em princípio, a assinatura eletrônica pode aparecer de maneiras completamente distintas e com regras diferentes daquelas para os PAHs. Tais regras podem também simplesmente não aparecer, indicando que a atividade biológica é dependente do processo biológico

em si e do meio em que ocorre. Estudos com efeitos de solventes para essas regras ainda devem ser realizados.

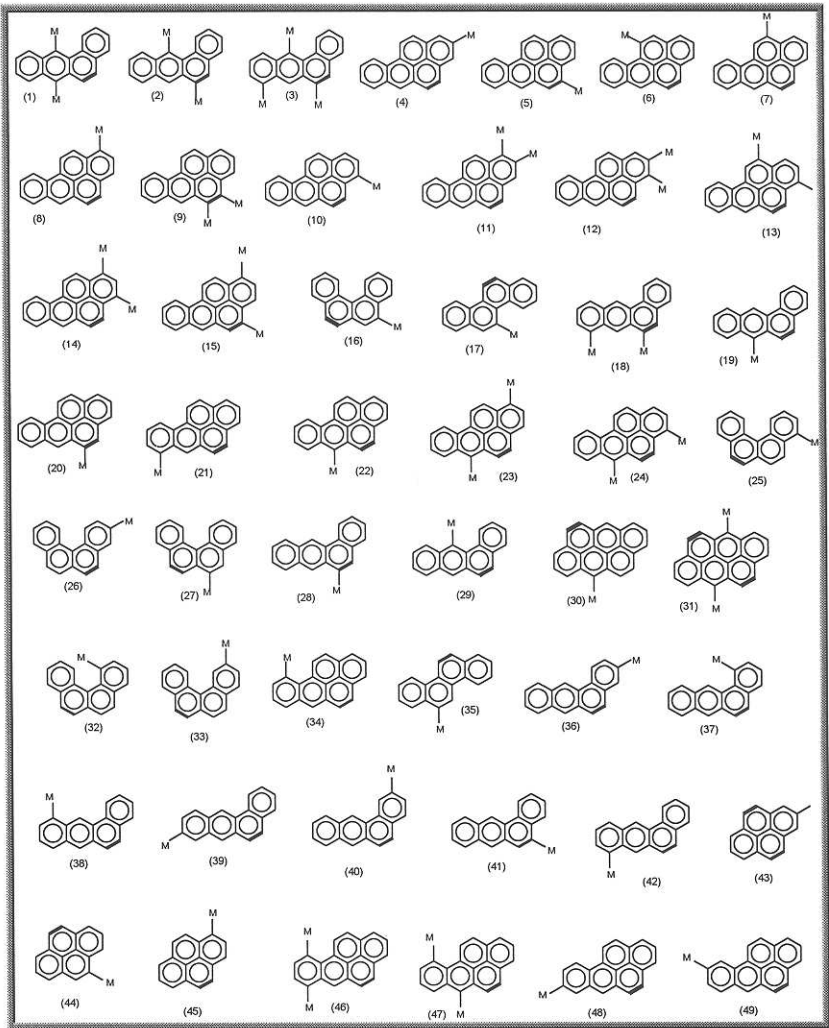


FIGURA 2.1: Alguns dos Hidrocarbonetos Aromáticos Policíclicos metilados estudados na extensão da teoria de índices eletrônicos.

A verificação da validade das regras para os PAHs com a densidade de estados calculadas através de cálculos *ab initio* também foi proposta. No entanto, temos motivos suficientes para acreditar que tais regras não são “fabricadas” pelos

métodos, já que Hückel e PM3 são bastante diferentes em suas implementações e forneceram resultados compatíveis neste estudo. Além disso, a estatística vem ficando mais confiável à medida que o número de compostos englobados na teoria e adaptações dela cresce consideravelmente.

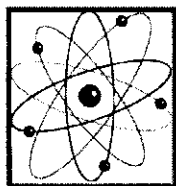
- ¹ Barone, P.M.V.B.; Camilo Jr., A. e Galvão, D.S., *Phys. Rev. Lett.* **1996**, 77, 1186
- ² Gelboin, H.V. e Ts' O, P.O.P., "*Polycyclic Hydrocarbons and Cancer*", Academic Press, New York, **1978**,2.
- ³ Coulson, C.A., *Adv. Cancer Res.* **1953**, 1, 1.
- ⁴ Schmidt, O., *Z. Physik Chem.* **1938**, 39, 59.
- ⁵ Svarthölm, N.V., *Arkiv Kemi Mineral. Geol.*, **1942**, A15, 1.
- ⁶ Pullman, A., *Bull. Soc. Chim. Fr.* **1954**, 21, 595.
- ⁷ Pullman, A. e Pullman, B., *Adv. Cancer Res.* **1955**, 3, 117.
- ⁸ Streitwieser, A., "*Molecular Orbital Theory*", Wiley, N.Y. **1961**.
- ⁹ Debnath, A.K. et al., *J. Med. Chem.* **1991**, 34, 786
- ¹⁰ Villemin, D.; Cherqaoui, D. e Mesbah, A., *J. Chem. Inf. Comput. Sci.* **1994**, 34, 1288.
- ¹¹ Nordén, U.E. e Svante, W., *Acta Chem. Scand.* **1978**, B32, 602.
- ¹² Szentpály, L.V., *J. Am. Chem. Soc.* **1984**, 106, 6021.
- ¹³ Song, X.-H.; Xiao, M. e Yu, R.-Q., *Comp. Chem.* **1994**, 18, 391.
- ¹⁴ Miyashita, Y. et al., *An. Chim. Acta*, **1982**, 143, 35.
- ¹⁵ Galvão, D.S.; dos Santos, D.A.; Laks, B. et al., *Phys. Rev. Lett.* **1989**, 63, 786.
- ¹⁶ Baughman, R.H. e Galvão, D.S., *Nature (London)* **1993**, 365, 735.
- ¹⁷ Dewar, J.S. e Dieter, K.M., *J. Am. Chem. Soc.* **1986**, 108, 8075.
- ¹⁸ Stewart, J.-J.P., *J. Comp. Chem.* **1989**, 10, 209.
- ¹⁹ Stewart, J.J.P., *J. Comp. Chem.* **1991**, 10, 209-221; MOPAC program, version 6.0, Quantum Chemistry Exchange Program No. 455.
- ²⁰ Barone, P.M.V.B., *Tese de Doutorado – IFGW, UNICAMP* **1996**.
- ²¹ Barone, P.M.V.B.; Camilo Jr., A. e Galvão, D.S., *J. Phys. Chem.*, submetido para publicação.

²² Braga, R.S.; Barone, P. M. V. B. e Galvão, D.S., *THEOCHEM*, "*Identifying Carcinogenic Activity of Polycyclic Aromatic Hydrocarbons*", submetido para publicação em Outubro de 1997.

²³ Braga, R.S.; Barone, P. M. V. B. e Galvão, D.S., em preparação.

²⁴ Lima, L.E.S., *THEOCHEM*, "*Structure-activity study of Indoquinones Bioreductive Alkylating Agents*", submetido para publicação em Outubro de **1997**.

²⁵ Paupitz, R., "*Estudo estrutura-atividade do Taxol e derivados*", *Tese de Mestrado*, **1998**, IFGW-UNICAMP.



Capítulo III

Problema e metodologia

III.1 Introdução

No capítulo anterior descrevemos a potencialidade da teoria de índices eletrônicos e nos propusemos a adaptá-la ao estudo de alguns inibidores da integrase do HIV. Como no caso da análise dos PAHs^[1], pretendemos investigar a existência de indicadores eletrônicos que possam ser correlacionados à presença ou ausência de atividade biológica desses compostos anti-HIV.

Devemos frisar que trata-se de um estudo teórico com resultado qualitativo. Para que fosse possível obter informação quantitativa a respeito do poder inibitório dos compostos, outros fatores deveriam ser levados em conta: solubilidade, hidratação, efeitos de solvente, interação enzima-substrato, etc., o que tornaria o problema muito mais difícil ou mesmo insolúvel do ponto de vista físico, dada a capacidade dos atuais computadores. Além disso, como já mencionado no capítulo 1, a interação enzima-substrato para a integrase é ainda desconhecida.

A razão da escolha por um processo enzimático como objeto de estudo, deve-se ao fato de a ligação entre a enzima e o substrato não alterar, *a priori*, a estrutura deste e de não haver processos de transferência de carga^[2]. Estes fatos possibilitam a especulação de que a estrutura eletrônica possa funcionar como indicador da atividade biológica dos compostos.

O sucesso da teoria de índices eletrônicos para o caso dos PAHs foi a grande motivação para o presente trabalho. Neste capítulo mostraremos quais foram os

passos necessários à adaptação da metodologia base, bem como alguns outros detalhes sobre os inibidores de integrase, que não foram discutidos nos capítulos anteriores, e uma pequena introdução sobre os métodos de cálculo comumente utilizados em química computacional.

III.2 Os compostos inibidores de integrase

Como discutido no capítulo 1, um dos focos da atual pesquisa anti-AIDS concentra-se na investigação de compostos capazes de inibir a enzima integrase (IN) do HIV. A IN tem a vantagem de não ser nativa ao corpo humano; a protease viral (PR), em contrapartida, possui similares nas células humanas que também são combatidas pelos inibidores anti-HIV^[3]. Além disso, por ser um “alvo” ainda muito pouco estudado, existe a perspectiva de que resultados mais eficientes possam ser obtidos.

Vários trabalhos experimentais vêm investigando intensivamente os requisitos necessários para se desenvolver inibidores de integrase mais potentes e menos tóxicos para um tratamento mais eficiente contra a AIDS^[3,4,5,6].

O compromisso entre a alta atividade biológica e a baixa toxicidade das drogas é um dos mais importantes tópicos no desenvolvimento de fármacos. Muitas vezes é necessário abrir mão de compostos menos tóxicos em prol de alta atividade biológica para que estes, ministrados em menores concentrações, produzam efeitos mais compensadores^[3].

O índice utilizado para medir a eficiência de um composto anti-HIV é chamado de IC_{50} . Este dá a concentração de inibidor necessária para inibir 50% de uma certa amostra de integrase^[3]. Sendo assim, quanto menor ele for, maior será a

atividade inibitória do composto a ele relacionado. O IC_{50} é medido através de uma interpolação de valores de concentração obtida por experiências de eletroforese^[3,5] e tem como unidade o micro molar (μM)*.

Dos muitos trabalhos experimentais que analisamos em busca de um conjunto de compostos para o nosso estudo^[3,4,5,6], um deles nos chamou muito a atenção^[3]: trata-se de um resumo dos últimos resultados da época, mostrando o aparecimento de um padrão ou modelo topológico presente na grande maioria dos compostos que apresentavam atividade biológica significativa. De maneira geral podemos descrevê-lo da seguinte maneira:

arila hidroxilada – conector – arila (hidroxilada),

como esquematizado na figura 3.1 a seguir, onde arila é um conjunto de anéis aromáticos ligados entre si direta ou indiretamente. O padrão dihidroxi-arila é chamado de catecol e é altamente tóxico para as células humanas^[3,6].

Os principais resultados daquele trabalho experimental^[3] podem ser resumidos da seguinte forma:

- A tentativa de se diminuir o poder tóxico dos compostos, eliminando o padrão catecol dos mesmos, conduziu a inibidores com pouca atividade inibitória;
- Como resultado de diversas sínteses intermediárias, descobriu-se 4 compostos entre 6 que, apesar de não se encaixarem no modelo topológico da figura 3.1, apresentavam atividade biológica bastante significativa. Estas moléculas podem ser vistas na figura 3.2 a seguir e passaremos a identificá-las por “monômeros”, pois possuem metade da topologia sugerida acima;

* $1\mu M$ de uma substância cujo peso molecular é dado por x, por exemplo, equivale a uma concentração da mesma de $(10^{-6}.x)$ g/l.

- Através de um processo de “dimerização” dos compostos da figura 3.2, chegou-se a 2 compostos de um conjunto de 6 (figura 3.3), que apresentavam atividade inibitória bastante alta. Daremos o nome de “dímeros” a estes compostos, mesmo não sendo a terminologia mais correta, como veremos;

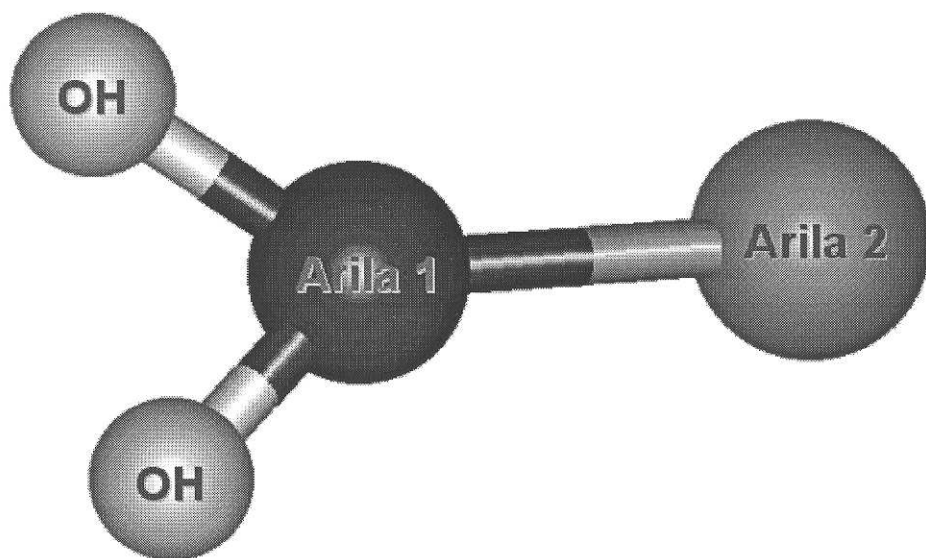
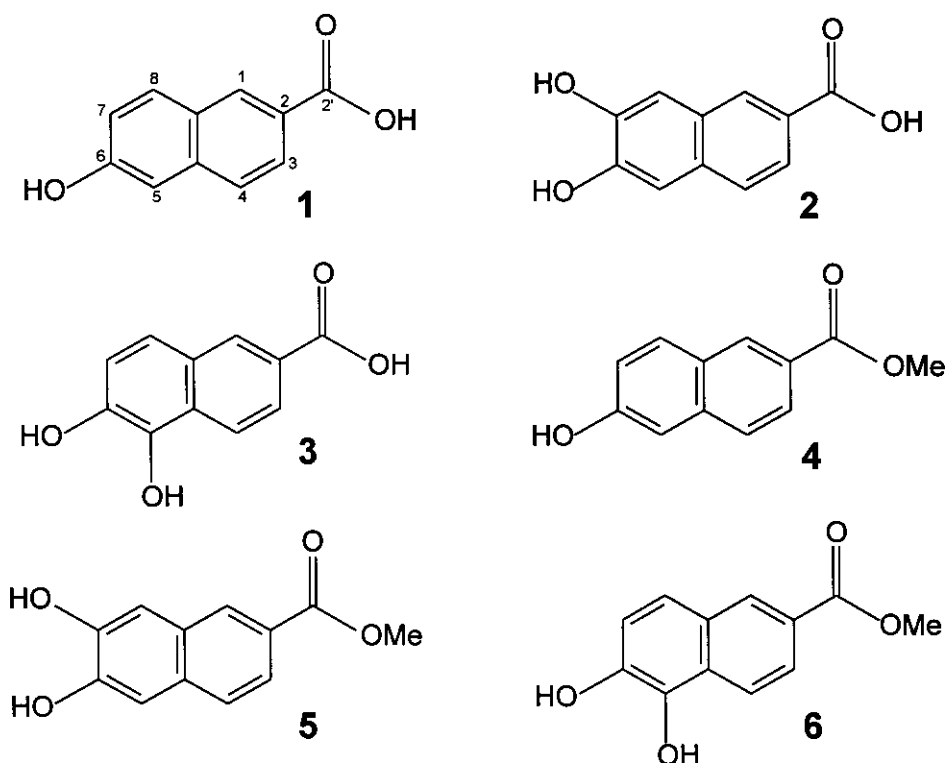


FIGURA 3.1: padrão topológico apresentado pela grande maioria dos compostos inibidores de integrase. Alguns exemplos de compostos que se encaixam nesse modelo foram dados na introdução deste trabalho.

- Tanto os “monômeros” quanto os “dímeros” possuem o padrão tóxico catecol. Entretanto, como alguns deles são muito ativos ($IC_{50} < 1\mu M$), tornam-se elementos mais interessantes que outros menos tóxicos e menos ativos;
- Os compostos com dihidroxilação são mais ativos que aqueles com hidroxilação simples, daí a dificuldade de se eliminar o padrão catecol mantendo a atividade biológica das moléculas;

- Se compararmos o composto **2** (muito ativo) com o composto **5** (inativo), chegaremos à conclusão de que a esterificação (substituição do hidrogênio por metil ao lado do oxigênio) tornaria um composto deste grupo inativo.
- A conclusão anterior é contrariada pela comparação dos compostos **3** e **6**. Neste caso a esterificação teve pouco ou quase nenhum efeito sobre a atividade do composto **3**. A atividade do composto **6** permaneceu sem explicação plausível. Comentou-se apenas que este fato seria um indício de que os inibidores estariam interagindo com a enzima de maneiras notadamente distintas.

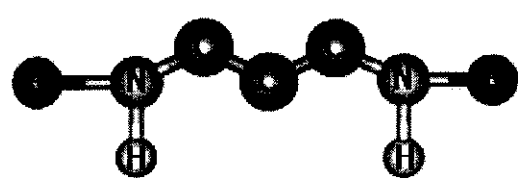


#	NOME	IC ₅₀ (μM)*	ATIVIDADE
1	6-hydroxy-2-naphtoic acid	172.3	Pouca
2	6,7-dihydroxy-2-naphtoic acid	4.7±0.6	Muita
3	5,6-dihydroxy-2-naphtoic acid	62.4±1.1	Média
4	6-hydroxy-2-naphtoic acid methyl ester	>200	Inativo
5	6,7-dihydroxy-2-naphtoic acid methyl ester	>200	Inativo
6	5,6-dihydroxy-2-naphtoic acid methyl ester	52.7±8.9	Média

*valores para o 3' processing.

FIGURA 3.2: Os “monômeros” inibidores de integrase.

Estrutura geral:



#	R	IC ₅₀ (μM)*	ATIVIDADE
7		>200	inativo
8		>200	inativo
9		>200	inativo
10		>200	inativo
11		0.89	alta
12		0.17	alta

*média dos valores para 3' processing e strand transfer.

Figura 3.3: Os “dímeros” inibidores de integrase

- Analisando os compostos da figura 3.3 conclui-se que, através de um processo de “dimerização”, a tendência apresentada pelos “monômeros” da figura 3.2 é amplificada: o composto **8**, que deriva de **1**, é completamente inativo, pois a atividade de **1** era baixa e muito provavelmente teria a tendência de ser inativo. Já os compostos **11** e **12**, que derivam de **2** e **3** respectivamente, têm atividade bastante amplificadas em relação os “monômeros” de que derivam.

Vale ressaltar que não se pode concordar totalmente com esta última análise, pois o conceito de “dimerização” não é aplicável de forma rigorosa neste caso. Tomemos por exemplo o composto **11** da figura 3.3; observando-se agora a figura 3.2, podemos dizer que ele poderia ser igualmente derivado dos compostos **2** e **5**, não sendo portanto a “dimerização” (com o conector da figura 3.3) de um desses compostos em particular. Além disso, o composto **2** é ativo enquanto o **5** é inativo.

Através das discussões anteriores, conclui-se que os compostos das figuras 3.2 e 3.3 devem ser tratados como doze compostos independentes apesar de possuírem muitas características estruturais em comum. São exatamente essas doze moléculas que compõem o objeto de nosso estudo.

Os “monômeros” se encaixam perfeitamente ao tipo de estudo realizado com os PAHs por serem também sistemas conjugados^[7], isto é, sistemas com alternância de ligações simples e duplas. Tal característica confere a essas moléculas algumas características importantes como planaridade (exceção para aquelas que possuem o grupo metil) e a separação dos orbitais σ e π , simplificando a análise eletrônica das mesmas.

Uma observação atenta dos compostos da figura 3.2 leva-nos a perguntar o porquê de possuírem atividades tão distintas se suas estruturas são tão semelhantes. Podemos ainda nos perguntar se existe alguma característica eletrônica que possa ser correlacionada a esse comportamento.

Sabemos que em um sistema conjugado, uma perturbação na nuvem eletrônica em algum ponto é transmitida por toda a molécula, alterando sua distribuição global de carga^[7]. Sendo assim, poderíamos dizer que essas pequenas alterações estruturais seriam responsáveis por alterações eletrônicas significativas no sítio ativo, resultando na atividade ou não atividade do composto. A intensidade da atividade seria, no entanto, determinada por muitos outros fatores químicos e biológicos, como já mencionado anteriormente.

Já os “dímeros” da figura 3.3 representam uma situação deveras diferente: não são compostos conjugados, pois o conector não é um elemento conjugado. Isto faz com que as duas unidades “monoméricas” fiquem quase isoladas, em termos de carga, uma da outra. Por este motivo esses compostos não são planares e, como veremos adiante, apresentam um espaço conformacional muito rico, o que torna a aplicação da teoria de índices eletrônicos não trivial e de estratégia difícil.

III.3 Métodos de Química Computacional

Nesta seção pretendemos fazer um rápido sumário sobre os principais métodos utilizados em química computacional, sem o intuito de nos aprofundarmos nos muitos detalhes envolvidos em cada um deles.

Dado um sistema molecular, ou seja, um conjunto de diversos átomos ligados entre si, podemos sobre ele realizar vários tipos de cálculos, cuja escolha irá

depende do tipo de estudo a que estamos nos propondo. Podemos dividi-los em dois grandes grupos genéricos: mecânica molecular (tratamento "clássico") e métodos mecânico-quânticos. Estes concentram ainda os métodos: empíricos (tipo Hückel), semi-empíricos e *ab initio*.

Os métodos supracitados consideram o sistema molecular em repouso e no vácuo. Para uma análise dinâmica deve-se utilizar os métodos de dinâmica molecular, que podem levar em conta os métodos de mecânica molecular, semi-empíricos ou *ab initio* para cada passo da evolução temporal do sistema. Não discutiremos os aspectos de dinâmica molecular, pois não analisamos resultados deste tipo de cálculo em nosso estudo.

Efeitos de solventes podem ser levados em conta direta ou indiretamente, seja via adição de moléculas de solvente ao sistema ou via parametrização, respectivamente. Neste trabalho não pretendemos estudar os efeitos de solventes, porém esta é uma extensão necessária que será levada a cabo brevemente, inclusive para o caso dos PAHs.

Se usarmos um dos métodos supracitados para calcularmos alguma propriedade do sistema quando este está em uma dada configuração de átomos (um ponto em sua hipersuperfície de energia), diremos que efetuamos um cálculo single point. Podemos ir mais longe e usar vários cálculos deste tipo para varrer o hiperespaço de fase, enquanto se altera a disposição dos átomos, em busca de regiões onde o gradiente é baixo e o sistema estável. Esta busca por mínimos de energia é chamada de otimização de geometria e utiliza métodos especiais que a tornam eficiente^[8,9]. A maioria desses métodos utiliza uma varredura sistemática na

direção contrária a do vetor gradiente, que sempre dá a direção de maior variação da função, até que a variação de energia seja menor que um limite predefinido.

Os cálculos de otimização de geometria são bastante custosos do ponto de vista computacional, dado o grande número de ciclos de cálculo que precisam ser realizados. São também muito importantes, pois na maioria dos casos não se tem a informação sobre a estrutura dos compostos que se quer estudar e uma análise conformacional[†] se faz necessária. Este é o caso do presente estudo, já que nenhuma informação sobre a geometria 3D das moléculas das figuras 3.2 e 3.3 está disponível na literatura.

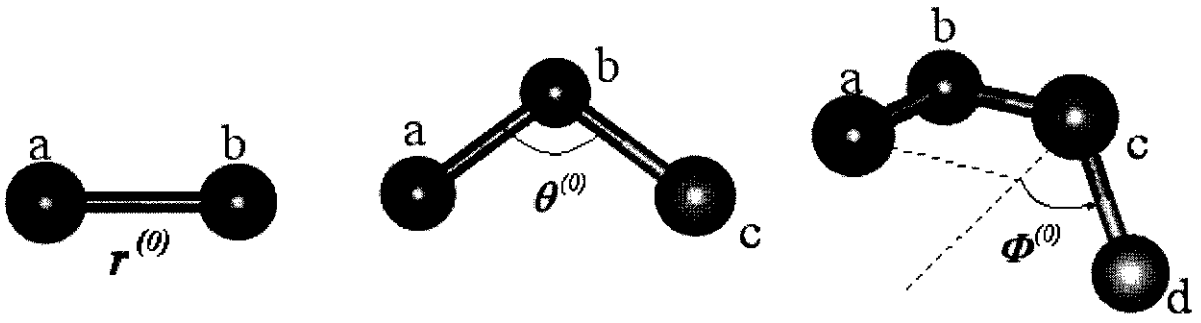
III.3.1 Mecânica Molecular

A filosofia básica dos métodos de mecânica molecular é usar funções analíticas parametrizáveis para descrever as diversas interações entre os átomos que compõem a molécula, ou seja, fundamenta-se na possibilidade de particionar a energia total do sistema em termos independentes. A forma de tais funções depende da interação que se quer descrever. Seus parâmetros são obtidos, dependendo dos átomos que participam da interação, pela comparação com dados experimentais.

A figura 3.4 a seguir mostra um exemplo de campo de força, como é chamado o conjunto de expressões parametrizáveis que compõem um método de mecânica molecular.

Em um campo de força genérico podemos, como no exemplo da figura 3.4, descrever a ligação entre dois átomos **a** e **b** por um oscilador harmônico

simples. Os parâmetros ajustáveis $K_r^{(0)}$ e $r^{(0)}$, representando respectivamente a constante de força e a distância de equilíbrio do oscilador, dependem dos átomos que estão participando da ligação e seus valores são obtidos através da comparação com dados experimentais, extraídos de um conjunto de moléculas escolhido como padrão.



Exemplo de Campo de Força:

$$E_{total} = V_1 + V_2 + V_3 + V_4 + V_5 + \dots + V_n$$

$$V_1 = V_{stretch} = \sum_{bonds} K_r^{(0)} (r - r^{(0)})^2$$

$$V_2 = V_{bend} = \sum_{angles} K_\theta^{(0)} (\theta - \theta^{(0)})^2$$

$$V_3 = V_{dihedrals} = \sum_{dihedrals} \frac{V_n^{(0)}}{2} [1 + \cos(n\Phi - \Phi^{(0)})]$$

$$V_4 = V_{vdW} = \sum_{i < j} \left[\frac{A_{ij}^{(0)}}{R_{ij}^{12}} - \frac{B_{ij}^{(0)}}{R_{ij}^6} \right]$$

$$V_5 = V_{El} = \sum_{i < j} \frac{q_i q_j}{\epsilon^{(0)} R_{ij}}$$

FIGURA 3.4: exemplo de um campo de força típico, mostrando cinco termos parametrizáveis da energia total. Os parâmetros ajustáveis são indexados com o valor (0) à direita e acima, por exemplo: $\epsilon^{(0)}$.

† Cada ponto estável na superfície de energia define um conformero.

O ângulo entre três átomos também pode ser descrito como um oscilador, como se pode ver pela expressão V_2 anterior. Os ângulos diedrais[‡] são apropriadamente descritos por séries truncadas de Fourier (V_3), pois estas os conferem o caráter de simetria e os vários mínimos sempre presentes. A interação eletrostática é dada pela expressão comum de Coulomb (V_3). A interação de van der Waals, que define os impedimentos estéricos, pode ser descrita de várias formas; no exemplo apresentado, usa-se o potencial de Lenard-Jones, que é relativamente fácil de se calcular e parametrizar (V_4). Muitas outras interações, inclusive termos cruzados, podem ser incluídas para tornar o campo de força mais preciso e abrangente^[10].

Um dado campo de força, em associação a um conjunto de parâmetros, define um método específico de mecânica molecular. Muitos métodos estão disponíveis hoje em dia, a maioria deles parametrizados para sistemas orgânicos, fornecendo ótimos resultados qualitativos^[11,12,13,14]. Os resultados serão tanto melhores quanto mais próxima estiver a estrutura da molécula que está sendo estudada daquelas usadas como padrão, já que em última análise os métodos de mecânica molecular podem ser descritos como procedimentos de *fitting*.

Um exemplo de campo de força muito difundido é o MM2^[11], desenvolvido com o intuito de descrever pequenos sistemas orgânicos. Um outro exemplo, aplicável principalmente a macromoléculas de interesse biológico, como proteínas e ácidos nucleicos, é o AMBER^[12].

[‡] O ângulo diedral formado pelos átomos a-b-c-d define o quanto este último distancia-se do plano formado pelos outros três.

A grande vantagem dos métodos de mecânica molecular é a rapidez com que podem ser processados, uma vez que a maioria das funções analíticas utilizadas é de cálculo trivial para os computadores de hoje em dia.

III.3.2 Métodos mecânico-quânticos

Os métodos baseados na mecânica quântica partem da formulação da equação de Schrödinger independente do tempo ($H\psi=E\psi$). Diversos tratamentos são dados a esta equação cuja solução é composta pelas funções de onda do sistema (ψ , ou orbital molecular) e as correspondentes energias (E), com as quais se pode extrair todas as informações de relevância química ou física sobre o sistema molecular.

O que difere um método de outro é a forma como simplificam as integrais decorrentes da formulação de Schrödinger. O hamiltoniano H também pode ser simplificado de várias formas, dependendo do tipo de interações que se considerem presentes entre núcleos e elétrons.

A aproximação que praticamente todos os métodos utilizam é a de Born-Oppenheimer^[15], que separa o Hamiltoniano nuclear do eletrônico, baseando-se no fato de os núcleos serem muito mais pesados que os elétrons. Estes podem, então, ser considerados como movendo-se no campo produzido pelos núcleos estacionários em cada instante de tempo em que a função de onda eletrônica é calculada.

Um hamiltoniano eletrônico típico inclui as interações elétron-núcleo, elétron-elétron e a energia cinética dos elétrons. O hamiltoniano molecular inclui

ainda as interações núcleo-núcleo e a energia cinética dos mesmos (veja apêndice A).

Entre os métodos baseados em mecânica quântica e utilizados para o estudo da estrutura eletrônica de compostos orgânicos, destacam-se os *ab initio* e os semi-empíricos.

O cálculo *ab initio* utiliza o hamiltoniano molecular e somente os valores das constantes físicas fundamentais, sendo que todas as integrais decorrentes da formulação de Schrödinger são calculadas. A única aproximação significativa é feita sobre a base que define a função de onda do sistema: usualmente expande-se os orbitais moleculares em termos de funções gaussianas. A qualidade das bases é medida pela forma e número de gaussianas utilizadas. Os métodos *ab initio* são muito custosos para moléculas de médio e grande portes, quando o cálculo computacional pode tornar-se inviável dado o grande número de integrais a resolver.

Em contraste, os métodos semi-empíricos utilizam um hamiltoniano mais simples e parâmetros cujos valores são ajustados baseados em dados experimentais. Os métodos semi-empíricos vêm sendo desenvolvidos nas últimas décadas e são capazes de produzir ótimos resultados quantitativos^[16]. Dependendo da forma como aproximam as integrais decorrentes da formulação de Schrödinger e dos tipos de parametrização utilizada, dão origem a diversos métodos. Como exemplo temos o MNDO (*Modified Neglect of Differential Overlap*), MINDO (*Modified Intermediate Neglect of Differential Overlap*), AM1 (*Austin Method 1*), PM3 (*Parametric Method 3*)^[17]. No apêndice

A, analisamos mais detalhadamente o método de Hartree-Fock e alguns dos métodos semi-empíricos mais utilizados.

O método de Hückel é muito mais simples que qualquer método semi-empírico. A simplificação feita sobre o hamiltoniano do sistema é tão grande, que não se calcula nenhuma integral, todos os valores sendo substituídos por parâmetros extraídos de resultados experimentais. Por este motivo é considerado um método empírico. Em contrapartida, o método de Hückel só pode ser utilizado para sistemas conjugados (planares), onde há uma grande separação em energia entre os orbitais σ e π . A metodologia não leva em conta todos os elétrons de valência, tratando apenas aqueles que estão ocupando orbitais π . Para os sistemas conjugados, onde as propriedades químicas são determinadas por esse tipo de elétrons^[7,18], isto é válido e simplifica muito a descrição do problema.

Um cálculo típico de Hückel envolve apenas uma construção e diagonalização de uma matriz bastante simples, que contém a informação topológica da conectividade atômica da molécula. Vamos descrever melhor o método de Hückel no capítulo 5 com o intuito de introduzir o cálculo de densidade local de estados (LDOS).

Os nossos cálculos foram realizados utilizando-se o programa MOPAC^[17] (*Molecular Package - Quantum chemistry Exchange Program n. 464*), um “pacote” de programas para o estudo de reações químicas e otimização de geometrias, onde os hamiltonianos MNDO, MINDO/3, AM1 e PM3 estão implementados de forma integrada. Existe um consenso de que tanto o AM1 quanto o PM3 sejam melhores que o MNDO e o MINDO/3 para

otimizações geométricas. O PM3 é basicamente uma melhoria, em termos de parametrização, do método AM1 (veja apêndice A) e por este motivo foi o método escolhido para os cálculos semi-empíricos desenvolvidos no presente trabalho.

Dada a complexidade do espaço de fase dos “dímeros”, como mostraremos a seguir, outros programas foram utilizados (HyperChem) e desenvolvidos (HyperSpin e Chem2Pac), como explicitaremos nos capítulos posteriores.

III.4 A adaptação da teoria de índices eletrônicos

Para o cálculo da densidade de estados sobre diversas regiões das moléculas, necessitamos dos autovetores da matriz de Fock/hamiltoniano, ou seja, dos coeficientes da expansão dos orbitais moleculares em termos de orbitais atômicos. Estes coeficientes podem ser obtidos por qualquer um dos métodos de mecânica quântica descritos na seção anterior.

O uso do método de Hückel foi impossibilitado pela não planaridade dos “dímeros”, pois neste caso não existe a separação dos orbitais σ e π . Mesmo para o caso dos “monômeros”, o uso do método de Hückel é impedido pela falta de parâmetros para o grupo $O - Me$ ou $O - CH_3$, que, a priori, poderia ser tratado como uma perturbação aos parâmetros dos átomos de carbono em que é ligado.

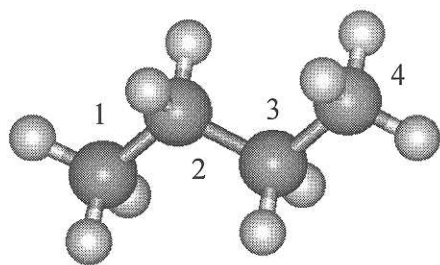
No outro extremo, o uso de um método *ab initio*, descrito com uma base boa, esbarra na capacidade dos computadores atualmente disponíveis. A título de exemplo, um único cálculo de otimização de geometria (base 6-31G*) para um dos “dímeros” pode levar vários dias de cálculo na máquina ODIN (a de melhor

desempenho na rede do IFGW). A busca pelo mínimo global de todas as 6 moléculas da figura 3.3, usando métodos sofisticados de primeiros princípios, levaria muito mais que o prazo de um ano e meio que tivemos para o desenvolvimento do projeto como um todo.

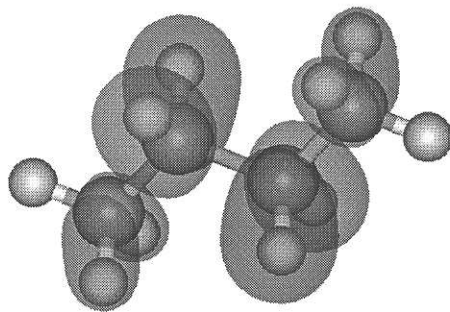
Como única alternativa restou-nos o uso dos métodos semi-empíricos. Dentre as várias possibilidades dentro desta categoria, escolhemos o PM3. Tal escolha deve-se ao ótimo resultado obtido para o caso dos PAHs e por se tratar de um dos mais sofisticados métodos de sua classe.

Os cálculos semi-empíricos são muito mais rápidos que os *ab initio* e estes nem sempre oferecem melhores resultados^[10]. Um cálculo de otimização de geometria, utilizando o PM3, para um dos “dímeros” não demoraria mais que alguns minutos na ODIN, levando menos de uma hora em um PC-486. Mesmo assim, para uma busca conformacional refinada, muito recurso computacional é requerido e decidimos utilizar uma pré-otimização via mecânica molecular para então otimizar usando o PM3. Mais detalhes e justificativas para tal procedimento serão dados no próximo capítulo.

Pretendemos dar agora uma idéia da complexidade envolvida numa busca conformacional. Em referência à figura 3.5 apresentada a seguir, vamos calcular a energia total da molécula de n-butano com respeito a apenas um de seus graus de liberdade (coincidentemente o mais importante): o ângulo diédral central (1-2-3-4).



(a)



(b)

FIGURA 3.5: (a) representação gráfica da molécula de n-butano C_4H_{10} .
(b) representação 3D de seu último orbital ocupado (HOMO)^[19].

A figura 3.6 a seguir mostra que a superfície unidimensional, obtida utilizando-se o método de mecânica molecular MM+^[10], está longe de ser simples: apresenta três mínimos, sendo que apenas um deles é o mínimo global. Além disso, uma vez que a molécula caia em um dos mínimos locais (B e F na figura), pode ser altamente improvável que ela “saia” desta região para atingir o mínimo global, necessitando para tal vencer uma alta barreira de potencial (no exemplo algo em torno de 2 Kcal/mol \cong 11meV). O resultado da análise conformacional para o n-butano concorda com o observado experimentalmente^[10].

Muitas vezes a molécula consegue obter do meio em que está imersa a energia necessária para atravessar barreiras conformacionais, mas há situações onde a possibilidade de se cair em um mínimo local profundo possa representar uma situação irreversível^[20].

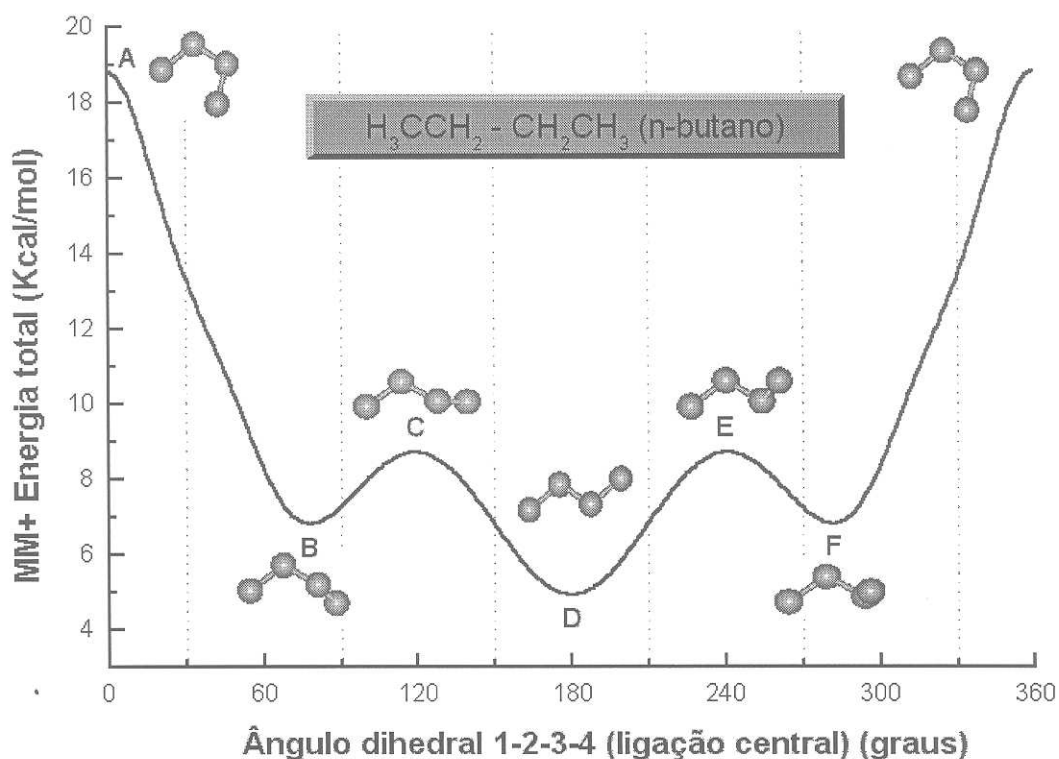


FIGURA 3.6: Análise conformacional de um dos graus de liberdade da molécula de n-butano vista na figura 3.4.

A complexidade das superfícies hiperdimensionais fica clara se imaginarmos uma molécula muito maior em tamanho e apresentando muito mais graus de liberdade que o n-butano. A exemplo, mostramos na figura 3.6 o “dímero” **7** (o mais simples dos seis), já em sua conformação de mínimo global, que obtivemos através de uma análise conformacional minuciosa e o método PM3.

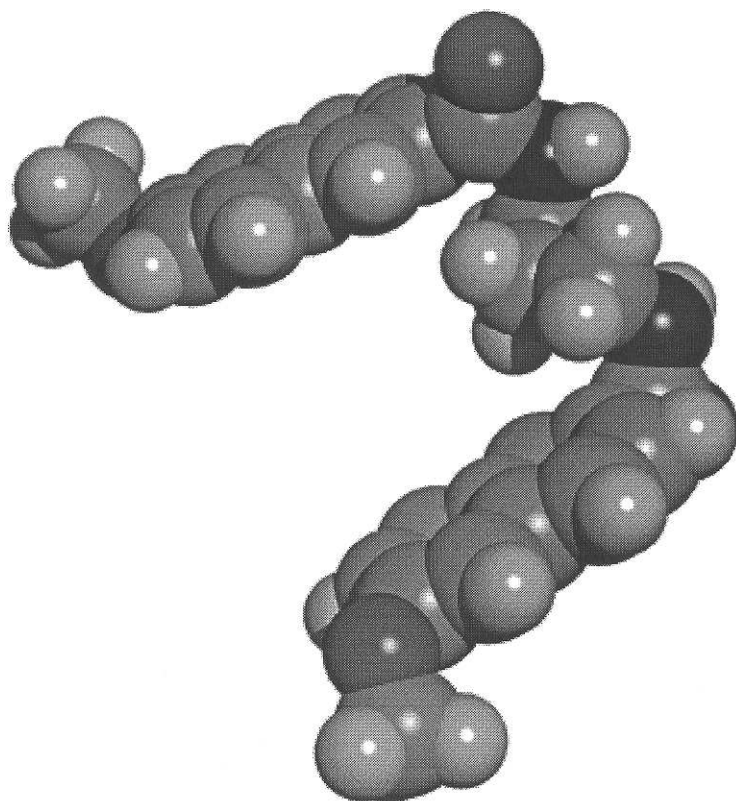
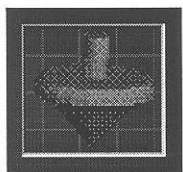


FIGURA 3.7: O “dímero” 7 da figura 3.3 visto em sua conformação de menor energia. Resultado obtido com o método PM3.

No próximo capítulo apresentaremos um software especialmente desenvolvido para facilitar, e muito, a busca conformacional de moléculas de porte médio e que foi fundamental para a realização desse projeto no curto espaço de tempo disponível. A otimização da molécula acima tornou-se muito simplificada com o uso deste programa, que permitiu selecioná-la como mínimo global dentre um grupo de mais de 50 confórmeros encontrados.

- ¹ Barone, P.M.V.B.; Camilo Jr., A. e Galvão, D.S., *Phys. Rev. Lett.* **1996**, 77, 1186.
- ² Marzzoco, A. e Torres, B.B. "*Bioquímica básica*", Editora Guanabara, **1990**.
- ³ Zhao, H.; Neamati, N.; Mazumder, A.; Sunder, S.; Pommier, Y. e Burke Jr., T.R., *J. Med. Chem.* **1997**, 40, 1186 e referências citadas.
- ⁴ Hong, H.; Neamati, N.; Wang, S. *et al.*, *J. Med. Chem.* **1997**, 40, 930-936.
- ⁵ Mazumder, A.; Neamati, N.; Sunder, S. *et al.*, *J. Med. Chem.* **1997**, 40, 3057-3063.
- ⁶ Zhao, H.; Neamati, N.; Hong, H. *et al.*, *J. Med. Chem.* **1997**, 40, 242-249 e referências citadas.
- ⁷ Salem, L., "*The molecular orbital theory of conjugated systems*", W.A. Benjamin, London **1972**.
- ⁸ Fletcher, R., "*Practical Methods of Optimization*", John Wiley & Sons, New York, **1980**.
- ⁹ Gill, P.E.; Murray, W. e Wright, M.H. "*Practical Optimization*", Academic Press, Inc., New York, **1981**
- ¹⁰ HyperChem™ *Computational Chemistry manual*, **1994**, © HyperCube inc.
- ¹¹ Allinger, N.L., *J. Am. Chem. Soc.* **1977**, 99, 8127
- ¹² Weiner, S.J.; Kollman, P.A. *et al.*, *J. Am. Chem. Soc.* **1984**, 106, 765-784.
- ¹³ Brooks, B.R.; Bruccoleri, R.E. *et al.*, *J. Comput. Chem.* **1983**, 4, 187-217.
- ¹⁴ Jorgensen, W.L. e Tirado-Rives, J., *J. Am. Chem. Soc.* **1988**, 110, 1657-1666.
- ¹⁵ Pople, J.A. e Beveridge, D.L., "*Approximate Molecular Orbital Theory*" McGraw-Hill, New York **1970**.
- ¹⁶ Stewart, J.J.P., "*Reviews of Computational Chemistry*", Lipkowitz, K. e Bouyd, D.B., Eds. VCM Publishers, New York, **1990**.
- ¹⁷ Stewart, J.J.P., *J. Comp. Chem.* **1991**, 10, 209-221; MOPAC program, version 6.0, Quantum Chemistry Exchange Program No. 455.
- ¹⁸ Barone, P.M.V.B., *Tese de Doutorado – IFGW, UNICAMP* **1996**.
- ¹⁹ A renderização 3D de orbitais moleculares foi efetuada com o programa HyperChem™ versão 5.01, © **1997** HyperCube, Inc. (<http://www.hyper.com>).
- ²⁰ Jensen, J.H. e Gordon, M.S., *J. Am. Chem. Soc.* **1991**, 113, 7917-7924.



Capítulo IV

Análise conformacional: software HyperSpin

IV.1 Introdução

No capítulo anterior descrevemos a dificuldade de se realizar uma análise conformacional conclusiva sobre moléculas de médio ou grande portes, pois nestes casos o número de graus de liberdade cresce rapidamente, tornando a hipersuperfície de fase bastante complexa. Neste capítulo apresentaremos um software especialmente desenvolvido com o objetivo de facilitar este trabalho. Ele possibilita efetuar a varredura automática de partes da superfície de energia e substituir um trabalho manual bastante repetitivo e exaustivo.

O programa, batizado de HyperSpin^[1], tem como versão final um módulo 32bits rodando no ambiente Windows95 e se comunicando com o software comercial HyperChem^[2], que atuará como seu servidor de dados. Além do módulo principal, vários outros itens foram implementados com o objetivo de facilitar a análise de resultados e a criação de futuros programas que interajam com o HyperChem.

Para que fiquem claros a importância dos programas desenvolvidos e a dificuldade de implementá-los de maneira eficiente, vamos apresentar primeiramente um resumo sobre o HyperChem e seus recursos. Logo em seguida, vamos descrever o trabalho que deveria ser manualmente realizado para o estudo conformacional das moléculas que pretendemos analisar. Por fim apresentaremos os detalhes do programa e dos outros componentes que com ele formam um pacote

de utilitários especiais para o HyperChem e que tem sido bastante utilizado por alguns membros de nosso grupo de pesquisa. Vale ressaltar que, a despeito de suas utilidades, ferramentas similares não estão comercialmente disponíveis.

IV.2 HyperChem

O programa comercial HyperChem é uma interface gráfica que permite ao usuário, de uma forma bastante simples e muitas vezes eficiente, desenhar moléculas e efetuar sobre elas cálculos de química computacional utilizando técnicas de mecânica e dinâmica molecular com métodos *ab initio* ou semi-empíricos. Pelo fato de operar em uma plataforma PC, os métodos semi-empíricos e principalmente *ab initio*, tornam-se bastante lentos e ineficientes. Foi por esse motivo que nossos cálculos com o PM3 foram realizados em estações de trabalho, através do pacote MOPAC^[3], apresentando eficiência várias ordens de grandeza maior que os mesmos cálculos com o HyperChem.

Por outro lado, dadas as facilidades de edição e visualização 3D e aos métodos de mecânica e dinâmica molecular que o HyperChem proporciona, este programa foi utilizado em muitas etapas deste trabalho, como nos seguintes procedimentos:

1. Desenho 2D das moléculas e modelo 3D inicial;
2. Pré-otimização de geometria das moléculas com mecânica molecular;
3. Uso de dinâmica molecular com a finalidade de gerar o maior número de conformeros possível, que mais uma vez terão suas geometrias otimizadas pela mecânica molecular;

4. Refinamento da análise conformacional das moléculas escolhidas durante o passo anterior, observando os graus de liberdade mais instáveis, como ângulos diedrais. É nesta etapa que o programa HyperSpin mostra a sua maior utilidade;
5. Preparo dos *inputs* para o MOPAC;
6. Visualização dos resultados do MOPAC (*outputs*) e desenho 3D de orbitais moleculares (como exemplificado na figura 3.5b do capítulo anterior), além de outras operações envolvendo a análise dos resultados.

A figura 4.1 a seguir mostra a tela principal do HyperChem após um cálculo *single point* para a molécula de antraceno. A renderização da molécula foi feita com esferas de van der Waals.

O processo de desenho, usando os recursos de edição do HyperChem é bastante simples, mesmo para moléculas de estrutura complexa, e não fosse a sua ineficiência nos cálculos mais avançadas, seria a principal ferramenta para os estudos realizados por nosso grupo. O seu uso, ainda sim, é fundamental aos passos 1-4 mencionados anteriormente e muito freqüente em nossos trabalhos.

O passo de número 4 é muito importante à localização do mínimo global. A busca pelo estado de menor energia das moléculas faz-se necessária ao nosso estudo, pois acreditamos que um inibidor enzimático possa ser caracterizado como ativo ou inativo quando se encontra nesta situação (desde que não haja transferência de carga no processo). Isto baseia-se no fato de que a grande maioria dos compostos de interesse biológico apresentam atividade quando em seus mínimos globais ou próximos deles^[4]. Lembramos que este poderia não ser o caso

se estivéssemos analisando proteínas, pois ainda não há um consenso de que elas não possam atuar biologicamente em conformações fora de seus mínimos globais^[5].

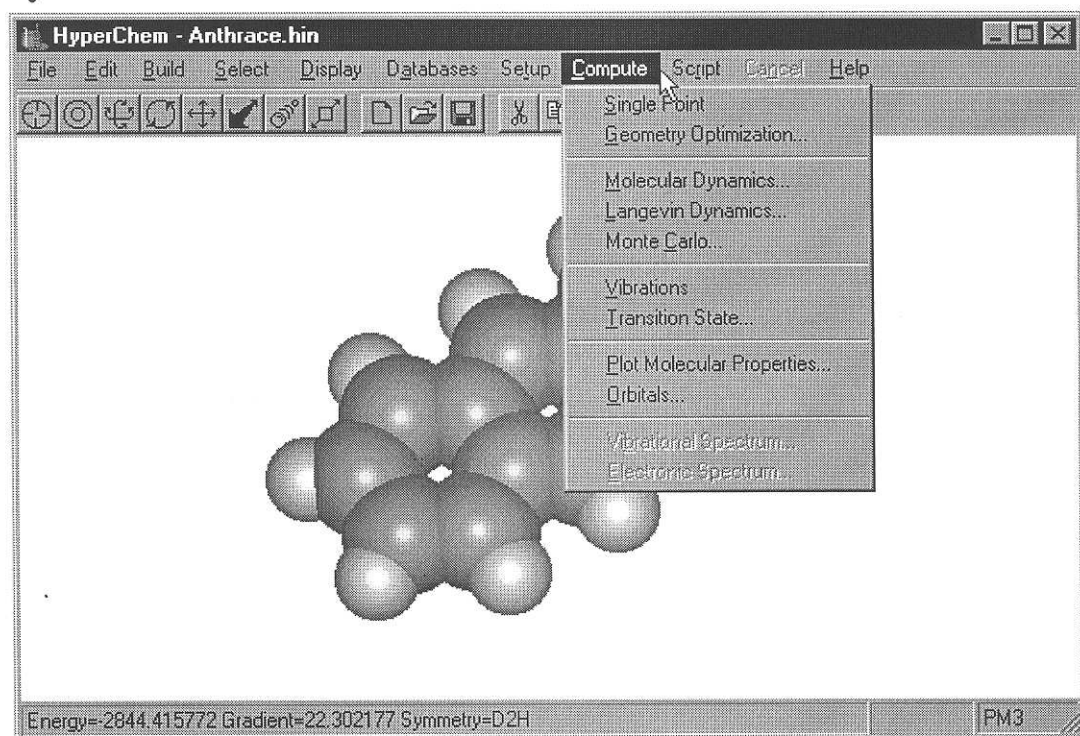


FIGURA 4.1: HyperChem, a interface gráfica que permite desenhar as moléculas e efetuar vários cálculos de química computacional (mecânica molecular, semi-empíricos, ab initio, Hückel Estendido, dinâmica, etc.).

IV.3 A busca conformacional refinada

Como dito anteriormente, os graus de liberdade mais susceptíveis são os ângulos diedrais, pois apresentam vários mínimos em que podem permanecer de maneira estável. Ângulos diedrais mal descritos podem levar a conformações distantes do mínimo global das moléculas e é por este motivo que devemos dar especial atenção a eles.

O processo de análise da variação da energia total com os ângulos diedrais foi inicialmente bastante exaustivo, pois utilizava os seguintes passos com o HyperChem*:

1. Colocar manualmente o diedro (definido por 4 átomos) na posição $\phi_{inicial}$;
2. “Setar” o cálculo para que não afetasse o valor do diedro (*restraint*);
3. Efetuar otimização de geometria (mecânica molecular ou semi-empírico);
4. Anotar o valor do ângulo e o resultado para a energia total do sistema definindo um ponto de cálculo;
5. Incrementar o valor de ϕ por um valor que depende da resolução da análise, tipicamente de 15 graus.
6. Reiniciar o processo até que $\phi = \phi + 2\pi$, ou, aproveitando a simetria que em muitos casos aparece, até que $\phi = \phi + \pi$.

Para cada ângulo diedral analisado separadamente (uma curva 1D sobre a hipersuperfície de energia), um mínimo de 30 pontos deve ser levantado de forma a caracterizá-lo apropriadamente. Uma molécula de porte médio pode possuir vários diedros e muitas vezes assimétricos, requerendo mais pontos.

O procedimento acima, para cada diedro, pode levar em média de uma a duas horas para se completar, dependendo da experiência do usuário com o HyperChem. Se quiséssemos analisar uma superfície 2D (dois diedros simultaneamente), o número de pontos cresceria rapidamente: $30 \times 30 = 900$ pontos (450 pontos se houver simetria), tornando inviável o levantamento manual. Análises de ordem superior seriam irrealizáveis.

* Este tipo de levantamento é chamado de cálculo de barreira de torção, pois relaciona a variação de

O programa HyperSpin foi escrito com a idéia de automatizar este processo e foi se desenvolvendo através das necessidades que foram surgindo durante a sua implementação, cujo histórico e mais detalhes apresentamos a seguir.

IV.4 HyperSpin: o programa para análise conformacional

A história deste programa iniciou-se em meados de Março de 1997. Sua primeira versão nascera de um programa bem simples escrito em PASCAL para o DOS (GetEnergy 1.0), que colhia alguns parâmetros fornecidos pelo usuário (método de cálculo, ângulo diédral, limites, etc.) e gerava um script, na linguagem que o HyperChem reconhece, com a finalidade de orientá-lo na rotação de um diédrico e no cálculo de sua barreira de torsão.

Ao final da rotação, permitia a leitura e extração de dados do arquivo de "log"[†] gerado pelo HyperChem e montava uma tabela que podia ser lida por outros aplicativos para a geração gráfica dos resultados (ângulo versus energia total da molécula). O layout da primeira versão do programa pode ser visto nas figuras 4.2(a e b) a seguir.

Apesar de muito simples e de todas as suas limitações, tal programa mostrou-se bastante útil, poupando muito tempo de trabalho repetitivo e sujeito a falhas humanas automatizando de forma muito simples os passos 1-6 citados anteriormente: o processo que levava cerca de uma hora já podia ser reduzido a alguns minutos.

energia do sistema com um ângulo diédral (também chamado ângulo de torsão).

[†] Os arquivos de "log" do HyperChem contêm grande parte das informações sobre os cálculos realizados.

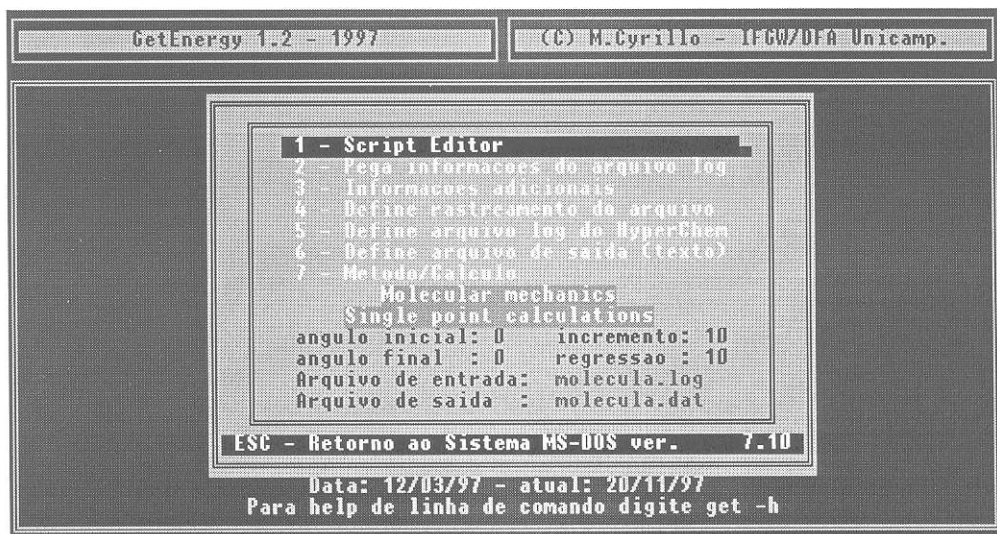


FIGURA 4.2a: Tela principal do programa que deu origem ao HyperSpin. Escrito em PASCAL e rodando em DOS.



FIGURA 4.2b: GetEnergy 1.0 e a tela onde o usuário entra com as opções do cálculo.

Ao final da rotação do dihedro, no entanto, este era mantido na posição final de busca, que na maioria das vezes não representava a situação de menor energia. Naturalmente surgiu a idéia de fazer com que o HyperChem recebesse de volta a informação sobre o ângulo que minimizava a energia total, obtido pela interpolação de todos os pontos lidos.

Algumas semanas foram investidas na tentativa de se fazer uma comunicação entre os dois programas, mas constatamos que isto seria impossível, dada a assincronia de execução de programas existente entre os sistemas MSDOS e Windows 3.11 ou 95. Isto se deve ao fato de problemas de design destes sistemas. O Windows deveria ser um ambiente independente, mas até a sua versão chamada de 95, nada ou pouco foi feito nesse sentido. Tem-se, então, um sistema operacional sofisticado, dependendo de outro bem mais simples (o DOS) e não podendo se comunicar com este de forma eficiente. Aquele que já tenha tentado sair do Windows com uma janela do DOS (onde um programa esteja rodando) aberta, entenderá melhor o que queremos dizer.

A única alternativa seria escrever o GetEnergy dentro do sistema Windows e para isso ter que mudar o programa e sua estrutura radicalmente. Escolhemos o Visual Basic (VB3.0), pois tornaria possível o desenvolvimento de programas amigáveis ao usuário através de seus recursos avançados de programação orientada a objetos (OOP ou *Object Oriented Programming*) e janelas. Alguns meses foram necessários para que confiança e conhecimentos básicos fossem adquiridos e fosse possível dar início ao desenvolvimento do programa.

O conceito básico e mais importante que foi aprendido, envolve o protocolo de comunicação DDE (*Dynamic Data Exchange*) do Windows, que permite a troca

dinâmica de dados entre os programas rodando sobre esse sistema operacional. O apêndice B traz exemplos de rotinas VB para estabelecer as conexões DDE com o HyperChem.

A versão zero do programa (obtida através do VB3.0 16-bit) e já rebatizada de HyperSpin, implementava muitos recursos adicionais em relação ao GetEnergy: a energia do sistema era minimizada em relação ao dihedro analisado e, via DDE, esta informação era reenviada ao HyperChem, que então colocava a molécula na conformação de menor energia e a salvava em disco.

Entretanto, a coleta de dados ainda acontecia pela análise dos arquivos de "log", o que poderia gerar incompatibilidades futuras com novas versões do HyperChem.

IV.5 Versões e componentes do HyperSpin.

IV.5.1 HyperSpin 1.0 (16-bit)

O HyperSpin 1.0 (16-bit) surgiu logo em seguida (uso do VB 4.0) e apresentava muitas novidades. Dentre elas podemos destacar:

- um módulo capaz de "plotar" a energia do sistema enquanto o HyperChem rodava o dihedro (HyperGraph), imitando as antigas placas de interface GPIB muito utilizadas pelos experimentais;
- um outro capaz de acionar variáveis e executar comandos do HyperChem remotamente (Energy Watcher);

- outros detalhes importantes como busca multidimensional e interface com o software MathematicaTM[6];
- a implementação de rotinas DDE mais eficientes permitiu a troca de todos os dados necessários entre o HyperChem e o HyperSpin, dispensando as leituras de arquivos de “log”. Ao final do cálculo um relatório era apresentado com os resultados, tudo acontecendo automaticamente.

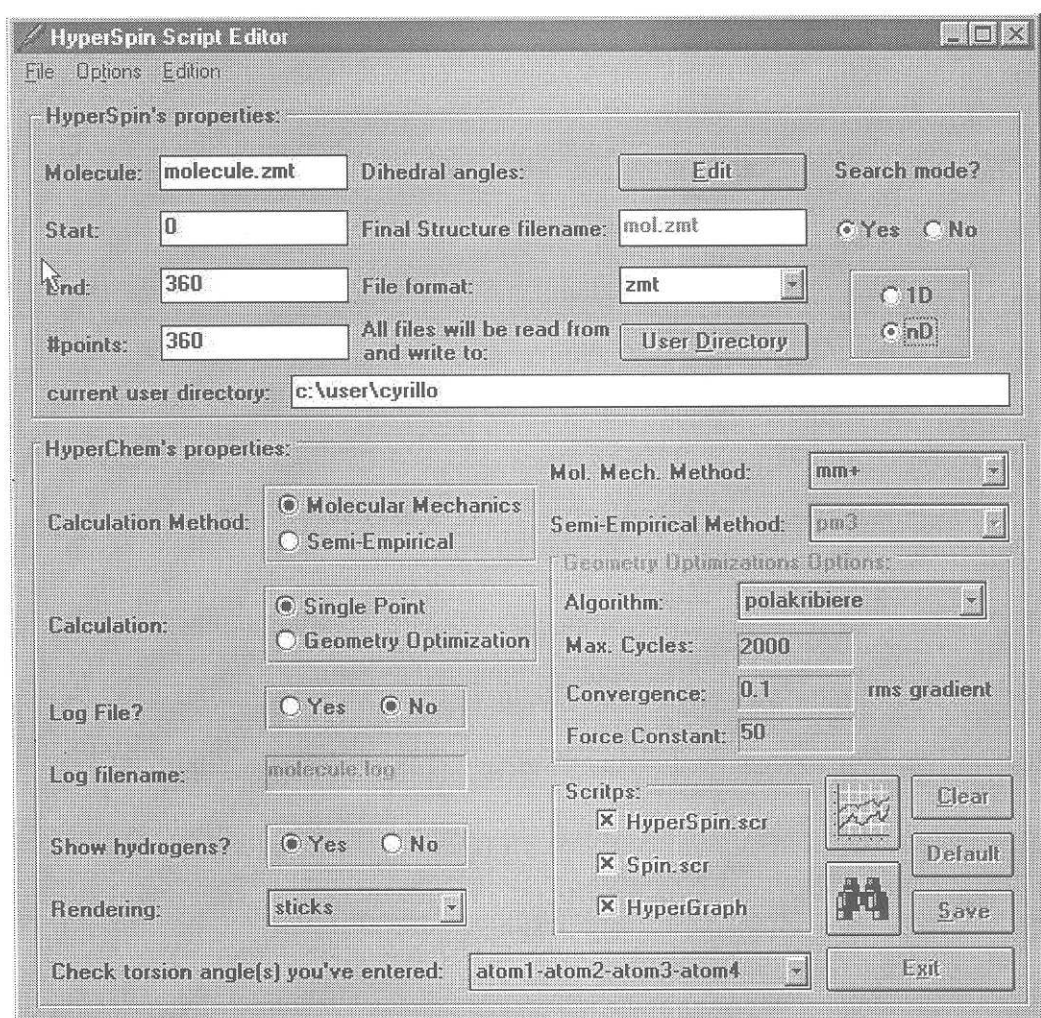


FIGURA 4.3a: HyperSpin 1.0 (16-bit) para Windows 3.11 ou 95, módulo responsável pela geração dos scripts de rotação de dihedros.

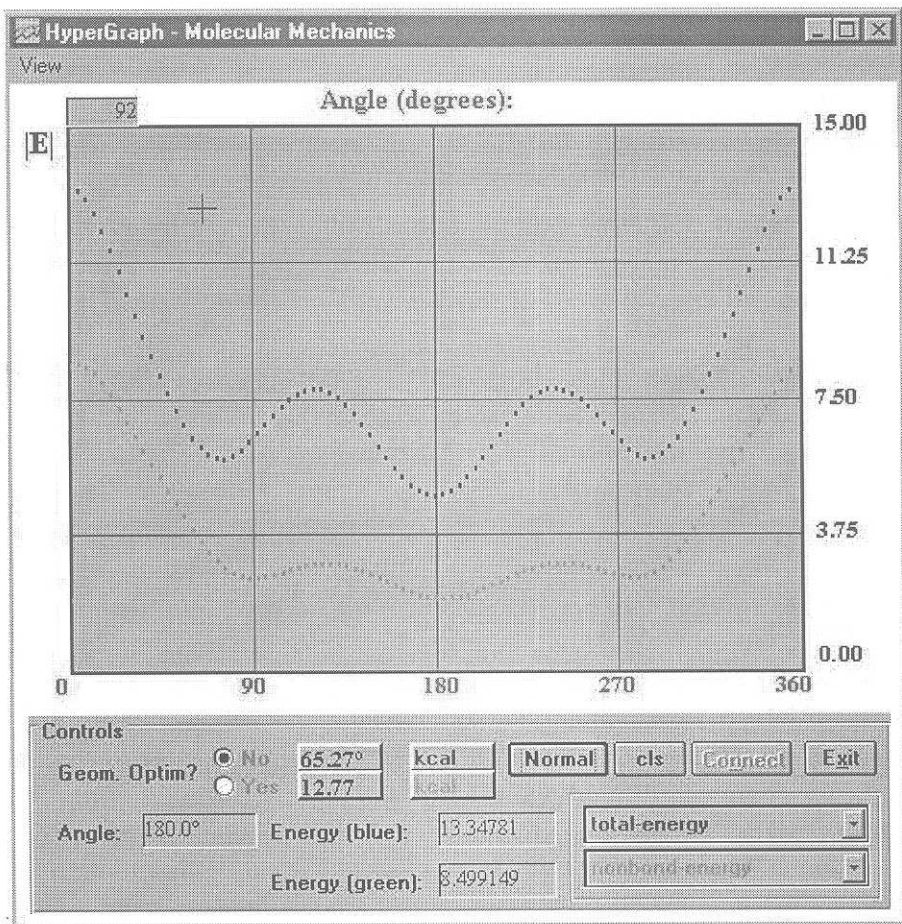


FIGURA 4.3b: HyperGraph 1.0 (16-bit), módulo capaz de plotar “on-line” a energia do sistema enquanto o HyperChem efetua os cálculos de barreira de torção, orientado pelo HyperSpin.

As figuras 4.3(a e b) anteriores, mostram os layouts do HyperSpin (o editor de scripts) e do HyperGraph, respectivamente. A janela da figura 4.3b mostra a barreira de torção obtida na análise conformacional do n-butano para o ângulo diédral central, como mostrado no capítulo anterior.

Esta é a versão final do programa até o momento. Em relação às versões anteriores, implementamos muitos recursos adicionais que tornaram a comunicação e troca de dados com o HyperChem mais eficiente e rápida. O uso da tecnologia ActiveX, implementada com o VB 5.0 *Professional Edition* e técnicas avançadas de OOP, conferiram um caráter profissional e enxuto ao código do programa. O modo de processamento com 32 bits tornou-se possível com o uso do novo compilador no ambiente Windows95.

Muitos detalhes de edição e apresentação de dados foram definitivamente incluídos, como por exemplo a geração de um *notebook* para o Mathematica™ capaz de desenhar superfícies 3D e de nível, quando na análise de dois ou três ângulos diedrais simultâneos (busca nD).

Até o presente momento, com todas as versões geradas e todas as rotinas necessárias, mais de 40000 linhas de código foram escritas. Isto sem contar as constantes alterações e o complexo processo de desenho das janelas do programa. Mais de 80% do programa é composto de código original, o restante tendo sido reaproveitado de rotinas de exemplo do VB da Microsoft.

A seguir vamos descrever o programa mais detalhadamente, apresentando suas opções e forma de operação. Como foi dito anteriormente, seu princípio básico de funcionamento consiste em gerar um script apropriado para o HyperChem, além das comunicações DDE necessárias à perfeita troca de dados entre os dois programas. Um script básico que o HyperChem reconhece é mostrado a seguir:

```

1. ;Script File generated by HyperSpin on Monday, 24 November 1997, 12:07:29 pm
2. ;Rotates dihedrals doing calculations and collecting data.
3. ;By M. Cyrillo - cyrillo@ifi.unicamp.br
4. ;-----
5. select-none
6. multiple-selections Yes
7. selection-target atoms
8. show-hydrogens Yes
9. render-method sticks
10. ;-----
11. calculation-method molecular-mechanics
12. molecular-mechanics-method mm+
13. ;-----
14. errors-are-not-omsgs
15. ;-----
16. query-response-has-tag (No)
17. select-atom 1,1
18. select-atom 2,1
19. select-atom 3,1
20. select-atom 4,1
21. name-selection dihedral
22. ;-----
23. write-file c:\hyper5\program\anth000.hin
24. select-name dihedral
25. set-bond-torsion 0.0
26. select-none
27. do-single-point
28. append-omsgs-to-file C:\HSPIN\Work\hspin.get
29. query-value named-selection-value
30. query-value total-energy
31. omsgs-not-to-file
32. ...
33. ;-----
34. select-name dihedral
35. set-bond-torsion ...
36. select-none
37. do-single-point
38. append-omsgs-to-file C:\HSPIN\Work\hspin.get
39. query-value named-selection-value
40. query-value total-energy
41. omsgs-not-to-file
42. ...
43. ;-----
44. select-name dihedral
45. set-bond-torsion 360.0
46. select-none
47. do-single-point
48. append-omsgs-to-file C:\HSPIN\Work\hspin.get
49. query-value named-selection-value
50. query-value total-energy
51. omsgs-not-to-file
52. ;-----
53. select-none
54. image-destination-file Yes
55. write-file c:\hyper5\program\anth000.hin
56. execute-client C:\HSPIN\hspin
57. ;-----
58. write-file c:\hyper5\program\anth000.hin
59. status-message "HyperSpin calculations done."
60. exit-script

```

Onde temos:

- linhas 1-22: "Setam" o tipo de cálculo e o ângulo dihedral, neste exemplo determinado pelos átomos 1, 2, 3 e 4 da molécula em questão;

- linhas 22-53: rodam o dihedro de 0 a 360 graus (ângulos intermediários omitidos), efetuam um cálculo *single-point* e coletam o resultado, que é colocado em um arquivo texto temporário;
- linhas 53-60: executam o programa de análise de dados que irá se comunicar com o HyperChem, analisar os dados e colocar a molécula no seu mínimo, além de apresentar um sumário com os resultados obtidos.

Como pode ser visto pela estrutura dos scripts do HyperChem, seu caráter seqüencial e de comandos sem estruturas de comparação e condições, limita e muito seu uso prático. A rotina para a busca multidimensional tornou-se possível graças a uma inteligente utilização de scripts recursivos, como exemplificado a seguir:

$$\text{Script2D} = \left\{ \begin{array}{l} \left\{ \text{Script_1}(\text{dihedro_1...pos} = \text{início}) \right\} \rightarrow \left[\begin{array}{l} \text{Script_2} \left(\begin{array}{l} \text{dihedro_2...pos} = \text{início} \\ \text{dihedro_2...pos} = \text{.....} \\ \text{dieddro_2...pos} = \text{fim} \end{array} \right) \end{array} \right] \\ \left\{ \text{Script_1}(\text{dihedro_1...pos} = \text{.....}) \right\} \rightarrow \left[\begin{array}{l} \text{.....} \\ \text{.....} \\ \text{.....} \end{array} \right] \\ \left\{ \text{Script_1}(\text{dihedro_1...pos} = \text{fim...}) \right\} \rightarrow \left[\begin{array}{l} \text{.....} \\ \text{.....} \\ \text{.....} \end{array} \right] \end{array} \right\}$$

Descobrimos que o esquema acima funcionava apesar de não haver nenhuma documentação sobre scripts recursivos presente no manual do HyperChem. O algoritmo de coleta de dados para o esquema acima é bastante complicado, mas foi implementado com sucesso utilizando múltiplos

arquivos de dados. Um algoritmo ainda mais complexo foi necessário para tratar o caso de três dihedros simultâneos. No momento, a busca nD limita-se a este caso e a extensão para ordens superiores é desencorajada pela imensa quantidade de arquivos que seriam necessários, já que a busca é sistemática: todo o espaço de fase dos dihedros selecionados é percorrido e o mínimo global deste subespaço encontrado.

Gostaríamos de frisar que rotinas de busca conformacional eficientes são raras em programas comerciais e nenhum deles oferece o tipo de busca implementado pelo HyperSpin. A melhor das possibilidades é oferecida pelo software SpartanTM^[7], um programa caro e de acesso restrito, que utiliza algoritmos genéticos ou busca sistemática para efetuar este trabalho, mas ainda assim não da forma clara e completa como a apresentada pelo HyperSpin, apesar de muito mais rápida.

Pode-se, através de um módulo comprado separadamente (ChemPlus), tornar o HyperChem capaz de efetuar buscas conformacionais. Este módulo, porém, é de uso complicado e ineficiente como constatamos ao adquiri-lo. Versões mais novas do HyperChem oferecem linguagens de script mais completas, incluindo instruções de condições, testes, etc., porém com recursos ainda muito limitados, impossibilitando a construção de um módulo como o HyperSpin.

Descreveremos a seguir as principais operações disponíveis com esta última versão do HyperSpin, que foi o nome dado a um pacote de programas composto pelo *Script Editor*, o *HyperGraph*, o *HyperRemote* e o *LogViewer*.

Muitos detalhes serão omitidos e poderão ser encontrados no manual do usuário ou no help on-line, que estão sendo escritos.



HyperSpin

IV.5.3 Script Editor

As figuras 4.4a, b e c a seguir mostram as três guias acessíveis a partir da janela principal do *Script Editor* em seu novo *layout* (apenas Windows95 ou superior). Este módulo é o principal, pois é onde o usuário define como será a busca conformacional. Nos campos definidos pela guia **Setup HyperSpin** **calculation** (figura 4.4a) o usuário seleciona a molécula, define o ângulo dihedral, seus limites, e os modos de operação que podem ser três:

- Modo-1 **search mode = no**: apenas um ângulo dihedral será analisado;
- Modo-2 **search mode = yes (1D)**: 999 dihedros podem ser analisados, mas a busca será feita um a um, ou seja, apenas uma curva 1D do espaço de fase será percorrida, cada dihedro sendo varrido quando seu predecessor tiver sido colocado em seu mínimo.
- Modo-3 **search mode = yes (nD)**: um máximo de 3 dihedros pode ser analisado simultaneamente, todo o subespaço de fase sendo percorrido. Além dos dados comuns às opções anteriores, *notebooks* auto-explicativos para o Mathematica versão 2.2, serão gerados para visualização 3D das subsuperfícies obtidas.

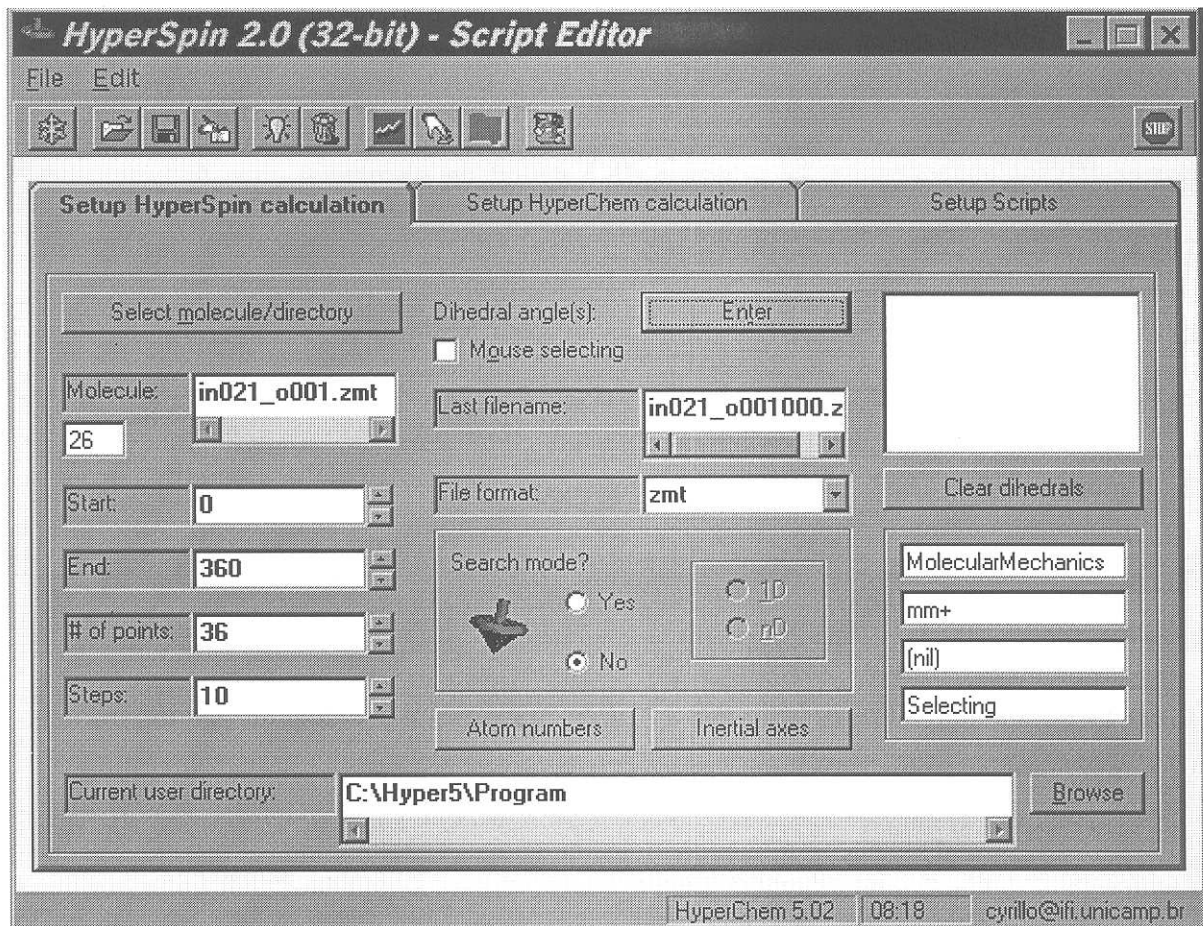


FIGURA 4.4a: A tela principal do *Script Editor*, o primeiro passo para definir a busca conformacional de um ou mais dihedros de uma molécula.

Durante uma busca no Modo-3 para dois dihedros, apenas uma superfície será obtida, correspondendo à energia do sistema para cada um dos valores dos dois ângulos dihedrais numa grade definida pelo range e número de pontos escolhidos pelo usuário. Para o caso de três dihedros, obtêm-se v superfícies de energia deste tipo para os dois últimos dihedros selecionados, sendo que cada uma delas corresponde a uma das v posições possíveis do primeiro dihedro. Este tipo de análise pode ser interessante, pois permite observar como a posição de um determinado dihedro pode alterar o

subespaço de fase de outros dois. Nenhum pacote comercial oferece este tipo de opção.

A guia mostrada na figura 4.4b a seguir **Setup HyperChem calculation** permite ao usuário selecionar o método de cálculo que será usado durante a busca conformacional e seus parâmetros mais importantes como critérios de convergência.

Além disso, o usuário pode testar se o(s) dihedro(s) selecionado(s) na guia anterior são próprios para rotação, já que, dependendo da definição do ângulo dihedral, ele pode ou não ser rotacionado sem deformar a estrutura da molécula. Dihedrais impróprios causam um erro no HyperChem, por isso é melhor testá-los antes de colocar os cálculos em andamento.

Outros detalhes menos importantes, como a renderização da molécula, a presença de um arquivo de “log” e scripts que giram a molécula somente para visualização, podem ser “setados” nesta guia. Não é aconselhável utilizar a renderização “esferas de van der Waals” para os cálculos com o HyperSpin, pois este procedimento torna o processo mais lento.

A figura 4.4c mostra a última guia do *Script Editor* (**Setup scripts**), representando também o último passo no preparo do HyperChem para a busca conformacional. Nela pode-se selecionar as variáveis que serão coletadas durante o processo e aquela que deve ser minimizada (seus pontos serão interpolados e o mínimo encontrado).

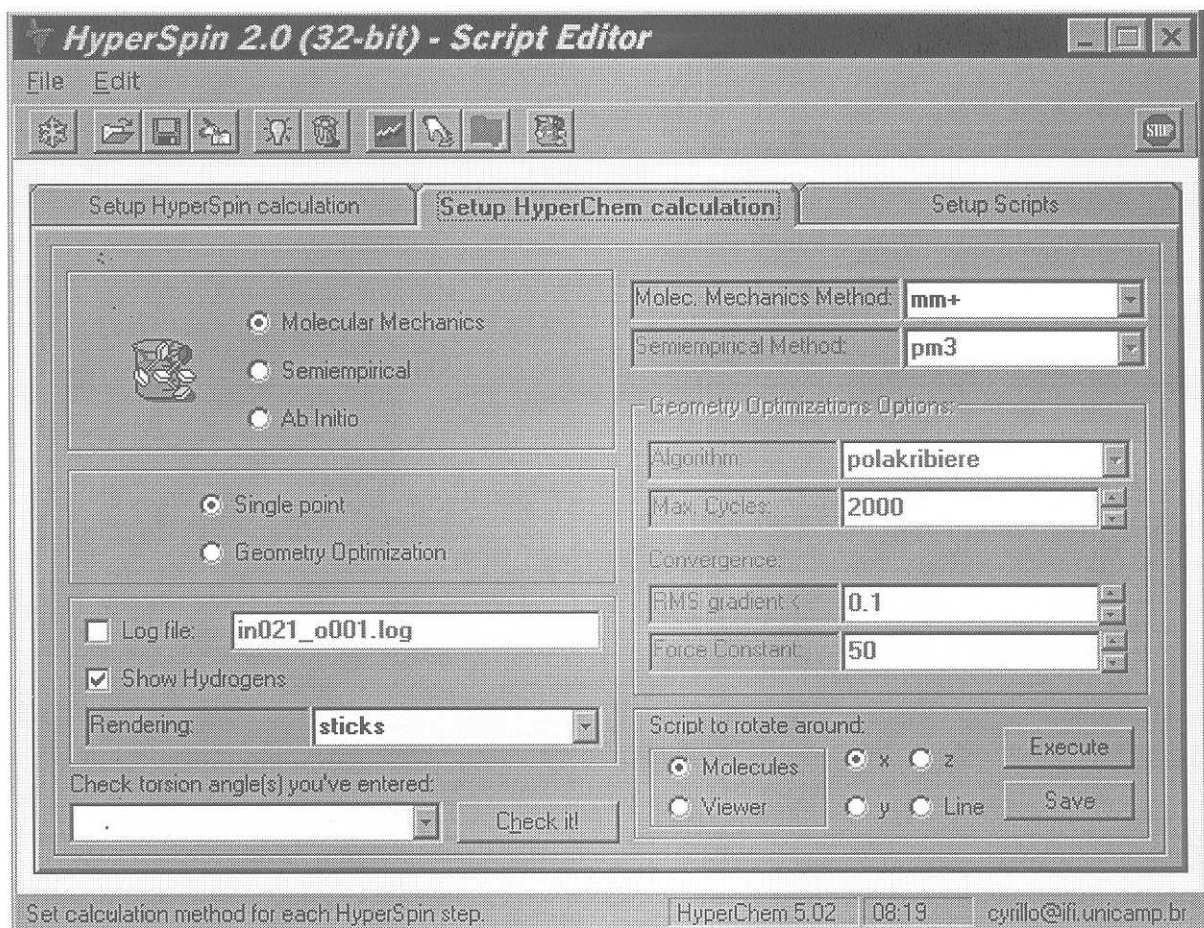


FIGURA 4.4b: Segundo passo: definindo o tipo de cálculo para a busca conformacional, bem como seus parâmetros específicos mais relevantes.

Há ainda as seguintes opções, como pode ser visto na terceira guia:

- **Reload starting molecule each step:** a molécula inicial e o valor inicial dos ângulos dihedrais serão carregados no *workspace* do HyperChem em cada passo da busca. Este procedimento é importante já que, no caso de otimizações de geometria, partes da molécula podem ser afetadas por uma posição específica de um dihedro, como no caso em que ocorrem impedimentos estéricos;

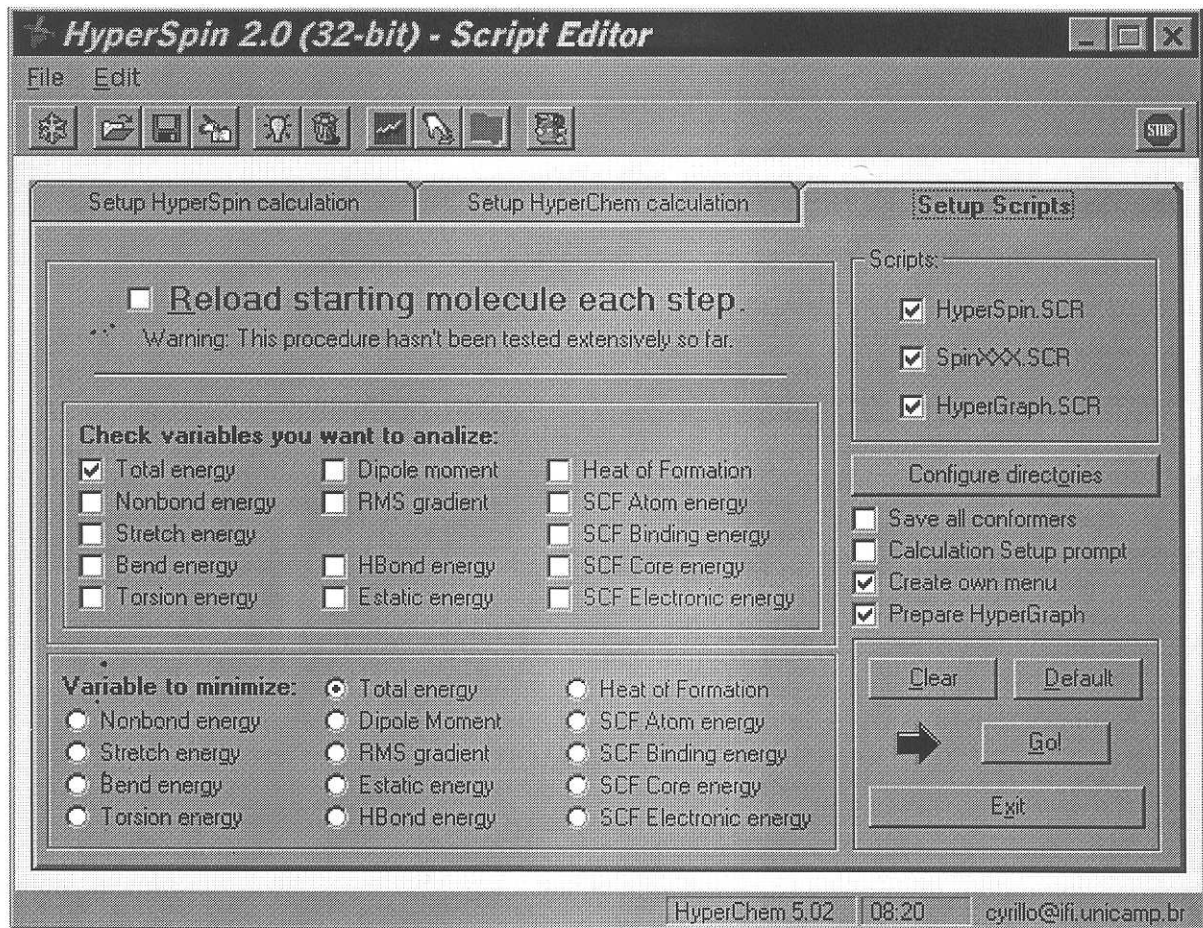


FIGURA 4.4c: Os últimos passos para definir a busca conformacional e preparar o HyperChem para incluir as opções oferecidas pelo HyperSpin.

- **Save all conformers:** todos os confórmeros gerados durante a busca serão salvos no diretório .\Conformers. Quando o programa termina é executado um módulo (figura 4.5 a seguir), que permite ao usuário selecionar duas moléculas para serem superpostas em um plano comum especificado. Esta pode ser uma ferramenta útil para a comparação de diferentes conformações obtidas da otimização da mesma molécula a partir de diferentes posições para um mesmo dihedro;

- **Calculation Setup prompt**: O script de busca irá solicitar ao usuário o método de cálculo antes de iniciar o processo. Importante quando a comparação dos resultados por diferentes métodos é desejada, não havendo a necessidade de se reeditar o script principal;
- **Create own menu**: Há duas formas de o HyperSpin incorporar suas facilidades ao HyperChem: gerando seu próprio menu chamado de HyperSpin, ou utilizando o menu Script do HyperSpin. A primeira possibilidade só é possível para o HyperChem 4.5 ou superior;

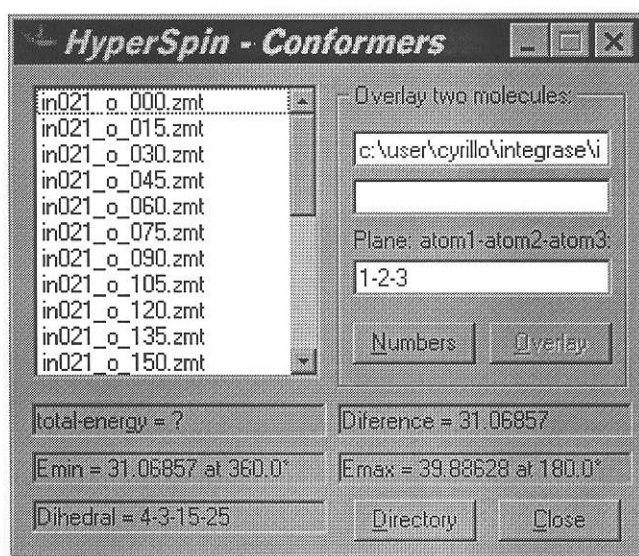
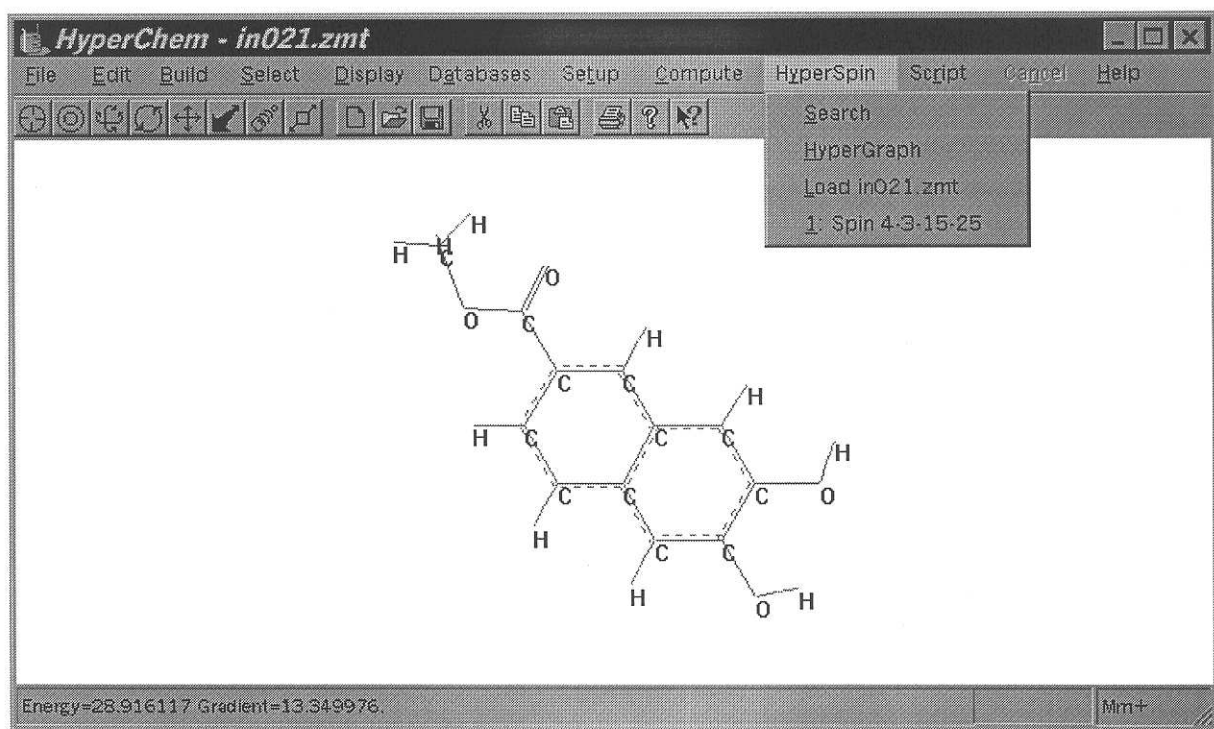


FIGURA 4.5: Módulo que permite a superposição de duas moléculas através de um plano comum. Esta janela aparecerá automaticamente se a opção **save all conformers** tiver sido selecionada.

- **Prepare HyperGraph**: O HyperGraph é acessível através do menu que o HyperSpin incorpora ao HyperChem. No entanto, através desta opção, o HyperGraph será automaticamente executado e preparado para coletar os dados assim que o usuário selecionar a busca conformacional e dê início ao processo.

Com apenas um clique no botão **GO!**, o HyperChem será preparado para conter as opções de busca conformacional em seus menus. As opções incluídas são as seguintes:



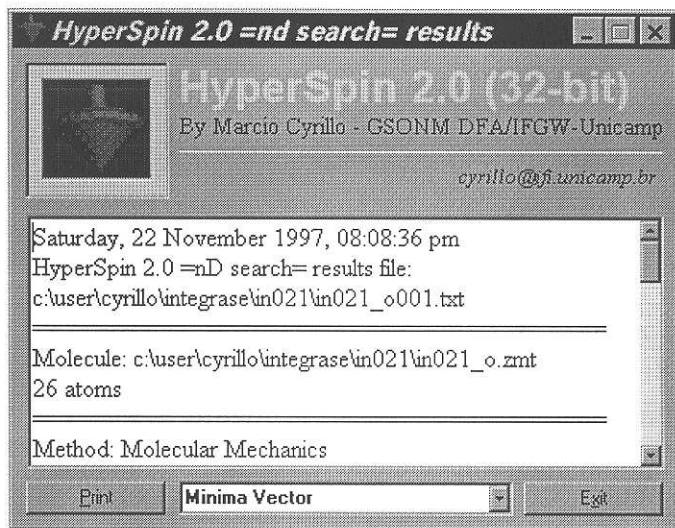
- **search**: inicia a busca conformacional;
- **HyperGraph**: carrega o módulo de visualização gráfica. Só será necessário se a opção **prepare HyperGraph** não tiver sido selecionada;

- **Load xxxx**: carrega a molécula em questão no workspace do HyperChem, onde xxxx é o seu nome;
- **Spin dihedral**: roda o dihedro dihedral para visualização (não efetua cálculos).

Um detalhe importante sobre o *Script Editor* é que a maioria de seus campos está ligada via DDE com suas variáveis correspondentes no HyperChem. Logo, se o usuário carregar uma molécula nova no HyperChem, seu nome aparecerá automaticamente no campo **molecule** do editor (figura 4.4a) ou vice-versa.

Uma edição particular pode ser salva em disco e recuperada em outra ocasião, permitindo a organização dos cálculos realizados. Além disso, é possível ao toque de um botão, recuperar a última edição.

Após a execução do script principal, acionado através da primeira opção do menu HyperSpin/Script do HyperChem, um módulo será automaticamente executado e efetuará a análise dos resultados além de apresentá-los. Este módulo pode ser visto na figura 4.6a a seguir, juntamente com um exemplo de arquivo texto de resultados.



```

Saturday, 22 November 1997, 08:08:36 pm
HyperSpin 2.0 =nD search= results file:
c:\user\cyrrillo\integrases\in021\in021_o001.txt
=====
Molecule: c:\user\cyrrillo\integrases\in021\in021_o.zmt
26 atoms
=====
Method: Molecular Mechanics
Molecular Mechanics method: mm+
Single Point Calculations
=====
Variable being minimized: Total energy
Global minimum found.
Emin = 31.06848
Emax = 39.88676
Mean: 35.31
=====
Minima vector:
4-3-15-25 = 8.537736e-007°
1-2-16-26 = 180°
There is another identical minimum
=====
HyperSpin set the molecule in the first global minimum.
Molecule in the minimum: c:\user\cyrrillo\integrases\in021\in021_o001.zmt
Additional files in c:\user\cyrrillo\integrases\in021:
inLst000.dat that is a Mathematica file.
Read it with the notebook c:\user\cyrrillo\integrases\in021\in021_o.ma.
==
inXYZ000.dat that is a 3-column-text file with the following meaning:
first column - x = dihedral angle 1-2-16-26
second column - y = dihedral angle 4-3-15-25
third column - z = Total energy of the molecule.
==

```

FIGURA 4.6: Tela do módulo que analisa e apresenta os resultados da análise conformacional, juntamente com um exemplo de texto apresentado após um cálculo de busca 2D.



A última versão do HyperGraph acompanhou o desenvolvimento do HyperSpin, lançando mão das comunicações DDE para um melhor desempenho na coleta *on-line* de dados. O seu uso limita-se a cálculos de mecânica molecular, pois são mais rápidos e justificam a visualização instantânea, o que não acontece nos casos de cálculos semi-empíricos ou *ab initio*.

O HyperGraph é capaz de coletar e visualizar as três formas de operação do HyperSpin:

- no Modo-1 os dados serão coletados e plotados para o dihedro que está sendo analisado pelo HyperSpin. Ao final do cálculo, o HyperGraph traça uma cruz marcando o ponto de mínimo global (se houver mais de um idêntico, marcará apenas o primeiro);
- no Modo-2, a operação é parecida com a do Modo-1, com o adendo de o HyperGraph detectar automaticamente a mudança de dihedros e mudar a cor do gráfico correspondente;
- no Modo-3 (somente para 2D) pode-se visualizar um dos dois ângulos separadamente, mas apenas um de cada vez. Uma janela extra coleta a energia para cada valor do vetor formado pelos dois dihedros. Ao final, pode-se com um clique em um dado vetor, ordenar o HyperChem para que coloque a molécula na configuração correspondente.

O programa foi dotado das opções mais comuns presentes em utilitários de visualização gráfica: mudança de escala e espessura das linhas, cursor que mostra as coordenadas dos pontos sobre os quais desliza, etc. Duas variáveis podem ser visualizadas de cada vez e devem ser selecionadas antes que a comunicação com o HyperChem se estabeleça.

A figura 4.7 a seguir mostra o HyperGraph logo após um cálculo típico de barreira de torsão de apenas um dihedro.

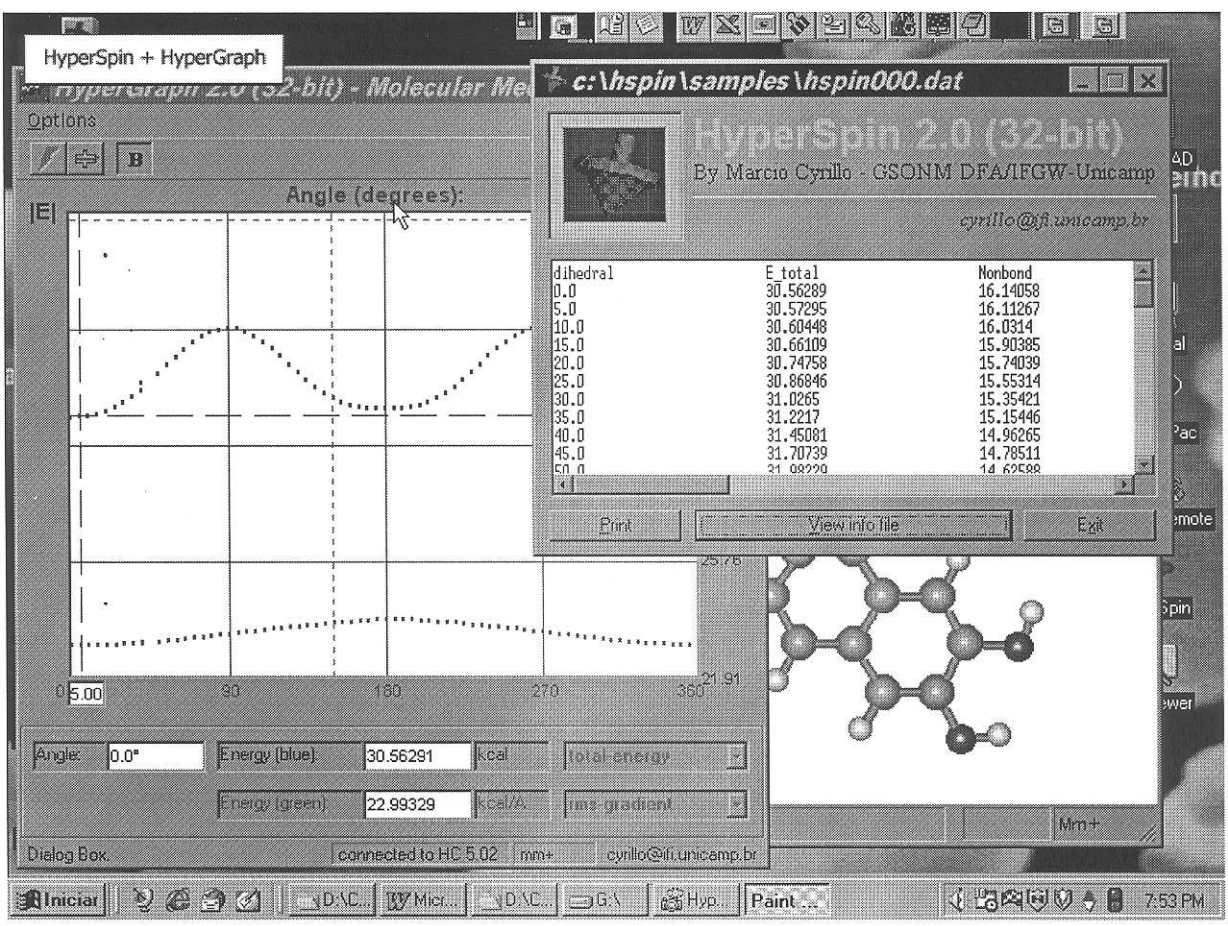


FIGURA 4.7: Tela típica do Windows95 mostrando o HyperChem, o módulo final do HyperSpin e o HyperGraph.



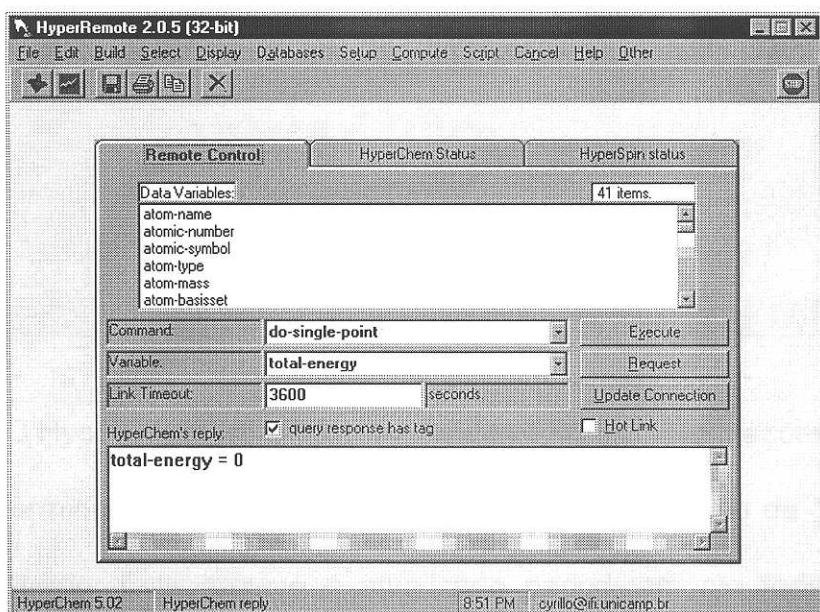
Durante o aprendizado do protocolo de comunicação DDE, bem como da linguagem de scripts do HyperChem, percebemos que muito tempo foi perdido na catalogação dos comandos e variáveis, bem como nos ajustes dos parâmetros ideais a uma comunicação mais eficiente e livre de erros. Pensando na possibilidade de tornar este trabalho mais simples a futuros desenvolvedores de interfaces para o HyperChem, construímos o programa *HyperRemote*, que substituiu o antigo *Energy Watcher*.

Como seu nome indica, o programa funciona como um controle remoto do HyperChem, permitindo executar um comando ou ler uma variável remotamente. Assim como o *Script Editor*, ele também foi dividido em três guias: as duas primeiras podem ser vistas na figura 4.8 a seguir. Na primeira delas, tem-se o controle remoto propriamente dito: os comandos e/ou variáveis podem ser acessados pelos menus que imitam os menus do HyperChem, possibilitando uma busca rápida e organizada ao item desejado.

Utilizar este controle remoto é uma forma eficiente e rápida de se conhecer todos os comandos e/ou variáveis do HyperChem, seus usos e suas sintaxes, poupando muito tempo a quem desejar construir novos programas que interajam com o HyperChem via DDE. No apêndice B encontram-se exemplos de rotinas DDE para comunicação com o HyperChem (versões do VB \geq 3.0), que podem ser adaptadas para a comunicação entre dois programas genéricos.

Na segunda guia temos campos com todas as variáveis de energia correspondentes aos cálculos mais comuns que podem ser feitos com o HyperChem, tornando possível a monitorização *on-line* das componentes de energia de um determinado cálculo. Pode ser um recurso interessante, pois o HyperChem só permite a visualização da energia total e do módulo de seu gradiente em seu *display* de *status*.

Na terceira guia temos o *status* da última configuração do HyperSpin, permitindo visualizar o(s) dihedro(s) e seu(s) valor(es) enquanto o HyperChem efetua a busca conformacional, bem como a maior e a menor energia lidas até o momento.



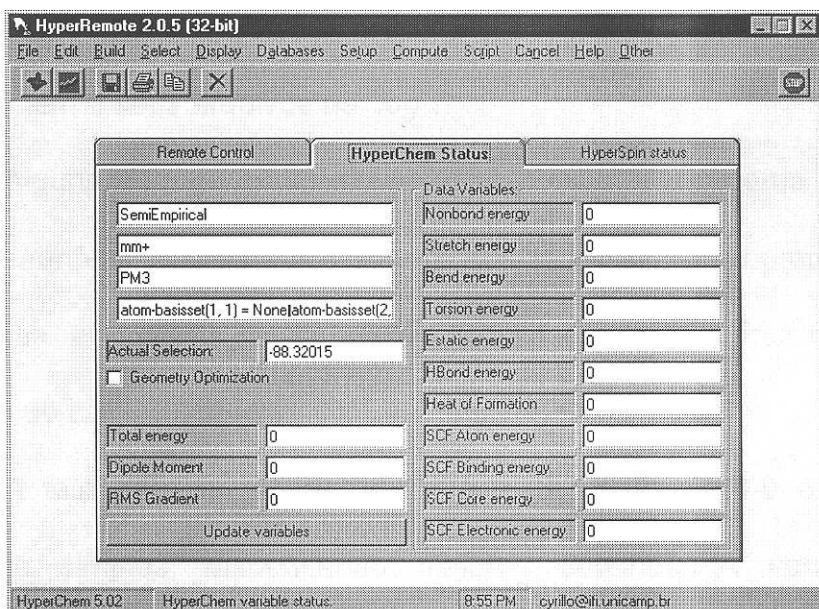


FIGURA 4.8: as duas telas mais importantes do HyperRemote onde o usuário tem completo acesso a todas as variáveis e comando do HyperChem e é capaz de observar “on-line” as mudanças das variáveis de cálculo.



LogViewer

IV.5.6 LogViewer

O HyperChem permite que grande parte das informações pertinentes a um determinado cálculo sejam colocadas em um arquivo de “log” escolhido pelo usuário. Este arquivo é atualizado periodicamente toda vez que um detalhe novo sobre o cálculo aparece, por exemplo a cada ciclo SCF de um cálculo semi-empírico. Apesar disso, estas informações só podem ser vistas pelo usuário ao final do cálculo, quando então é possível a edição do arquivo.

Com o objetivo de suprir esta deficiência, surgiu a idéia do programa *LogViewer*, que é capaz de mostrar o conteúdo de um arquivo enquanto este

é manipulado por um outro aplicativo, ajustando-se perfeitamente ao caso do HyperChem e seus arquivos de “log”.

Alguns truques foram necessários para que o sistema não acusasse erros durante a leitura de um arquivo aberto por outro programa e para que ele fosse atualizado rapidamente, permitindo ao usuário observar cada detalhe do cálculo “on-line”.

A tela principal do programa é vista na figura 4.9 e o seu uso é bastante simples, dispensando maiores detalhes. A versão 1.0.0 foi rapidamente atualizada para a 1.1.1 com novos recursos de edição.

Este programa encerra em si uma idéia muito simples e preenche uma das lacunas presentes no HyperChem.

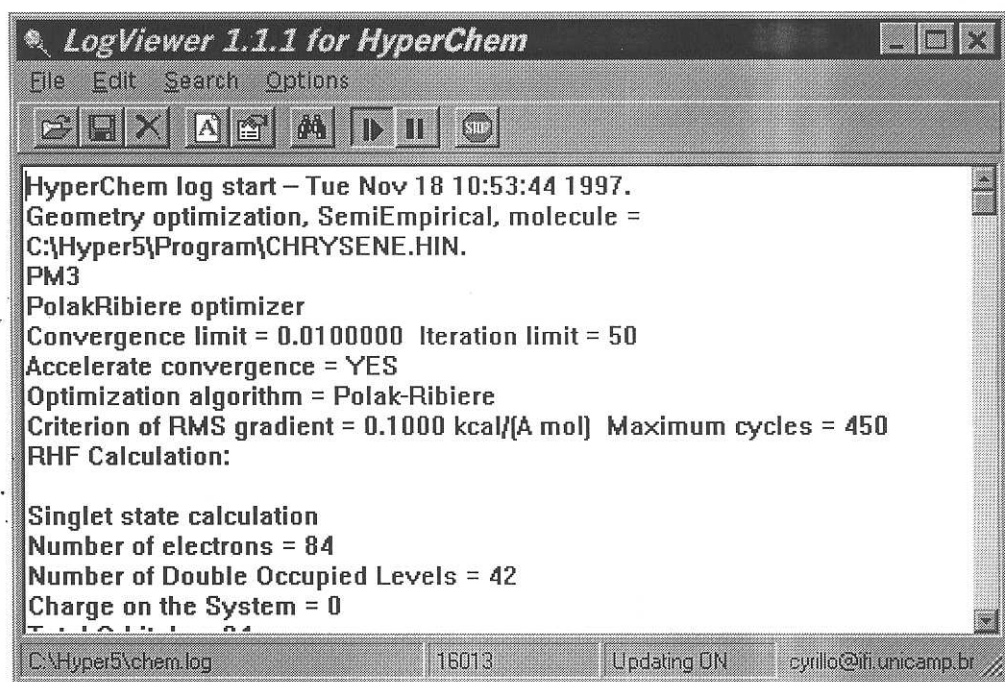


FIGURA 4.9: Tela principal do *LogViewer*, programa capaz de mostrar o conteúdo do arquivo “log” do HyperChem durante o processo de cálculo, suprimindo uma deficiência deste software comercial

Mencionamos anteriormente que mostraríamos a validade de se usar mecânica molecular para a pré-otimização de geometria das moléculas antes de submetê-las aos métodos semi-empíricos, muito mais custosos do ponto de vista computacional. A figura 4.10a a seguir mostra o monômero número 5 do grupo de moléculas que estamos estudando: enfocando o problema conformacional, uma das perguntas que podemos fazer é a respeito da posição relativa dos dois grupos OH (16-26 e 15-25) que podem adquirir, num mesmo plano, quatro configurações distintas (por razões de simetria, não há motivos para estas estruturas estarem em planos distintos).

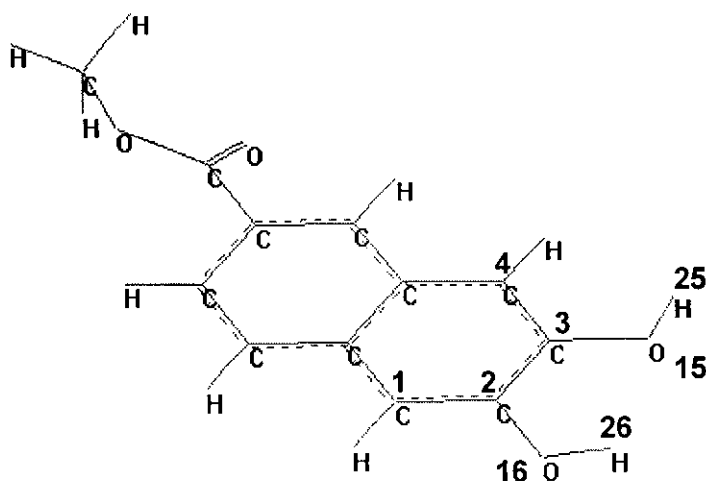


FIGURA 4.10a: O monômero número 5 e dois ângulos diedrais críticos (4-3-15-26 e 1-2-16-26) para a análise conformacional.

Vamos utilizar o HyperSpin e uma busca 2D para descobrir qual delas é a configuração de menor energia. Os diedrais que posicionam as duas hidroxilas

são: 1-2-16-26 e 4-3-15-25. Escolhemos um *grid* de 24x24 pontos, o que é mais do que suficiente para uma boa análise qualitativa.

A escolha de um método de mecânica molecular ou semi-empírico é crucial, dependendo dos objetivos do cálculo. De fato, pretendemos mostrar que, qualitativamente, os resultados são semelhantes, justificando o uso de mecânica molecular no caso de análises prévias.

As figuras 4.10(b,c) a seguir mostram os resultados obtidos com os programas HyperSpin e Mathematica: na figura 4.10a usando mecânica molecular (MM+) e na figura 4.10c, o semi-empírico PM3.

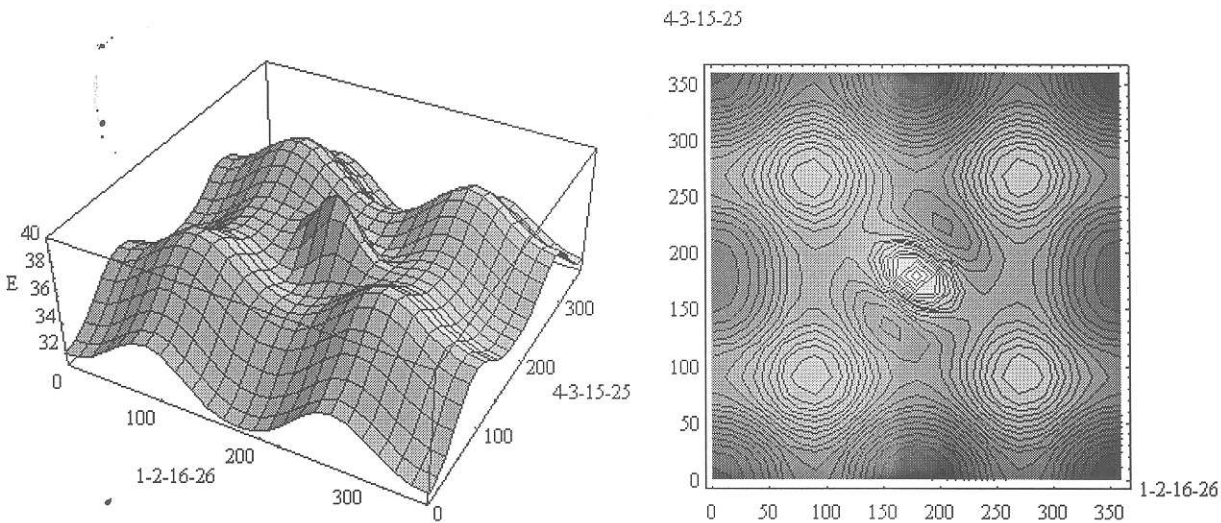
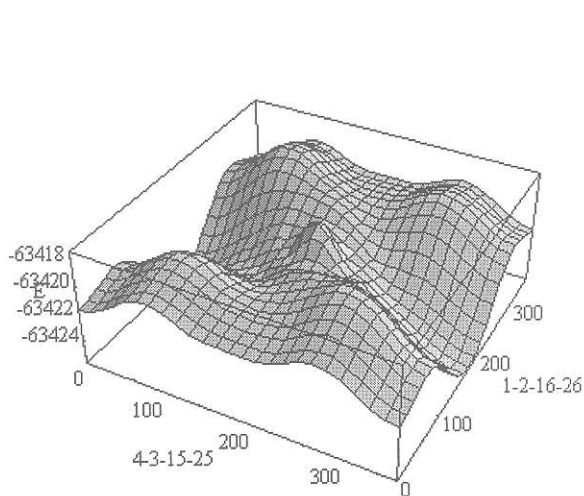


FIGURA 4.10b: Superfícies de energia obtidas com o programa HyperSpin e o método MM+ (mecânica molecular).



1-2-16-26

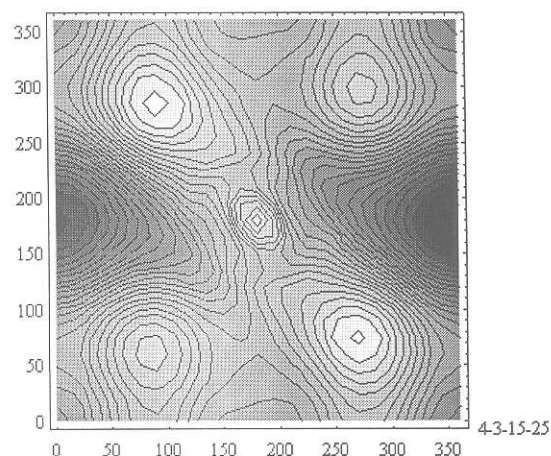


FIGURA 4.10c: Superfícies de energia obtidas com o programa HyperSpin e o método semi-empírico PM3.

A observação, tanto das superfícies 3D quanto das superfícies de nível, revela que o resultado qualitativo é praticamente o mesmo, mostrando a configuração de mínimos e máximos.

É por este motivo que nos sentimos confiantes ao usarmos a mecânica molecular como método para gerar as estruturas iniciais para os cálculos semi-empíricos, ganhando com isso muito tempo na realização dos cálculos e preparo das moléculas.

Já que o programa HyperSpin incorpora muitas facilidades não disponíveis com o HyperChem e outros programas comerciais, achamos por bem anunciá-lo e torná-lo gratuitamente disponível à comunidade científica. Este processo foi gradual, o programa tendo sido distribuído a algumas poucas pessoas que se dispuseram a testá-lo.

Um dos usuários escolhidos gentilmente enviou para lista de usuários do HyperChem uma mensagem entusiástica falando sobre sua experiência com o programa. Esta mensagem pode ser vista na página seguinte.

Algumas semanas depois decidimos colocar o programa disponível a qualquer usuário interessado. Primeiramente apresentamos o programa ao Dr. Krassimir Stravrev da empresa HyperCube Inc., responsável pela assistência técnica aos usuários do HyperChem, que elogiou muito o programa e tornou possível o seu *download* através da página oficial do HyperChem: <http://www.hyper.com>

Os links para *download* são:

- http://www.hyper.com/support/software/Other/other_index.html

e

- <ftp://ftp.hyper.com/pub/other/hspin.zip>

Basta descompactar o arquivo “zip” e rodar o programa SETUP.EXE que guiará a instalação automática do programa.

Procuramos também deixar o programa disponível em um *link* mais rápido, confeccionando uma pequena e simples *homepage* local que deverá ser atualizada e incrementada em breve:

- <http://www.ifi.unicamp.br/~cyrillo/hyperspin/hspin.html>

Somente a partir deste *link*, cerca de 390 pessoas já efetuaram o *download* desde Dezembro do ano passado até o momento. Cerca de 80% delas, dos mais diversos pontos do mundo (mais de 15 países), enviaram-nos um registro pessoal dizendo-se interessados em receber mais informações e/ou novas versões do programa.

From: Sandy Yates <prrls08@holyrood.ed.ac.uk>

To: hyperchem@hyper.com

Subject: Add-on to HyperChem

Date: Sat, 13 Dec 1997 00:59:10 +0000 (GMT)

Hi HC users,

I have had the chance to use HyperSpin from Marcio (cyrillo@ifi.unicamp.br). It is easy to install and functions very smoothly! Within a few minutes I was able to calculate the rotamer conformational energies for butane. This program graphed these data at my selected angular step values. This program is a delight to use. I was excited by the watching the molecules being manipulated and the energy being calculated before my eyes. This third-party HyperChem add-on is capable of plotting a wide range of parameters (dipole moment etc etc) as a function of the dihedral angle. I used this program with HyperChem 5.02 under Win95. Basically there might be a few minor refinements that I could suggest, but all in all the program made it very easy visualize this type of calculation. The script necessary for the scripts used in the calculation could be edited separately as they are stored to disk. I have asked the author if he could develop the program further by allowing other type of adjustments to be included. If you have the opportunity to download this program and try it I strongly recommend that you do. It would be nice to see other such programs being developed. Personally, this program provided me with just the type of "live" graphing facilities which I was looking for. It allowed me to track the conformational energies. I can imagine that it would be a great teaching aid. It looks likely that HyperCude may make this add-on available from their server in their FTP support directory

Sandy Yates / University of Edinburgh / Dept of Chemistry (Physical)

Joseph Black Building / West Mains Road / EH9 3JJ Tlf: (+44) 131

650 4734 or 4799 / Fax: (+44) 131 650 4743 email: S.Yates@ed.ac.uk

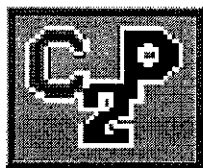
Muitos usuários estão contribuindo para o desenvolvimento dos programas, sugerindo algumas modificações e novas implementações. O programa *LogViewer* foi bastante elogiado pela sua simplicidade, originalidade e utilidade.

O mais importante é que até o momento poucos *bugs* foram reportados, indicando que o programa, além de sua utilidade, foi bastante bem construído. Apesar de não estar documentado em detalhes, qualquer usuário com experiência em Visual Basic será capaz de alterá-lo de forma a atender necessidades específicas.

Os usuários interessados em obter uma cópia do código fonte do programa, ou parte dele, devem entrar em contato com o seguinte endereço eletrônico:

cyrillo@ifi.unicamp.br

- ¹ Cyrillo, M. e Galvão, D.S., "*HyperSpin conformational search program, version 2.0 for Windows 95*", em preparação.
- ² HyperChem™ *Molecular Visualization and Simulation program*, © HyperCube inc.
(<http://www.hyper.com>)
- ³ Stewart, J.J.P., *J. Comp. Chem.* **1991**, 10, 209-221; *MOPAC program, version 6.0, Quantum Chemistry Exchange Program No. 455*.
- ⁴ Marzzoco, A. e Torres, B.B., "*Bioquímica básica*", Editora Guanabara, **1990**.
- ⁵ *Enzyme and Protein*, **1993**, 47; **1994/1995**, 48 e **1996**, 49, IB - UNICAMP.
- ⁶ Mathematica™ 2.2 para Windows, © 1993 Wolfram Research Inc.
- ⁷ Spartan™ for UNIX Workstations, © Wavefunction, Inc. (<http://www.wavefun.com>)



Cálculo da densidade de estados: software Chem₂Pac

V.1 Introdução

Neste capítulo vamos formalizar a definição de densidade de estados eletrônicos, bem como apresentar um programa especialmente desenvolvido para efetuar esse tipo de cálculo: o Chem2Pac. Assim como o HyperSpin^[1], esse programa tem sido muito útil ao nosso grupo de pesquisa, tendo sido uma das principais ferramentas em alguns de nossos trabalhos realizados recentemente^[2,3]

Iniciaremos com uma breve descrição do método de Hückel, que servirá como ponte para introduzirmos o conceito de densidade local de estados eletrônicos (LDOS). Apesar de não termos empregado este método diretamente em nosso estudo, acreditamos que ele seja a maneira mais simples de se introduzir a filosofia dos métodos de química computacional.

Em seguida introduziremos a parte prática dos cálculos LDOS, enfocando os principais problemas computacionais encontrados e as soluções aplicadas.

Por fim, descreveremos o programa Chem2Pac e suas facilidades, que vêm crescendo dia a dia através de novas implementações sugeridas por seus usuários. Futuras implementações e projetos serão também sugeridos.

Descreveremos aqui o método de Hückel simplificado (HMO ou *Hückel Molecular Orbital Theory*), na forma usada por Pullman e Pullman em seus trabalhos com os PAHs^[4] e discutiremos os conceitos envolvidos nesse método empírico e nos cálculos LDOS.

V.2.1 O Método HMO

Dado um sistema molecular, nosso objetivo é resolver a equação de Schrödinger independente do tempo:

$$\mathbf{H}\psi = E\psi, \quad (5.1)$$

porém de forma bastante aproximada, como mostraremos a seguir.

O hamiltoniano de Hückel é escrito como a soma dos hamiltonianos monoelétrônicos efetivos para os elétrons π do sistema*, ou seja, pressupõe uma separação σ - π e, conseqüentemente, limita-se a sistemas planares. Dentro da aproximação de Born-Oppenheimer, podemos escrever então:

$$\mathbf{H} = \sum_i H^\pi(i). \quad (5.2)$$

Os elétrons σ são explicitamente ignorados[†] e suas interações com os elétrons π são levadas em conta apenas implicitamente, via parametrização.

* Dentro da idéia da expansão dos orbitais moleculares em termos de orbitais atômicos, os elétrons π são aqueles que ocupam os orbitais moleculares formados a partir de orbitais atômicos do tipo p_z .

† Os elétrons σ ocupam orbitais moleculares formados por orbitais atômicos de valência do tipo s ou sp

Aplicamos a seguir a aproximação LCAO (*Linear Combination of Atomic Orbitals*), que estabelece a expansão dos orbitais moleculares em termos de orbitais atômicos dos átomos que formam o sistema molecular:

$$\psi_j = \sum_k c_{jk} \varphi_k ,$$

(5. 3)

onde ψ_j refere-se aos orbitais moleculares, φ_k aos orbitais atômicos e c_{jk} aos coeficientes da expansão. A somatória estende-se sobre todos os elétrons π presentes. Os coeficientes c_{jk} podem ser obtidos através do método variacional que minimiza a energia eletrônica do sistema:

$$\frac{\partial E}{\partial c_x} = 0 ,$$

(5. 4)

onde c_x é um coeficiente fixo. A energia do estado fundamental, sendo dada por

$$E_0 = \frac{\langle \psi_0 | H^\pi | \psi_0 \rangle}{\langle \psi_0 | \psi_0 \rangle} ,$$

(5. 5)

pode ser escrita então como

$$E_0 = \frac{\sum_r \sum_s c_{0r} c_{0s} \int \varphi_r H^\pi \varphi_s d^3r}{\sum_r \sum_s c_{0r} c_{0s} \int \varphi_r \varphi_s d^3r} .$$

(5. 6)

Derivando a expressão (5.6) acima em relação a c_x e igualando o resultado a zero, chegamos à seguinte expressão:

$$\sum_m \left[(H_{km}^\pi - S_{km} E_l) c_{ml} \right] = 0,$$

(5. 7)

e para que haja solução não trivial, devemos ter:

$$\det(H_{km}^\pi - S_{km} E_l) = 0$$

(5. 8)

Resolvendo a equação secular acima, teremos não somente os valores dos coeficientes da expansão (c_{ml}), mas também os autovalores de energia (E_l).

Nas equações acima, as seguintes definições foram utilizadas:

$$H_{km}^\pi = \int \varphi_k^* H^\pi(i) \varphi_m d^3r;$$

(5. 9)

$$S_{km} = \int \varphi_k^* \varphi_m d^3r,$$

(5. 10)

ou seja, H_{km} são os elementos de matriz do hamiltoniano e S_{km} são as integrais de *overlap*, que medem o grau de recobrimento dos orbitais atômicos φ_k e φ_m .

A forma de aproximação das integrais provenientes da equação secular (8) define o método de química computacional. Se o hamiltoniano do sistema incluir vários termos, como por exemplo interações entre os elétrons, elas tornam-se rapidamente complexas e numerosas.

Quando as integrais decorrentes são resolvidas integralmente, dão origem aos métodos *ab initio*, quando aproximadas via parametrização, aos métodos semi-empíricos (em ambos os casos todos os elétrons de valência

são levados em consideração e o método autoconsistente de Hartree-Fock utilizado)[‡]. Este não é o caso do HMO: as aproximações são radicais, considerando que todas as integrais envolvendo os elétrons π possam ser parametrizadas através de resultados experimentais e não diretamente resolvidas. As demais integrais são desprezadas.

Temos:

- $H_{kk} = \alpha$ para todo átomo de carbono. Será uma fração de α no caso de outros átomos (são os parâmetros do tipo 1 ou integrais de Coulomb);
- $H_{km} = \beta$ para dois átomos de carbono ligados. Será uma fração de β no caso de outras combinações entre átomos (são os parâmetros do tipo 2 ou integrais de ressonância);
- $H_{km} = 0$ para átomos não ligados;
- $S_{km} = \delta_{km}$ (Delta de *Kronecker*), para a matriz de recobrimento.

Os parâmetros utilizados têm caráter empírico, pois são determinados através da comparação com resultados experimentais para certas características de sistemas moleculares escolhidos como padrão, como por exemplo potenciais de ionização, eletronegatividades, eletroafinidades, etc. Vários trabalhos apresentam a calibração de parâmetros para diversos átomos em pequenos sistemas orgânicos^[4,5].

Apesar de sua simplicidade, o método de Hückel e teorias similares são ainda muito úteis, pois são capazes de fornecer informações físicas

[‡] o hamiltoniano, rescrito de maneira conveniente, é chamado de operador de Fock (**F**). No apêndice A

relevantes para uma boa análise qualitativa de propriedades eletrônicas e estruturais de compostos orgânicos^[6,7].

V.2.2 Definições de DOS e LDOS

A densidade de estados eletrônicos (**DOS**) é definida como o número de estados eletrônicos por unidade de energia. Para um conjunto de níveis discretos não degenerados (E_i) de um dado sistema molecular, por exemplo, teríamos simplesmente:

$$DOS(E) = \delta(E - E_i).$$

(5. 11)

O conceito de densidade local de estados (**LDOS**) é introduzido para que se possa descrever a distribuição espacial dos estados sobre o sistema que se está estudando. A contribuição de cada átomo para um dado nível eletrônico pode ser ponderada de diversas formas. A definição de “peso” mais usada pelos químicos leva em conta a densidade de probabilidade calculada no sítio em questão^[8], com o fator sendo dado, por:

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

(5. 12)

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

Nossa metodologia leva em conta um conceito um pouco diferente: a modulação pelos elementos da matriz densidade, dados simplesmente pela

soma dos quadrados dos coeficientes que compõem o orbital molecular em questão:

$$W'_\mu(E_i) = \sum_\nu c_i^{\ast\nu} c_i^\nu = \sum_\nu (c_i^\nu)^2, \quad (5.13)$$

já que os coeficientes serão reais para os métodos semi-empíricos.

Vale ressaltar que, dentro da formulação de Hückel, as duas funções de peso acima serão idênticas, pois neste método a matriz de *overlap* **S** é uma matriz n-identidade. Temos ainda que os pesos *W* são sistematicamente menores que *W'*, pois os elementos da matriz **S**, para os métodos semi-empíricos, são sempre menores que 0.25^[9].

Ambas as definições para a LDOS podem fornecer informações sobre as contribuições relativas de regiões específicas das moléculas para a reatividade química, reposta óptica, etc. e conseqüentemente para seus comportamentos bioquímicos^[10,11].

Para tornar o resultado visualmente mais próximo do que se espera fisicamente, podemos ainda incluir um envelope Lorentziano, o que leva à seguinte expressão para a densidade local de estados (de acordo com nossa definição):

$$LDOS(E) = \sum_{l=1}^{n_c} \left[\frac{\gamma^2}{(E - E_l)^2 + \gamma^2} \sum_{m=n_l}^{n_f} c_{ml}^2 \right]. \quad (5.14)$$

O parâmetro γ é muito importante, pois controla a largura dos envelopes. Se mal dimensionado, pode causar resultados artificiais com a

superposição indevida de níveis muito próximos. Um valor típico é $\gamma=0.01\beta$ ^[12] para o método de Hückel, onde β vale aproximadamente 2.4eV.

A teoria de índices eletrônicos tem como objetivo analisar os valores críticos da LDOS, como definida acima, e tentar relacioná-los à presença ou ausência de atividade em moléculas orgânicas de interesse biológico. No primeiro e mais importante trabalho que a utilizou^[12] para estudar a atividade carcinogênica dos PAHs, foram empregados o método de Hückel e a modulação com funções Lorentzianas.

O presente trabalho foi um dos primeiros na extensão do método dos índices eletrônicos a sistemas mais complexos que os PAHs utilizando métodos semi-empíricos e por isso esteve sujeito a vários problemas na adaptação desta teoria. O desenvolvimento paralelo do programa Chem2Pac possibilitou examinar de maneira mais eficiente as dificuldades que foram surgindo.

Notadamente, o primeiro grande problema detectado foi a superposição indesejável dos níveis eletrônicos mais próximos devida ao envelope Lorentziano. Observamos que era difícil encontrar um valor ideal de γ para evitar este problema e ao mesmo tempo permitir uma boa visualização dos níveis.

Concluimos que deveríamos analisar os espectros discretos da LDOS para as moléculas mais complexas e para os métodos que levassem em conta todos os elétrons de valência:

$$LDOS(E) = \sum_{V_{inicial}}^{V_{final}} c_{Vi}^2 \delta(E - E_i).$$

(5. 15)

Alguns trabalhos que haviam estudado os espectros alargados foram refeitos com os espectros discretos e melhoraram seus resultados^[3,13].

V.3 O cálculo LDOS do ponto de vista computacional

O cálculo LDOS é extremamente simples: consiste apenas em somas sobre os elementos da matriz de coeficientes provenientes de um método de cálculo apropriado, como por exemplo o PM3. Estas somas podem ser moduladas por funções lorentzianas ou mantidas em seu caráter discreto (expressões (5.14) e (5.15) respectivamente).

Computacionalmente, tal tarefa resume-se em ler os autovetores da matriz de Fock/hamiltoniano e efetuar os cálculos observando os parâmetros relevantes: sítio inicial, final, opção de cálculo discreto ou contínuo, valor de γ , etc.

A figura 5.1 a seguir mostra um pedaço de uma matriz de coeficientes. As somas LDOS varrem esta matriz da seguinte forma:

- nos cálculos discretos: um cálculo LDOS (eq. 5.15) é realizado para cada nível de energia molecular, varrendo os sítios atômicos[§] selecionados para o cálculo. Um gráfico LDOS versus energia orbital é então montado; A figura 5.2 mostra um exemplo desta situação.

[§] Sítios neste caso referem-se aos orbitais atômicos.

	(MO=1)E= -42.113	(MO=2)E= -40.876
AO = 1	0.27930	-0.26169
AO = 2	-0.01889	-0.00560
AO = 3	0.02737	-0.03888
AO = 4	-0.00031	0.00059
AO = 5	0.19354	-0.21703
AO = 6	-0.04032	0.03713
AO = 7	0.03393	-0.04403
AO = 8	-0.00096	0.00107
AO = 9	0.16779	-0.20526
AO = 10	-0.04764	0.05511
AO = 11	0.00451	0.00295

FIGURA 5.1: parte de uma matriz de coeficientes. As linhas mostram os diferentes orbitais atômicos e as colunas os diferentes orbitais moleculares e suas energias.

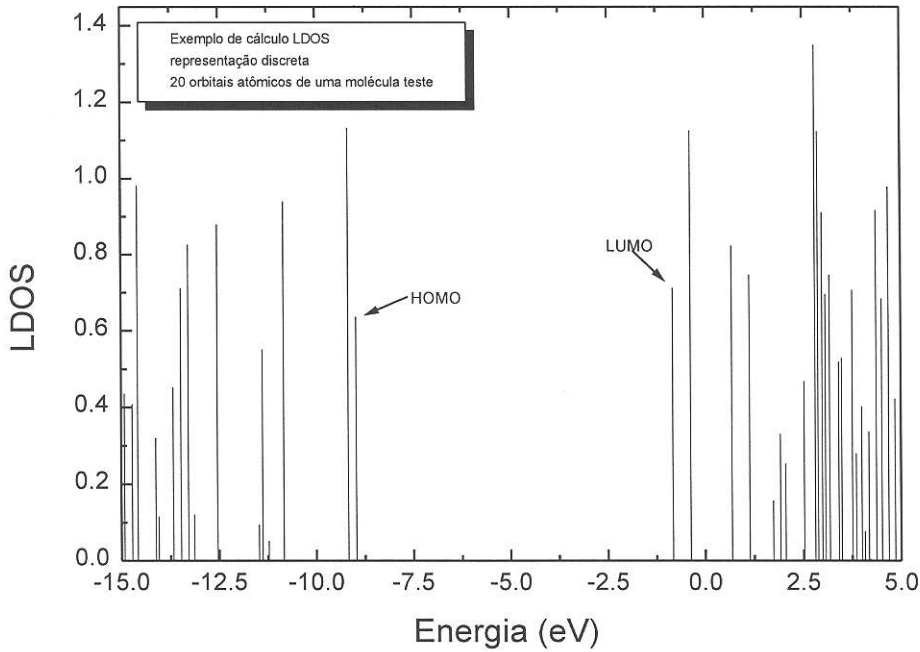


FIGURA 5.2: Exemplo de cálculo LDOS no modo discreto. A unidade da LDOS não é arbitrária, pois ela pode assumir um valor entre 0 e 2, que é o número máximo para a ocupação de um orbital molecular.

- no contínuo: A expressão (5.14) é utilizada e gera um gráfico suave de LDOS versus energia orbital. O parâmetro n_c determina os níveis moleculares que serão varridos e n_i e n_f os orbitais atômicos envolvidos. A figura 5.3 a seguir mostra um exemplo.

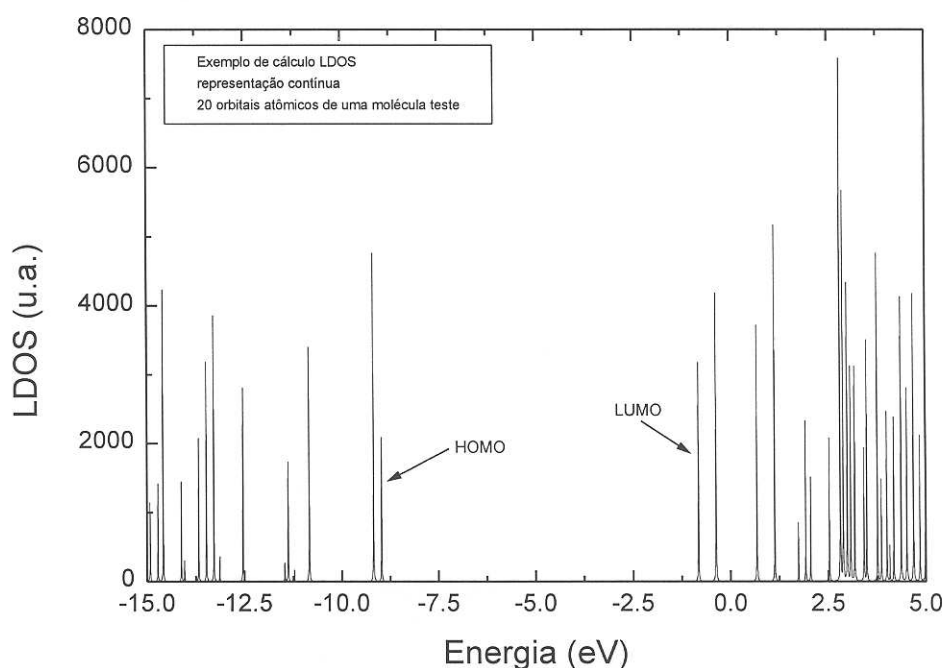


FIGURA 5.3: Exemplo de cálculo LDOS no modo contínuo. Neste caso a LDOS é medida em unidades arbitrárias, pois não há limites para seu valor, que dependerá das somas efetuadas, do número de pontos escolhidos, etc.

A dificuldade na implementação de um algoritmo computacional para efetuar estas tarefas aparece quando observamos uma parte da saída de dados do programa MOPAC^[14], mostrado a seguir.

EIGENVECTORS

ROOT NO.			1	2	3	4	5	6
			-42.13669	-40.18544	-38.48902	-37.41224	-35.75382	-33.87896
S	C	1	-0.25143	-0.07783	0.08807	-0.12457	0.06753	0.26326
PX	C	1	-0.02839	-0.03741	-0.04021	-0.03979	-0.00414	0.04876
PY	C	1	-0.05172	0.00145	0.04745	0.02224	0.02127	0.06562
PZ	C	1	-0.00001	0.00000	0.00001	0.00000	0.00002	0.00004
S	C	2	-0.29720	-0.19252	-0.09830	-0.24603	0.01359	0.17449
PX	C	2	-0.00335	-0.03757	-0.08064	-0.03750	-0.03092	-0.10932
PY	C	2	-0.01078	0.02331	0.05676	0.16499	0.03311	0.13454
PZ	C	2	0.00000	0.00001	0.00002	-0.00002	0.00003	0.00004
{...}								
ROOT NO.			73	74				
			6.12902	7.18668				
S	C	1	0.04345	-0.00127				
PX	C	1	0.10811	0.00650				
PY	C	1	0.26797	-0.00126				
PZ	C	1	0.00007	-0.00007				
S	C	2	0.00143	-0.00094				
PX	C	2	0.13885	0.00350				
PY	C	2	-0.31556	0.00166				
PZ	C	2	0.00019	0.00013				
{...}								

** outros programas como o SPARTAN têm saída de dados semelhante.

As diferentes colunas correspondem aos diferentes níveis de energia, cujos valores em eV são mostrados abaixo do número seqüencial que identifica as energias moleculares (ROOT NO.):

- a primeira coluna mostra o tipo de orbital atômico envolvido: s, p_x, p_y ou p_z, etc.;
- a segunda, o símbolo do átomo correspondente;
- a terceira, o número do átomo: note que para um átomo de carbono temos quatro orbitais de valência;
- as demais colunas mostram os coeficientes para diferentes níveis de energia.

O fato de o número de colunas não ser fixo complica o algoritmo de leitura de dados. O exemplo acima mostra que este número variou de seis para dois no fim do

relatório. Algumas tentativas foram feitas, mas nenhum programa que cobrisse todas as possibilidades pôde ser desenvolvido usando a linguagem FORTRAN.

Um outro problema aparece devido ao fato de os orbitais atômicos aparecerem em ordem seqüencial na saída de dados. Como as somas nas expressões (5.14) e (5.15) varrem desde o sítio inicial até o sítio final, um algoritmo especial deve ser implementado para permitir a inclusão de regiões distantes das moléculas, que correspondem a conjuntos de orbitais atômicos fora de seqüência.

Além disso as moléculas são desenhadas utilizando o HyperChem^[15] e o processo de desenho influencia a disposição dos átomos: a organização deles aparecerá na seqüência em que forem desenhados. Isto implica em um cuidado especial para manter regiões de interesse dentro de uma seqüência contínua. O algoritmo citado anteriormente resolveria este problema evitando que uma mesma molécula tivesse que ser desenhada diversas vezes para que todas as regiões de interesse pudessem ser analisadas. Um programa em FORTRAN para lidar com todas as possibilidades, se factível, seria de imensa complexidade.

Aproveitando a experiência adquirida no desenvolvimento do software HyperSpin, desenvolvemos o programa Chem2Pac. Este permite realizar os cálculos LDOS de maneira simples, eficiente e livre dos problemas enfrentados nos trabalhos iniciais, que utilizavam a teoria de índices eletrônicos e programas em FORTRAN para a leitura dos autovetores, que tinham de ser editados manualmente antes de serem lidos.

Descreveremos a seguir as principais funções do programa Chem2Pac, pois muitas delas são de uso fácil e imediato, obedecendo os principais padrões encontrados no sistema Windows95.

A figura 5.4 a seguir mostra a janela inicial do Chem2Pac. O programa foi desenvolvido em ambiente VB 5.0 e possui todas as características de um editor de textos do nível do WordPad (componente do sistema Windows).

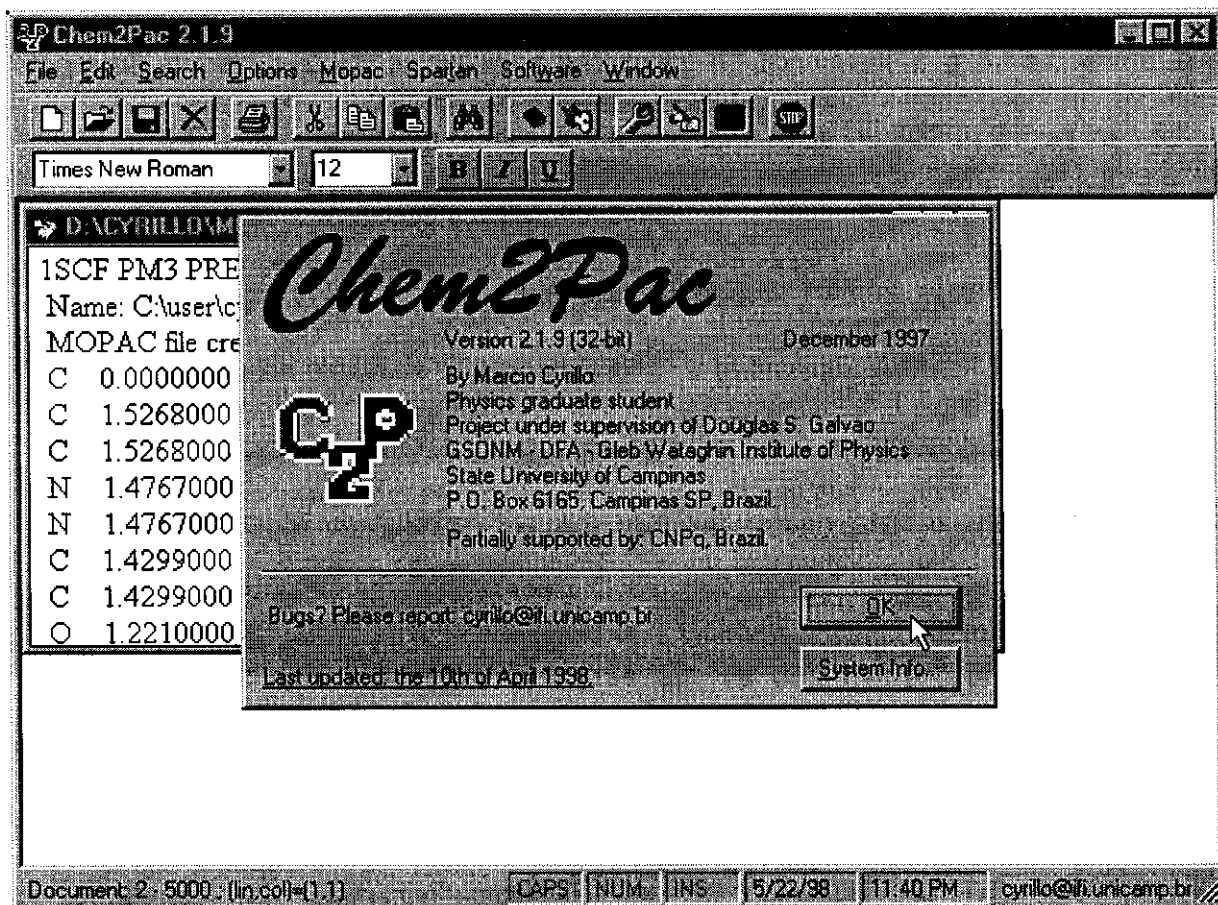


FIGURA 5.4: A tela principal do programa Chem2Pac, um editor de textos voltado para os cálculos LDOS e preparo de *inputs* para o programa MOPAC.

A idéia foi construir um sistema integrado de edição com utilitários que auxiliassem o preparo dos arquivos provenientes do programa HyperChem para que estes pudessem ser processados pelo pacote MOPAC. Além disso, outras opções estariam disponíveis para a análise das saídas de dados, com a finalidade de automatizar a coleta e organização dos resultados provenientes dos cálculos semi-empíricos e efetuar automaticamente os cálculos LDOS.

O histórico de desenvolvimento do programa foi bastante parecido com o do programa HyperSpin: várias versões estiveram disponíveis (16 e 32-bit) até que a versão final ficasse pronta em Março deste ano com mais de 20000 linhas de código tendo sido escritas e testadas exaustivamente.

A maior parte do programa é composta por rotinas simples que utilizam inteligentemente os recursos disponíveis com a tecnologia ActiveX incorporada nas linguagens visuais. No entanto, algumas rotinas para a manipulação de arquivos texto razoavelmente complexas tiveram que ser desenvolvidas, pelo fato de não estarem disponíveis na linguagem VB como acontece com a linguagem C.

Todas as dificuldades encontradas em se desenvolver um programa em FORTRAN que fosse capaz de ler as saídas de dados do MOPAC foram resolvidas e uma interface amigável foi construída a fim de facilitar as especificações e a execução dos cálculos LDOS.

Explicitaremos a seguir o funcionamento do programa dentro das duas etapas em que é mais utilizado: preparo de *inputs* para o MOPAC com arquivos provenientes do HyperChem (onde as moléculas são desenhadas) e cálculos LDOS utilizando as saídas de dados dos cálculos semi-empíricos.

O sistema molecular pode ser manualmente desenhado com a utilização do programa HyperChem, como vimos no capítulo 4. O HyperChem possibilita então a geração de um arquivo no formato Matriz-Z (coordenadas internas), que o MOPAC reconhece. Porém, o usuário deve editar este arquivo colocando nele palavras-chaves que definirão o tipo de cálculo que será efetuado e os parâmetros pertinentes ao mesmo, como critérios de convergência, algoritmos de otimização, etc.

Mostramos a seguir um arquivo do tipo zmt (*MOPAC Z-Matrix*) para uma das moléculas que estudamos neste trabalho:

```
PRECISE VECTORS PM3 NOINTER NOXYZ T=36H EF GNORM=0.001
Name: C:\USER\CYRILLO\MESTRADO\INTEGR-1\INO21.ZMT
MOPAC file created on 25/8 13:40:17 1997 by HYPERCHEM
C 0.0000000 0 0.000000 0 0.000000 0 0 0 0 -0.1347
C 1.3782537 1 0.000000 0 0.000000 0 1 0 0 0.0951
C 1.4378400 1 119.882936 1 0.000000 0 2 1 0 0.0370
C 1.3756210 1 120.548417 1 0.000000 0 3 2 1 -0.1512
C 1.4206290 1 119.835643 1 0.000000 0 4 3 2 -0.0509
C 1.4195266 1 120.033065 1 0.000000 0 1 2 3 -0.0050
C 1.4206828 1 120.907361 1 180.000000 0 6 1 2 -0.0983
C 1.3685653 1 120.628122 1 180.000000 0 7 6 1 -0.0607
C 1.4184707 1 120.199629 1 0.000000 0 8 7 6 -0.1264
C 1.3758862 1 120.298387 1 0.000000 0 9 8 7 -0.0234
C 1.4861869 1 120.688214 1 180.000000 0 9 8 7 0.4220
O 1.2167875 1 127.137100 1 -180.000000 0 11 9 8 -0.3833
O 1.3687035 1 114.115695 1 00.000000 0 11 9 8 -0.2639
C 1.4114407 1 118.525780 1 0.000000 0 13 11 12 0.0763
O 1.3723590 1 116.034265 1 180.000000 0 3 2 1 -0.2422
O 1.3669158 1 116.448679 1 180.000000 0 2 1 6 -0.2202
H 1.0972312 1 120.703113 1 180.000000 0 1 2 3 0.1282
H 1.0975415 1 121.409838 1 0.000000 0 4 3 15 0.1167
H 1.0961783 1 118.808781 1 0.000000 0 7 6 1 0.1108
H 1.0960188 1 120.235675 1 -180.000000 0 8 7 6 0.1169
H 1.0978463 1 120.668716 1 180.000000 0 10 9 8 0.1191
H 1.0929348 1 102.289893 1 180.000000 0 14 13 11 0.0439
H 1.0952185 1 111.839463 1 -62.125763 1 14 13 11 0.0357
H 1.0952159 1 111.891810 1 61.416210 1 14 13 11 0.0355
H 0.9487793 1 107.708423 1 180.000000 0 15 3 2 0.2116
H 0.9483270 1 108.848467 1 0.000000 1 16 2 3 0.2116
```

A primeira linha contém uma lista das palavras-chaves para um cálculo particular. Neste caso temos um cálculo PM3 com os autovetores sendo

gerados na saída de dados (VECTORS), além de outras opções. Estas palavras-chaves devem iniciar na primeira linha e ocupar um máximo de 80 colunas. Se o limite for ultrapassado, sinais de “+” devem ser incorporados no finais das linhas para dizer ao programa que mais uma linha de parâmetros deve ser lida, num limite máximo de três linhas;

A segunda e terceiras linhas são comentários sobre a molécula. Neste caso introduzidos pelo programa HyperChem onde ela foi desenhada;

Da quarta linha em diante temos as definições dos átomos em termos de coordenadas internas, que obedecem ao seguinte esquema:

A	bond	X1	angle	X2	dihedral	X3	A1	A2	A3	Q
---	------	----	-------	----	----------	----	----	----	----	---

onde temos:

- ❑ A, A1, A2, A3: átomos (símbolo e numerações respectivamente);
- ❑ bond, angle, dihedral: graus de liberdade associados ao átomo A;
- ❑ X1, X2 e X3: dígitos de controle dos graus de liberdade;
- ❑ Q: carga líquida sobre o átomo (opcional).

com os seguintes significados:

- A: símbolo do átomo em questão;
- bond: distância em Ângstrons entre os átomos A e A1 (a numeração obedece a ordem contínua, o primeiro átomo A da lista sendo o número 1);
- X1: se 1, bond será otimizada, senão será mantida fixa;
- angle: ângulo formado pelos átomos A, A1 e A2;
- X2: se 1, angle será otimizada, senão será mantida fixa;
- dihedral: ângulo dihedral formado pelos átomos A, A1, A2 e A3 nesta ordem;
- X3: como X1 e X2, para a variável dihedral.

Se o usuário desejar controlar as variáveis que serão otimizadas e as que serão mantidas fixas, deve editar os campos X1, X2 e X3 para cada átomo. Se um destes campos tiver o valor -1 (somente um em cada *input* pode ter este valor), a variável correspondente será tratada como coordenada de reação. O MOPAC lerá então uma linha adicional, que deve ser colocada no final do arquivo, com os valores que esta variável irá possuir em cada passo do cálculo de caminho de reação.

Os valores para *bond*, *angle* e *dihedral* que vemos no exemplo anterior são apenas valores iniciais que se modificarão durante o cálculo de otimização de geometria. Como em todo método numérico, as condições iniciais influenciam sua eficiência e convergência.

O programa Chem2Pac possibilita trabalhar sobre o arquivo texto de *input* (ou uma coleção deles)** de maneira bastante simples e rápida. As opções estão acessíveis do menu MOPAC, como vemos na figura 5.5 a seguir.

Deste menu muitas opções para o preparo de *inputs* e tratamento dos *outputs* estão acessíveis: geração de arquivos de lote para trabalhar com várias moléculas, conversão de *output* em *input*, leitura de dados de coordenadas de reação†† (com a geração automática da tabela no formato coordenada versus calor de formação), etc.

Pode-se ainda acessar o editor de palavras-chaves e de travamento que auxiliam a edição desses parâmetros.

** O programa Chem2Pac é um editor do tipo MDI (*Multiple Document Interface*), que permite a manipulação/visualização simultânea de um grande número de arquivos texto.

†† Coordenadas de reação podem ser usadas para utilizar o MOPAC em buscas conformacionais.

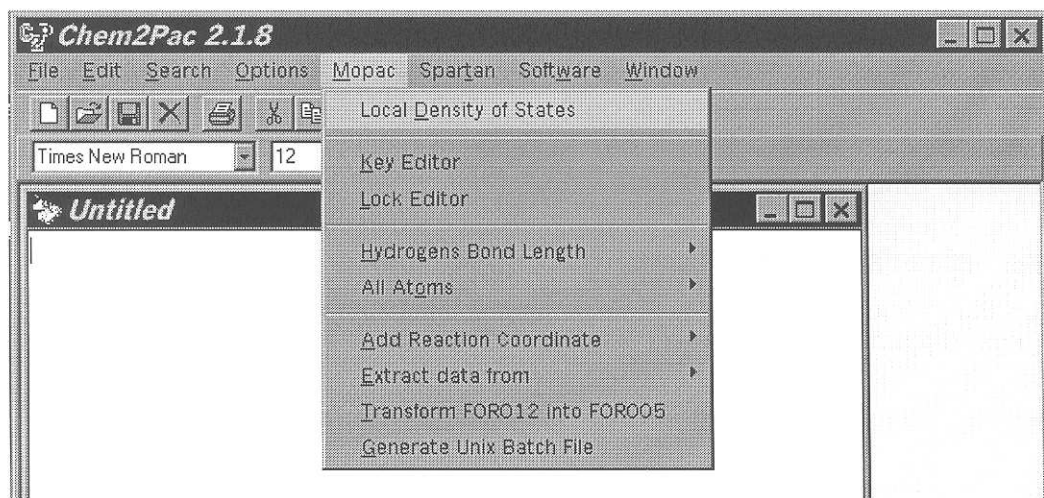


FIGURA 5.5: O menu MOPAC do Chem2Pac que oferece várias ferramentas comumente utilizadas no preparo de *inputs* para esse programa.

A figura 5.6 a seguir mostra o editor de palavras-chaves. Ele permite ao usuário o acesso a todas as *keywords* presentes na versão 6.0 do MOPAC, que podem ser selecionadas em menus organizados quanto as suas funções.



FIGURA 5.6: O editor de palavras-chaves, que permite montar a linha de palavras-chaves do MOPAC automaticamente no arquivo de *input*.

Ao se selecionar um item que requeira um ou mais argumentos, o editor pedirá que usuário os digite. Não é preciso se preocupar em como colocá-las no arquivo texto: logo após selecionar o conjunto desejado, ao simples toque do botão **Apply**, isto será automaticamente efetuado, colocando-as em ordem alfabética no lugar correto e adicionando os sinais de “+” quando necessário. Quaisquer alterações posteriores, podem ser aplicadas no texto já editado.

A opção **Apply to all opened documents** faz com que todos os arquivos abertos no ambiente do Chem2Pac sejam afetados pela edição. Isto permite que um grande conjunto de moléculas possa ser preparado rapidamente.

Como dito anteriormente, para se controlar os graus de liberdade que serão otimizados e aqueles que serão mantidos fixos, é necessário editar manualmente o arquivo de *input* e “setar” os valores dos campos X1, X2 e X3 nos átomos correspondentes.

Um exemplo de utilização deste recurso consiste em travar os comprimentos de ligação dos hidrogênios em moléculas orgânicas no valor de 1.1Å (valor mais provável), pois tal procedimento reduz o número de graus de liberdade, n , em prol da velocidade do cálculo, que diminui com n^3 [16].

Para moléculas grandes o procedimento de edição do arquivo de entrada pode ser trabalhoso e sujeito a erros. Com o objetivo de torná-lo simples e rápido, algumas opções deste tipo foram incluídas no menu MOPAC. Além disto um editor de travamento/destravamento automático de graus de liberdade foi criado, como vemos na figura 5.7 a seguir.

Basta escolher o tipo de átomo (H, C, all, etc.), os campos que devem ser afetados pela operação e clicar no botão correspondente. A opção de atuação em todos os documentos abertos também está disponível. Os arquivos serão automaticamente manipulados e reformatados. Pode-se aplicar este procedimento em qualquer arquivo que obedeça o formato de coordenadas internas.

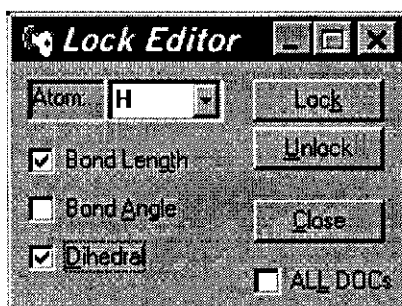


FIGURA 5.7: O editor de opções de otimização, que permite travar ou destravar os graus de liberdade que devem ou não devem ser otimizados durante o cálculo semi-empírico.

Com a opção **Local Density of States** do menu MOPAC o usuário seleciona um de seus arquivos de saída (normalmente referenciado por FOR006), que será lido e processado automaticamente. O programa verifica se os autovetores estão presentes e aborta o processo em caso negativo.

É possível acompanhar a leitura dos blocos de dados, compostos por conjuntos de 4 orbitais moleculares, e uma barra móvel mostra o *status* do processo até sua finalização. Um arquivo contendo os resultados será então automaticamente aberto. Ele conterá todas as informações que puderem ser extraídas do arquivo FOR006: calor de formação, partes da energia total, o espectro de energia em forma de tabela, o tempo de cálculo entre outros.

Além disso o programa colhe algumas informações que julgamos pertinentes à teoria de índices eletrônicos e que se relacionam à estrutura eletrônica das moléculas: posição do HOMO (H) e LUMO (L), orbitais atômicos que mais contribuem para H e L, várias diferenças em energia (H-L, H-(H-1), etc), maior e menor separação entre níveis, etc.

Uma representação gráfica do espectro de energia é produzida e apresentada, como pode ser visto na figura 5.8 a seguir; é possível deslizar o mouse sobre o diagrama e ler o valor da energia correspondente em uma caixa de texto.

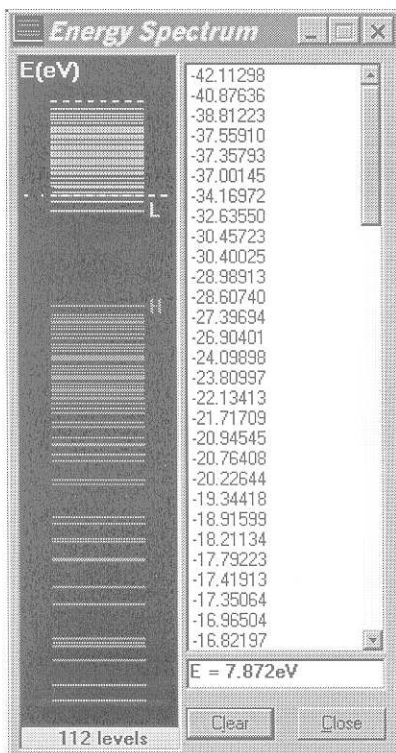


FIGURA 5.8: Espectro de energia obtido da leitura do arquivo de saída de um cálculo semi-empírico. O Chem2Pac permite navegar sobre seus níveis e salvá-lo em disco.

A janela que permite dar continuidade ao cálculo LDOS (figura 5.9) também é apresentada. O programa utiliza as informações extraídas do arquivo FOR006 para montar uma estrutura facilmente manipulável pelo usuário: basta selecionar, usando o mouse, os orbitais atômicos que definem a região onde se quer calcular a densidade de estados e clicar os botões **Add** e **LDOS**, para adicioná-los a lista de sítios e dar continuidade ao cálculo respectivamente. Neste passo só é possível selecionar orbitais atômicos que estejam em ordem contínua.

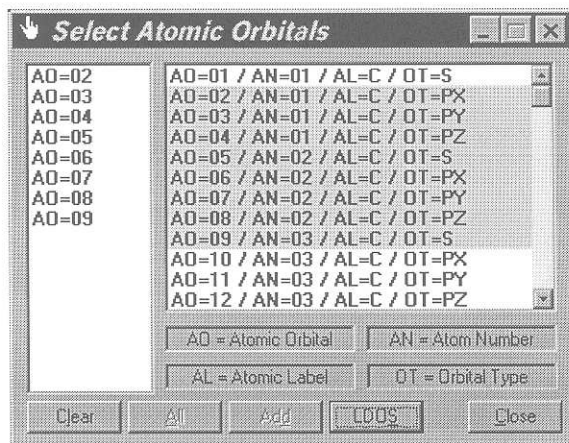


FIGURA 5.9: Janela onde o usuário pode selecionar os orbitais atômicos que entrarão no cálculo LDOS. A subjanela à esquerda apresenta os orbitais que já foram selecionados na subjanela central.

A janela da figura 5.10 é então apresentada. Nela é possível adicionar mais orbitais atômicos manualmente, digitando seus números ou range (ex. 1, 2, 55-66), bastando para isso selecionar **Enter more Atomic Orbitals**. Isto permite que a deficiência da janela anterior, de só permitir a seleção de itens que estejam continuamente dispostos, seja corrigida.

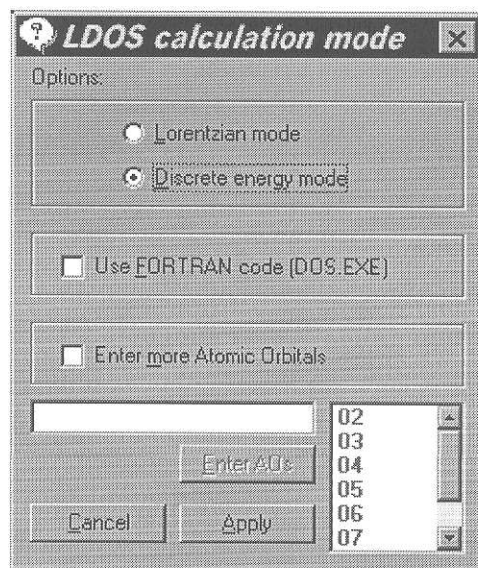


FIGURA 5.10: Janela onde o usuário escolhe como será o cálculo LDOS (contínuo ou discreto) e pode adicionar mais orbitais atômicos aos cálculos, permitindo que átomos fora de sequência possam ser analisados.

Além disso, deve-se escolher nesta janela o tipo de cálculo LDOS: contínuo (alargamento Lorentziano) ou discreto. Pode-se ainda selecionar a rotina de cálculo: se a opção `Use FORTRAN code` estiver selecionada uma rotina nesta linguagem será executada via *shell* para o DOS, caso contrário a rotina interna em VB será utilizada. O uso da rotina em FORTRAN é aconselhável para cálculos com alargamento, pois é muito mais rápida. Vale ressaltar que o algoritmo para o código em FORTRAN torna-se trivial, porque o VB efetua toda a conversão de dados para um formato fácil de ser lido.

Um clique no botão `Apply` leva ao último passo do cálculo, quando mais uma janela aparece (figura 5.11 a seguir).

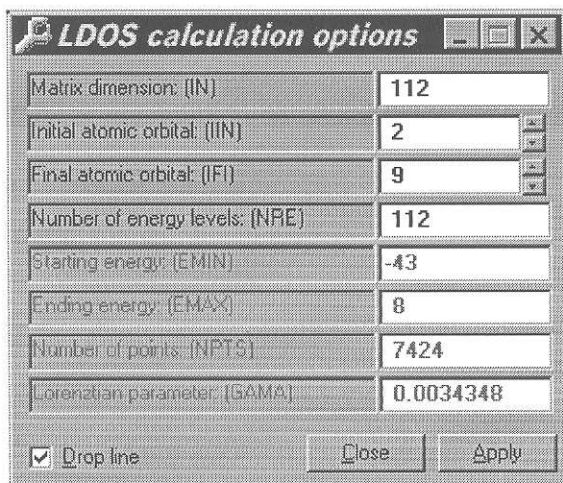


FIGURA 5.11: Janela onde o usuário define os parâmetros do cálculo LDOS. Os quatro últimos campos só serão habilitados no caso de cálculos LDOS com modulação Lorentziana.

As quatro últimas opções só estarão disponíveis se um cálculo com alargamento tiver sido previamente escolhido e são parâmetros que dizem respeito à forma do envelope (número de pontos, fator γ , energia inicial e final).

As outras opções permitem que o usuário monitore as seleções que fez e possa alterá-las neste ponto final sem precisar voltar para fazer uma nova seleção (sítio inicial e final).

O botão **Apply** finaliza o processo: o programa pede que o usuário escolha o nome do arquivo de dados. Este será então gerado no formato energia versus LDOS, podendo ser lido por aplicativos como o Microcal Origin para a confecção dos gráficos. A opção **Drop line** permite que o arquivo, para o

caso discreto, seja preparado de forma especial para o Origin tornando mais fácil a produção de gráficos como o da figura 5.2.

O usuário pode ainda pedir que o Chem2Pac apresente a matriz de coeficientes. Uma planilha no padrão do Excel será então exibida como na figura 5.12 a seguir. É possível navegar sobre suas células e calcular somas de quadrados de coeficientes. Tal recurso pode ser útil para uma visualização compreensiva dos resultados dos cálculos semi-empíricos.

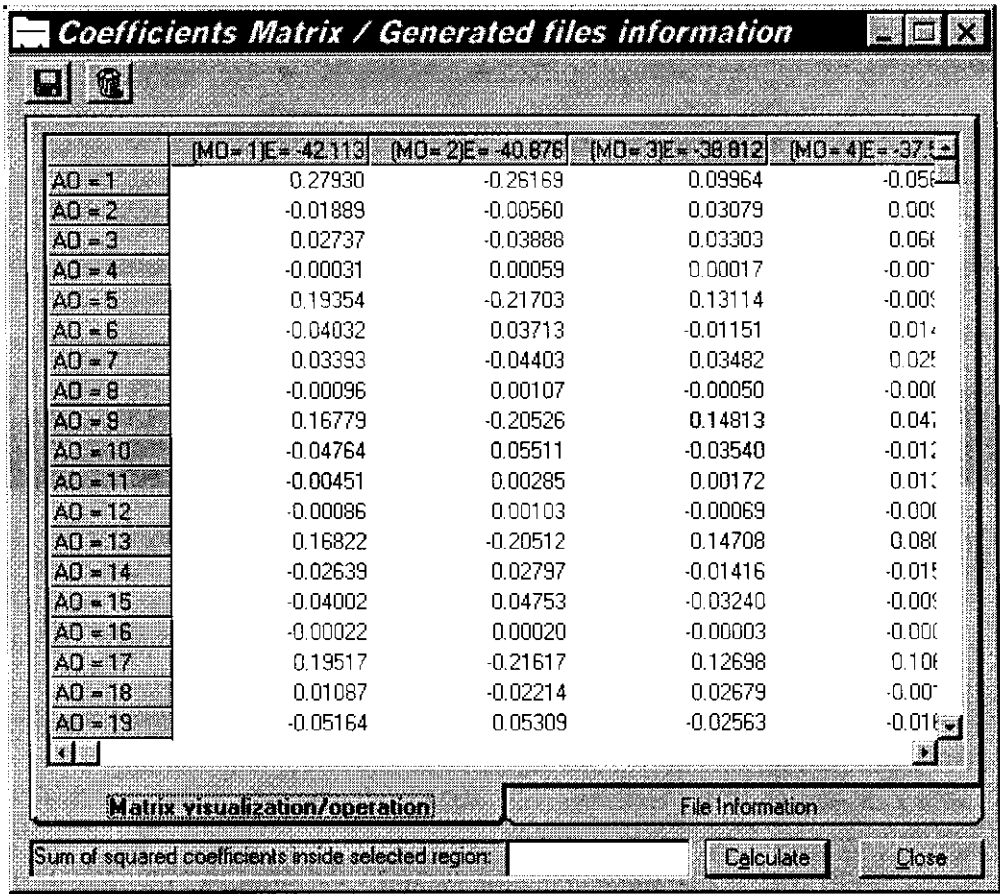


FIGURA 5.12: A matriz de coeficientes disponível ao usuário na forma de uma planilha padrão do Excel. A guia **File Information** contém informações sobre todos os arquivos gerados pelo programa.

Recentemente implementamos um módulo capaz de efetuar o mesmo processo descrito acima, mas analisando os arquivos de saída do programa SPARTAN^[17]. Este passo foi fundamental para o estudo dos efeitos de solventes sobre as regras derivadas pela teoria de índices eletrônicos, pois cálculos deste tipo estão disponíveis com este pacote comercial. No entanto, este estudo foi adiado pelo fato de estarmos enfrentando problemas de hardware com a máquina onde o programa estava instalado e a licença de uso se restringia a esta máquina em específico.

V.5 Usuários e implementações futuras

Ao contrário do programa HyperSpin, o programa Chem2Pac não foi distribuído livremente pela internet. Os testes exaustivos foram feitos por alguns membros de nosso grupo, que passaram a utilizá-lo com bastante frequência. Isto permitiu que a maioria de seus problemas fossem solucionados rapidamente e que novas idéias fossem implementadas.

Apesar do interesse na continuidade de seu desenvolvimento, não pretendemos tornar o Chem2Pac um programa de domínio público, reservando-o somente à utilização interna.

Porém, a idéia de uma “central de edição para química computacional” continua sendo investigada no sentido de se desenvolver um software ainda mais útil para a comunidade científica da área. O programa capaz de implementá-la já

está sendo desenvolvido e chama-se Bab2Pac. Utiliza grande parte da estrutura do Chem2Pac e um programa gratuito chamado Babel^{††}, além de outros módulos.

Pretendemos fazer com que o Bab2Pac dê suporte a programas comerciais bastante utilizados como o MOPAC^[14], HyperChem^[15], SPARTAN^[17], Gromos^[18], Gaussian^[19], etc. e que inclua rotinas de compressão/descompressão de arquivos, FTP, visualização de moléculas em 3D, entre outras, possibilitando a centralização das tarefas mais comuns em um só programa. Será bastante útil também a um novo segmento de pesquisa de nosso grupo que utiliza o programa Gromos no estudo de biomembranas.

A continuidade no desenvolvimento do Chem2Pac reside no fato de torná-lo apto à leitura dos autovetores de outros programas que não o MOPAC e SPARTAN. Esta tarefa poderá ser efetuada por qualquer pessoa com um conhecimento mínimo da linguagem VB.

Pretendemos ainda incluir a opção de leitura dos elementos da matriz de overlap **S**, de tal forma a implementar a definição de LDOS proveniente da equação (12), para efeitos de comparação com nossa definição.

A criação de um manual detalhado está em fase inicial, mas é parte de nossos objetivos efetuar a boa documentação de todo software desenvolvido por nosso grupo para que possam ser futuramente alterados de acordo com as necessidades de pesquisa.

^{††} O programa Babel permite traduzir entre os formatos mais comuns utilizados pelos programas comerciais e públicos: zmt do MOPAC, hin do HyperChem, etc.

- ¹ Cyrillo, M. e Galvão, D.S., "HyperSpin conformational search program, version 2.0 for Windows 95", em preparação.
- ² Paupitz, R., "Estudo estrutura-atividade do Taxol e derivados", Tese de Mestrado, **1998**, IFGW-UNICAMP.
- ³ Braga, R.S.; Barone, P. M. V. B. e Galvão, D.S., THEOCHEM, "Identifying Carcinogenic Activity of Polycyclic Aromatic Hydrocarbons", submetido para publicação em Outubro de 1997.
- ⁴ Pullman, A. e Pullman, B., *Adv. Cancer Res.* **1955**, 3, 117.
- ⁵ Streitwieser, A., "Molecular Orbital Theory", Wiley, N.Y. **1961**.
- ⁶ Galvão, D.S.; dos Santos, D.A.; Laks, B. *et al.*, *Phys. Rev. Lett.* **1989**, 63, 786.
- ⁷ Baughman, R.H. e Galvão, D.S., *Nature (London)* **1993**, 365, 735.
- ⁸ Sansores, L.E.; Valladares, R.M. *et al.*, *Journal of Non-Crystalline Solids* **1992**, 143, 232-240.
- ⁹ Pople, J.A. e Beveridge, D.L., "Approximate Molecular Orbital Theory" McGraw-Hill, New York **1970**.
- ¹⁰ Schmidt, O., *Z. Physik Chem.*, **1938**, 39, 59 e
Svarthölm, N.V., *Arkiv Kemi Mineral. Geol.*, **1942**, A15, 1.
- ¹¹ Barone, P.M.V.B., Tese de Doutorado – IFGW, UNICAMP **1996**.
- ¹² Barone, P.M.V.B.; Camilo Jr., A. e Galvão, D.S. *Phys. Rev. Lett.* **1996**, 77, 1186.
- ¹³ Lima, L.E.S., THEOCHEM, "Structure-activity study of Indoquinones Bio-reductive Alkylating Agents", submetido para publicação em Outubro de **1997**.
- ¹⁴ Stewart, J.J.P., *J. Comp. Chem.* **1991**, 10, 209-221; MOPAC program, version 6.0, Quantum Chemistry Exchange Program No. 455.
- ¹⁵ HyperChem™, *Molecular Visualization and Simulation program*, © HyperCube inc.
- ¹⁶ HyperChem™, *Computational Chemistry manual*, © **1994** HyperCube inc.
- ¹⁷ Spartan™ for UNIX Workstations, © Wavefunction, Inc. (<http://www.wavefun.com>)
- ¹⁸ Gromos 96™ Biomolecular Software Package, (<http://igc.ethz.ch/gromos/gromos.html>)
- ¹⁹ Gaussian 92™ System, © **1992-1993** Gaussian Inc.



Capítulo VI

Resultados e Conclusões

VI.1 Introdução

Neste capítulo apresentaremos os resultados da pesquisa realizada e algumas conclusões a respeito da adaptação da teoria de índices eletrônicos para sistemas moleculares complexos. Discutiremos também sobre a continuidade natural deste trabalho e perspectivas futuras, bem como limitações a que estivemos sujeitos.

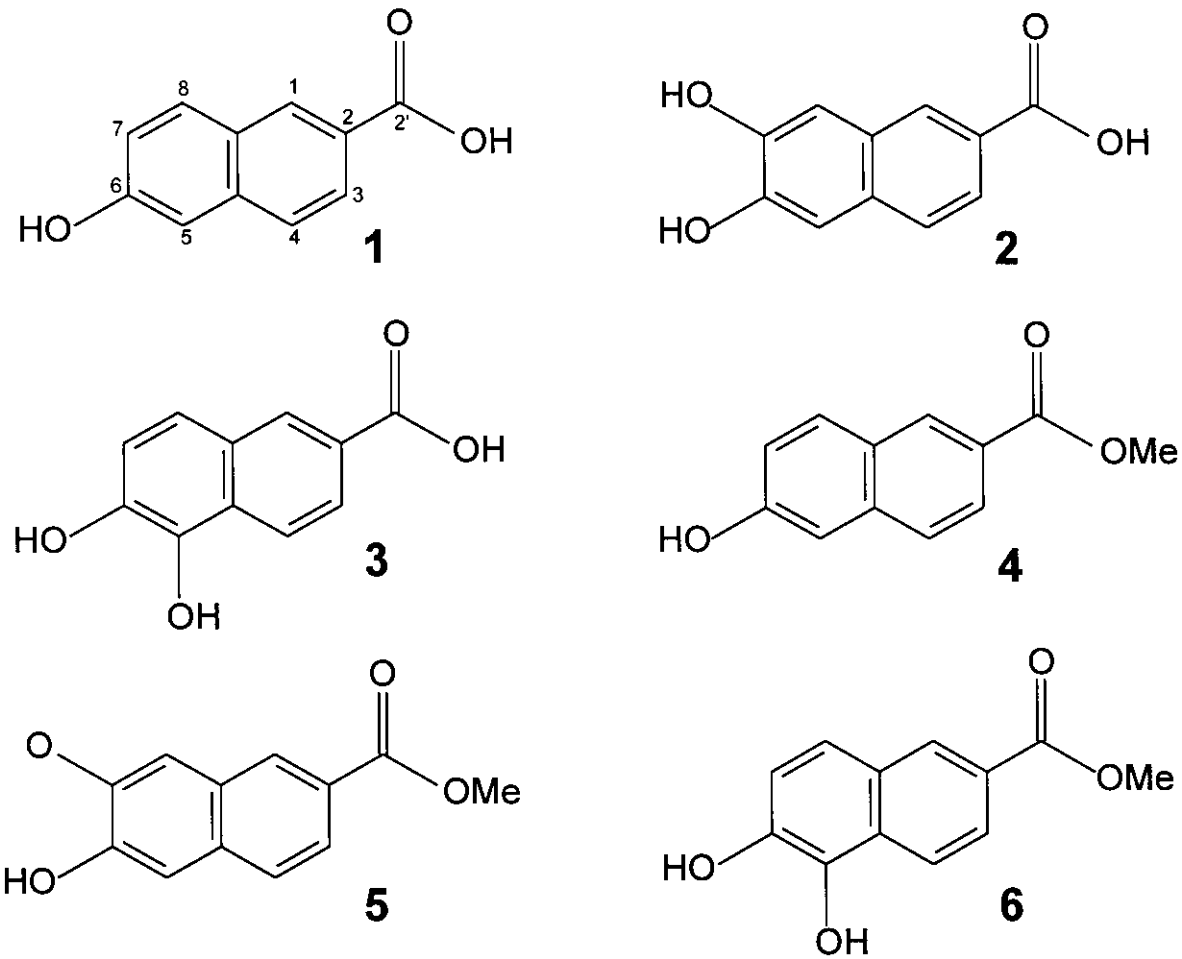
A apresentação dos resultados obedecerá a ordem cronológica: o trabalho com os “monômeros” (figura 3.2) e posteriormente a extensão ao caso dos “dímeros” (figura 3.3). Em ambos os casos a utilização dos programas HyperSpin^[1] e Chem2Pac (capítulo anterior) foi muito importante no sentido de facilitar e agilizar muitos dos procedimentos realizados.

Alguns dos conceitos já abordados em capítulos anteriores serão novamente explicitados com o objetivo de contextualizar cada tópico discutido.

VI.2 Os trabalhos com os monômeros

As moléculas que compõem o grupo que chamamos de “monômeros” são apresentadas novamente na figura 6.1 a seguir. Apesar de suas estruturas serem muito semelhantes, suas manifestações biológicas (medidas pelo índice IC_{50} , ver capítulo 3), apresentam notável variação. Temos, por exemplo, duas moléculas

completamente inativas (4,5) e quatro que apresentam atividade variando desde tênue até muito intensa (1, 6, 3 e 2).



#	NOME	IC ₅₀ (μM)	ATIVIDADE
1	6-hydroxy-2-naphtoic acid	172.3	pouca
2	6,7-dihydroxy-2-naphtoic acid	4.7±0.6	muita
3	5,6-dihydroxy-2-naphtoic acid	62.4±1.1	média
4	6-hydroxy-2-naphtoic acid methyl ester	>200	inativo
5	6,7-dihydroxy-2-naphtoic acid methyl ester	>200	inativo
6	5,6-dihydroxy-2-naphtoic acid methyl ester	52.7±8.9	média

FIGURA 6.1: O conjunto de moléculas chamadas de “monômeros” e que foram sintetizadas e estudadas quanto à capacidade de inibir a enzima integrase do vírus HIV-1^[2].

Pelo fato de possuírem alternância de ligações simples e duplas, estes compostos são classificados como sistemas conjugados. Isto os confere a característica de serem essencialmente planares. Além disso, a nuvem eletrônica situada em orbitais π pode sofrer perturbações a partir de qualquer ponto das moléculas, pois a carga consegue fluir livremente pelos átomos^[3].

As características acima citadas colocam este grupo de moléculas em situação semelhante à encontrada com os PAHs^[4]. Esta foi a grande motivação que levou-nos a submetê-las ao teste da teoria de índices eletrônicos com o objetivo de estabelecer regras que as separassem quanto ao fato de possuírem ou não atividade biológica.

Além disso uma pergunta interessante surge da apreciação da figura 6.1: observamos que as moléculas **4**, **5** e **6** são versões “esterificadas” das moléculas **1,2** e **3** respectivamente, ou seja, diferem entre si apenas pela presença do grupo OMe (ou OCH₃) no lugar de um OH. Observamos que a esterificação tornou as moléculas **1** e **2** inativas (**4** e **5** respectivamente), mas não afetou a atividade da molécula de número **6**; por que? Responder a esta pergunta foi um de nossos objetivos, e com isso preencher uma lacuna deixada no trabalho que caracterizou experimentalmente estas moléculas^[2].

VI.2.1 O preparo das moléculas

O desenho 2D das moléculas e a produção do modelo 3D inicial com o HyperChem^[5] já são capazes de revelar a estabilidade planar destas estruturas, com pequenas elevações dos grupos laterais em relação aos anéis. Apesar disso, aplicamos várias seções de dinâmica molecular

(variando parâmetros como temperatura, tempo de aquecimento e esfriamento, etc.)* com o objetivo de verificar a presença de possíveis conformeros não planares. No entanto, somente pequenas variações de conformação foram encontradas, o que atestou a estrutura quase planar dos compostos.

Como exemplo de problema conformacional, uma pergunta surge naturalmente a respeito da posição relativa dos grupos OH duplos presentes nas moléculas **2,3,5** e **6**. Em um mesmo plano pode haver quatro configurações distintas como mostrado na figura 6.2 a seguir. A utilização do programa HyperSpin (modo 2D) mostrou-se útil e eficiente na busca pelo mínimo global, gerando superfícies como mostradas nas figuras 4.10a e b, que auxiliaram muito a classificação dos mínimos, e, neste exemplo, apontando a configuração **D** como sendo a mais estável.

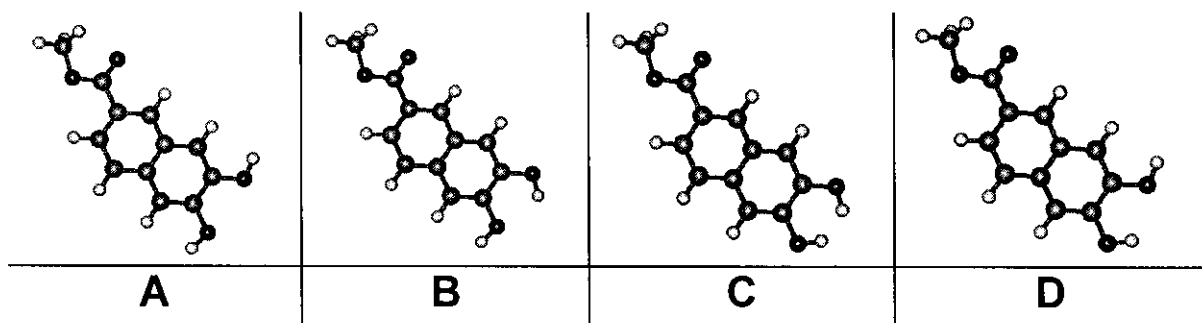


FIGURA 6.2: As quatro conformações possíveis para os grupos OH presentes nas moléculas **2,3,5** e **6** (no exemplo a número **5**). Cálculos com o programa HyperSpin e MOPAC^[6] concordaram que a configuração mostrada em **D** é a mais estável.

* Processo conhecido como *quenched dynamics*.

Buscas 3D foram também realizadas para incluir o grupo lateral formado pelo oxigênio com ligação dupla e, no caso das moléculas **4**, **5** e **6**, ainda o grupo OCH_3 .

Todos os confórmeros (as moléculas nas diversas configurações que poderiam ser os mínimos globais) foram testados usando o MOPAC, o que confirmou a eficiência do programa HyperSpin para a busca conformacional, pela total concordância dos resultados e automatização de todo o processo de busca.

Para todas as moléculas da figura 6.1, encontramos os mínimos globais (fase gasosa) e preparamos seus *inputs* para submetê-las a processos de cálculo PM3 com o MOPAC rodando nas estações de trabalho. Estes arquivos foram gerados utilizando alguns dos recursos do programa Chem2Pac, como foi descrito no capítulo 5.

VI.2.2 Cálculos e resultados

A reotimização de geometria efetuada com o MOPAC teve duplo objetivo: refinar os cálculos para a geometria e obter a matriz de coeficientes, necessária para os cálculos LDOS. Pequenas variações na geometria e no calor de formação foram observados e devem-se ao fato de o algoritmo de otimização implementado no MOPAC ser mais eficiente que aquele disponível com o HyperChem.

A saída de dados contendo a matriz de coeficientes é pedida com a palavra-chave "VECTORS", que deve estar presente no *input* correspondente. Os coeficientes aparecerão em uma formatação especial em meio a outras

informações também presentes na saída de dados, como por exemplo a geometria final e informações sobre todos os ciclos de cálculo.

O programa Chem2Pac foi então utilizado para extrair as informações relevantes dos arquivos de saída do MOPAC: espectro energético, energias dos orbitais HOMO (H) e LUMO (L), calor de formação (C_F), momentos de dipolo, diferenças em energia entre os orbitais mais importantes para definir a reatividade química (H, H-1, H-2, L, L+1, L+2), etc.

Até este momento não havíamos realizado nenhum cálculo LDOS, pois a assinatura eletrônica da atividade biológica poderia estar contida em algum destes resultados, como aconteceu parcialmente em alguns trabalhos realizados por nosso grupo^[4,7,8], que serviram de guia para as estratégias de análise que empregamos.

A tabela 6.1 a seguir apresenta um sumário com os resultados do cálculo PM3 para as moléculas 1-6 mostradas na figura 6.1.

#	C_F	H (eV)	L (eV)	Δ_1 (eV)	Δ_2 (eV)	IC_{50} (μM) ^[2]
1	-94.417	-9.039	-0.884	0.521	0.443	172.3
2	-137.622	-9.024	-0.914	0.203	0.467	4.7 \pm 0.6
3	-137.019	-8.737	-0.970	0.910	0.412	62.4 \pm 1.1
4	-86.227	-8.985	-0.796	0.526	0.431	>200
5	-129.451	-8.972	-0.823	0.208	0.451	>200
6	-128.872	-8.689	-0.874	0.910	0.386	52.7 \pm 8.9

TABELA 6.1: Resultados dos cálculos semi-empíricos (PM3) para os monômeros da figura 6.1. Para cada uma destas moléculas, temos que C_F dá o seu calor de formação, H a energia de seu HOMO, L a energia de seu LUMO, Δ_1 a separação em energia entre HOMO e HOMO-1 e Δ_2 entre LUMO e LUMO+1. O índice IC_{50} mede a intensidade de sua atividade biológica.

Os valores para o calor de formação (C_F), as energias para os orbitais HOMO e o LUMO (H e L respectivamente) e as energias de separação entre HOMO e HOMO-1 (Δ_1) e entre LUMO e LUMO+1 (Δ_2) são apresentadas. Na última coluna mostramos novamente o valor do índice IC_{50} . Observamos que nenhum destes parâmetros pode ser utilizado separadamente como indicador da atividade biológica dos compostos.

Além destes parâmetros, analisamos também os valores para outras diferenças em energia ($H - H-2$, $H - L$, $L+2 - L$, etc.) e para os módulos dos momentos de dipolo, bem como o formato dos orbitais de fronteira usando o HyperChem para a renderização 3D (como na figura 3.5b). Nenhum resultado conclusivo pôde ser obtido, ou seja, o uso dessas grandezas isoladas não nos permitiu estabelecer uma correlação direta com a atividade biológica das moléculas.

Passamos então para a análise da densidade local de estados eletrônicos. Ainda guiados pelo trabalho realizado com os PAHs^[4], calculamos a LDOS para os anéis contendo a maior ordem de ligação. Procuramos correlações entre a atividade biológica dos compostos e os valores críticos das contribuições dos orbitais HOMO e HOMO-1 para a LDOS, observando uma possível dependência com as energias Δ_1 ou Δ_2 . Além dos anéis de maior ordem de ligação, repetimos os cálculos para os anéis terminais.

Como nenhum padrão pôde ser observado nestas regiões, procedemos à análise sobre outras regiões comprovadamente importantes para a atividade anti-HIV^[2]: os átomos de oxigênio (notadamente aquele com a ligação dupla, pois está presente na grande maioria dos inibidores de

integrase^[2,9,10,11]), os grupos OH e os átomos de carbono a eles ligados. Novamente, nenhuma correlação envolvendo as contribuições para o HOMO e HOMO-1 foi encontrada.

Entretanto, quando analisamos o resultado para a LDOS calculada sobre o átomo de carbono 2' (figura 6.1), levando em consideração os valores críticos para as contribuições dos orbitais LUMO e LUMO+1 (em vez de HOMO e HOMO-1), pudemos observar um padrão que pôde ser perfeitamente correlacionado à presença ou ausência de atividade dos compostos. Este padrão, que deu origem às regras que enumeraremos a seguir, incluiu todas as seis moléculas, inclusive a de número 6, que havia surpreendido os experimentais e cuja atividade havia ficado sem explicação plausível^[2].

É interessante frisar que as regras derivadas para os PAHs com o método de Hückel^[4] podem ser rescritas em termos das contribuições relativas da LDOS para os orbitais LUMO e LUMO+1 (como em nosso trabalho), já que no método de Hückel, para o caso dos PAHs, a matriz de coeficientes é simétrica em relação ao *gap* HOMO/LUMO. Isto, *a priori*, não aconteceria com os métodos semi-empíricos. Surpreendentemente, as regras para os PAHs usando o PM3^[12] também poderiam ser rescritas em regras similares envolvendo os orbitais LUMO e LUMO+1, como constatamos posteriormente.

A tabela 6.2 a seguir mostra os resultados obtidos para as contribuições para a LDOS, calculada sobre o átomo de carbono 2', dos orbitais LUMO (L) e LUMO+1 (L+1) e a diferença entre seus valores (δ_{LDOS}).

#	LDOS(L)	LDOS(L+1)	δ_{LDOS}
1	0.67596	0.66362	0.01234
2	1.57808	1.55998	0.01810
3	1.28644	1.24922	0.03722
4	0.05016	0.06066	-0.01050
5	0.35162	0.35665	-0.00503
6	0.95748	0.94937	0.00811

TABELA 6.2: Contribuições para a LDOS, calculada sobre o átomo de carbono 2' das moléculas da figura 6.1, dos orbitais moleculares LUMO (L) e LUMO+1 (L+1). A diferença **LDOS(L) – LDOS(L+1)** também é apresentada (δ_{LDOS}) e é um índice que separa estas moléculas quanto às suas atividades biológicas.

A figura 6.3 mostra uma representação gráfica das contribuições para a LDOS de todos os orbitais moleculares contidos no espectro de energia das moléculas. A abscissa (LDOS) não contém seus valores numéricos por razões estéticas, pois, como se pode observar pela tabela 6.2, são números muito pequenos. Este fato é justificável, já que a contribuição de um orbital molecular para a LDOS discreta, podendo variar de 0 a no máximo 2 (quando calculada sobre todos os átomos da molécula), pode ser bem pequena quando analisada em um único átomo ou sítio[†]. As diferenças entre os valores críticos das contribuições do LUMO e LUMO+1 são, por consequência, menores ainda. Estas diferenças (da ordem de 10^{-3}) são, entretanto, justificáveis por estarem dentro da barra de erro do método PM3 para os valores dos coeficientes, que utilizam seis dígitos significativos.

[†] No nosso caso estamos calculando a LDOS sobre os quatro orbitais atômicos de um átomo de carbono (s, p_x, p_y e p_z).

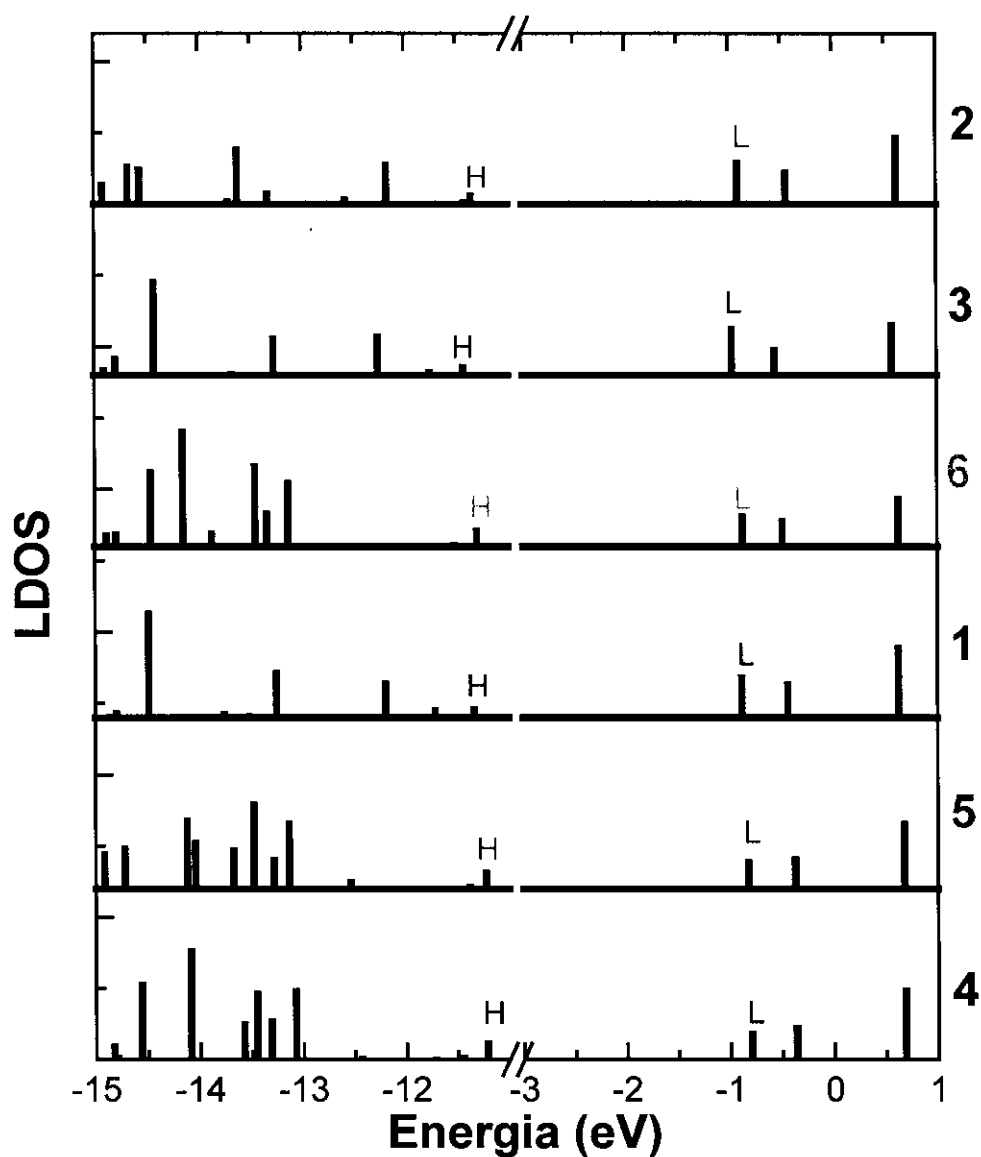


FIGURA 6.3: Densidade local de estados eletrônicos (LDOS) calculada sobre o átomo de carbono 2' (figura 6.1) para as moléculas de **1-6** (monômeros). Resultados provenientes do método semi-empírico PM3. Os símbolos H e L correspondem ao último orbital ocupado (HOMO) e o primeiro desocupado (LUMO), respectivamente.

As regras que separam os monômeros quanto ao fato de apresentarem ou não atividade anti-HIV podem ser derivadas pela observação dos resultados da tabela 6.2 ou da figura 6.3:

- Se a contribuição para a LDOS do LUMO for maior que a contribuição do LUMO+1 ($\delta_{\text{LDOS}} > 0$), a molécula apresentará atividade anti-HIV. Esta pode ser tênue, moderada ou alta ($\text{IC}_{50} < 200 \mu\text{M}$);
- Se, ao contrário, a contribuição para a LDOS do LUMO for menor que a do LUMO+1 ($\delta_{\text{LDOS}} < 0$), a molécula será inativa ($\text{IC}_{50} > 200 \mu\text{M}$).

Como vemos, não precisamos de nenhuma correlação dos valores de δ_{LDOS} com as energias Δ_1 e Δ_2 , como aconteceu para o caso dos PAHs, para chegar a estas regras. Em outros trabalhos do grupo este tipo de correlação apareceu^[7,8] parcialmente.

Apesar da estatística pequena, não acreditamos que a regra anteriormente mencionada seja acidental. Além disso, o átomo em que a LDOS é calculada é importante, pois está ligado a um componente que está presente em quase a totalidade dos compostos inibidores da integrase (o oxigênio com ligação dupla). Alguns pesquisadores mencionaram que a presença deste oxigênio simula um sítio de água importante para a ligação com a enzima integrase^[9].

Poderíamos especular dizendo que o átomo de carbono 2' funcionaria, então, como um *gate* de carga entre um sítio ativo (o átomo de oxigênio) e o resto da molécula e que a densidade de estados sobre ele poderia ser uma medida de quão efetiva seria a atuação deste sítio quando na ligação com a enzima. Devemos frisar, no entanto, que não se espera transferências de carga entre os sítios ativos e a enzima.

Estes resultados foram submetidos para publicação durante o IX Simpósio de Química Teórica (SBQT) realizado em Caxambú em Novembro de 1997 sob a forma de um artigo^[13]. Naquele trabalho mencionamos que um resultado mais conclusivo poderia ser obtido com a extensão das regras ao conjunto de moléculas formadas pelas versões “dimerizadas” das moléculas da figura 6.1 (ver figura 6.4 a seguir). Este trabalho até então estava em estágio inicial e sabíamos que muitas dificuldades apareceriam ao se tratar sistemas não planares como os “dímeros” e que se regras fossem encontradas, poderiam diferir daquelas obtidas para os monômeros.

VI.3 Os trabalhos com os dímeros

O conjunto de moléculas que chamamos de “dímeros” é mostrado novamente na figura 6.4. Estas moléculas são compostas basicamente pela junção de dois pedaços, com a estrutura dos monômeros, interligados por um elemento conector razoavelmente complexo: não conjugado e contendo átomos de nitrogênio[‡].

[‡] Os átomos de nitrogênio não possuem parâmetros semi-empíricos tão bons quanto aqueles para os átomos de carbono^[6].

Como vimos no capítulo 3 (seção 3.2) estas moléculas não podem ser consideradas versões dimerizadas dos monômeros e apesar das semelhanças estruturais com estas, devem ser tratadas como moléculas independentes.

Como o conector é um elemento não conjugado, as moléculas como um todo não serão sistemas conjugados (não planares). *A priori* podemos dizer que a carga não fluirá livremente entre as porções monoméricas, pois não poderá “passar” facilmente pelo conector não conjugado. Deste fato conclui-se que perturbações em uma extremidade destas moléculas não causarão, a princípio, alterações significativas no extremo oposto. Esta característica nos tira da situação ideal para a aplicação da teoria de índices eletrônicos, que foi apresentada pelos monômeros.

A figura 3.6 deu-nos uma idéia da complexidade conformacional apresentada por estes compostos. Devemos levar em consideração ainda que o número bem maior de átomos (mais que o dobro), em relação aos monômeros, trazem dificuldades na convergência dos métodos semi-empíricos, seja do ponto de vista de tempo de cálculo ou de falhas em certos algoritmos de otimização.

Concluimos, então, que o cenário apresentado por estas 6 novas moléculas é bem diferente que aquele obtido para os monômeros:

- não são propriamente dímeros, no sentido estrito;
- possuem um espaço conformacional muito mais complexo;
- não são sistemas conjugados.

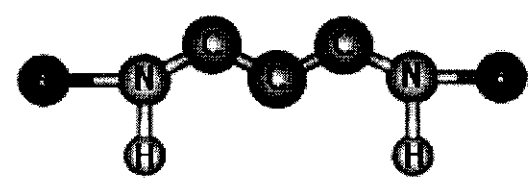
Precisávamos traçar uma estratégia de análise para estes compostos e procuramos levar em consideração os seguintes aspectos:

- os “dímeros” são formados pela duplicação de estruturas similares aos monômeros; Isto pode causar degenerescências (ou quase degenerescências) entre níveis próximos como HOMO e HOMO-1, LUMO e LUMO+1, etc. o que ocorreria pelo princípio comum de formação de polímeros (os níveis moleculares vão se aproximando com a união de unidades similares);
- as regiões interessantes para a análise não estão relacionadas a átomos específicos, mas às funções que eles podem exercer. De acordo com o resultado obtido para os monômeros, devemos procurar por átomos que possam funcionar como *gates* de carga para sítios ativos das moléculas.
- Devemos encontrar as estruturas que representem os mínimos de energia globais das moléculas, mas observar os efeitos de diferentes conformações para a LDOS calculadas em sítios de interesse.

Os cálculos de otimização de geometria para os dímeros foi trabalhoso e delicado. Utilizamos os pedaços otimizados dos monômeros, que foram interligados pelo conector também previamente otimizado, e realizamos análises conformacionais extensivas e detalhadas.

Procuramos analisar os ângulos diedrais mais flexíveis. Devido ao tamanho das moléculas, a utilização do programa HyperSpin foi limitada pela baixa eficiência apresentada pelos cálculos semi-empíricos no HyperChem. Por este motivo fomos obrigados a fazer análises conformacionais preliminares utilizando métodos de mecânica molecular, procedimento este justificado no capítulo 4.

Estrutura geral:



#	R	IC ₅₀ (μM)	ATIVIDADE
7		>200	Inativo
8		>200	Inativo
9		>200	Inativo
10		>200	Inativo
11		0.89	Alta
12		0.17	Alta

FIGURA 6.4: O conjunto de moléculas que chamamos de “dímeros” e que foram sintetizados e estudados quanto à capacidade de inibir a enzima integrase do vírus HIV-1^[2].

Os resultados dos cálculos PM3 para as estruturas que acreditamos ser os mínimos podem ser vistos na tabela 6.3 abaixo:

#	C _f	H (eV)	L (eV)	Δ ₁ (eV)	Δ ₂ (eV)	IC ₅₀ (μM) ^[2]
7	-73.513	-8.928	-0.784	0.00158	0.00067	>200
8	-88.024	-8.990	-0.819	0.00167	0.00068	>200
9	-143.401	-9.075	-0.863	0.00041	0.00055	>200
10	-139.005	-9.069	-0.894	0.00069	0.00061	>200
11	-175.183	-9.068	-0.867	0.00137	0.00068	0.89
12	-171.314	-8.681	-0.765	0.00044	0.00074	0.17

TABELA 6.3: Resultados dos cálculos semi-empíricos (PM3) para os “dímeros” da figura 6.4. Para cada uma destas moléculas, temos que C_f dá o seu calor de formação, H a energia de seu HOMO, L a energia de seu LUMO, Δ₁ a separação em energia entre HOMO e HOMO-1 e Δ₂ entre LUMO e LUMO+1. O índice IC₅₀ mede a intensidade de sua atividade biológica.

Os valores obtidos para Δ₁ e Δ₂ mostram que, como esperávamos, temos uma quase degenerescência entre os níveis HOMO/HOMO-1 e LUMO/LUMO+1. Como no caso dos monômeros, nenhum destes índices eletrônicos pode ser utilizado separadamente para distinguir as moléculas quanto a atividade biológica.

Partimos, então, para os cálculos LDOS. Vale ressaltar que, apesar do detalhamento aplicado durante a análise conformacional, não podemos garantir que os mínimos globais foram efetivamente encontrados, dada a altíssima complexidade das hipersuperfícies de energia. O manual do programa HyperChem cita bons exemplos que dão uma idéia de como estas superfícies podem ser complexas^[14].

De posse dos *outputs* contendo os autovetores da matriz de Fock para cada molécula da figura 6.4, utilizamos novamente o programa Chem2Pac para os cálculos LDOS.

O primeiro procedimento foi verificar se as regras apareceriam com a LDOS calculada no átomo 2', mas neste ponto já surge um problema: para os dímeros há dois átomos 2' correspondentes, um de cada lado do conector. Em qual deles calcular? Testamos todas as possibilidades: calculamos de um lado, do outro e na soma dos dois. No entanto, nenhuma regra que separasse as moléculas pôde ser derivada. Neste ponto resolvemos estudar a sensibilidade dos cálculos LDOS em relação à geometria das moléculas e constatamos que variações significativas poderiam ocorrer no espectro LDOS devido a variações conformacionais da ordem de 0.5 Kcal/mol. Esta característica sugere que devemos ser bastante criteriosos durante as buscas conformacionais.

Convencidos de termos encontrado as geometrias de mínimos globais, concluímos, então, que deveríamos procurar por outras regiões que poderiam estar atuando como *gates* de carga para os átomos de oxigênio ligados aos carbonos 2.

Quando o sistema perde a característica de ser conjugado, é natural que concentrações de carga estejam mais localizadas pelas moléculas[§]; se, para o caso dos monômeros, o átomo 2' atuava como *gate* para a transferência da carga entre o oxigênio com a ligação dupla e o resto das moléculas, esta situação poderia ter sido modificada pela "dimerização".

[§] Este fato foi verificado utilizando o HyperChem para calcular e "plotar" a densidade de carga sobre os dímeros.

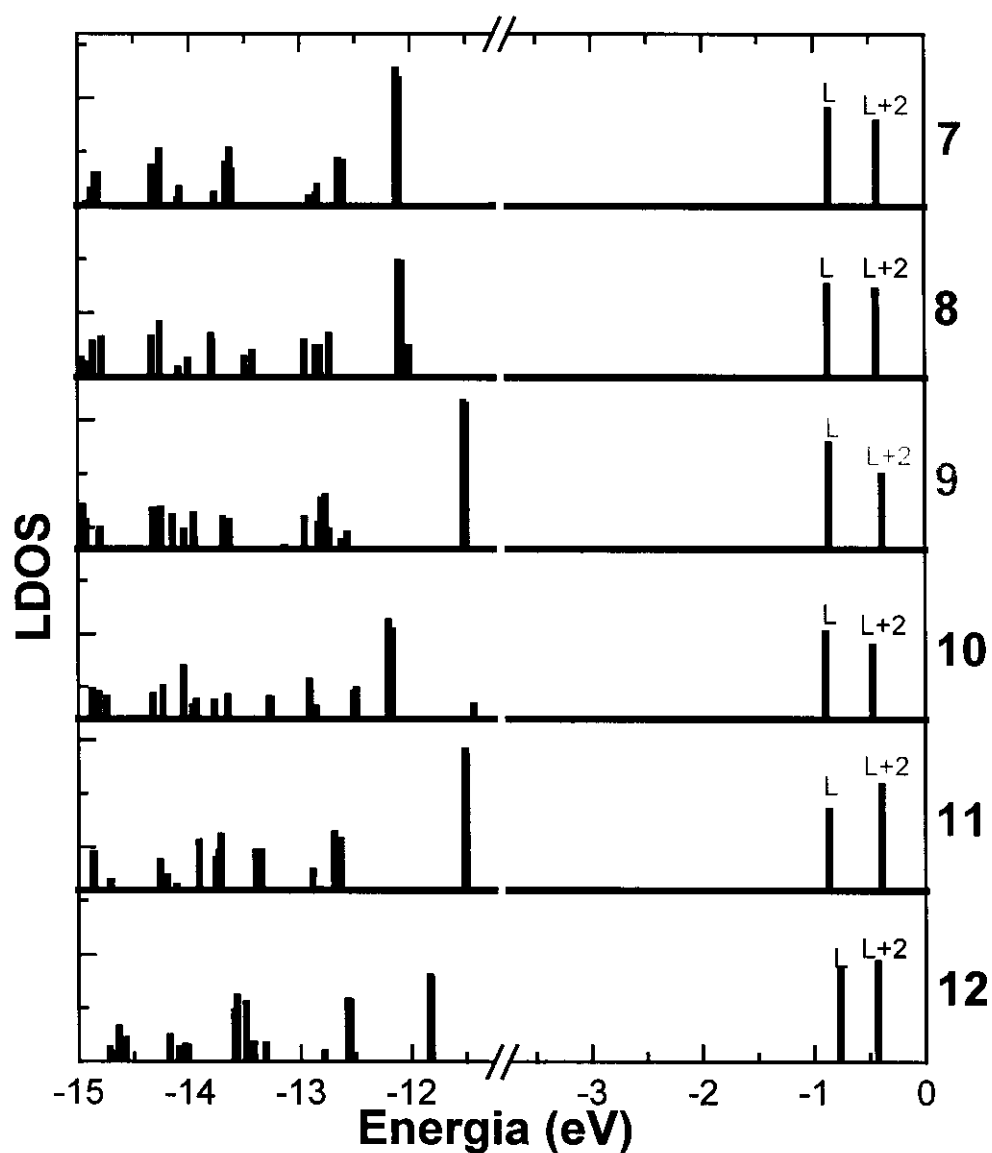


FIGURA 6.5: Densidade local de estados eletrônicos (LDOS) nos átomos de carbono 2 (figura 6.4) para as moléculas de 7-12 (dímeros). Resultados provenientes do método semi-empírico PM3. Os símbolos L e L+2 correspondem ao primeiro e ao terceiro orbitais desocupados respectivamente.

O candidato natural a ser investigado seria o átomo de carbono 2 (ver figura 6.4 elemento 8), pois faz parte do primeiro trecho conjugado ligado ao conector não conjugado e consequentemente ao átomo de oxigênio. Calculamos a LDOS para o átomo de carbono 2 de um lado, do outro e para a soma dos dois. Em todos os casos, o padrão obtido é basicamente o mesmo, e está esquematizado pela figura 6.5.

A análise do padrão apresentado pela figura 6.5 mostra que, similarmente ao caso dos monômeros, podemos encontrar uma regra simples que separa as moléculas quanto ao fato de possuírem ou não atividade biológica. Desta vez a relação aparece entre os orbitais LUMO+1 e LUMO+2 e a regra se inverte em relação à obtida para as moléculas 1-6. A tabela 6.3 a seguir mostra estes resultados.

#	LDOS(L+1)	LDOS(L+2)	δ'_{LDOS}
7	2.23035	2.20151	0.02884
8	1.82099	1.81139	0.00960
9	1.45235	1.3803	0.07205
10	1.00972	0.97874	0.03098
11	0.59359	0.65083	-0.05724
12	0.21574	0.23072	-0.01498

TABELA 6.3: Contribuições para a LDOS, calculada sobre a soma dos dois átomos de carbono 2 das moléculas da figura 6.4, dos orbitais moleculares LUMO+1 (L+1) e LUMO+2 (L+2). A diferença **LDOS(L+1) – LDOS(L+2)** também é apresentada (δ'_{LDOS}) e é um índice capaz de separar estas moléculas quanto às suas atividades biológicas.

As novas regras são:

- Se a contribuição para a LDOS do LUMO for menor que a contribuição do LUMO+2 ($\delta'_{\text{LDOS}} < 0$), a molécula apresentará atividade anti-HIV ($\text{IC}_{50} < 200 \mu\text{M}$);
- Se, ao contrário, a contribuição para a LDOS do LUMO for maior que a do LUMO+2 ($\delta'_{\text{LDOS}} > 0$), a molécula será inativa ($\text{IC}_{50} > 200 \mu\text{M}$).

Como já mencionamos anteriormente, não esperávamos encontrar regras idênticas às obtidas para as moléculas **1-6**, pois sabíamos que os sistemas moleculares, apesar de semelhantes, conservavam características bastante distintas. Felizmente uma regra bastante similar pôde ser observada, reforçando a idéia de que a teoria de índices eletrônicos é uma ferramenta útil na classificação de substâncias orgânicas semelhantes com atividades biológicas distintas.

Além do mais, estes resultados mostram que uma teoria estática bastante simples pode ser capaz de explicar um dos comportamentos de um processo dinâmico bastante complexo, no nosso caso um processo enzimático. Como já mencionamos anteriormente neste trabalho, imaginamos que a presença ou ausência de atividade biológica de certas moléculas possa ser determinada por uma assinatura eletrônica contida nas mesmas. Tal fato torna-se plausível quando lembramos que todos os processos biológicos acontecem dentro de uma janela de energia de 1 a 2eV e que durante processos enzimáticos, pouca ou nenhuma mudança conformacional dos substratos são observadas.

Novamente seria possível afirmar que a estatística ainda é pequena e que poderíamos estar observando um fenômeno de incrível coincidência. O teste de prova seria analisar mais substâncias estruturalmente semelhantes e verificar a

validade das regras ou pequenas variações delas. Infelizmente a literatura nos ofereceu apenas moléculas muito mais complexas^[9] que estas que estudamos neste trabalho e não tivemos tempo hábil para analisá-las. No entanto, temos a intenção de continuar submetendo moléculas com poder anti-HIV ao teste de regras similares, para que uma visão mais global sobre o método e suas estratégias possam ser derivadas. Recentemente tivemos a idéia de combinar diferentes unidades monoméricas para formar “dímeros” como os da figura 6.4 e utilizar as mesmas regras para prever se seriam ou não ativos.

Obtidas as regras, procuramos analisar sua sensibilidade à geometria das moléculas. Observamos que variações conformacionais significativas (maiores que 0,5 Kcal/mol) levavam ao desaparecimento do padrão LDOS da figura 6.5. Concluimos que este fato seria muito importante para a caracterização das moléculas quando em solução: uma porcentagem delas adquiriria conformações que levariam à modificação da atividade (ativa \Leftrightarrow inativa), o restante ainda estaria dentro do limite conformacional observado. Como resultado teríamos que o poder inibitório de um certo composto seria determinado pela quantidade percentual de elementos em uma determinada conformação.

Um possibilidade como esta poderia explicar, por exemplo, porque a molécula de número 1, apesar de ser classificada como ativa por nossa regra, possui uma atividade muito tênue, estando perto do limite de atividade mensurável: ela poderia possuir uma maioria de confórmeros não ativos e aqueles com atividade não estariam bem representados quando em solução.

Este trabalho possuiu duplo objetivo: o primeiro deles foi investigar de maneira criteriosa os aspectos mais relevantes na adaptação da teoria de índices eletrônicos a sistemas moleculares complexos, repletos de graus de liberdade; em segundo plano situou-se o desenvolvimento de ferramentas que ajudassem este trabalho e possibilitassem otimizar os procedimentos nele envolvidos. Além disso, colocaríamos disponíveis recursos que poderiam ser utilizados por outros membros de nosso grupo de pesquisa e que auxiliariam seus trabalhos.

Dado o curto intervalo de tempo em que este trabalho se realizou, (pouco menos de um ano e meio) concluímos que seus resultados foram satisfatórios, pois um conjunto muito útil de programas foi desenvolvido, o que tornará rápidas e simples as futuras investigações com a teoria de índices eletrônicos.

Além do mais, analisamos de perto as dificuldades na extensão desta teoria a sistemas complexos, conseguindo obter regras que separaram um conjunto de 12 moléculas quanto à presença/ausência de atividade biológica anti-HIV. Como resultado conseguimos abordar um aspecto muito complexo, de um processo também muito complexo, através de uma metodologia bastante simples.

Concluimos que, idealmente, a teoria de índices eletrônicos deve ser aplicada a sistemas conjugados e que, se possível, as moléculas devem ter características estruturais semelhantes. No caso de sistemas não conjugados, isto se torna fundamental, pois caso contrário ficaria muito difícil estabelecer regiões estratégicas para os cálculos LDOS.

Resultados experimentais são muito importantes, pois podem guiar as investigações da teoria para as regiões mais relevantes. No nosso caso, estas informações foram de grande ajuda^[2]; em outro trabalho do grupo, fundamental^[7].

A característica mais delicada observada durante este trabalho foi a forte dependência dos resultados para a LDOS com a geometria dos sistemas. Sabemos, por exemplo, que as moléculas estão em solução (dentro de um organismo por exemplo) e elas podem assumir conformações totalmente distintas daquelas encontradas em fase gasosa (as conformações que obtivemos). Chegamos à conclusão que o procedimento mais correto e minucioso seria analisar a consistência das regras para um conjunto de confórmeros dentro de uma distribuição de Boltzmann para cada molécula, ou seja, montar um conjunto significativo de confórmeros para cada molécula e observar como as regras se afetariam para cada um deles.

No entanto, um estudo como este é muito trabalhoso e ainda não dispomos de todas as ferramentas necessárias para colocá-lo em prática, como programas para gerar os confórmeros automaticamente dentro da distribuição de Boltzmann; além disso, não tivemos disponível neste semestre o programa Spartan, que é capaz de realizar cálculos incluindo efeitos de solventes. O programa Chem2Pac, no entanto, já está preparado para ler as suas saídas de dados.

Muitos detalhes para estudos mais completos que envolvam a teoria de índices eletrônicos precisam ser ainda investigados, mas um fato incontestável é a capacidade desta teoria extremamente simples de já ter classificado mais de uma centena de compostos orgânicos com atividades biológicas diversas^[4,7,8,12].

Teorias simples como esta, capazes de apontar *a priori* se um composto possui ou não atividade biológica, podem ser de grande utilidade. Milhões de dólares são gastos anualmente em sínteses e testes de substâncias com atividade biológica variada, com a finalidade de aumentar suas potências ou reduzir suas toxicidades. Como a pesquisa em bancos de dados de substâncias orgânicas¹⁵ pode ser feita trivialmente^[15], pode-se economizar tempo e dinheiro, apontando as moléculas mais promissoras para testes *in vitro* e *in vivo*.

Perspectivas futuras

Pretendemos ainda investigar mais características sobre a teoria de índices eletrônicos, principalmente sobre a sua utilização em sistemas não planares. O efeito de solventes será uma das primeiras investigações, juntamente com a verificação das regras para os PAHs usando métodos *ab initio*.

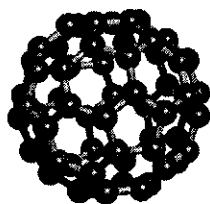
Dentro da impossibilidade de tornar mais explícitos os mecanismos que asseguram a funcionalidade desta teoria, pretendemos, todavia, estabelecer um quadro bastante claro de suas potencialidades e resultados positivos nos mais variados problemas biológicos de interesse, que seriam muito complicados ou mesmo intratáveis através de ferramentas mais sofisticadas.

Em suma, similarmente ao observado para classes de compostos com atividade biológica tão diversa como cancerígenas (PAHs)^[4], anti-cancerígenas (Taxol^[7], Tamoxifeno^[16]), antibióticos (Mitomicina^[17]), antiinflamatórios (Talidomida^[16]) e presentemente aspectos imunológicos (HIV^[13]), a teoria de índices eletrônicos é capaz de correlacionar indicadores eletrônicos à atividade biológica, usando regras

¹⁵ O banco do NCI (*National Cancer Institute*) contém mais de 200mil moléculas catalogadas.

extremamente simples. É intrigante que uma teoria estática como esta seja capaz de esclarecer alguns aspectos biológicos sabidamente dinâmicos. Mais pesquisa se faz necessária para uma melhor compreensão deste fato, mas parece-nos fora de dúvida, em face das dezenas de compostos até aqui investigados, que essa teoria pode ser usada de forma eficiente no desenvolvimento de pesquisa de mais e melhores compostos de ação farmacológica e/ou biológica.

- ¹ Cyrillo, M. e Galvão, D.S., "*HyperSpin conformational search program, version 2.0*", em preparação.
- ² Zhao, H.; Neamati, N.; Mazumder, A.; Sunder, S.; Pommier, Y. e Burke Jr., T.R., *J. Med. Chem.* **1997**, 40, 1186 e referências citadas.
- ³ Salem, L., "*The molecular orbital theory of conjugated systems*", W.A. Benjamin, London **1972**.
- ⁴ Barone, P.M.V.B.; Camilo Jr., A. e Galvão, D.S., *Phys. Rev. Lett.* **1996**, 77, 1186.
- ⁵ HyperChem™, *Molecular Visualization and Simulation program*, © HyperCube inc.
(<http://www.hyper.com>)
- ⁶ Stewart, J.J.P. *J. Comp. Chem.* **1991**, 10, 209, 10, 221; MOPAC program, version 6.0, Quantum Chemistry Exchange Program No. 455.
- ⁷ Paupitz, R., "*Estudo estrutura-atividade do Taxol e derivados*", Tese de Mestrado, **1998**, IFGW-UNICAMP.
- ⁸ Lima, L.E.S., THEOCHEM, "*Structure-activity study of Indoquinones Bio-reductive Alkylating Agents*", submetido para publicação em Outubro de **1997**.
- ⁹ Hong, H.; Neamati, N.; Wang, S. *et al.*, *J. Med. Chem.* **1997**, 40, 930-936.
- ¹⁰ Mazumder, A.; Neamati, N.; Sunder, S. *et al.*, *J. Med. Chem.* **1997**, 40, 3057-3063.
- ¹¹ Zhao, H.; Neamati, N.; Hong, H. *et al.*, *J. Med. Chem.* **1997**, 40, 242-249.
- ¹² Braga, R.S.; Barone, P. M. V. B. e Galvão, D.S., em preparação.
- ¹³ Cyrillo, M. e Galvão, D.S., THEOCHEM, "*Structure-activity relationship study of some new inhibitors of HIV-1 integrase*", submetido para publicação em Outubro de **1997**.
- ¹⁴ HyperChem™ *Computational Chemistry manual*, © **1994** HyperCube inc.
- ¹⁵ Milne, G.W.A.; Nicklaus, M.C. *et al.*, "*The NCI Drug Information System 3D Database*" *J. Chem. Inf. Comput. Sci.* **1994**, 34, 1219-1224.
- ¹⁶ Lima, L.E.S., Tese de Doutorado, IFGW-UNICAMP, em preparação.
- ¹⁷ Barone, P.M.V.B., Tese de Doutorado, IFGW- UNICAMP **1996**.



Apêndice A

Métodos de Mecânica Quântica

A.1 Aproximação de Hartree-Fock

À exceção de sistemas moleculares simples como o íon H_2^+ , todos os outros possuem vários núcleos e elétrons que interagem entre si. Moléculas constituem sistemas quânticos de muitos corpos bastante complexos, para os quais não há solução analítica exata.

Diversos métodos aproximados para lidar com este problema foram e vêm sendo desenvolvidos com o propósito de obter propriedades moleculares de interesse e, em especial, para encontrar o estado fundamental de moléculas. Uma das aproximações mais usadas foi desenvolvida por Hartree, Fock e Roothan^[1] utilizando o conceito de campo médio.

Para encontrar o estado fundamental de uma molécula devemos resolver a equação de Schrödinger independente do tempo para a mesma:

$$\mathbf{H}|\psi_0\rangle = E_0|\psi_0\rangle,$$

(A. 1)

onde \mathbf{H} é o hamiltoniano para o sistema molecular e $|\psi_0\rangle$ é a função de onda para o estado fundamental do mesmo, cuja energia é E_0 .

Desprezando efeitos relativísticos, \mathbf{H} pode ser escrito como:

$$\mathbf{H} = \mathbf{H}_{el} + \mathbf{H}_{núcleos} = -\sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} + \sum_{A=1}^M \frac{1}{2M_A} \nabla_A^2 + \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{R_{AB}},$$

(A. 2)

para um sistema de M núcleos e N elétrons, em unidades atômicas.

Desta equação temos:

- M_A é a razão da massa do núcleo A pela massa do elétron;
- Z_A é o número atômico do núcleo A ;
- os operadores Laplacianos, ∇_i^2 e ∇_A^2 envolvem diferenciações em relação às coordenadas do i -ésimo elétron e A -ésimo núcleo, respectivamente;
- o primeiro e o quarto termo representam as energias cinéticas dos elétrons e núcleos, também respectivamente;
- o segundo termo representa a energia de interação entre núcleos e elétrons (atração de Coulomb) e o terceiro termo a energia de repulsão entre os elétrons.

Podemos simplificar o problema aplicando a aproximação de Born-Oppenheimer, que separa o problema eletrônico do problema nuclear, admitindo que isto pode ser feito pelo fato de os núcleos serem muito mais pesados que os elétrons. Estes movem-se, então, no campo produzido pelos núcleos em cada instante. Dentro desta aproximação, teremos que $|\psi_0\rangle = |\psi_{el}\rangle |\psi_{nuclear}\rangle$, onde $|\psi_{el}\rangle$ satisfará a equação de Schrödinger eletrônica:

$$\mathbf{H}_{el} |\psi_{el}\rangle = E_{el} |\psi_{el}\rangle .$$

(A. 3)

Aplicamos a seguir a aproximação de campo médio, onde escrevemos a função de onda $|\psi_{el}\rangle$ como um produto antissimetrizado de funções de onda de uma partícula, ou seja, como se os elétrons não interagissem.

O spin do elétron também deve ser levado em conta. Isto é feito através de duas funções ortonormais, $\alpha(\omega)$ e $\beta(\omega)$ (spin $+1/2$ e $-1/2$, respectivamente). A função que descreverá tanto a função de onda espacial como a de spin é denominada spin-orbital, $\chi(\mathbf{x})$, onde \mathbf{x} representa todas as coordenadas espaciais e de spin.

Teremos:

$$\chi_{2i-1}(\mathbf{x}) = \psi_i(\mathbf{r})\alpha(\omega) \text{ e } \chi_{2i}(\mathbf{x}) = \psi_i'(\mathbf{r})\beta(\omega),$$

(A. 4)

onde $i=1,2,\dots,K$.

Pela condição de ortonormalidade, os spins-orbitais, χ 's, devem satisfazer:

$$\langle \chi_i | \chi_j \rangle = \int \chi_i^*(\mathbf{x}) \chi_j(\mathbf{x}) d\mathbf{x} = \delta_{ij}.$$

(A. 5)

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

$$h(i)\chi_j(\mathbf{x}) = \varepsilon_j \chi_j(\mathbf{x}).$$

(A. 6)

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

$$\Psi^{HP}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) = \chi_i(\mathbf{x}_1) \chi_j(\mathbf{x}_2) \dots \chi_k(\mathbf{x}_N),$$

(A. 7)

que é uma autofunção de \mathbf{H}_1 , portanto teremos:

$$\mathbf{H}_1 \Psi^{HP} = E_1 \Psi^{HP},$$

(A. 8)

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

$$E_1 = \varepsilon_i + \varepsilon_j + \dots + \varepsilon_k.$$

(A. 9)

A função de onda aproximada Ψ^{HP} é chamada de produto de Hartree e podemos observar dois aspectos a respeito da mesma:

- 1) não inclui os efeitos de repulsão elétron-elétron;
- 2) não respeita a antissimetria requerida para uma autofunção eletrônica, ou seja, para a troca de coordenadas entre dois elétrons não há a necessária troca de sinal para a função de onda para férmions.

A idéia é utilizar Ψ^{HP} como função de onda aproximada para o sistema de elétrons interagentes. Para antissimetrizá-la corretamente utilizaremos o chamado determinante de Slater, onde colocaremos os spins-orbitais alinhados em linhas e colunas. Com isso, a troca de duas linhas ou colunas (coordenadas de dois elétrons) causará a troca de sinal do determinante (da função de onda). De forma compacta podemos escrever a função de onda do estado fundamental como o seguinte determinante de Slater:

$$|\psi_0\rangle = |\chi_1 \chi_2 \dots \chi_N\rangle.$$

(A. 10)

Com o hamiltoniano eletrônico total sendo dado por

$$\mathbf{H}_{el} = -\sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}},$$

(A. 11)

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

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

(A. 12)

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

Através destas expressões, podemos chegar, após algumas manipulações algébricas, à seguinte expressão para a energia do estado fundamental:

$$E_0 = \langle \psi_0 | \mathbf{H} | \psi_0 \rangle = \sum_i^N \langle i | \mathbf{h} | i \rangle + \frac{1}{2} \sum_i^N \sum_j^N \langle ii | jj \rangle - \langle ij | ji \rangle.$$

(A. 13)

Onde foram usadas as seguintes definições:

$$\langle i | \mathbf{h} | j \rangle = \int d\mathbf{x}_1 \chi_i^*(\mathbf{x}_1) \mathbf{h}(\mathbf{r}_1) \chi_j(\mathbf{x}_1),$$

(A. 14)

as integrais de um elétron e

$$\langle ij | kl \rangle = \int d\mathbf{x}_1 d\mathbf{x}_2 \chi_i^*(\mathbf{x}_1) \chi_j(\mathbf{x}_1) \frac{1}{r_{12}} \chi_k^*(\mathbf{x}_2) \chi_l(\mathbf{x}_2),$$

(A. 15)

as integrais de dois elétrons.

O caminho que nos leva às equações de Hartree-Fock consiste em minimizar a energia do sistema $E_0(\{\chi_i\})$ (A. 13) em relação aos spins-orbitais, sendo que a ortonormalidade dos mesmos funcionará como condição de vínculo. Escrevemos então a função de Lagrange $L(\{\chi_i\})$ para a energia do estado fundamental:

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

(A. 16)

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

Utilizaremos ainda duas definições:

- **Operador de Troca:**

$$K_j(1)\chi_i(1) = \left[\int d\mathbf{x}_2 \chi_j^*(2) \frac{1}{r_{12}} \chi_i(2) \right] \chi_j(1);$$

(A. 17)

- **Operador de Coulomb:**

$$J_j(1)\chi_i(1) = \left[\int d\mathbf{x}_2 \chi_j^*(2) \frac{1}{r_{12}} \chi_j(2) \right] \chi_i(1).$$

(A. 18)

O Operador de Coulomb $J(1)$ representa a interação entre o elétron 1 e a densidade eletrônica dada por $|\chi_j(2)|^2$, já o Operador de Troca não possui uma interpretação física similar. É um efeito puramente quântico que surge da

necessidade de a função de onda do sistema ser antissimétrica em relação a troca das coordenadas de dois elétrons. Podemos observar pela expressão (A. 17), que quando $K_j(1)$ opera sobre $\chi_i(1)$, ocorre a troca entre o elétron 1 e o elétron 2 do lado direito de $\frac{1}{r_{12}}$. O resultado desta operação depende do valor do spin-orbital

$\chi_i(1)$ em todo o espaço.

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

Chegamos, assim, a uma expressão chamada de **Forma não Canônica das Equações de Hartree-Fock**, por não ser uma equação de autovalores na forma mais comum:

$$\left[\mathbf{h}(1) + \sum_{j=1}^N J_j(1) - K_j(1) \right] \chi_i(1) = \sum_{j=1}^N \lambda_{ij} \chi_j(1),$$

(A. 19)

com $i=1,2,3,\dots,N$. Podemos definir o operador de Fock

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

(A. 20)

para rescrever (A.18) como

$$\mathbf{F} | \chi_i \rangle = \sum_{j=1}^N \lambda_{ij} | \chi_j \rangle.$$

(A. 21)

Consideremos agora um novo conjunto de spins-orbitais $\{\chi'_i\}$, que são obtidos a partir de uma transformação unitária aplicada ao conjunto antigo $\{\chi_i\}$:

$$\chi'_i = \sum_j \chi_j U_{ji} ,$$

(A. 22)

que será também um conjunto ortonormal. Podemos verificar que tal transformação, uma rotação no sistema de coordenadas, não afeta o significado físico da função de onda, pois teremos

$$|\Psi'_0\rangle = e^{i\varphi} |\Psi_0\rangle ,$$

(A. 23)

nem tampouco a forma dos operadores de Troca e de Coulomb. Do ponto de vista dos multiplicadores de Lagrange, λ_{ij} , uma expressão matricial envolvendo a nova e velha matriz pode ser escrita:

$$\lambda' = \mathbf{U}^\dagger \lambda \mathbf{U} .$$

Podemos verificar que é sempre possível encontrar uma matriz unitária \mathbf{U} tal que a transformação acima diagonaliza λ . Não nos preocuparemos em como obtê-la, mas somente com o fato de que ela existe e é única. Podemos então escrever a

Forma Canônica das Equações de Hartree-Fock:

$$\mathbf{F} |\chi_i\rangle = \lambda_i |\chi_i\rangle .$$

(A. 24)

A essência destas equações consiste em substituir o complicado problema multieletrônico, por vários problemas de um elétron, de tal maneira que a repulsão eletrônica possa ser tratada através de um campo médio.

Podemos limitar os spins-orbitais de tal maneira que a parte espacial dos mesmos possa ser igual ou diferente para elétrons com spins. O primeiro caso denomina-se restrito, o segundo irrestrito. Trataremos apenas o caso restrito ($\psi_i'(\mathbf{r}) = \psi_i(\mathbf{r})$ na equação (A. 4)), que foi empregado no nosso trabalho.

Temos um sistema molecular de N elétrons, N par, com todos eles estando emparelhados dois a dois em cada um dos $n=N/2$ orbitais espaciais. Ou seja, restringimos a discussão a sistemas no estado fundamental e de camada fechada. O formalismo para os estados fundamentais de camada aberta difere bastante deste e não será tratado aqui.

Desejamos converter os spins-orbitais gerais da equação de Hartree-Fock a uma equação de autovalores espacial, vista a seguir, onde cada um dos orbitais moleculares espaciais está duplamente ocupado. Para tal, faz-se necessário integrar sobre as funções de spin^[2].

$$\mathbf{F}(1)\Psi_j(1) = \varepsilon_j \Psi_j(1),$$

(A. 25)

com o operador de Fock para camada fechada podendo ser escrito como:

$$\mathbf{F}(1) = \mathbf{h}(1) + \sum_i^{N/2} 2J_i(1) - K_i(1),$$

(A. 26)

onde as definições de operador de Coulomb (J) e de Troca (K) são totalmente análogas àquelas derivadas para os spins-orbitais, com a mudança $\chi \rightarrow \Psi$ no integrando. Além disso, temos o fator 2 à frente do operador de Coulomb, que está ligado à degenerescência de spin.

Uma vez eliminadas as coordenadas de spin, o cálculo de orbitais moleculares torna-se equivalente ao problema de resolver a seguinte equação espacial íntegro-diferencial:

$$\mathbf{F}(\mathbf{r}_1)\Psi_j(\mathbf{r}_1) = \varepsilon_j\Psi_j(\mathbf{r}_1).$$

(A. 27)

Em vez de tentar resolver diretamente a equação acima, o que não seria um procedimento simples, Roothaan introduziu um conjunto de funções de base espaciais de forma a transformar um problema de equação íntegro-diferencial num problema matricial. Para tal, a função de onda espacial deve ser projetada em uma base de dimensão finita ξ . Se esta base fosse completa, a função de onda seria descrita perfeitamente.

$$\Psi_i = \sum_{\mu=1}^{\xi} C_{\mu i} \phi_{\mu},$$

(A. 28)

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

Substituindo a equação (A. 28) acima na equação (A. 27) e utilizando as seguintes definições:

- matriz de recobrimento*: $S_{\mu\nu} \equiv \int d^3r_1 \phi_{\mu}^*(1) \phi_{\nu}(1);$

(A. 29)

- matriz de Fock: $F_{\mu\nu} \equiv \int d^3r_1 \phi_{\mu}^*(1) \mathbf{F}(1) \phi_{\nu}(1),$

(A. 30)

podemos chegar às equações de Roothaan:

$$\sum_{\nu=1} C_{\nu} \int d^3 r_1 \phi_{\mu}^*(1) \mathbf{F}(1) \phi_{\nu}(1) = \varepsilon_i \sum_{\nu=1} C_{\nu} \int d^3 r_1 \phi_{\mu}^*(1) \phi_{\nu}(1),$$

(A. 31)

ou

$$\sum_{\nu} F_{\mu\nu} C_{\nu} = \varepsilon_i \sum_{\nu} S_{\mu\nu} C_{\nu},$$

(A. 32)

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

A equação de Roothaan pode ser escrita na forma matricial:

$$\mathbf{FC} = \mathbf{SC}\varepsilon.$$

(A. 33)

Para determinarmos o conjunto de orbitais moleculares espaciais $\{\Psi_i\}$ de Hartree-Fock e as energias orbitais ε_i , teremos que resolver a equação matricial acima. Para simplificar o processo de diagonalização precisamos de uma maneira de eliminar a matriz \mathbf{S} desta equação. Antes, porém, vamos definir o conceito de matriz densidade.

A.2.2 A densidade de carga

A função distribuição de probabilidade (densidade de carga) de um elétron é dada pela função de onda $\Psi_i(\mathbf{r})$. Uma molécula de camada fechada descrita por uma função de onda total de determinante único, onde cada orbital molecular é ocupado por dois elétrons, possui a densidade de carga total (em unidades atômicas $e=1$) dada por:

* Também denominada matriz de *overlap*.

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

(A. 34)

onde a densidade de carga total $\rho(\mathbf{r})d^3r$ é a probabilidade de se encontrar um elétron qualquer em um volume d^3r centrado em \mathbf{r} . Integrando-se sobre todo o espaço teremos o número total (N) de elétrons.

A expressão final para a densidade de carga é obtida substituindo a equação (A.28) na equação (A.34) acima, o que nos leva a:

$$\rho(\mathbf{r}) = \sum_{\mu\nu} P_{\mu\nu} \phi_{\mu}(\mathbf{r}) \phi_{\nu}^*(\mathbf{r}) ,$$

(A. 35)

onde usamos a definição de matriz densidade ou matriz de ordem de ligação:

$$P_{\mu\nu} = 2 \sum_i^{N/2} C_{\mu i} C_{\nu i}^* .$$

(A. 36)

A partir da expressão (A. 35), podemos observar que dado um conjunto conhecido de funções de base $\{\phi_{\mu}\}$, bem como os seus coeficientes, podemos obter a matriz \mathbf{P} , que especifica completamente a densidade de carga $\rho(\mathbf{r})$ da molécula. Essa matriz está diretamente relacionada aos coeficientes de expansão (matriz \mathbf{C}) através da equação (A. 36) acima.

Em um procedimento auto consistente (SCF ou *Self Consistent Field*), cujo esquema será apresentado mais adiante, os coeficientes da matriz **C** serão modificados até que se possa encontrar uma energia total convergente, quando os elementos da matriz **P** também convergem.

Podemos escrever o operador de Fock para a camada fechada em termos da matriz densidade:

$$\mathbf{F}(\mathbf{r}_1) = \mathbf{h}(\mathbf{r}_1) = \frac{1}{2} \sum_{\lambda\sigma} P_{\lambda\sigma} \left[\int d\mathbf{r}_1 \phi_{\sigma}^*(\mathbf{r}_2) (2 - p_{12}) \phi_{\lambda}(\mathbf{r}_2) \right],$$

(A. 37)

onde p_{12} é o operador que permuta os índices 1 e 2.

Um outro modo de escrever a equação acima, explicitando os operadores de Coulomb e de Troca foi visto na equação (A. 26). Podemos utilizá-la para escrever uma nova expressão final para os coeficientes da matriz de Fock:

$$F_{\mu\nu} = H_{\mu\nu} + G_{\mu\nu} = (T_{\mu\nu} + V_{\mu\nu}) + G_{\mu\nu},$$

(A. 38)

utilizando as seguintes definições:

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

(A. 39)

- $V_{\mu\nu} \equiv \int d^3r_1 \phi_{\mu}^*(1) \left[-\sum_A \frac{Z_A}{r_{1A}} \right] \phi_{\nu}(1);$

(A. 40)

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

(A. 41)

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

Os dois primeiros termos da expressão (A. 38) envolvem apenas um elétron e são fixos para um dado conjunto de funções de base $\{\phi_\mu\}$. A expressão para $G_{\mu\nu}$ é, no entanto, um termo de dois elétrons e depende da matriz densidade \mathbf{P} , além de um conjunto de integrais de dois elétrons. Estas constituem a maior dificuldade do cálculo Hartree-Fock, pela grande quantidade de cálculos complexos que envolvem.

As equações de Roothaan ($\mathbf{FC}=\mathbf{SC}\epsilon$) teriam somente a forma simplificada $\mathbf{FC}=\mathbf{C}\epsilon$ se o conjunto de base fosse ortonormal, para os quais a matriz de *overlap* \mathbf{S} seria uma matriz N -identidade.

Como na maioria dos casos o conjunto de funções de base $\{\phi_\mu\}$ não é ortonormal, precisaremos de uma matriz de transformação \mathbf{X} não unitária que nos leve a um conjunto $\{\phi_\mu'\}$ ortonormal:

$$\mathbf{X}^\dagger \mathbf{S} \mathbf{X} = \mathbf{1};$$

(A. 42)

$$\mathbf{U}^\dagger \mathbf{S} \mathbf{U} = \mathbf{s},$$

(A. 43)

onde \mathbf{S} é uma matriz hermitiana, \mathbf{U} unitária e \mathbf{s} diagonal. De (A. 42) temos que $\mathbf{X} = \mathbf{S}^{-1/2} = \mathbf{U} \mathbf{s}^{-1/2} \mathbf{U}^\dagger$ (usando a inversão de (A. 43)).

Teremos:

$$\mathbf{C}' = \mathbf{X}^{-1} \mathbf{C} \text{ ou } \mathbf{C} = \mathbf{X} \mathbf{C}';$$

(A. 44)

$$(X^{\dagger}FX)C'=(X^{\dagger}SX)C'\varepsilon;$$

(A. 45)

$$F'\equiv X^{\dagger}FX;$$

(A. 46)

$$F'C'=C'\varepsilon,$$

(A. 47)

Como a matriz **S** é diagonalizada através da matriz unitária **U**, a matriz diagonal **s** será formada pelos autovalores de **S**. O método de Hartree-Fock-Roothaan resume-se em diagonalizar a matriz **F'** dada por (A. 47).

Através das ferramentas matemáticas descritas até o momento podemos implementar um método numérico para aplicar o método de Hartree-Fock-Roothaan. A figura A.1 a seguir mostra o algoritmo computacional para efetuar esta tarefa.

PROCEDIMENTO SCF — Hartree-Fock-Roothaan

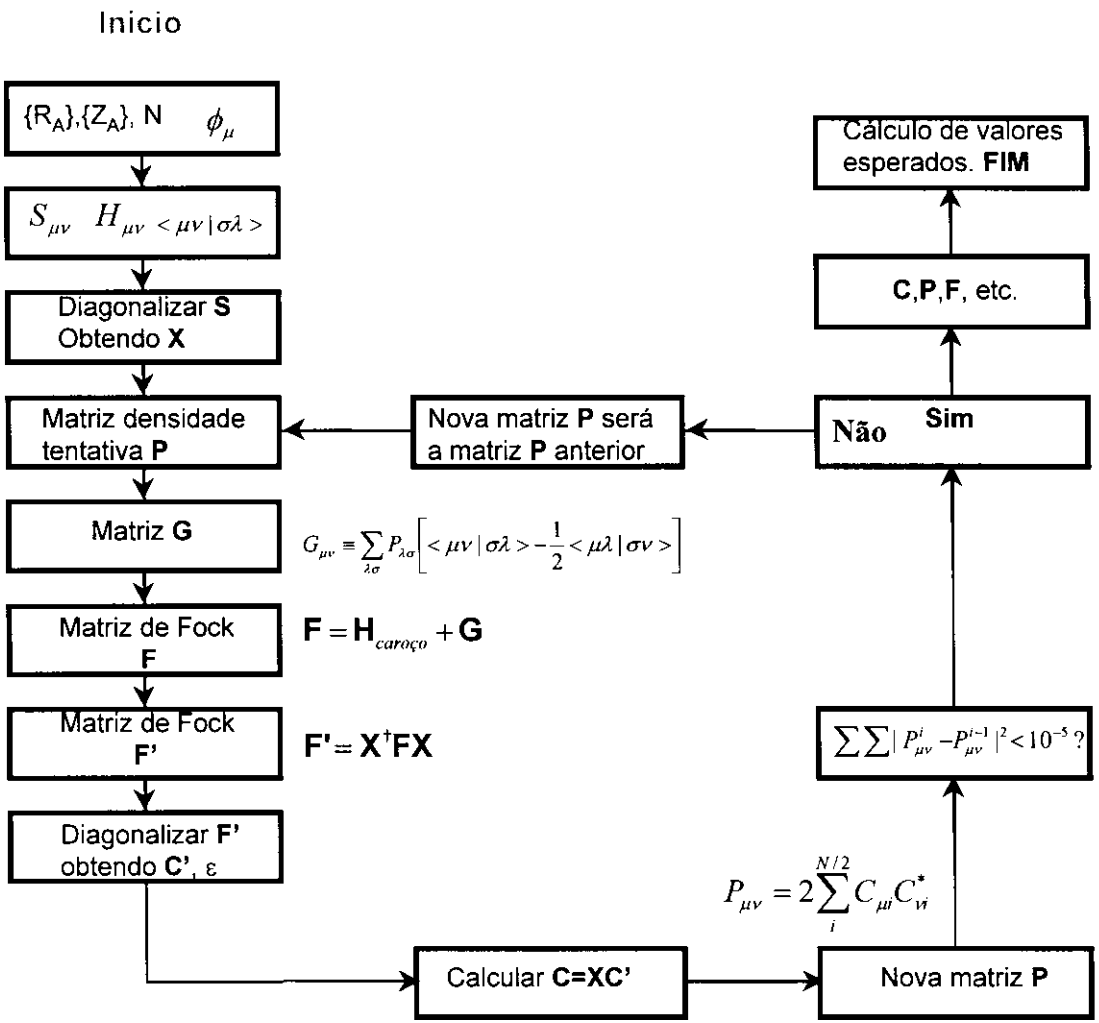


FIGURA A.1: O algoritmo computacional para a implementação do método de Hartree-Fock-Roothaan.

O Cálculo *ab initio* (primeiros princípios) descrito até o momento torna-se muito custoso do ponto de vista computacional quando aplicado a moléculas de médio e grande portes. O número de integrais de dois elétrons (elementos de $G_{\mu\nu}$) torna-se muito grande e o seu cálculo inviável para os atuais recursos computacionais disponíveis.

A forma de contornar este problema é aplicar mais aproximações ao método de Hartree-Fock-Roothaan. Dentre as diversas possibilidades, existe a técnica de incorporar parametrização empírica aos valores de algumas integrais que deveriam ser resolvidas. Este procedimento deu origem aos chamados métodos semi-empíricos, que vêm sendo desenvolvidos nas três últimas décadas e produzindo ótimos resultados quantitativos em muitos casos de interesse^[4].

Descreveremos a seguir alguns dos métodos semi-empíricos mais utilizados. Vale ressaltar que destes, utilizamos no presente trabalho apenas o método PM3.

A.3.1 O método de Hückel Estendido (EHT)

A teoria de Hückel estendida (EHT – *Extended Hückel Theory*) foi desenvolvida principalmente por R. Hoffmann e colaboradores da Universidade de Cornell^[5]. Este método é rápido e qualitativamente bom para investigar a estrutura de orbitais de sistemas moleculares, mas não é utilizado em procedimentos de otimização de geometria.

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

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

(A. 48)

Construímos assim uma matriz H , onde os orbitais moleculares Ψ_i e suas energias orbitais ε_i são soluções do problema de autovalores padrão $HC=SC\varepsilon$, demonstrável pelo método da variação linear de funções^[6]. A matriz C contém os coeficientes da combinação linear que descreve os orbitais moleculares em termos de orbitais atômicos (aproximação LCAO):

$$\Psi_i = \sum_{\mu=1} C_{\mu i} \phi_{\mu} ,$$

(A. 49)

e ε é uma matriz diagonal contendo as energias orbitais.

Este método requer ainda que sejam especificados o conjunto de base de orbitais atômicos $\{\phi_{\mu}\}$, e os elementos de matriz do hamiltoniano ($H_{\mu\nu}$) projetados sobre ele. As integrais de recobrimento $S_{\mu\nu}$ são computadas utilizando o conjunto de orbitais $\{\phi_{\mu}\}$.

Os métodos semi-empíricos mais comuns desprezam os orbitais atômicos e elétrons de camadas mais internas das moléculas e usam um conjunto de base mínima de orbitais de Slater. Temos, por exemplo, que os átomos de Carbono (2 elétrons na camada mais interna) são descritos por um conjunto de base mínima composto por 4 orbitais: 2s, 2p_x, 2p_y e 2p_z.

Estes orbitais, quando são descritos por funções hidrogenóides, apresentam nodos radiais inconvenientes do ponto de vista numérico. O uso de funções de Slater simplifica o problema:

$$\phi_v(r, \theta, \varphi) = N_{n,l,m} r^{n-1} e^{-\xi r} Y_{l,m}(\theta, \varphi),$$

(A. 50)

onde N é uma constante de normalização, ξ é o expoente orbital e Y é um harmônico esférico. Temos ainda que $l=0,1,2,3,\dots$, etc. e m varia entre $-l$ e l .

Os expoentes orbitais são usualmente os mesmos para os orbitais ns e np , mas diferentes para nd ^[6].

Os elementos da matriz diagonal $H_{\mu\mu}$ para um orbital atômico ϕ_μ representam aproximadamente a energia necessária para remover um elétron deste orbital, denominada energia de ionização do estado de valência (VSIE), que pode ser determinada experimentalmente. Esta aproximação é conhecida como teorema de Koopman:

$$H_{\mu\mu} = -\text{VSIE}.$$

(A. 51)

Os elementos fora da diagonal, $H_{\mu\nu}$, com $\mu \neq \nu$, representam os efeitos de ligação entre átomos, que estão relacionados proporcionalmente com as integrais de recobrimento $S_{\mu\nu}$. A chamada aproximação de Mulliken, por exemplo, trata o recobrimento diferencial da seguinte maneira:

$$\phi_{\mu}\phi_{\nu} = \frac{1}{2}S_{\mu\nu}(\phi_{\mu}^2 + \phi_{\nu}^2),$$

(A. 52)

o que pode ser usado para estimar os elementos fora da diagonal:

$$H_{\mu\nu} = \frac{1}{2}kS_{\mu\nu}(H_{\mu\mu} + H_{\nu\nu}).$$

(A. 53)

A constante de Hückel k é escolhida em função do tipo de orbitais envolvidos no cálculo.

A.3.2 Aproximações NDO

A maior dificuldade computacional do método de Hartree-Fock é o número de integrais de dois elétrons (A. 41) a serem resolvidas, que é da ordem de K^4 , onde K é o número de funções de base ou orbitais atômicos. Para contornar este problema, são utilizados métodos aproximativos para estas integrais. A aproximação mais simples que se pode empregar, despreza completamente o recobrimento diferencial (CNDO – *Complete Neglect of Differential Overlap*):

$$\phi_{\mu}\phi_{\nu} = \delta_{\mu\nu}\phi_{\mu}\phi_{\mu}.$$

(A. 54)

O sufixo NDO (*Neglect of Differential Overlap*) foi utilizado subsequente em outros métodos que utilizam parte deste conceito. Uma segunda aproximação básica consiste em desprezar o recobrimento

diferencial diatômico, o chamado NDDO (*Neglect of Diatomic Differential Overlap*):

$$\phi_{\mu}^A \phi_{\nu}^B = \delta_{AB} \phi_{\mu}^A \phi_{\nu}^A.$$

(A. 55)

Estas aproximações levam a uma grande simplificação das equações gerais de Hartree-Fock-Roothaan. O negligenciamento intermediário do recobrimento diferencial foi incorporado em outros métodos como o INDO (*Intermediate Neglect of Differential Overlap*), ZINDO (*Zerner INDO*), MNDO (*Modified NDO*), AM1 (*Austin Method 1*), PM3 (*Parametric Method 3*), etc.

O método INDO apresenta diversos tipos de parametrização, sendo que uma das mais populares deu origem ao MNDO, este, através de melhorias na forma de parametrização de algumas integrais e nos parâmetros utilizados deu origem aos métodos mais recentes AM1 e PM3.

A.3.3 Aproximações MNDO, AM1 e PM3

O MNDO é um método semi-empírico, sofisticado, autoconsistente e adequado ao estudo de geometrias moleculares e caminhos de reação.

Os orbitais moleculares da camada de valência (Ψ_i) são representados por combinações lineares de um conjunto de base mínima de camada de valência (ϕ_{ν}) na aproximação LCAO (A. 49). Os coeficientes da expansão são obtidos pelas equações de Roothaan, que na aproximação NDDO é escrita como:

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

(A. 56)

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

A energia eletrônica E_{el} será dada por:

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

(A. 57)

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

Utilizando-se a notação ϕ_{μ} e ϕ_{ν} para os orbitais atômicos centrados em um átomo A e ϕ_{λ} e ϕ_{σ} para aqueles centrados no átomo B ($A \neq B$ genéricos), os elementos da matriz de Fock no formalismo MNDO serão dados por:

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

(A. 58)

para os termos da diagonal e

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

(A. 59)

para os termos fora da diagonal. Temos ainda:

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

(A. 60)

onde:

- $P_{\mu\nu}$ é a matriz de ordem de ligação (*bond order*);

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

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

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

(A. 61)

O calor de formação molecular C_F será dado por:

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

(A. 62)

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

Na aproximação MNDO, os vários termos da matriz de Fock e as repulsões $E_{AB}^{caroço}$ não são determinados analiticamente, mas sim determinados através de dados experimentais ou de expressões semi-empíricas que contêm parâmetros ajustáveis aos dados experimentais. A introdução destes parâmetros compensará, em princípio, as deficiências

básicas da função de onda total (que despreza a correlação eletrônica) e erros inerentes às simplificações implementadas pelo método.

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

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

(A. 63)

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

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

(A. 64)

onde $f_i(R_{ij})$ é uma expressão semi-empírica apropriada.

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

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

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

(A. 65)

onde:

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

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

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

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

Por fim, o método **PM3** difere do AM1 somente nos valores de parâmetros que foram derivados pela comparação de um número muito maior de propriedades moleculares computadas versus experimento^[7]. Existe um consenso de que o AM1 e o PM3 são ambos melhores que o MNDO, mas um longo debate sobre os méritos relativos desses métodos foi e vem sendo travado^[8,9]. No entanto, o método PM3 mostra-se melhor que o AM1 em muitos casos, principalmente quando interações de pontes de hidrogênio estão presentes^[7,8,9,10,11].

- ¹ Szabo, A. e Ostlund, N.S., "*Modern Quantum Chemistry*", McGraw-Hill, New York **1989**, pág. 64.
- ² Dewar, M.J.S.; Zebisch, E.G.; Healy, E.F. e Stewart, J.J.P., *J. Am. Chem. Soc.*, **1985**, 107, 3902.
- ³ Pople, J.A. e Beveridge, D.L., "*Approximate Molecular Orbital Theory*" McGraw-Hill, New York **1970**.
- ⁴ Michael, J.S.; Dewar, M.J.S. e Thiel, W., *J. Am. Chem. Soc.*, **1977**, 95, 15.
- ⁵ Hoffman, R., "*An Extended Hückel Theory*", *J. Chem. Phys.* **1963**, 39(6), 1397-1412.
- ⁶ Levine, I.N., "*Quantum Chemistry*", Prentice-Hall, New Jersey **1991**, seção 8.5.
- ⁷ Stewart, J.J.P., "*Reviews of Computational Chemistry*", Lipkowitz, K. e Bouyd, D.B., Eds. VCM Publishers, New York, **1990**.
- ⁸ Zerner, M.C., "*Reviews in Computational Chemistry II*", Lipkowitz K.B. e Boyd D.B. , Eds., VCH Publishers, **1991**, p.313.
- ⁹ Scano, P. e Thompson, C., *J. Comp. Chem.* **1991**, 12, 172.
- ¹⁰ Zoos, Z.G. e Galvão D.S., *J. Phys. Chem.* **1994**, 98, 1029.
- ¹¹ Galvão, D.S.; Zoos, Z.G.; Ramasesha, S. e Etemad, S., *J. Chem. Phys.* **1993**, 98, 3016.



Apêndice B

Rotinas DDE para o HyperChem

Projeto .vbp de exemplo

Como mencionamos no capítulo 4 desta tese, utilizamos rotinas especiais para a efetuar a comunicação entre os programas HyperSpin^[1] e HyperChem^[2]. Tais rotinas foram implementadas através do protocolo DDE (*Dynamic Data Exchange*) do Windows.

Neste apêndice pretendemos exemplificar o uso de algumas rotinas DDE através de um programa exemplo bastante simples: uma caixa de texto será conectada ao HyperChem e, ao clique de um botão, conterá o valor da energia para a molécula que estiver no seu *workspace*.

Através deste exemplo simples, o usuário será capaz de desenvolver rotinas bem mais complexas para uma perfeita e eficiente troca de dados entre os dois programas.

Este pequeno projeto foi feito em Visual Basic versão 5.0 (VB 5.0), mas funcionará em qualquer versão anterior e muito provavelmente em qualquer versão futura com pequenas alterações.

A figura B.1 a seguir mostra parte do ambiente de desenvolvimento proporcionado pelo VB 5.0. Nela podemos ver ainda um pedaço de nosso projeto exemplo, com um formulário (Form1), uma caixa de texto (Text1) e um botão (Command1) necessários para o teste do programa. Pode-se alterar suas formas e

suas propriedades (tamanho, *captions*, etc.), mas para que o exemplo funcione, mantenha seus nomes como no exemplo (Form1, Text1 e Command1).

Para desenhar estes componentes, insira um novo formulário em um projeto vazio e utilize as ferramentas para desenhar caixas de texto e botões.

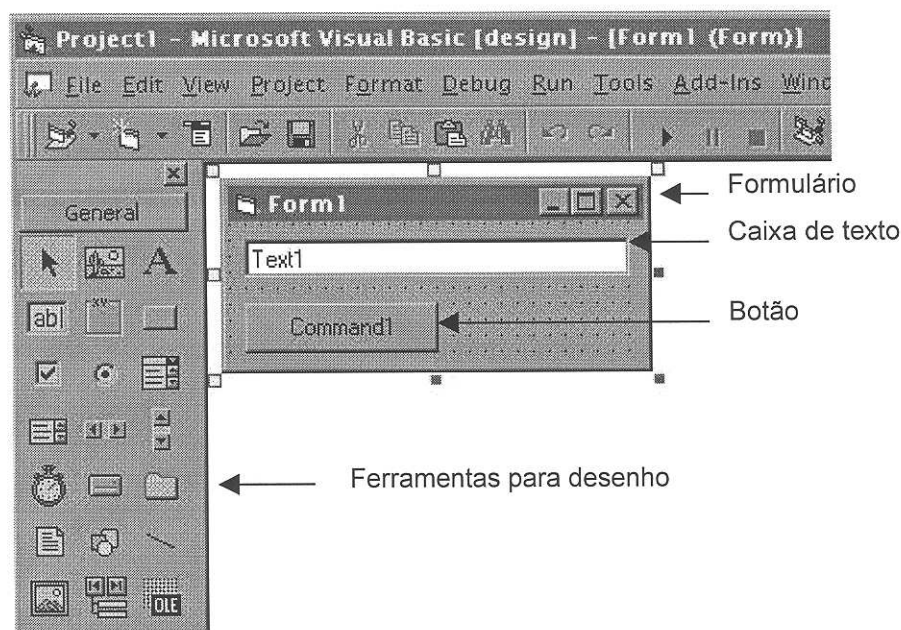


FIGURA B.1: Ambiente de desenvolvimento de programas do VB 5.0 contendo o projeto de exemplo: Project1.vbp.

Introduzindo código aos objetos

Precisamos introduzir código de programa aos componentes desenhados e para tal basta clicar duas vezes sobre eles. Neste caso vamos fazê-lo apenas para o formulário e para o botão, pois eles serão os responsáveis pela conexão da caixa de texto com a variável de energia total do HyperChem (`total-energy` é nome que o HyperChem reconhece).

Quando o programa iniciar, o formulário será aberto, e exibirá seus componentes (a caixa de texto e o botão). Aproveitamos o seu evento `Form1_Load`* para introduzir o código que conecta a caixa de texto ao HyperChem. Assim, quando o programa iniciar, a caixa será automaticamente conectada.

Dois cliques no corpo do formulário abrirão a seguinte porção do código:

```
Private Sub Form1_Load()  
End Sub
```

Colocamos então o código de conexão e o evento `Form1_Load` fica:

```
Private Sub Form1_Load()  
    Conecta Text1, "total-energy", vbLinkManual  
End Sub
```

Para VB < 5.0 use:

```
Private Sub Form1_Load()  
    Conecta Text1, "total-energy", 2  
End Sub
```

A seguir, clique no botão e introduza o seguinte código no seu evento

`Command1_Click` (rotina que será executado ao clique do mouse):

```
Private Sub Command1_Click()  
    Text1.LinkRequest  
End Sub
```

Para finalizar, falta ainda criar um módulo com as rotinas DDE necessárias ao funcionamento do programa.

As rotinas DDE

- **Rotina “Conecta”**

→ *Função*: é a rotina principal; deve ser chamada para conectar uma caixa de texto à uma variável do HyperChem.

* Os nomes em negrito correspondem aos nomes das classes ou objetos.

→ Parâmetros: *objeto*, *item*, *modo*

onde:

- *objeto*: é o objeto que será conectado. Normalmente só pode ser uma caixa de textos;
- *item*: é o item para a “conversa”. São as variáveis do HyperChem, como por exemplo `total-energy`, `rms-gradient`, etc.;
- *modo*: é o modo de conexão. Use 1 (`vbLinkAutomatic`) ou 2 (`vbLinkManual`); a diferença é que no modo 1 as alterações da variável referenciada por *item* serão automaticamente notificadas ao *objeto*. No modo 2, estas notificações devem ser “manualmente” requeridas, via evento comando `LinkRequest`.

→ *Posição no projeto*: pode ser colocada em um módulo qualquer. Para padronizar a posição das outras rotinas, vamos chamá-lo de `HyperChem.bas`;

→ *Como criar*: abra um novo módulo, chame-o de `HyperChem.bas` e digite o texto abaixo em seu interior.

```
Sub Conecta(ct1 As Control, Item As String, LinkMode As Integer)

    hyper_on = True
    Select Case MakeConnection(ct1, Item, LinkMode)
        Case 0
            ConnectState True

        Case NO_APP_RESPONDED
            If MsgBox("HyperChem doesn't seem to be running. Should I
start it?", vbQuestion + vbYesNo, App.Title) = vbYes Then
                If StartHyp() Then
                    Select Case MakeConnection(ct1, Item, LinkMode)
                        Case 0
                            ConnectState True
                            hyper_on = False
                        Case NO_APP_RESPONDED
                            MsgBox "Sorry, still can't connect.",
vbCritical, App.Title
                    End Select
                End If
            End If
        End Select
    End Select

End Sub
```

Definições de Variáveis e Procedimentos

Para que a rotina acima funcione precisamos ainda definir algumas variáveis globais e escrever os procedimentos que ela chama.

• Variáveis

→ *Funções*: apenas controle e organização;

→ *Posições no projeto*: devem ser digitas no início do módulo HyperChem.bas.

```
Global connected As Integer
Global hyper_on As Integer
Global ChemDir as string

' DDE constants
Global Const NONE = 0
Global Const NO_APP_RESPONDED = 282
Global Const DDE_REFUSED = 285
```

• Procedimentos

→ *Funções*: rotinas que separam a conexão com o HyperChem em vários passos e permitem detectar o status da conexão.

→ *Posições no projeto*: digitá-las no módulo principal HyperChem.bas

```
Function MakeConnection(ctl As Control, Item As String, clientlinkmode As Integer) As Integer

Dim ConnectTxt As Integer
ConnectTxt = CreateLink(ctl, "HyperChem", "System", Item, clientlinkmode)
If ConnectTxt = NO_APP_RESPONDED Then
    MakeConnection = NO_APP_RESPONDED
ElseIf ConnectTxt = 0 Then
    MakeConnection = 0
End If

End Function

Function CreateLink(ctl As Control, appname As String, topic As String, Item As String, LinkType As Integer) As Integer
On Error Resume Next
    ctl.LinkMode = NONE
    ctl.LinkTopic = appname & "|" & topic
    ctl.LinkItem = Item
    ctl.LinkMode = LinkType
    CreateLink = Err
End Function
```

```

Sub ConnectState(state As Integer)
    connected = state
End Sub
Function StartHyp() As Integer
On Error Resume Next
    StartHyp = (Shell(ChemDir$ & "\chem", vbMinimizedNoFocus) > 31)
    If Err <> 0 Then MsgBox "Couldn't start HyperChem", vbCritical,
App.Title
    StartHyp = True
    frmedit.SetFocus
End Function

```

Estas rotinas, quando implementadas, permitem o funcionamento correto do programa. Na função `StartHyp` mostrada acima, usa-se a variável `ChemDir`, que define o caminho para se encontrar o executável do HyperChem. Pode-se usar uma rotina que lê este valor diretamente do arquivo `c:\windows\win.ini` ou substituir um valor fixo dentro do comando `Shell` visto acima.

Rodando o programa exemplo

Para rodar o programa deve-se primeiramente carregar uma molécula no HyperChem e efetuar um cálculo qualquer. No caso de um cálculo MM+ (otimização de geometria) para a molécula Antraceno[†], deve-se obter o resultado mostrado na figura B.2 a seguir (logo após um clique no botão `Command1`).

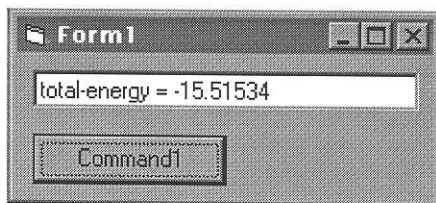


FIGURA B.2: o programa exemplo rodando e conectado ao HyperChem.

[†] No HyperChem 5.02, pode ser encontrada em `c:\hyper5\samples\anthrace.hin`.

Muitas rotinas adicionais podem ser implementadas:

- Por exemplo, temos o código necessário para desconectar um objeto, previamente conectado, do HyperChem:

```
Sub Disconnect(ct1 As Control)
Dim tempTimeOutVal
On Error Resume Next
    tempTimeOutVal = ct1.LinkTimeout
    ct1.LinkTimeout = 1
    ct1.LinkMode = NONE
    ct1.LinkTimeout = tempTimeOutVal
End Sub
```

No nosso exemplo usaríamos: `Disconnect Text1`

- Um outro exemplo: para executar um comando no HyperChem, basta utilizar um objeto já conectado e a rotina abaixo:

```
Sub Executa(comando As Variant)
On Error Resume Next
'-----=
'Simplesmente monta string DDE e executa comando
'-----=
    Form1.Text1.LinkExecute "[" & comando & "]"
End Sub
```

Se, por exemplo, quisermos efetuar um cálculo *single point* para a molécula que estiver no workspace do HyperChem, basta utilizar a seguinte linha de programa: `Executa "do-single-point"`.

- Para receber os valores das variáveis sem as “tags”, que o HyperChem envia juntamente com elas (a string “total-energy=” que vem antes do valor da energia, por exemplo), poderíamos utilizar o seguinte comando:

```
Executa "query-reponse-has-tag No"
```

Para aquele que estiver interessado em aprender mais sobre as rotinas DDE e os comandos e variáveis do HyperChem, que podem ser acessados através delas, aconselhamos o uso do programa HyperRemote descrito no capítulo 4.

Infelizmente, nenhuma documentação concisa sobre estas rotinas está presente no manuais do VB. Versões mais atualizadas do HyperChem oferecem exemplos de conexão deste programa com o VB, Visual C++, Fortran, etc, mas dentro de um nível de detalhe que deixa muito a desejar^[3].

¹ Cyrillo, M. e Galvão, D.S. "*HyperSpin conformational search program, version 2.0 for Windows 95*", em preparação.

² HyperChem™, *Molecular Visualization and Simulation program*, © HyperCube inc.
(<http://www.hyper.com>)

³ HyperChem™, *Chemists Developer Kit manual*, © **1997** HyperCube inc.