

UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA QUÍMICA
DESENVOLVIMENTO DE PROCESSOS QUÍMICOS

IMPLEMENTAÇÃO E ANÁLISE DE FERRAMENTAS DE
QUÍMICA COMPUTACIONAL APLICADA AO
DESENVOLVIMENTO DE PROCESSOS

MSc. Jefferson Ferreira Pinto

Prof. Dr. Rubens Maciel Filho
(ORIENTADOR)

Campinas, fevereiro de 2006

UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA QUÍMICA
DESENVOLVIMENTO DE PROCESSOS QUÍMICOS

IMPLEMENTAÇÃO E ANÁLISE DE FERRAMENTAS DE
QUÍMICA COMPUTACIONAL APLICADA AO
DESENVOLVIMENTO DE PROCESSOS

Tese de Doutorado apresentada à Faculdade de Engenharia Química da Universidade Estadual de Campinas, como parte dos requisitos exigidos para a obtenção do título de Doutor em Engenharia Química.

MSc. Jefferson Ferreira Pinto

Prof. Dr. Rubens Maciel Filho
(ORIENTADOR)

Campinas, fevereiro de 2006

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA ÁREA DE ENGENHARIA E ARQUITETURA - BAE -
UNICAMP

P658i Pinto, Jefferson Ferreira
Implementação e análise de ferramentas de química
computacional aplicada ao desenvolvimento de
processos / Jefferson Ferreira Pinto.–Campinas,SP:
[s.n.], 2006.

Orientador: Rubens Maciel Filho.
Tese (doutorado) - Universidade Estadual de
Campinas, Faculdade de Engenharia Química.

1. Modelagem de dados. 2. Processos químicos. 3.
Métodos de simulação. 4.Computação de alto
desempenho. 5. Programação paralela (Computação). 6.
Termoquímica. I. Maciel Filho, Rubens. II.
Universidade Estadual de Campinas. Faculdade de
Engenharia Química. III. Título.

Título em Inglês: Implementation and analysis of computational chemistry tools applied to
the process development

Palavras-chave em Inglês: Modeling, Simulation, Computational chemistry, Parallel
computing

Área de concentração: Desenvolvimento de Processos Químicos

Titulação: Doutor em Engenharia Química

Banca examinadora: César Benedito Batistella, João Sinézio de Carvalho Campos ,
Ronaldo Luiz Mincato e Valdir Apolinário de Freitas

Data de defesa: 22/02/2006

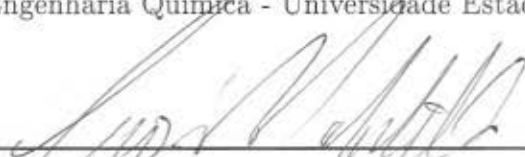
Programa de Pós-graduação: Engenharia Química

Tese de Doutorado "Implementação e Análise de Ferramentas de Química Computacional Aplicada ao Desenvolvimento de Processos", defendida por Jefferson Ferreira Pinto e aprovada em 22 de fevereiro de 2006 pela banca examinadora constituída por:



Prof. Dr. Rubens Maciel Filho - Orientador

Faculdade de Engenharia Química - Universidade Estadual de Campinas



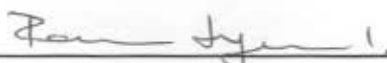
Pesq. Dr. César Benedito Batistella

Faculdade de Engenharia Química - Universidade Estadual de Campinas



Prof. Dr. João Sinézio de Carvalho Campos

Faculdade de Engenharia Química - Universidade Estadual de Campinas



Prof. Dr. Ronaldo Luiz Mincato

Faculdade de Química - Pont. Universidade Católica de Campinas



Pesq. Dr. Valdir Apolinário de Freitas

Centro de Pesquisas de Paulínia - Rhodia S.A.

200730526

Atesto que este exemplar corresponde à versão final da Tese de Doutorado:

“Implementação e Análise de Ferramentas de Química Computacional
Aplicada ao Desenvolvimento de Processos”

defendida por Jefferson Ferreira Pinto em 22 de fevereiro de 2006.

Prof. Dr. Rubens Maciel Filho - Orientador

À minha avó Anna, cuja ausência sempre é sentida.

À Deus, por tudo, e principalmente por ter colocado minha esposa e meu filho em minha vida. Sem eles eu jamais teria forças para continuar existindo. À eles também dedico esta tese, por tudo que fizeram e pela paciência que tiveram.

Agradecimentos!

É impossível citar todas as pessoas que contribuíram, direta ou indiretamente, para o desenvolvimento desta tese, mas mesmo assim gostaria de expressar os meus mais sinceros agradecimentos:

- Aos amigos Rubens Maciel Filho e Maria Regina Wolf Maciel, pela confiança e ajuda, o meu muito obrigado de forma especial;
- Aos professores com os quais tive o prazer de conviver durante o doutorado, que muito contribuíram na minha formação profissional e acadêmica;
- Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo auxílio financeiro;
- Aos colegas de classe, de laboratório e convívio geral, antigos e novos, pela amizade e companheirismo;
- Aos funcionários da UNICAMP, em especial aos da Faculdade de Engenharia Química e do Sistema de Bibliotecas;
- Aos Técnicos de Informática Rafael e Rodrigo (*in memoriam*);
- Aos voluntários, programadores ou não, que contribuem para a existência e disponibilização gratuita dos inúmeros *softwares* utilizados ao longo desta tese;
- E por último, mas não menos importante, aos meus pais, por sempre estarem presentes em minha vida.

Prefácio

Este trabalho foi escrito com o objetivo prático de introduzir o leitor no universo da química computacional, apresentando os conceitos quânticos, clássicos e correlacionais. Seu principal objetivo foi projetar, implantar e utilizar diversas ferramentas computacionais para obtenção de propriedades de interesse para o desenvolvimento de processos químicos.

Nele está apresentada uma breve introdução teórica sobre os assuntos envolvidos, uma vez que um maior aprofundamento pode ser obtido junto aos inúmeros livros sobre o assunto, sendo que diversos deles estão citados nas referências bibliográficas.

Basicamente, foram delineadas as ferramentas computacionais disponíveis, apresentando como elas funcionam, e indicando brevemente como estas ferramentas computacionais podem auxiliar no desenvolvimento de projetos envolvendo modelagem, simulação e otimização de processos químicos, foco de atuação do Laboratório de Otimização, Projetos e Controle Avançado, onde esta tese foi desenvolvida.

E como o assunto é computacional, foi incluído também um projeto conceitual para criação de um ambiente computacional paralelo de baixo custo e alto desempenho (*cluster de PC's*), e também um comparativo de desempenho entre diferentes sistemas operacionais. Desta forma foi possível idealizar e avaliar qual o melhor ambiente computacional para os objetivos almejados neste trabalho.

Espero que gostem!

Campinas, fevereiro de 2006

Jefferson F. Pinto

IMPLEMENTAÇÃO E ANÁLISE DE FERRAMENTAS DE QUÍMICA COMPUTACIONAL APLICADA AO DESENVOLVIMENTO DE PROCESSOS

Resumo

As indústrias vêm mudando profundamente nos últimos anos, principalmente para redução de consumo energético, melhoria na qualidade dos produtos e adequação às leis ambientais. Estas mudanças podem ser auxiliadas pelas técnicas de modelagem e simulação, incluindo o detalhamento do modelo em nível atômico, quando então recebe o nome de química computacional. Diversas ferramentas abrangendo todas as áreas de química computacional, em sua maioria gratuitas ou de domínio-público, foram implementadas em um microcomputador e analisadas para aplicação no desenvolvimento de processos. Foi analisado também o desempenho computacional em função do sistema operacional, que apresentou diferenças no desempenho de até 353% para cálculos de ponto flutuante, 18% para acesso a memória RAM e 67% para acesso a disco. Para melhorar o desempenho computacional, foi elaborado o projeto de um ambiente computacional paralelo de alto desempenho, no qual o custo ficou limitado à aquisição de hardware, de fácil disponibilidade no mercado, sendo que os softwares utilizados são gratuitos ou de domínio-público.

IMPLEMENTATION AND ANALYSIS OF COMPUTATIONAL CHEMISTRY TOOLS APPLIED TO THE PROCESSES DEVELOPMENT

Abstract

Industries are changing in the last years, mainly for reduction of energy consumption, improvement in the product quality and adequacy to the environmental laws. These changes can be assisted by the modeling and simulation techniques, including the detailing of the model in atomic level, when then it receives the name of computational chemistry. Several tools enclosing all the areas of computational chemistry, in its mainly free or of public-domain, had been implemented in a microcomputer and analyzed for application in the processes development. The computational performance in function of the operational system was also analyzed, that presented differences in the performance of up to 353% for floating-point calculations, 18% for access the RAM memory and 67% for access the hard disk. To improve the computational performance the project of high performance computer system was elaborated, in which the cost was limited the acquisition of the hardware, of easy availability in the market, being that software used is free or of public-domain.

Sumário

Lista de Figuras	xxii
Lista de Tabelas	xxiv
Lista de abreviações	xxvi
Lista de símbolos	xxxi
1 Introdução	1
1.1 Objetivos	2
1.2 Organização da tese	3
2 Revisão bibliográfica	5
2.1 Definição da Química Computacional	5
2.2 Química quântica	13
2.2.1 Origem	13
2.2.2 A formulação da mecânica quântica	14
2.2.3 Solução pelo método de Hartree-Fock	17
2.3 Simulação molecular	25
2.3.1 Mecânica estatística	26
2.3.2 Dinâmica molecular	27
2.3.3 Simulação de Monte-Carlo	28
2.4 Métodos correlacionais	29

3	Ambiente computacional paralelo	33
3.1	Introdução	33
3.2	O ambiente computacional paralelo	34
3.2.1	Classificação das arquiteturas	34
3.2.2	Topologia de rede	35
3.2.3	Sistema operacional	37
3.2.4	Armazenamento de dados	37
3.2.5	Linguagens de programação	38
3.2.6	Aplicativos para ambiente de alto desempenho	40
3.3	O <i>cluster</i> proposto	40
3.4	Conclusões	42
4	Análise de desempenho de Sistemas Operacionais	45
4.1	Introdução	45
4.2	Fatores relevantes para o desempenho computacional	47
4.2.1	Análise de desempenho de ponto flutuante	47
4.2.2	Análise de desempenho de acesso à memória	50
4.2.3	Análise da velocidade de acesso aos discos	52
4.3	Conclusões	54
5	Métodos de química computacional	55
5.1	Introdução	55
5.2	Recursos computacionais	55
5.3	Arquivos de entrada	56
5.3.1	Coordenada interna	56
5.3.2	Coordenada cartesiana	57
5.3.3	Exemplos	58
5.4	Geometria molecular	59

5.5	Métodos quânticos	61
5.5.1	Conjunto de base	61
5.5.2	Métodos <i>ab-initio</i>	66
5.5.3	Termoquímica através de cálculos <i>ab-initio</i>	71
5.5.4	Métodos da Teoria do Funcional de Densidade	75
5.5.5	Método semi-empírico	80
5.5.6	Termoquímica através de cálculos semi-empíricos	84
5.6	Métodos clássicos	88
5.6.1	Equações de movimento	89
5.6.2	Agrupamentos (<i>Ensembles</i>)	95
5.6.3	Potenciais de força intermoleculares	103
5.6.4	Métodos de Monte-Carlo	108
5.6.5	Métodos de Dinâmica Molecular	109
5.6.6	Aplicação em equilíbrio de fase	110
5.7	Metodologia correlacional	111
5.8	Descrição do processo	111
5.8.1	Etapas do processo	112
5.9	Conclusões	114
6	Ambiente computacional criado	117
6.1	Introdução	117
6.2	Programas disponibilizados	117
6.3	Banco de dados	118
6.4	Conclusões	118
7	Avaliação das potencialidades e aplicabilidade	123
7.1	Introdução	123
7.2	Recursos computacionais	123

7.2.1	Tempo de processamento	124
7.2.2	Interação do usuário com o programa	124
7.3	Correção da vibração molecular	125
7.4	Exatidão dos resultados computacionais	126
7.5	Estruturas de Transição	127
7.6	Constante de equilíbrio	130
7.7	Auto-difusão	133
7.8	Avaliação dos métodos	134
7.8.1	Métodos <i>ab-initio</i>	134
7.8.2	Métodos semi-empíricos	135
7.8.3	Métodos de simulação molecular	136
7.9	Conclusão	136
8	Conclusões	137
	Índice Remissivo	153
	Apêndice - Programas analisados	A-1

Lista de Figuras

2.1	Relações entre experimento, modelagem e simulação computacional. . .	5
2.2	Escala dos modelos utilizados em engenharia química.	6
2.3	Esquema para cálculos termodinâmicos para moléculas em fase gasosa.	9
3.1	O <i>cluster</i> proposto.	41
3.2	Demonstração da flexibilidade dos <i>clusters</i> : o da esquerda possui cinco nós, enquanto o da direita possui acima de 1500 nós, ambos do Laboratório Sandia.	43
5.1	Ordem atômica sugerida para representação molecular por coordenadas internas: a) seleção inicial e b) sequência de seleções.	58
5.2	Molécula de ácido fórmico utilizada para representação das coordenadas.	58
5.3	Etapas de otimização geométrica molecular quando são utilizadas: a) coordenadas internas; e b) coordenadas cartesianas.	60
5.4	Aproximação da função de Slater através de três funções gaussianas. . .	62
5.5	Efeito variacional do método de <i>HF</i>	67
5.6	Possíveis resultados para várias correlações de Møller-Plesset.	68
5.7	Geometria molecular do N_2H_2 obtida através do método <i>CCSD(T)</i> no nível cc-pVQZ	70
5.8	Duplicação da caixa de simulação	95
5.9	<i>Ensemble</i> de Gibbs.	110
5.10	Modelagem QSAR	112

7.1	Coordenada de reação de hidrogenação do CO em catalisadores de Pd e de Ni	132
-----	---	-----

Lista de Tabelas

2.1	Propriedades obtidas por diferentes cálculos de <i>QC</i>	10
2.2	Massa de produtos secundários por produto desejado para diferentes indústrias químicas.	13
3.1	Tabela comparativa entre processadores da Intel e da AMD.	36
4.1	Evolução do desempenho dos processadores da arquitetura <i>x86</i>	46
4.2	Número e tipo de operações de ponto flutuante para cada série do programa <i>flops</i>	48
4.3	Desempenho em <i>MFLOPS</i> de diversos sistemas operacionais medidos pelo programa <i>flops</i>	49
4.4	Velocidade de acesso à memória medida em <i>MBytes/s</i> com o programa <i>STREAM</i>	51
4.5	Velocidade de acesso à disco medido pelo programa Bonnie.	54
5.1	Equações das teorias de Hartree-Fock e do Funcional de Densidade . . .	77
5.2	Principais <i>ensembles</i> termodinâmicos	96
6.1	Programas de estrutura eletrônica.	118
6.2	Programas de mecânica molecular.	119
6.3	Programas correlacionais.	119
6.4	Ferramentas de apoio.	120
6.5	Fornecedores de banco de dados de informações químicas	121
7.1	Estimativa do tempo computacional em função do método.	125

7.2	Fator de correção de frequência vibracional	126
7.3	Resultados computacionais <i>versus</i> resultados experimentais	128
7.4	Resultados computacionais <i>versus</i> resultados experimentais (continuação).	129
7.5	Resultados computacionais <i>versus</i> resultados experimentais (continuação).	130
7.6	Valores teóricos e experimentais do $\ln K_{eq}$ para a reação $NO_2 \rightleftharpoons NO + 1/2O_2$ em diferentes temperaturas.	132
7.7	Valores teóricos e experimentais do $\ln K_{eq}$ para a reação $1/2N_2 + 3/2H_2 \rightleftharpoons NH_3$ em diferentes temperaturas.	133
7.8	Coeficiente D_i^*) para a α Ribonuclease em água, em diferentes intervalos de tempo.	134

Lista de Abreviações

AM - *Austin Model*

AMBER - *Assisted Model Building with Energy Refinement*

ASAD - *A Self-contained Atmospheric chemistry coDe*

ASOG - *Analytical Solution of Groups*

ATM - *Asynchronous Transfer Mode*

B3LYP - *Becke 3-Parameter (Exchange), Lee, Yang and Parr*

CC - *Coupled Cluster*

CFF - *Consistent Force Field*

CHARMM - *Chemistry at Harvard Macromolecular Mechanics*

CHEAT - *Carbohydrate Hydroxys Represented by External Atoms*

CI - *Configuration Interaction*

CISC - *Complex Instruction Set Computer*

CNDO - *Complete Neglect of Differential Overlap*

CPU - *Central Processing Unit*

DFT - *Density Functional Theory*

ECEPP - *Empirical Conformational Energy Program for Peptides*

ECP - *Effective Core Potential*

EFF - *Empirical Force Field*

EP - *Effective Potentials*

FDDI - *Fiber Distributed Data Interface*

FLOPS - *Floating Point Operations per Second*

GROMOS - *Gronigen Molecular Simulation*

GTO - *Gaussian Type Orbital*

GVB - *Generalized Valence Bond*

HF - *Método de Hartree-Fock*

HPF - *High Performance Fortran*

IBM-PC - - *Arquitetura de microcomputadores pessoais desenvolvida pela International Business Machine*

IDE - *Integrated Disk Electronics*

IRC - *Intrinsic Reaction Coordinate*

IUPAC - *International Union of Pure and Applied Chemistry*

LCAO - *Linear Combination of Atomic Orbitals*

LD - *Langevin Dipoles*

LDA - *Local Density Approximation*

MBPT - *Many-Body Perturbation Theory*

MC - *Método de Monte-Carlo*

MCSCF - *Multi-Configurational Self-Consistent Field*

MD - *Molecular Dynamics*

MIMD - *Multiple Instruction/Multiple Data*

MINDO - *Intermediate Neglect of Differential Overlap*

MM - *Molecular Mechanics*

MMFF - *Merck Molecular Force Field*

MNDO - *Modified Neglect of Diatomic Overlap*

MPI - *Message Passing Interface*

MP x - Método de Møller-Plesset, onde x pode ser de 2 a 4

NMR - *Nuclear Magnetic Resonance*

PES - *Potential Energy Surface*

PFortran - *Parallel Fortran*

PM3 - *Parameterization Method 3*

PVM - *Parallel Virtual Machine*

QC - Química Computacional

QM/MM - *Quantum Mechanics/Molecular Mechanics*

QRRK - *Quantum Statistical Rice-Ramsperger-Kassel*

QSAR - *Quantitative Structure/Activity Relationships*

QSPR - *Quantitative Structure/Properties Relationships*

RAM - *Random Access Memory*

RISC - *Reduced Instruction Set Computer*

SATA - Serial ATA

SCF - *Self-Consistent Field*

SCSI - *Small Computer System Interface*

SIMD - *Single Instructions/Multiple Data*

SPEC - *Systems Performance Evaluation Cooperative*

STO - *Slater Type Orbital*

TLSER - *Theoretical Linear Solvation Energy Relationship*

TPB - *Tetraphenylborate*

UFF - *Universal Force Field*

UNIFAC - *Universal Quasichemical Functional Group Activity Coefficients*

VWN - Modelo Vosko, Wilks e Nusair

ZINDO - *Zerner Intermediate Neglect of Differential Overlap*

Lista de Símbolos

α - Expoente da função gaussiana

α_P - Coeficiente de expansão térmica

β_s - Compressibilidade adiabática

β_T - Compressibilidade isotérmica

ϕ - Coordenada esférica

ϕ_μ - μ_{simo} orbital molecular

γ_V - Coeficiente de pressão térmica

μ^{ex} - Excesso de potencial químico

ν^{HF} - Campo médio de partículas

ν_i - Frequências vibracionais

ψ - Função de onda eletrônica

σ - Número de simetria

θ - Coordenada esférica

ζ - Expoente da função de Slater,

$1/r_{AB}$ - Coordenadas de núcleos A e B

$1/r_{iA}$ - Coordenadas do elétron i e núcleo A

$1/r_{ij}$ - Coordenadas dos elétrons i e j

$\mathbf{A}$ - Número real

C_v - Capacidade calorífica molar em volume constante

$c_{\mu i}$ - Coeficiente de combinação linear

D - Coeficiente de difusão

D_i^* - Coeficiente de auto-difusão

D_x - Determinante dos spins eletrônicos x

E - Energia total

E_{cin} - Energia cinética

E_{pot} - Energia potencial

$f(r_{ij})$ - Efeito de correlação eletrônica

G - Função de Gibbs

g - Função gaussiana

H - Operador *Hamiltoniano*

h - Número de Planck, igual a $6,626 \times 10^{-34} J s$

$\hat{h}_i$ - Operador mono-eletrônico

I_w - Momento de inércia de rotação no eixo w

i - Elemento de um conjunto

j - Elemento de um conjunto

k - Elemento de um conjunto

Mb - Milhões de bytes

MHz - Milhões de ciclos por segundo

m - Um número qualquer

N - Número de orbitais moleculares

n - Um número qualquer

Q - Função de partição

R - Constante dos gases

r_m - Coordenadas dos m núcleos

r_n - Coordenadas dos n elétrons

r^{n-1} - Fator pré-exponencial

S - Entropia

X - Número real

x86 - Arquitetura de processadores utilizados em microcomputadores

Y - Número real

Y_{lm} - Função que descreve o momento angular

Capítulo 1

Introdução

As indústrias, notadamente a química, vêm sofrendo mudanças profundas nos últimos anos. Os principais aspectos destas mudanças dizem respeito ao consumo de energia, qualidade do produto, segurança operacional, depreciação de equipamentos, custos de manutenção, adequação às leis ambientais e ciclo de vida do produto no mercado consumidor.

A tendência atual, e também em um futuro próximo, deve ser aumentar a automatização na tomada de decisões e no projeto de novos produtos e processos industriais. Uma ferramenta importante para ambas as tendências são as técnicas de modelagem, simulação e otimização, que continuamente podem lidar com problemas maiores e mais complexos, graças a constante evolução computacional, tanto de *hardware* quanto de *software*, e ao desenvolvimento de novos modelos químicos [1, 2].

A modelagem e simulação de processos podem ser realizadas independente da dimensão física do sistema estudado. Entretanto, a engenharia química tradicionalmente divide os modelos em macroscópicos, microscópicos e atomísticos. Cada um destes modelos possui suas particularidades, e não é raro um modelo estar incluso em outro como estratégia para melhor detalhamento e confiabilidade. Os modelos podem ser desenvolvidos baseados em constantes físico-químicas e modelos matemáticos detalhados, de cada um dos componentes do sistema (modelos determinísticos), ou através da análise probabilística dos dados relativos às variáveis de entrada e de saída do sistema (modelos estocásticos) [3, 4].

O objetivo da otimização, que pode existir em ciências, engenharia e economia, é maximizar ou minimizar um ou mais parâmetros através de uma correta adequação dos elementos envolvidos no problema, por métodos quantitativos eficientes. A implantação

de uma boa metodologia de otimização deve considerar: **a)** uma análise crítica do processo ou planejamento; **b)** quais são os objetivos reais a serem atingidos; e **c)** as experiências anteriores.

Pode-se perceber, pelo exposto nos parágrafos anteriores, que *a priori* não existem limites quanto ao que está sendo modelado ou otimizado. É claro que, na prática, a dimensão e complexidade do modelo matemático, e conseqüente custo computacional requerido para a solução deste, são os limitantes entre o que é ou não viável para um procedimento de otimização, principalmente quando o objetivo da otimização é seu uso em tempo real [1].

Com a crescente preocupação ambiental, as indústrias vêm sendo forçadas a adotar uma política de minimizar as agressões ao meio ambiente, seja pela diminuição na demanda energética, pela reformulação ou mesmo alteração de processos químicos, ou pela substituição de produtos ambientalmente agressivos por outros mais ecológicos [5].

Outro fator importante é a crescente preocupação industrial em fornecer produtos de acordo com a demanda e o interesse do consumidor, o que aumenta ainda mais a necessidade de plantas flexíveis quanto ao volume e especificação do que é produzido. Tradicionalmente, grande parte dos trabalhos envolvidos para obtenção destes objetivos eram realizados exclusivamente em laboratórios ou em plantas pilotos. Atualmente, com o constante aperfeiçoamento de *softwares* na área de química computacional, parte dos trabalhos que antes só podiam ser realizados experimentalmente, podem ser realizados via cálculos computacionais, permitindo a exploração de uma maneira não imaginável até alguns anos atrás.

Inicialmente, a química computacional foi desenvolvida e utilizada apenas por especialistas com larga experiência no uso de ferramentas computacionais, normalmente de difícil entendimento e utilização. Hoje, ela pode ser facilmente utilizada por pesquisadores e profissionais de outras áreas de conhecimento, graças a grande diversidade de *softwares* disponíveis com interface gráfica e livros que descrevem os princípios da química computacional em todos os níveis de detalhamento, assim como uma ampla gama de periódicos voltados especificamente para este assunto.

1.1 Objetivos

O principal objetivo desta tese foi avaliar as potencialidades e aplicabilidades das diversas técnicas de química computacional, em níveis quânticos, atômicos e

correlacionais, no desenvolvimento de processos químicos.

Para obtenção deste objetivo as metas cumpridas foram:

1. Levantamento bibliográfico sobre o tema química computacional;
2. Projeto conceitual um ambiente computacional paralelo de alto desempenho e baixo custo, que pode ser utilizado em outras áreas, como simulação, otimização e controle avançado de processos;
3. Avaliação do desempenho real de microcomputadores em função do sistema operacional instalado, utilizando diversas ferramentas para este tipo de estudo;
4. Criação de um ambiente computacional disponibilizando diversos *softwares* de química computacional, com ênfase na utilização de programas de baixo custo e código aberto;
5. Avaliação, através de resultados computacionais e dados experimentais, das técnicas de química computacional para o desenvolvimento de processos químicos.
6. Sugestão de trabalhos futuros para prosseguimento da linha de pesquisa iniciada nesta tese.

A obtenção computacional de parâmetros como constantes e taxa de reação, perfis de equilíbrio de fases, valores de energia de ativação e de solvatação, além de modelos matemáticos que correlacionem propriedades ou atividades com estrutura molecular, permite ampliar a automatização na elaboração de novos projetos de processos químicos e/ou otimização de plantas químicas existentes.

1.2 Organização da tese

Para uma melhor compreensão do conteúdo deste trabalho, os tópicos abordados foram separados de acordo com a seguinte descrição:

O Capítulo 1 é o presente capítulo e introduz a tema principal da tese, os objetivos almejados e sua organização;

O Capítulo 2 apresenta a revisão bibliográfica sobre o desenvolvimento e utilização de técnicas de química computacional, para que o leitor tenha um panorama das principais potencialidades e do estágio atual de desenvolvimento;

- O **Capítulo 3** apresenta o projeto conceitual de um ambiente computacional paralelo de alto desempenho e baixo custo, que pode ser utilizado em qualquer área que exija alto poder de processamento;
- O **Capítulo 4** descreve os procedimentos e resultados da análise de desempenho de diferentes sistemas operacionais;
- O **Capítulo 5** contém a metodologia para desenvolvimento de trabalhos em química computacional, descrevendo a preparação dos arquivos de entrada, uma breve síntese teórica dos diversos procedimentos de cálculo e seus modelos teóricos;
- O **Capítulo 6** é uma breve descrição das características e potencialidades dos *softwares* que foram adquiridos, compilados e disponibilizados para uso ao longo do desenvolvimento deste trabalho;
- O **Capítulo 7** apresenta uma avaliação das potencialidades e limitações do emprego de técnicas de química computacional no desenvolvimento de processos químicos;
- O **Capítulo 8** possui as conclusões deste trabalho e as sugestões para prosseguimento desta linha de pesquisa;
- O **anexo A** contém o detalhamento dos *softwares* avaliados.

Capítulo 2

Revisão bibliográfica

2.1 Definição da Química Computacional

Todas as pesquisas realizadas em Química Computacional (*QC*) partem de um modelo matemático para simular o comportamento do objeto de estudo, que pode ser a estrutura eletrônica de um átomo ou um sistema formado por milhares de moléculas. É importante destacar que um modelo é uma representação da realidade, e o termo modelo pode possuir diferentes significados em função dos objetivos específicos do observador [4, 3]. A figura 2.1 apresenta as relações existentes entre experimento, modelagem e simulação computacional [6].

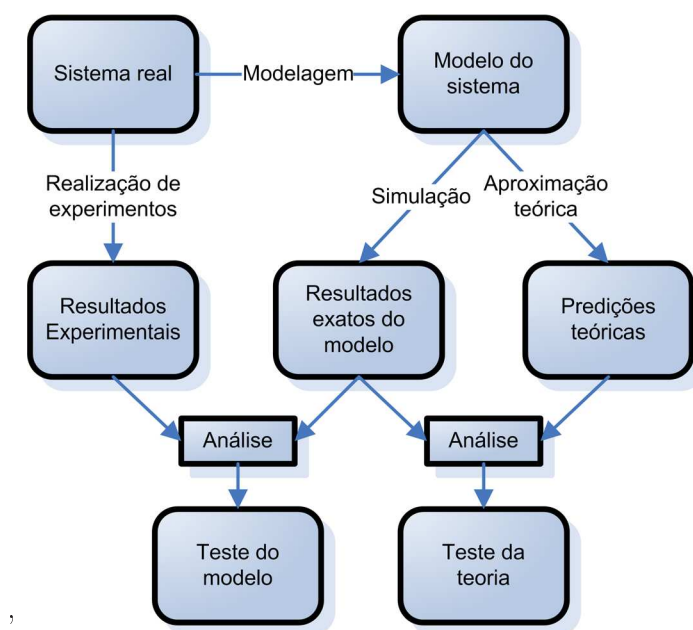


Figura 2.1: Relações entre experimento, modelagem e simulação computacional.

Tradicionalmente a engenharia química divide o universo em estudo em função das dimensões físicas envolvidas. Desta forma, temos os modelos macroscópicos, os microscópicos e os atomísticos (figura 2.2) [3]:

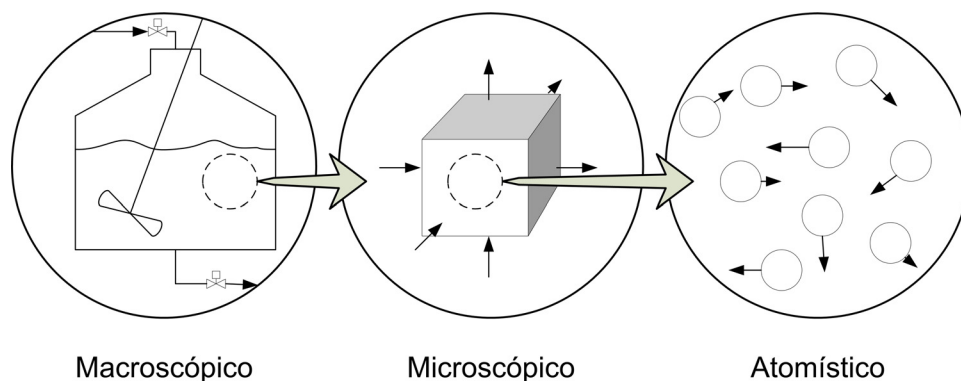


Figura 2.2: Escala dos modelos utilizados em engenharia química.

modelos macroscópicos São modelos que utilizam um conjunto de equações, denominadas de balanços macroscópicos, para descrever como a massa, o movimento e a energia de um sistema variam em função da introdução e/ou remoção destas entidades no sistema, via fluxos de entrada e/ou saída do processo. Na verdade, os modelos macroscópicos podem ser utilizados mesmo no caso de descontinuidade espacial, como no caso de uma cadeia produtiva realizada em diversas unidades fabris diferentes.

modelos microscópicos São modelos que descrevem, através de um conjunto de equações denominadas equações de mudança, o que está acontecendo em uma pequena região do processo, com o objetivo de obtenção dos perfis de velocidade, temperatura, pressão e concentração ao longo do sistema estudado.

modelos atomísticos São modelos que utilizam propriedades dos átomos, moléculas e grupos funcionais para descrever o transporte de massa, momento e energia, sendo particularmente interessante à engenharia química para descrever processos onde existam moléculas complexas, grandes faixas de temperatura e/ou pressão e reações químicas.

O desenvolvimento e solução de modelos atomísticos recebe o nome de *QC* [7]. Este campo de atuação foi apontado como uma das tecnologias chaves para a indústria química, no sentido de tornar possível o desenvolvimento de processos e produtos de forma totalmente automatizada até o ano de 2020. Ou seja, a expectativa é projetar

computacionalmente, a partir dos átomos envolvidos, processos e produtos químicos economicamente lucrativos e ambientalmente benignos [8].

A *QC* é, portanto, uma área multidisciplinar da ciência, rompendo barreiras que tradicionalmente separam as áreas de química, biologia, física, engenharia e matemática, que utiliza o computador como principal ferramenta de cálculo. O termo *QC* é normalmente utilizado quando um método matemático da química teórica é implementado em um programa de computador, e por sua vez o termo química teórica pode ser definido como uma descrição matemática da química [9].

A *QC* é um procedimento bem aceito pela comunidade científica, uma vez que muitas informações úteis podem ser derivadas dela [10]. A padronização de suas técnicas e seus termos foi sugerida pela União Internacional de Química Pura e Aplicada (*IUPAC*)¹ e os termos e acrônimos sugeridos serão utilizados ao longo desta tese [11, 12]. A *QC* também é dividida em função do detalhamento teórico utilizado em seus programas, resultando nas seguintes abordagens:

métodos quânticos ou de estrutura eletrônica Tratam de forma direta a configuração eletrônica do sistema em estudo. Tradicionalmente estes métodos são divididos em *ab-initio*², *DFT*³ ou semi-empíricos⁴;

métodos clássicos ou de mecânica molecular Empregam potenciais (campos de força) empíricos para resolver a equação de movimento do sistema em estudo. A referência passa a ser o ponto central do núcleo dos átomos, sendo que toda e qualquer influência da configuração eletrônica molecular deve estar contida no potencial empírico utilizado. São divididos em Dinâmica Molecular⁵ e Simulação de Monte-Carlo⁶;

métodos correlacionais ou de ajuste Permitem estimar propriedades físico-químicas, bioquímicas, mecânicas, etc., a partir de descritores moleculares⁷

¹Do inglês *International Union of Pure and Applied Chemistry*.

²Do latim, indicando do princípio. Estes métodos utilizam apenas abordagens teóricas em seu desenvolvimento.

³Do inglês *Density Functional Theory*. A propriedade desejada é obtida a partir do estudo da densidade da distribuição eletrônica, e não dos elétrons isolados.

⁴Utilizam correlações obtidas a partir de observações experimentais e/ou resultados obtidos por métodos *ab-initio* ou *DFT*.

⁵Os deslocamentos atômicos são estimados a partir da energia potencial e cinética.

⁶O deslocamento é realizado de forma aleatória, sendo estimada a energia global do sistema.

⁷Descritores moleculares são informações utilizadas para identificar ou quantificar uma determinada característica molecular.

e procedimentos matemáticos de ajuste de parâmetros;

métodos híbridos Resultam da união entre métodos quânticos e métodos de mecânica molecular, sendo conhecidos também como métodos *QM/MM*⁸ [13, 14, 15, 16, 17, 18, 19].

Cada um destes métodos possui suas potencialidades, vantagens e desvantagens, sendo que ao longo desta tese serão apresentados maiores detalhes.

As predições de propriedades moleculares a partir de programas computacionais tem apresentado um grande crescimento nas diversas áreas da química. Graças ao desenvolvimento de algoritmos⁹ e da constante evolução dos computadores, cálculos de química quântica se tornaram mais tratáveis e rotineiros [20, 21, 22, 23].

Os cálculos para predição de propriedades termoquímicas envolvem o uso de estatística quântica, mas estes cálculos são realizados utilizando métodos estatísticos clássicos. Os cálculos clássicos envolvem a estrutura molecular e as frequências vibracionais, enquanto os cálculos quânticos permitem obter as propriedades termoquímicas em função das contribuições rotacionais, translacionais e vibracionais da estrutura eletrônica molecular, conforme figura 2.3 [20].

Com a ajuda de análise termodinâmica, é possível obter as entalpias de reação na temperatura de interesse e derivar desta as constantes de equilíbrio. Graças a estas constantes de equilíbrio, todas as possíveis reações, que constituem o modelo cinético, podem ser consideradas na análise do mecanismo reacional [20].

A tabela 2.1 apresenta as principais propriedades que podem ser obtidas em cálculos de *QC*, sendo que muitos destes valores podem ser estimados em condições não disponíveis rotineiramente em laboratório, seja por motivos físicos ou por questões de segurança.

A exatidão na estimativa de diferentes propriedades termofísicas e termoquímicas possuem diferentes impactos na predição do custo de várias operações unitárias de uma planta química [24]. Por exemplo:

- 20% de erro no resultado de densidade leva a um erro de 16% na predição de custo/tamanho do equipamento, uma vez que a densidade é utilizada para converter as unidades de projeto (massa, moles) em unidade volumétrica;

⁸Do inglês *Quantum Mechanics/Molecular Mechanics*.

⁹Um algoritmo é uma sequência não ambígua de instruções que é executada até que uma determinada tarefa seja concluída.

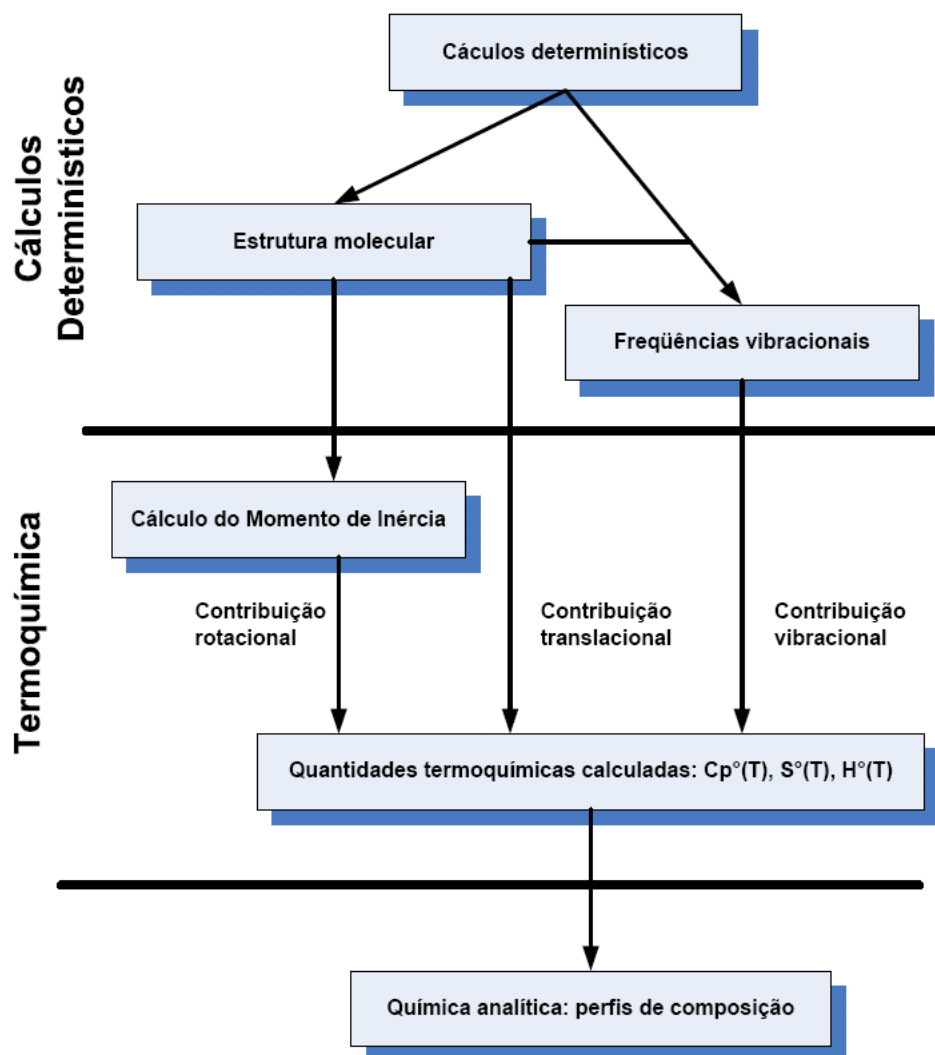


Figura 2.3: Esquema para cálculos termodinâmicos para moléculas em fase gasosa.

- 20% de erro no resultado de difusividade leva a um erro de 4% na predição do custo/tamanho do equipamento;
- 20% de erro no coeficiente de atividade resulta em erro desprezível no custo/tamanho de equipamentos quando a mistura for facilmente separável, mas pode resultar em erros acima de 200% quando a mistura for de difícil separação (volatilidade relativa $< 1,1$);
- 1 kcal/mol de erro na energia livre de Gibbs para reações à temperatura ambiente pode resultar em grandes diferenças nas predições do rendimento de reações.

Considerando-se que a química é por excelência uma ciência experimental, os resultados obtidos computacionalmente só se tornaram aceitos quando passaram

Tabela 2.1: Propriedades obtidas por diferentes cálculos de QC .

Categoria	Propriedade
Termodinâmicas:	equações de estado equilíbrio de fases constantes químicas
Termoquímicas:	energias de formação energias de reação mecanismos de reação
Espectroscópicas:	momentos dipolo espectros vibracionais espectro Raman
Mecânica:	<i>stress</i> mecânico elasticidade
Transporte:	viscosidade condutividade térmica coeficiente de difusão
Morfológicas:	estrutura cristalina contorno de sítios ativos

a possuir o mesmo nível de incerteza que os resultados experimentais.

Entre as principais áreas beneficiadas pelos estudos de QC podemos citar [25]:

físico-química Cálculo de propriedades termodinâmicas de gases e líquidos, interpretação de espectros vibracionais, determinação de propriedades estruturais, obtenção de energia conformacional e de barreira de energia rotacional, caracterização de estados de transição, energia de solvatação, etc.;

química orgânica Cálculo da estabilidade relativa de isômeros, caracterização de intermediários para o estabelecimento e entendimento de mecanismos de reação, obtenção e análises de espectros de Ressonância Magnética Nuclear (NMR)¹⁰, etc.;

química inorgânica Cálculos aplicáveis a teoria do campo ligante, estudo de íons de complexos de metais de transição, catálise homogênea e heterogênea, processos de adsorção, etc.;

¹⁰Do inglês *Nuclear Magnetic Resonance*.

química analítica Interpretação de métodos analíticos espectroscópicos (frequências e intensidades), estudo de compostos de interesse ambiental, etc.;

bioquímica análise conformacional de grandes sistemas moleculares de importância biológica, estudo de interação enzima-substrato, planejamento racional de fármacos, etc.;

química ambiental desenvolvimento de novos processos com o objetivo de atingir o conceito de *Green Chemistry*¹¹ [5, 25].

Muitos processos químicos, como reações de adições de radical em múltiplas ligações, oxidação de compostos nitrogenados ou hidrocarbonetos, reações fotoquímicas, reações de oxidação térmica (combustão), envolvem complexas reações em cadeia onde são produzidos intermediários de alta energia, sendo muito difícil medir experimentalmente os parâmetros cinéticos envolvidos. A cinética é realizada com a avaliação de constantes de reação, usando métodos nos quais a energia de ativação é estimada a partir de dados termodinâmicos e as frequências vibracionais são ajustadas para corresponder aos dados experimentais. As barreiras de ativação, as frequências vibracionais e as estruturas no estado de transição são difíceis de prever. É muito importante calcular a superfície de energia potencial (*PES*)¹². Quando necessária, a identificação de cada estrutura do estado de transição pode ser estimada através das coordenadas intrínsecas de reação (*IRC*)¹³ [16].

A pesquisa por novos compostos com propriedades controladas pode ser realizada por um ou mais dos seguintes procedimentos [26]:

experimentação intuitiva Utiliza informações conhecidas e disponíveis para o projeto de novos compostos e também para testar métodos de síntese já conhecidos;

seleção auxiliada por computador Utiliza um sistema inteligente para seleção de compostos através de técnicas de reconhecimento de padrões, sendo que neste caso as fórmulas dos compostos são representadas como um conjunto de descritores; e

¹¹Nomenclatura sugerida pela *IUPAC* que pode ser definida como o projeto e implantação de processos e produtos químicos com o objetivo de reduzir e/ou eliminar o uso ou geração de substâncias químicas nocivas à saúde humana e ao ambiente.

¹²Do inglês *Potential Energy Surface*.

¹³Do inglês *Intrinsic Reaction Coordinate*.

projeto auxiliado por computador Utiliza diretamente descritores e métodos correlacionais para o projeto de novos compostos.

Um dos desafios da química dos materiais é projetar estruturas com propriedades ótimas de algumas propriedades moleculares sujeitas a restrições práticas, como estabilidade térmica e bandas óticas. Trabalhos recentes para a construção de modelos quânticos consistentes e preditivos, além do aumento de informações disponíveis, têm levado a relações estruturas-funcionalidades úteis em muitas áreas. O grande problema é que estas relações são limitados a pequenas famílias estruturais e/ou funcionais [27].

O desenvolvimento de correlações estrutura-funcionalidade a partir de informações quânticas pode ser realizada de três maneiras distintas, a saber [27]:

estratégia direta A estratégia direta para o projeto de novos compostos começa com a estrutura molecular. Através de cálculos de elétrons independentes é possível construir a matriz eletrônica associada a esta estrutura. A diagonalização desta matriz gera n^{14} autovalores de energia ϵ_k com as correspondentes funções de onda $\psi_k = \sum_i^N C_i^k \phi_i$. Com o aumento do número de grupos aumenta a dificuldade de identificação dos mínimos na combinação linear dos orbitais atômicos (*LCAO*)¹⁵, podendo exigir o emprego de técnicas de amostragem estatística;

estratégia indireta A propriedade de interesse é otimizada através da manipulação dos coeficientes da função de onda, mantendo fixo o espectro de energia. Os coeficientes das funções de onda do *LCAO* para cada um dos orbitais moleculares são construídos a partir de um conjunto ortonormal das funções de onda, e a otimização destes coeficientes permite a obtenção dos autovetores ótimos da equação $\psi_k = R \cdot P_R \{\phi_i\}$;

estratégia mista A propriedade desejada é a variável a ser otimizada: a matriz dos elementos e os coeficientes da função de onda podem ser obtidos para a propriedade de interesse a partir da auto-consistência, e o valor ótimo da propriedade é obtida através de procedimento de minimização do valor desta propriedade respeitando a auto-consistência.

A indústria química, em muitos casos, produz uma quantidade maior do que o aceitável de rejeitos, e a *QC* permite estudar novas rotas de reação com o objetivo

¹⁴Um número qualquer

¹⁵Do inglês *Linear Combination of Atomic Orbitals*.

de minimizar a formação de produtos indesejáveis. Por exemplo, para cada quilo de metil-hidroquinona¹⁶ produzido são gerados 18 quilos de rejeitos de compostos aromáticos e sais de cromo, além da utilização de uma quantidade significativa de solventes orgânicos no processo de purificação. Estima-se que a razão entre a massa de produtos secundários pela massa de produto desejado, representado como fator E , aumenta significativamente com a especialização do produto final, conforme exemplificado na tabela 2.2) [28].

Tabela 2.2: Massa de produtos secundários por produto desejado para diferentes indústrias químicas.

Atividade Industrial	Produção (ton/ano)	Fator E
Refino de petróleo	Dezenas de milhões	0, 1
Química pesada	Centenas de milhões	1a5
Química fina	Mihares	5a50
Química farmacêutica	Centenas	25a100

2.2 Química quântica

2.2.1 Origem

No final do século *XIX* os cientistas acreditavam que a maioria das leis fundamentais já haviam sido esboçadas. Eles esperavam apenas pequenos refinamentos que pudessem levar a um maior grau de precisão. Contudo ainda existiam algumas observações que não podiam ser explicadas através da mecânica clássica (definição atual do conjunto teórico introduzido no século *XVII* por Isaac Newton) e que levaram a formulação de uma nova teoria, a mecânica quântica. Podemos citar, por exemplo, a chamada “catástrofe da luz ultravioleta” e a observação do “efeito fotoelétrico” como problemas sem solução na mecânica clássica [29].

A catástrofe da luz ultravioleta

O corpo negro é um objeto idealizado que absorve e emite todas as frequências. A teoria clássica pode ser usada para derivar a equação que descreve a intensidade do corpo negro como função da frequência em uma temperatura fixa, resultando na lei de

¹⁶Matéria-prima para a obtenção da vitamina K-3.

Rayleigh-Jeans. Contudo, esta lei se aplica apenas para baixas frequências, divergindo com um valor do quadrado da frequência, fenômeno este conhecido como catástrofe da ultravioleta.

Max Planck formulou uma nova explicação para a radiação do corpo negro em 1900, assumindo que a intensidade deste deveria ser proporcional a múltiplos inteiros da frequência, obtendo uma equação similar a lei de Rayleigh-Jeans, mas com um parâmetro ajustável h ¹⁷. Contudo, Planck não conseguiu explicar esta discretização nos valores de energia.

Efeito fotoelétrico

Em 1886 Reinrich Hertz descobriu que a luz ultravioleta poderia causar a ejeção de elétrons na superfície de um metal. De acordo com a teoria clássica, a intensidade da luz determinaria a velocidade, e conseqüente energia cinética eletrônica. Os experimentos demonstravam que na verdade a energia cinética dos elétrons dependia da frequência da luz incidente, e que a intensidade influenciava apenas no número de elétrons ejetados.

O problema do efeito fotoelétrico foi resolvido por Einstein em 1905, assumindo a equação desenvolvida por Planck, o que também levava à discretização energética, com o valor de h próximo ao obtido por Planck.

Dois anos depois, Einstein demonstrou que as vibrações atômicas também possuem características discretas. A teoria clássica prediz que o calor específico molar em volume constante (C_V) de um cristal é $3R$, onde R é a constante dos gases. Esta predição é válida para altas temperaturas, mas para baixas temperaturas o valor encontrado para C_V tende para o zero. Einstein explicou este resultado assumindo que as oscilações dos átomos em torno das posições de equilíbrio são definidas de acordo com a expressão $E = nh\nu$, a mesma desenvolvida por Planck.

2.2.2 A formulação da mecânica quântica

A observação de discretização de energia começou a se firmar por volta de 1926, quando começaram a surgir os conceitos e equações fundamentais para descrever o comportamento de partículas subatômicas, originando desta forma a mecânica quântica. Nesta época foi formulada uma revisão fundamental em dois conceitos

¹⁷Este parâmetro é conhecido como número de Planck, igual a $6,626 \times 10^{-34} Js$.

básicos sobre a natureza do mundo. Pelo efeito fotoelétrico pode-se perceber que a radiação eletromagnética, que a física tratava até então como onda apenas, também apresentava algumas características de partícula. Por outro lado, experimentos com difração de elétrons, tratados até então exclusivamente como partículas, mostraram que estes poderiam se comportar como onda.

Para explicar a dualidade partícula-onda, a mecânica quântica substituiu o conceito clássico de trajetória pelo conceito de função de onda, representada por ψ , que possui as seguintes características:

- é uma função matemática que pode ser larga em uma região, pequena em outras e ainda assumir valor nulo;
- possui informações para determinar a posição e o deslocamento das partículas que descreve;
- se ψ for grande em uma posição, a partícula possui uma alta probabilidade de ser encontrada neste ponto, e se ψ for zero, a partícula não pode ser encontrada neste ponto;
- a derivada de ψ é proporcional a energia cinética das partículas que descreve.

Em 1929, Paul A. M. Dirac¹⁸ emite a seguinte opinião sobre as potencialidades da química quântica [30]:

“As leis necessárias para uma teoria matemática englobando grande parte dos fenômenos físicos e toda a química são agora completamente conhecidas. A dificuldade para aplicação destas leis é que elas se apoiam em equações matemáticas muito complexas para serem resolvidas.”

Esta opinião de Dirac, que à primeira vista pode parecer conservadora, tem motivado reflexões e discussões na comunidade acadêmica. A opinião de Dirac possui uma tradução simples para a química: informações completas sobre o comportamento de sistemas moleculares podem ser obtidas resolvendo-se a equação de Schrödinger correspondente, acoplando-se os resultados obtidos aos modelos formulados com auxílio da mecânica estatística [31].

¹⁸Um dos pioneiros da teoria quântica, tendo sido laureado com o Prêmio Nobel de Física em 1933, junto com Erwin Schrödinger.

A procura de metodologias para resolver a equação de Schrödinger tem sido um tema central para a química teórica. Trabalhos iniciais de Heitler e London em 1927 sobre a molécula de H_2^3 , Hylleraas em 1928-1930 sobre o átomo de Hélio, foram importantes para a afirmação da nova metodologia, ao mesmo tempo que descortinaram as dificuldades para abordagem de sistemas multielementares. O primeiro trabalho é ainda um marco importante para a compreensão da ligação química, ao mostrar um elo entre a nova teoria quântica e a sugestão de G. N. Lewis, formulada em 1916, sobre a formação de ligação química através do compartilhamento de elétrons [31, 32].

A partir destes avanços iniciais surge uma nova abordagem para a química, a química quântica, cujas origens estão sintetizadas em diversas publicações, com destaque para o livro '*The Chemical Bond*', escrito por Linus Pauling em 1939. As potencialidades da mecânica quântica na elucidação de propriedades moleculares, são apresentadas no livro '*Valence*' de C. C. Coulson (1952).

Tentativas para resolver a equação de Schrödinger para sistemas multieletrônicos levaram a proposição do Modelo das Partículas Independentes, que introduz simplificações que permitem o tratamento matemático para estes sistemas. As aproximações necessárias podem ser compreendidas inspecionando-se a equação de Schrödinger para um sistemas de m núcleos e n elétrons. Na formulação independente do tempo e em unidades atômicas, esta equação de autovalores é escrita na seguinte forma:

$$H\psi(r_m, r_n) = E\psi(r_m, r_n) \quad (2.1)$$

onde r_m e r_n representam as coordenadas dos m núcleos e n elétrons, respectivamente, E é a energia total, ψ é a função de onda que corresponde à solução da equação diferencial, e H é o operador *Hamiltoniano*¹⁹ para o sistema, expresso na forma:

$$H = -\sum_{i=1}^n \frac{1}{2} \nabla_i^2 - \sum_{i=1}^m \frac{1}{2m_A} \nabla_i^2 - \sum_{i=1}^n \sum_{A=1}^m \frac{Z_A}{r_{iA}} + \sum_{i=1}^n \sum_{j>i}^n \frac{1}{r_{ij}} + \sum_{A=1}^m \sum_{B>A}^m \frac{Z_A Z_B}{r_{AB}} \quad (2.2)$$

Na equação 2.2 os termos com notação $1/2\nabla_\alpha^2$ são operadores diferenciais, representantes da energia cinética. Os outros termos estão relacionados com as interações eletrostáticas que ocorrem no sistema. Na expressão para o operador

¹⁹Este termo pode assumir dois significados distintos: em Mecânica Quântica é o operador correspondente a energia total do sistema; e em Mecânica Clássica é a função que descreve o estado mecânico do sistema em termos de posições e velocidades.

Hamiltoniano observam-se a existência de termos $1/r_{ij}$ onde coordenadas dos elétrons i e j estão acoplados, termos $1/r_{AB}$ onde coordenadas de núcleos A e B estão acoplados e termos $1/r_{iA}$ onde aparece o acoplamento de coordenadas do elétron i e núcleo A .

Os elétrons e os núcleos, interpretados como partículas clássicas, possuem velocidades diferentes: o núcleo, sendo mais pesado, possui velocidade muito menor que os elétrons. Assim, sob o efeito de uma perturbação admite-se que os elétrons respondam imediatamente, sendo acompanhados posteriormente pela relaxação dos núcleos. Este conceito é traduzido na Aproximação de Born-Oppenheimer, que separa o movimento nuclear do movimento eletrônico (equação 2.3), sendo o núcleo da Teoria de Partículas Independentes.

$$H = H_{\text{núcleos}} + H_{\text{elétrons}}(r_m; r_n) \quad (2.3)$$

Com esta separação de variáveis, a função $\psi(r_m, r_n)$ pode ser fatorada em um produto do tipo:

$$\psi(r_m, r_n) = \Theta_{\text{núcleos}}(r_m) \Theta_{\text{elétrons}}(r_m; r_n) \quad (2.4)$$

O problema pode então ser tratado em dois níveis, resolvendo-se a equação eletrônica e a equação nuclear separadamente. A equação eletrônica assume uma dependência paramétrica com relação à posição dos núcleos, e sua solução é uma aproximação para a interação entre os n'_s elétrons do sistema.

Quando o sistema atinge o mínimo de energia em relação as forças de atração e repulsão de todos os elétrons em relação aos próprios elétrons e aos núcleos fixos dizemos que a molécula atingiu um Campo de Força Auto-consistente (*SCF*)²⁰. No procedimento de simulação computacional quântica o núcleo é deslocado iterativamente, seguido por novos cálculos de *SCF*, até ser encontrada a energia mínima do sistema. Este procedimento é chamado de minimização energética ou otimização de geometria e permite prever características estruturais e eletrônicas das moléculas.

2.2.3 Solução pelo método de Hartree-Fock

Descartando-se os termos $1/r_{ij}$ no operador *Hamiltoniano* (equação 2.2), a função eletrônica de um sistema contendo n elétrons pode ser fatorada na forma:

²⁰Do inglês *Self-Consistent Field*.

$$\phi_{elétrons}(r_1, \dots, r_n; r_m) = A[\phi_1(r_1), \dots, \phi_n(r_n)] \quad (2.5)$$

onde as $\phi(r_k)$ são funções de um elétron, que recebem o nome de orbitais²¹. Na equação acima A é um operador que introduz a anti-simetria, ou seja, a função eletrônica deve mudar de sinal quando duas coordenadas são intercambiadas. A anti-simetria é um requisito do princípio da exclusão de Pauli²². A função escrita neste formato recebe o nome de Determinante de Slater.

O procedimento matemático para obtenção da função $\phi_{elétrons}(r_1, \dots, r_n; R)$ foi desenvolvida por Douglas Hartree e Vladimir Fock, sendo denominado Método Hartree-Fock (*HF*). Com o método *HF* é possível obter uma solução aproximada para a equação de Schrödinger, ou seja, um produto anti-simétrico de funções de um elétron ou orbitais. Neste modelo a interação entre elétron k e demais $n - 1$ elétrons é substituída pela interação deste elétron com o campo médio gerado pelos elétrons restantes. Aplicando-se o método variacional sobre a função proposta, minimizando a energia E do sistema em relação aos spin-orbitais, obtêm-se as equações de Fock:

$$f(k)\phi_i[(r_m)_k] = \varepsilon_i\phi_i[(r_n)_k] \quad (2.6)$$

ou seja, cada função de um elétron $\phi_i[(r_n)_k]$ é uma auto função do operador de Fock, definido como:

$$f(k) = -\frac{1}{2}\nabla_k^2 - \sum_{A=1}^M \frac{Z_A}{r_{kA}} + \nu^{HF}(k) \quad (2.7)$$

Nesta equação o termo $\nu^{HF}(k)$ representa o potencial médio gerado pelos $n - 1$ elétrons do sistema. Este termo depende das demais funções ϕ_j relativas a estes elétrons, existindo portanto acoplamento neste sistema de equações.

As equações de Hartree-Fock podem ser resolvidas por uma metodologia auto-consistente: o processo é iniciado com uma solução aproximada, conhecida como de ordem zero, utilizada para calcular o potencial médio $\nu_0^{HF}(k)$; a solução destas equações

²¹Para átomos, este produto é obtido quando a regra de distribuição de elétrons na sequência $1s, 2s, 2p, 3s, \dots$, etc., é utilizada. Obtem-se a disonal do determinante de Slater.

²²O operador *Hamiltoniano* não possui dependência com o *spin* do elétron. Assim, o número quântico de *spin* não aparece na resolução da equação de Schrödinger. A dependência com a variável *spin* é obtida multiplicando-se as funções espaciais $\phi(r_k)$ por uma função de *spin*, χ ou β , obtendo-se *spin*-orbitais $\phi(r_k)\chi(k)$ ou $\phi(r_k)\beta(k)$. Por simplicidade, estes *spin*-orbitais serão representados apenas por $\phi(r_k)$.

geram uma segunda solução, de ordem um, utilizada para calcular um novo potencial médio; o processo é repetido até obtenção da auto-consistência, ou seja, quando a solução de ordem $n+1$ for igual a solução de ordem n . Mesmo com estas simplificações as equações resultantes ainda eram muito complexas para serem resolvidas rotineiramente na investigação de propriedades moleculares [33, 20, 34].

Duas contribuições importantes para o tratamento deste problema apareceram no início dos anos 50: C. C. Roothraan sugere que a função de uma partícula, o orbital, seja expandido como uma combinação linear de funções de base. Para moléculas, esta proposta é o embrião da combinação linear de orbitais atômicos para formar os orbitais moleculares; S. S. Boys mostra que as funções gaussianas possuem o comportamento matemático adequado para expandir as funções de uma partícula²³.

Utilizando estas duas sugestões, o orbital $\phi[(r_n)_k]$ pode ser escrito na forma

$$\phi_i(r_k) = - \sum_{\mu=1}^K C_{\mu i} g_{\mu}(r_k) \quad (2.8)$$

onde g_{μ} é uma função gaussiana:

$$g_{\mu}(r) = (B\alpha^L/\pi)f \exp(-\alpha r^2) \quad (2.9)$$

onde B e L são constantes e f uma função que caracteriza o tipo de função gaussiana (s , p , d , f , etc.).

Com estas sugestões a resolução das equações de um elétron obtidas com o método *HF* é transformada em uma álgebra de matrizes²⁴, ou seja, requer a determinação do conjunto de coeficientes ($C_{\mu i}$), que pode, por exemplo, ser obtido pelo método auto-consistente. Esta modificação matemática é extremamente importante: as funções $\phi_i(r_k)$ possuem um comportamento radial oscilante, o que dificulta o cálculo de integrais utilizando métodos numéricos. A expansão destas funções em bases gaussianas possibilita que as integrais sejam obtidas analiticamente, reduzindo assim eventuais problemas numéricos. Esta nova versão do problema possui uma característica muito importante: é adequada para ser resolvida utilizando computadores.

²³A teoria mostra que o comportamento assintótico correto para a função radial é do tipo $\exp(-\alpha r)$ e não o gaussiano do tipo $\exp(-\alpha r^2)$. Funções gaussianas são utilizadas pela imensa facilidade que introduzem no cálculo de integrais.

²⁴Para uma molécula, as matrizes resultantes geralmente possuem de 10^5 à 10^9 elementos, sendo que cada um destes elementos matriciais pode ser o resultado para uma integral em um espaço de seis dimensões, correspondentes as três dimensões dos dois elétrons envolvidos.

A utilidade de métodos computacionais *ab initio* para a pesquisa química está condicionado a capacidade destes em produzirem informações sobre propriedades atômicas e moleculares. Gradativamente foram desenvolvidos métodos para extrair da função de onda informações importantes para a compreensão destas propriedades [24]. A partir do reconhecimento que uma geometria de equilíbrio corresponde a um ponto de mínimo na superfície de energia do sistema, algoritmos sofisticados foram desenvolvidos para a determinação precisa destes pontos. Assim, a química quântica tornou-se capaz de fornecer informações confiáveis sobre a geometria molecular [35].

Constantes de força necessárias para a reprodução de espectros vibracionais estão relacionadas com derivadas de segunda ordem da superfície de energia potencial. Diferentes algoritmos foram desenvolvidos para calcular estas derivadas, possibilitando que a teoria produza informações valiosas para a interpretação de espectros no infravermelho e localização dos estados de transição em reações químicas. Os resultados finais obtidos são dependentes das funções gaussianas utilizadas para expandir a função de onda. Atualmente existem funções de base para quase todos os elementos da tabela periódica, adequadas para transformar o formalismo teórico em uma ferramenta útil e confiável para a pesquisa química.

O método de *HF*, ao considerar a interação entre partículas através do campo médio ν^{HF} , não inclui a correlação instantânea existente entre os elétrons, introduzindo um erro da ordem de 1% à 2%, diferença esta conhecida como energia de correlação. Para obter superfícies de energia potencial para análise do mecanismo de reações químicas, foram introduzidos métodos conhecidos como pós-*HF* para obtenção da energia de correlação. Destaca-se a introdução das correções de efeitos de muitos corpos utilizando a teoria de Møller-Plesset, que permite que superfícies de potencial sejam calculadas com erros inferiores a $2,0 kcal/mol$. Em conjunto com o formalismo da mecânica estatística, os resultados obtidos com programas de química quântica fornecem dados confiáveis para a termoquímica, inclusive em situações não acessíveis ao trabalho experimental.

Os métodos de mecânica quântica disponíveis para avaliação incluem os métodos *ab-initio* ou os métodos semi-empíricos. O custo computacional cresce aproximadamente n^4 para as técnicas *ab-initio*, onde n é o número de funções matemáticas necessárias para descrever cada orbital atômico (que são usados para gerar os orbitais moleculares), enquanto para os métodos semi-empíricos a escala é n^3 . Nos métodos semi-empíricos os orbitais moleculares são descritos com um número menor de funções matemáticas, resultado direto do emprego de dados experimentais. Devido a

estas características, os métodos semi-empíricos são mais rápidos computacionalmente que os métodos *ab-initio*, e este ganho computacional se torna mais evidente com o aumento da molécula considerada. Algumas abordagens *ab-initio* mais rigorosas, que incluem parâmetros para correção dos efeitos de correlação eletrônica, omitidos pelo método de *HF*, como os métodos *CI*²⁵, *MBPT*²⁶, *MCSCF*²⁷, entre outros, possuem um escalonamento na ordem de N^5 à N^{10} .

A utilização de métodos de cálculos baseados na química quântica computacional tem-se tornado uma ferramenta de grande importância para o estudo de problemas de estrutura molecular, estabilidade e mecanismos de reação. As técnicas fundamentais, como a teoria do orbital molecular, datam dos primórdios da teoria quântica, há quase um século. Contudo, aplicações quantitativas em larga escala só foram possíveis recentemente em função do explosivo desenvolvimento dos computadores (*hardware*) associado ao desenvolvimento e implementação de algoritmos matemáticos eficientes.

Em um estudo típico desta área, Fukuda e colaboradores (1999) desenvolveram novos mecanismos para formação de derivados de 2-Mercapto Carboxilatos (intermediário na síntese de inúmeros pesticidas, fármacos e compostos heterocíclicos) a partir de cálculos *ab-initio*. Com os resultados obtidos foi possível propor uma rota para produção de ácido 2-pirimidiniltio carboxílico em um procedimento único (*one-pot*, com rendimento de 92% [36]).

As reações de substituição nucleofílica são de grande importância sintética e mecanística. Quando grupos funcionais inerentemente de saídas pobres, como os encontrados nas aminas, éteres ou tioéteres, são incorporados em um anel heterocíclico de três membros, é observado um enorme incremento na reatividade, sendo esta reação de grande importância na rota de síntese de aminas, álcoois e tióis. Mesmo sabendo que quase todas as sínteses químicas importantes são conduzidas em fase condensada, uma investigação computacional considerando fase gasosa é atrativa, não apenas por representar um menor custo computacional, mas também por que as simulações são praticamente livres de efeitos de interação complexas com moléculas vizinhas ou de contribuições de fatores críticos na determinação de taxas de reação relativas possíveis. Banks e White (2001) estudaram a reação de tiranos com amônia e aminas através de métodos quânticos e os resultados obtidos permitiu a avaliação dos efeitos de polaridade

²⁵Do inglês *Configuration Interaction*.

²⁶Do inglês *Many-Body Perturbation Theory*.

²⁷Do inglês *Multi Configuration Self-Consistent Field*.

e estéricos dos reagentes no estado de transição [22].

Rossi e colaboradores (1999) estudaram o mecanismo de regeneração do radical nitroxil a partir de aminoéteres, sendo este um dos mais eficientes estabilizadores de polímeros conhecidos. Foi empregado o modelo híbrido de Funcional de Densidade *B3LYP*²⁸, com diferentes conjuntos de base, o que permitiu obter os mecanismos de reação mais relevantes para regeneração do radical nitroxil [37].

Os carbenos, mesmo sendo os mais energéticos intermediários de reação, apresentam alta seletividade, e evidências indicam que esta é decorrente da geometria dos carbenos. Em um estudo para alterar o comportamento químico do carbeno terc-butilmetileno, Amstrong e colaboradores (1998) estudaram as consequências de inserção de um grupo fenil por métodos quânticos e observaram redução na diferença entre barreira intermolecular de inserção $C-H$ e $C-C$, efeito este decorrente em parte da estabilização do carbeno promovida pela inserção do grupo fenil, e em parte devido a grande alteração geométrica, permitindo uma correlação desta última com a reatividade dos carbenos [21].

Stern e colaboradores (1999) utilizaram diversas ferramentas de química quântica *ab-initio* para a parametrização de um campo de força para utilização futura em métodos semi-empíricos ou mesmo clássicos, uma vez que o custo computacional destes últimos é bem menor que dos métodos *ab-initio*. Os resultados mostraram-se promissores para emprego em fase gasosa e condensada [38].

Muitas reações químicas, como adição de radicais em múltiplas ligações, oxidação de hidrocarbonetos ou compostos de nitrogênio, envolvem cadeias reacionais complexas com intermediários de alta energia, sendo comuns em reações fotoquímicas atmosféricas e de combustão. Normalmente, o estudo cinético destas reações não pode ser realizado facilmente em laboratório, sendo comum o emprego de modelos teóricos, com determinação dos efeitos energéticos e entrópicos. Com este objetivo, Bohr e Heron (1998) utilizaram diversos métodos quânticos para predição de propriedades termodinâmicas, que dependem fortemente dos parâmetros energéticos e geométricos, que em última análise determinam a estrutura eletrônica. Os resultados obtidos apresentaram boa concordância com os dados obtidos experimentalmente, mesmo para condições de pressão próximas da não idealidade dos gases [32].

Devido ao desenvolvimento de computadores mais rápidos e melhores

²⁸Maiores detalhes sobre a Teoria do Funcional de Densidade podem ser observados no capítulo sobre metodologias de *QC*.

algoritmos já é possível calcular a superfície de energia potencial (*PES*)²⁹ através de métodos computacionais. Uma vez que cadeias reacionais complexas envolvem uma ou mais etapas de isomerização, não é sempre significativo predizer suas possibilidades considerando simplesmente uma sobreposição das etapas reacionais simples. Nguyen e colaboradores (1996) utilizaram o método *QRRK*³⁰, acoplado a idéia de processo de ativação químico e *PES*, para calcular constantes de reação em função da temperatura e pressão. O processo de ativação químico descreve quando os complexos ativados energizados podem produzir situações de não-equilíbrio onde as moléculas podem adquirir energia em excesso da energia térmica colisional, podendo originar diversas reações não possíveis originalmente. As estruturas de estados de transição foram determinadas pelas Coordenadas Intrínsecas de Reação (*IRC*)³¹. Os resultados obtidos pelo procedimento estão de acordo com os resultados experimentais, e algumas informações importantes do mecanismo reacional não obtidas experimentalmente foram explicadas através de resultados teóricos [16].

Labanowski (1997) apresentou uma boa revisão sobre conjunto de bases, incluindo aspectos como Orbitais do Tipo Slater (*STO*)³², Orbitais do Tipo Gaussiana (*GTO*)³³, Primitivas Gaussianas, Contrações de Primitivas, Conjunto de Bases Segmentadas, Polarização, Funções Difusas, Contrações Gerais e Potenciais Internos Efetivos [39].

O mecanismo de reações de adição para aldeídos através de ácidos tem sido objeto de um longo debate na comunidade de química orgânica, e com o objetivo de elucidar este mecanismo reacional Bottoni e colaboradores (1997) aplicaram dados experimentais e computacionais para elucidação do mecanismo reacional de Sakurai. Os cálculos foram realizados pela Teoria do Funcional de Densidade através do programa Gaussian, versão 94. Apesar dos resultados não serem mais precisos do que os outros relatados, os autores encontraram novos estados de transição [33].

Em um estudo híbrido com métodos *ab-initio* e dinâmica molecular, Peräkylä e Kollman (1997) partiram de estruturas obtidas por espectroscopia de raios X e simularam o mecanismo catalítico da enzima Aspartilglucosaminidase. Os resultados permitiram detalhar as similaridades com outros mecanismos enzimáticos, a interação eletrônica entre resíduos e sítios ativos, e a contribuição de pontes de hidrogênio com o

²⁹Do inglês *Potential Energy Surface*.

³⁰Do inglês *Quantum Statistical Rice-Ramsperger-Kassel*.

³¹Do inglês *Intrinsic Reaction Coordinate*.

³²Do inglês *Slater Type Orbital*.

³³Do inglês *Gaussian Type Orbital*.

meio reacional aquoso [40].

Zavodinsky e colaboradores (1999) estudaram, através de método semi-empírico, diversos parâmetros de partículas de sílica nanométricas, como distância interatômica e densidade de estados eletrônicos, sendo que este último parâmetro não possui valores experimentais. Os resultados permitiram determinar as principais diferenças de partículas de sílica nanométricas e partículas de sílica maiores [41].

Engkvist e colaboradores (2000) desenvolveram um potencial semi-empírico, a partir de resultados quânticos, baseado na teoria de perturbação intermolecular, onde nas longas distâncias os parâmetros do potencial são obtidos por funções de onda, que apresentou um bom compromisso entre acurácia e tempo computacional, além de permitir facilmente a inclusão de novos parâmetros [17].

Dixon e Feller (1999) realizam o estudo da degradação do tetrafenilborato (*TPB*)³⁴, utilizado como sequestrante de *Cs* e *Sr* no tratamento de efluentes líquidos de usinas nucleares. Os resultados permitiram o entendimento físico e químico dos processos envolvidos, através das cinéticas de formação do benzeno, que é o produto de degradação [7].

Dufresne e colaboradores (2000) apresentam o planejamento de polímeros conjugados com determinadas características termocrômicas e solvatocrômicas, utilizando técnicas *ab-initio*. Os resultados obtidos computacionalmente e experimentalmente apresentam boa concordância, mostrando que as características estudadas estão intimamente relacionadas com as interações intra-moleculares estéricas e interações atrativas inter-cadeias poliméricas [42].

O estudo de cinética de reação envolve a consideração de todos os estados de transição disponíveis, sendo elaborada a superfície de energia potencial. Entre os trabalhos que utilizaram esta metodologia podemos citar:

- estudo dos mecanismos de reação do agente refrigerante e anti-chama CF_3Br com átomos de oxigênio [43];
- estudos reacionais de enolatos organometálicos de estanho (*Sn*) [44];
- estudo do mecanismo de reação de eliminação redutiva de alcanos por complexos de platina [45].

³⁴Do inglês *Tetraphenylborate*.

2.3 Simulação molecular

O termo simulação molecular, conhecido também como mecânica molecular (MM)³⁵ é o procedimento que utiliza os fundamentos da mecânica clássica para descrever o movimento dos elétrons, englobando as técnicas computacionais de Simulação de Monte-Carlo (MC)³⁶ e de Dinâmica Molecular (MD)³⁷. O objetivo primário das técnicas de MM é a obtenção de resultados exatos para problemas de mecânica estatística, ao invés de uma solução aproximada [46, 6]. As coordenadas moleculares do sistema são desenvolvidas de acordo com cálculos rigorosos das forças ou energias intra/intermoleculares, permitindo a obtenção de propriedades macroscópicas, cuja acurácia depende do modelo teórico utilizado [47].

Nas técnicas MC a determinação das propriedades moleculares é realizada independentemente da dinâmica de movimento, uma vez que durante a simulação são geradas randomicamente diferentes configurações, sendo calculadas as interações intermoleculares destas configurações. Estima-se probabilisticamente quais configurações devem ser descartadas de acordo com um critério estabelecido. Já nas técnicas MD são resolvidas as equações de movimento do sistema em função das forças intermoleculares de cada átomo e/ou molécula, permitindo desta forma a obtenção de características dinâmicas do processo, sendo esta uma técnica totalmente determinística [48]. Uma diferença importante entre estas técnicas é que a MD usa forças intermoleculares para evolução do sistema, enquanto que na MC a simulação envolve o cálculo de mudanças na energia intermolecular.

O histórico das simulações moleculares se confunde com a história dos computadores. O primeiro programa de MM de um líquido foi realizada por Metropolis e colaboradores, em 1953, com a introdução do método de Monte-Carlo. Posteriormente, Alder e Wainright, em 1957, introduziram o método de MD .

Os problemas de mecânica molecular podem ser quase que totalmente tratados como um problema de N corpos, formulado da seguinte maneira: dado N pontos de massa, que interagem de acordo com as leis de forças de atração e repulsão, resolver as N equações de movimento do sistema. Vários sistemas podem ser abordados desta forma, desde aglomerados de galáxias até sistemas químicos soluto-solvente, e a solução exata é conhecida apenas quando N é igual a 2 [49].

³⁵Do inglês *Molecular Mechanics*.

³⁶Método onde as novas configurações são geradas aleatoriamente.

³⁷Do inglês *Molecular Dynamics*. Método onde as novas configurações são geradas através de métodos de otimização.

Spinnato e colaboradores (2001) implementaram um modelo para avaliar ambientes computacionais híbridos a partir da solução de problemas de N corpos, destacando as vantagens e complexidades na utilização de um ambiente computacional paralelo [49].

McLeod (1999) apresenta simulações *MC* para reações de hidrogenação de hidrocarbonetos em catalisadores heterogêneos com o objetivo de avaliar a atividade de uma série de modelos representativos de superfície de catálise e para reações onde um dos reagentes é fornecido pelo próprio suporte de catálise, através da adsorção de um dos reagentes. Os resultados mostram que diversas variáveis dos processos, assim como suas correlações, podem ser avaliadas pela metodologia proposta, incluindo a taxa de reação, absorção de reagentes, tamanho de partícula, distribuição de tamanho de partículas, entre outras [50].

2.3.1 Mecânica estatística

A energia cinética pode ser obtida a partir dos momentos das partículas. A energia potencial é a somatória das interações interpartículas. Tipicamente, o cálculo considera todos os pares de moléculas distintas, embora outras interações multicorpos podem estar envolvidas.

A temperatura (T) pode ser obtida pela Teoria do Virial, que em termos gerais pode ser expressa como:

$$\left\langle p_k \frac{\partial H}{\partial p_k} \right\rangle = kT \quad (2.10)$$

Para N átomos, cada um com três graus de liberdade, obtém-se a relação:

$$T = \frac{2\langle E_{cin} \rangle}{3Nk} \quad (2.11)$$

Alternativamente podemos considerar T como a média das contribuições instantâneas de energia (t):

$$t = \frac{2E_{cin}}{3Nk} \quad (2.12)$$

É importante salientar que aproximadamente 90% do tempo computacional gasto em cálculos de *MD* corresponde ao cálculo de contribuições de energia e forças

entre os pares atômicos ligados ou não aos átomos de hidrogênio. Deste modo, o emprego de técnicas de computação paralela podem ser utilizadas para distribuir a lista de pares de átomos sobre um grande número de processadores, somando-se ao final de cada ciclo as contribuições para as energias e forças.

O modelo de dipolos de Langevin (*LD*)³⁸ provavelmente representa a primeira tentativa quantitativa de obtenção de estimativas de energia de solvatação de moléculas em líquidos, pois os modelos anteriores consideravam as interações entre um soluto esférico e o solvente³⁹. O método *LD* utiliza parâmetros atômicos, como raio de van der Waals, calibrados através de observações reais de energia de solvatação, sendo desta forma um meio efetivo para tratar solutos de contorno arbitrário[51].

2.3.2 Dinâmica molecular

Em dinâmica molecular o movimento dos átomos é descrito pelas leis da mecânica clássica ou newtoniana: $F_i(t) = m_i a_i = m_i \ddot{r}_i(t)$ onde $F_i(t)$, $a_i(t)$ e $r_i(t)$ são a força, aceleração e posição do átomo i no tempo t , respectivamente. A força atuante no átomo i é o gradiente negativo do potencial que é obtido a partir do mesmo campo de força empírico (*EFF*)⁴⁰ usado em mecânica molecular.

Para integrar as equações de movimento, vetores de posição são determinados utilizando estas forças em conjunto com as posições atômicas anteriores. As velocidades e a temperatura também são calculadas. Com estes ingredientes, a trajetória, que é simplesmente o histórico dos movimentos do sistema por um período de tempo de interesse, pode ser gerada.

Como a *MD* possui tanto as energias cinéticas quanto potenciais, pode ser aplicada uma mudança na energia cinética (temperatura), fazendo com que a simulação se mova pela *PES*, amostrando desta forma muitos estados conformacionais ao longo da simulação. Infelizmente, o custo computacional é bastante intenso, e os tempos de simulação são tipicamente da ordem de pico segundos (10^{-12} s), de forma que apenas processos extremamente rápidos, como rotações nas ligações de baixa energia, podem ser estudados.

De forma simplificada, pode-se dizer que os métodos de *MD* simulam os

³⁸Do inglês *Langevin Dipoles*.

³⁹Como eram atribuídos valores incertos para os raios das esferas nos cálculos de energia de solvatação, os métodos anteriores eram bastante limitados na qualidade dos resultados obtidos.

⁴⁰Do inglês *Empirical Force Field*.

movimentos do átomos para um dado período de tempo, permitindo determinar agrupamentos médios que, com uso de ferramentas da termodinâmica estatística, permitem a conexão entre os mundos microscópicos e macroscópicos.

2.3.3 Simulação de Monte-Carlo

Este método emprega os mesmos *EFF's* que as técnicas de *MD*. Inicialmente, parte-se de uma coleção de partículas e calcula-se a energia do sistema, E_1 , para a configuração inicial escolhida. Uma ou mais partículas são então selecionadas e deslocadas randomicamente para criar uma segunda configuração, sendo novamente calculada a energia do novo sistema (E_2). A nova configuração é aceita se $E_2 < E_1$, com uma probabilidade definida por $p = \exp[(E_2 - E_1)/kT]$. Um grande número (na ordem de milhões) de movimentos randômicos são realizados.

Elcock e colaboradores (1999) empregaram a *MC* no estudo de associações proteína-proteína, e os resultados obtidos mostraram que as constantes de taxa para proteínas mutantes foram corretamente reproduzidas, embora os valores absolutos tenham sido superestimados, devido principalmente ao fato da simulação desconsiderar efeitos de solvatação resultantes da interação de solvente de alta constante dielétrica com as proteínas de baixa constante dielétrica [48].

O processo de produção de compósitos poliméricos é uma das apostas para a produção de novos materiais com baixo custo, pois as propriedades variam em função da quantidade e tipo dos polímeros utilizados, influenciando desde o desempenho do composto final, passando por melhorias na condição de processamento do material. Devido estes fatores, Lyatskaya e colaboradores (1996) estudaram o planejamento de compatibilizadores para redução da tensão interfacial em compostos poliméricos, empregando a teoria de Campo de Força Auto-Consistente (*SCF*)⁴¹ e a técnica de simulação *MC*. Os resultados permitiram obter as características do material estudado, principalmente na região interfacial, sendo que somente com *MC* foi possível simular o comportamento conformacional das cadeias poliméricas individuais [52]. Nas simulações, a interface entre os dois homopolímeros foi representada por um plano permeável no meio de um cubo de vértices $32 \times 32 \times 128$, onde foi introduzido o seguinte potencial para diferenciar as duas regiões:

⁴¹Do inglês *Self-Consistent Mean Field*. Teoria onde um problema complexo pode ser reduzido para variáveis independentes que sofrem ação de um potencial de valor médio, resultado da ação do conjunto de variáveis envolvidas. As equações resultantes podem ser resolvidas numericamente, sendo auto-consistentes.

$$U_A(z) = \begin{cases} 0 & z > 64 \\ \chi_{AB} & z < 64 \end{cases} \quad (2.13)$$

e

$$U_B(z) = \begin{cases} \chi_{AB} & z > 64 \\ 0 & z < 64 \end{cases} \quad (2.14)$$

onde χ_{AB} representa a energia de interação entre os polímeros.

O potencial tem o efeito de introduzir uma penalidade energética dos elementos de campo A entrando em B e vice-versa. Como consequência, quando $\chi_{AB} > 0$, o monômero A prefere a região A e o monômero B prefere a região B . Condições periódicas foram impostas nas direções x e y , enquanto a direção z possui camadas impermeáveis em $z = 1$ e $z = 128$.

2.4 Métodos correlacionais

A modelagem correlacional, ou de adição de grupos, pode ser realizada usando procedimentos de ajuste, ou através da aplicação direta da teoria fundamental. Os procedimentos de ajuste geralmente são tentativas racionais de correlacionar estrutura molecular com propriedades físico-químicas (*QSPR*)⁴², atividade biológica (*QSAR*)⁴³ ou atividade biológica associada à estrutura espacial molecular (*QSAR-3D*).

A solução final é uma equação de regressão que considera um conjunto de características moleculares para estimativa da propriedade de interesse. Desta forma, pode-se prever com certa exatidão as propriedades de uma molécula desconhecida, ou projetar uma nova molécula com propriedades desejadas, calculando e substituindo os descritores moleculares do modelo. Não existe regra definida sobre quais descritores são usados, mas estes geralmente incluem informações sobre o tamanho e forma molecular, efeitos eletrônicos e lipoficidade.

Este tipo de modelagem requer o uso de dados existentes para criar o modelo e permitir a interpolação ou extrapolação de novas propriedades e/ou atividades de uma molécula desconhecida desta mesma classe de compostos. *QSAR* e *QSPR* têm apresentado grande influência para o desenvolvimento de novos agentes quimioterápicos.

⁴²Do inglês *Quantitative Structure Property Relationships*.

⁴³Do inglês *Quantitative Structure Activity Relationships*.

Para a elaboração de modelos correlacionais quantitativos é necessário que um número estatisticamente significativo de dados experimentais estejam disponíveis para que seja proposto um modelo de correlação com os dados obtidos computacionalmente [35].

Talvez os melhores exemplos de métodos correlacionais utilizados em engenharia química sejam os métodos *UNIFAC*⁴⁴ e *ASOG*⁴⁵ para predição de equilíbrio líquido-vapor através de contribuição de grupos [53].

Selves e colaboradores (1998) classificaram o solvente em função de interações soluto-solvente, empregando descritores moleculares e quebrando o sistema soluto-solvente em três estágios: **a)** quebra de interações solvente-solvente para abrigar o soluto (processo endotérmico); **b)** as moléculas do solvente são reorganizadas ao redor da cavidade para recebimento do soluto; e **c)** recebimento do soluto e formação de interações soluto-solvente (processo exotérmico). A metodologia permitiu a obtenção de diversas propriedades cromatográficas para os sistemas estudados [53].

Uma breve introdução do histórico do desenvolvimento de *QSAR* é apresentado por Bevan (2004), juntamente com um exemplo envolvendo Anidrase Carbônica, que é catalisador das reações de hidratação de alguns aldeídos e cetonas e das reações de hidrólise de alquil e aril ésteres [54].

Um caso real bem detalhado de um emprego de *QSAR* foi reportado pelo Instituto Sandoz de Pesquisas Médicas, com o desenvolvimento de uma nova série de agentes analgésicos, sendo que os compostos selecionados foram sintetizados e testados em cobaias, apresentando resultados de atividade condizentes com o modelo matemático elaborado [55].

Katritzky e colaboradores (1997) desenvolveram uma metodologia *QSPR* implementada no *software* CONDESSA, onde é empregada uma grande variedade de descritores relacionados à estrutura molecular, juntamente com descritores combinados. Os dados são então tratados com técnicas estatísticas avançadas como regressões múltiplas e não lineares, análise de fatores e ainda regras heurísticas, o que permite a consideração de fatores multidimensionais obtidos teoricamente, não sendo necessários dados empíricos. A potencialidade da metodologia desenvolvida fica evidente quando são obtidas correlações de respostas cromatográficas de 152 compostos utilizando apenas cinco descritores [56].

⁴⁴Do inglês *Universal Quasichemical Functional Group Activity Coefficients*.

⁴⁵Do inglês *Analytical Solution of Groups*.

Germachev e colaboradores (2004) observaram o crescente desenvolvimento de métodos correlacionais para o projeto de produtos químicos ao longo dos últimos anos, onde os novos compostos são projetados a partir de combinações e recombinações. Eles também realizaram a validação de uma metodologia correlacional para o projeto de produtos a serem utilizados como aditivos em polímeros compósitos (retardadores de fogo, estabilizantes e plastificantes), metodologia esta que pode ser expandida para uso em sistemas de inteligência artificial [26].

A Relação Teórica Linear de Energia de Solvatação (*TLSER*)⁴⁶ e *QSAR* são empregadas por Famini e Wilson (2004) no estudo a ação da enzima Citocromo P450 na intoxicação por cianeto a partir de nitrilas e também na correlação de índices de retenção de cromatografia gasosa, em ambos os casos com excelentes resultados quando comparados com dados experimentais [57].

Através de um procedimento empírico de determinação do calor de combustão de aproximadamente 140 compostos foi possível estimar a contribuição aditiva de grupos funcionais. A partir destas estimativas Walters (2001) obteve uma correlação que permite a obtenção de valores de calores de combustão com um desvio padrão de 2,5% dos valores experimentais [58].

Maranas (1996) apresenta casos de estudo para o planejamento de polímeros cujas funcionalidades estrutura/propriedades são não-lineares. Inicialmente é apresentada uma metodologia para transformar o problema original em um problema linear inteiro-misto, que permite a obtenção da classificação das diversas configurações ótimas. Os resultados mostram aplicabilidade no planejamento de diversos produtos químicos e que o sucesso depende apenas do conhecimento de interação entre os diversos grupos (monômeros, no caso deste trabalho) utilizados [59].

⁴⁶Do inglês *Theoretical Linear Solvation Energy Relationship*.

Capítulo 3

Ambiente computacional paralelo - *Cluster*

3.1 Introdução

Todas as áreas de conhecimento citadas nesta tese não possuem restrição teórica quanto à dimensão do problema estudado, mas na prática o esforço computacional e o tempo necessário para obtenção dos resultados devem ser considerados.

Uma maneira eficiente de se aumentar o desempenho computacional é a utilização de processamento em paralelo, que consiste basicamente em dividir o problema original em diversos segmentos, sendo estes segmentos tratados em diferentes unidades centrais de processamento.

Já se encontram disponíveis um grande número de ambientes computacionais paralelos comerciais (Cray Research [60], Fujitsu [61], IBM [62], etc.), mas uma tendência mundial crescente é a utilização de vários microcomputadores da linha *IBM-PC*, interligados entre si, ambiente este doravante denominado simplesmente de *cluster*¹.

Pelo exposto acima pode-se perceber que a construção de um ambiente computacional de alto desempenho e baixo custo é de grande interesse para todas as áreas que exijam alto poder de processamento, e no caso específico desta tese, para realização de pesquisas utilizando ferramentas de química computacional. Este é o assunto discutido ao longo deste capítulo.

¹Termo da língua inglesa, indicando um conglomerado de elementos similares ou relacionáveis.

3.2 O ambiente computacional paralelo

Neste trabalho foi realizado um estudo visando a criação de um ambiente computacional paralelo de alto desempenho, com utilização de componentes disponíveis comumente no mercado.

Os seguintes aspectos foram abordados e justificados para a criação de um ambiente computacional paralelo de alto desempenho:

- classificação das arquiteturas;
- topologias de rede;
- sistemas operacionais;
- sistemas de armazenamento de dados;
- linguagens de programação;
- programas aplicativos para ambiente de alto desempenho.

3.2.1 Classificação das arquiteturas

Existem dois grandes grupos de computadores paralelos a saber:

SIMD ² Onde todos os processadores trabalham efetuando instruções idênticas em diferentes blocos de dados, como nos computadores vetoriais;

MIMD ³ Onde o sistema computacional é formado por diversas Unidades Centrais de Processamento (CPU's)⁴ realizando cálculos variados. Este último grupo pode ainda ser dividido em computadores com memória compartilhada ou memória distribuída.

Os sistemas *MIMD* com memória compartilhada possuem um número determinado de processadores operando sob o controle de um fluxo de instruções simples distribuídas pelo sistema operacional, ocorrendo o compartilhamento dos módulos de memória, sendo a comunicação de dados realizada pelo próprio barramento⁵ da máquina

²Do inglês *Single Instructions/Multiple Data*.

³Do inglês *Multiple Instruction/Multiple Data*.

⁴Do Inglês *Central Processing Unit*.

⁵Termo que define o conjunto de regras de transferência de dados pelos componentes do sistema

com grande taxa de transferência de dados. No caso dos sistemas *MIMD* de memória distribuída, cada processador está localizado em um computador completo, possuindo sistema operacional e módulos de memória próprios, e a transferência de dados é realizada por uma rede externa de conexão entre os diversos computadores.

Nas máquinas de memória compartilhada, o paralelismo necessita de muitas instruções (ciclos) de *CPU*, quando comparadas com máquinas de memória distribuída, mas é realizado automaticamente pelo sistema operacional, ocorrendo de forma automática do ponto de vista do usuário. O conceito de paralelismo de memória compartilhada está presente na última geração dos processadores, onde uma unidade física é formada por várias unidades lógicas de processamento. O ganho de desempenho, neste caso, depende do aplicativo executado [63].

Os processadores comumente empregados nas estações de trabalho e microcomputadores podem ser divididos em dois grandes grupos: os processadores de arquitetura *RISC*⁶ e os *CISC*⁷. Os processadores do tipo *RISC* são os mais empregados pelos grandes fornecedores comerciais como IBM [62], SUN [64] e HP [65], apresentando características específicas de cada fabricante ou de cada arquitetura, enquanto os processadores *CISC* são os empregados nos microcomputadores da linha *IBM-PC*, que apresentam características padrões conhecidas como *x86*⁸. Estes processadores são produzidos por diversos fabricantes, como destaque para Intel [66], AMD [67] e Via [68], mas apresentam compatibilidade lógica (características como barramento e pinos de conexão podem variar entre os fabricantes ou entre gerações de processadores).

É importante salientar que um programa criado para a arquitetura *x86* deve rodar em microcomputadores da linha *IBM-PC* independente do fabricante do processador utilizado, salvo diferenças entre as gerações de processadores e do sistema operacional. A tabela 3.1 apresenta dados comparativos entre diversas características de várias famílias destes processadores [67].

3.2.2 Topologia de rede

A topologia de rede, ou arranjo envolvido na interconexão dos processadores, determina o desempenho do sistema como um todo e pode ser dividida em topologia estática ou dinâmica.

⁶Do inglês *Reduced Instruction Set Computer*.

⁷Do inglês *Complex Instruction Set Computer*.

⁸Nome genérico dada à família de processadores que seguem o conjunto de instruções baseadas no processador Intel 8086.

Tabela 3.1: Tabela comparativa entre processadores da Intel e da AMD.

	Athlon XP/MP ^a	Opteron ^b	Pentium 4 ^c	Xeon ^d
Operações por Ciclo	9	9	5	6
Instruções Inteiras	3	3	2	4
Inst. Ponto Flutuante	3	3	1	2
Decodificadores x86	3	3	1	1
Cache L1 (Kbytes)	128	128	32	16 + 16
Cache L2 (Kbytes)	512	1024	2048	1024
Barramento (MHz)	266-400	1400-2800	400-800	400-800

Processadores: ^a AMD de 32 bits, ^b AMD de 64 bits, ^c Intel de 64 bits e ^d Intel de 32 bits.

A topologia de rede estática apresenta conexões (*links*) entre os elementos de processamento, como as arquiteturas em árvores, anéis, hipercubos, etc., com a comunicação realizada através de protocolos específicos. Na topologia de rede dinâmica as comunicações entre os diversos processadores podem ser estabelecidas de duas maneiras: **a)** via canais comuns de comunicação, através de protocolos específicos ou **b)** via rede com interfaces de comunicação baseadas em protocolos padronizados, como Ethernet, *ATM*⁹, *FDDI*¹⁰, Token-ring, etc.).

No caso da interface de comunicação do tipo Ethernet é necessário ainda um equipamento que faça a distribuição dos dados de forma conveniente, que pode ser uma *switch* ou um *hub*. No primeiro as mensagens são transferidas por canais exclusivos estabelecidos entre os processadores envolvidos, suportando comunicação denominada *full-duplex*¹¹, enquanto que no último, para que uma mensagem seja enviada de um processador para outro é necessário um *broadcast*¹² para todos os computadores da rede, prejudicando o desempenho global do sistema tanto em termos de velocidade de transferência quanto de utilização de recursos computacionais.

⁹Do inglês *Asynchronous Transfer Mode*.

¹⁰Do inglês *Fiber Distributed Data Interface*.

¹¹Os dados enviados e recebidos trafegam por pares de cabos específicos.

¹²O envio ou recebimento de dados é realizado por um par de cabos, enquanto outro par é utilizado para controle de fluxo.

3.2.3 Sistema operacional

O sistema operacional, em conjunto com programas aplicativos, forma a base de toda a disponibilidade e funcionalidade do ambiente computacional. Existem diversas opções comerciais disponíveis, sendo algumas específicas para determinadas arquiteturas de *hardware*, como Microsoft Windows XP [69] ou IBM OS/2 [62], enquanto outros estão disponíveis para mais de uma arquitetura, como Apple Mac OS X [70] ou o Sun Solaris [64]. A arquitetura *x86* é a que apresenta a maior variedade de sistemas operacionais disponíveis.

Os sistemas operacionais Unix, e os baseados nestes, são padronizados com características multiusuário e multitarefa, facilidade de uso em rede e possuir um bom conjunto de utilitários, além de apresentarem grande estabilidade. Existem diversas distribuições comerciais (IBM AIX [62], HP Ultrix [65], Sun Solaris [64], Cray UnicOS [60], etc.) e de domínio público baseadas no *kernel*¹³ Linux (Mandriva [71], Debian [72], Red Hat [73], etc.) ou no *kernel* *BSD (FreeBSD [74], NetBSD [75], Darwin [76]) que rodam praticamente em todas as versões de arquitetura de *hardware* existentes.

Um sistema operacional, também disponibilizado gratuitamente, que pode ser aplicado em computação paralela é o PLAN9 [77]. Este sistema apresenta características que o distinguem de todos os outros sistemas operacionais, como por exemplo:

- recursos são nomeados e acessados como arquivos na hierarquia do sistema de arquivos;
- possui um protocolo específico, denominado 9P, para acesso destes recursos; e
- a hierarquia disjunta proporcionada por diferentes serviços são agrupadas em um único espaço hierárquico privado.

3.2.4 Armazenamento de dados

Os dispositivos de armazenamento magnéticos utilizados em microcomputadores podem ser classificados em uma das seguintes famílias:

¹³O *kernel* corresponde ao coração do sistema operacional, sendo responsável pela alocação de recursos necessários, interface com dispositivos e fluxo de processamento.

IDE ¹⁴ também conhecido como *ATA*, é o padrão de armazenamento mais utilizado no mundo, possuindo também o maior suporte por parte dos sistemas operacionais. Em sua última versão (*ATA133*) possui capacidade de transferência de dados de $133Mb/s$ ¹⁵ através de uma interface paralela;

SATA ¹⁶ é a evolução do padrão *IDE* que utiliza uma interface serial para transferência de dados. A primeira geração deste padrão possui capacidade de transferência de dados de $150Mb/s$;

SCSI ¹⁷ é o padrão recomendado para sistemas com grande demanda de acesso a discos por processos diferentes, mantendo alta confiabilidade, sendo normalmente adotada nos grandes servidores. A última especificação possui capacidade de transferência de dados de $320Mb/s$.

Nos itens acima a velocidade de transferência de dados corresponde ao processo de gravação/leitura interna dos discos, e não a velocidade de troca de dados com o restante do sistema.

Os padrões *SATA* e *SCSI* requerem normalmente a instalação de aplicativos específicos para permitir sua utilização, o que pode impedir sua utilização por sistemas operacionais menos utilizados, devido ao suporte reduzido de *hardware* nestes sistemas. O padrão *SCSI* possui um custo de aquisição bem mais elevado que os outros, sendo que o padrão *IDE* e *SATA* possuem custo praticamente equivalentes.

3.2.5 Linguagens de programação

Com o surgimento de novas arquiteturas de *hardware* houve também a necessidade de avanços das linguagens de programação para um melhor aproveitamento dos novos recursos. Basicamente, as linguagens de programação empregadas em paralelismo estão divididas em:

programação em paralelo utilização de sistemas constituídos de vários processadores alocados em uma região, com uma pequena distância entre eles, sendo que o fluxo de informação entre os diversos cálculos é realizado pelo

¹⁴Do inglês *Integrated Disk Electronics*.

¹⁵Milhões de bits por segundo.

¹⁶Do inglês *Serial ATA*.

¹⁷Do inglês *Small Computer System Interface*.

sistema operacional utilizando o próprio barramento interno (alta velocidade de transferência de dados)

programação distribuída a comunicação entre os processadores é realizada de forma explícita no código-fonte¹⁸, com a utilização de bibliotecas específicas para implementação do paralelismo, como por exemplo *MPI*¹⁹ [78, 79, 80], *PVM*²⁰ [81] e *LINDA* [82];

programação concorrente desenvolvida com o propósito de introduzir execução condicional ou repetitiva de declarações, e também para agrupar seqüências de declarações em procedimentos ou subrotinas²¹. Entre estas linguagens destacam-se *HPF*²² [83], *PFortran*²³, *Maisie* [84] e *Alef* [77].

O desenvolvimento de programas visando a utilização de recursos computacionais paralelos é uma área de grande atividade atualmente. Existem projetos onde são exploradas as melhores características de cada linguagem, sendo alguns deles apresentados a seguir.

O projeto OpenMOSIX [85] aplica o conceito de programação paralela para sistemas de memória distribuída. A implementação é realizada no *kernel* do sistema operacional e permite que várias estações de trabalho sejam utilizadas como um grande computador de memória compartilhada, permitindo o desenvolvimento de aplicativos sem a necessidade de programação distribuída.

O projeto ADAPTOR [86] oferece um sistema de compilação para a linguagem *HPF* com a geração de um código-fonte em *FORTRAN*²⁴ 77 ou 90/95 com chamadas para as bibliotecas *MPI* ou *PVM*, sendo que neste caso a implementação do paralelismo acontece de forma transparente para o usuário. O projeto SHPF [87] oferece um sistema de compilação semelhante, porém o código-fonte é gerado exclusivamente em *FORTRAN* 90/95.

Existe um grupo informal dedicado a tornar a computação paralela realizada via *cluster* prática em termos de computação de alto desempenho, agrupando diversas

¹⁸Código-fonte é o nome dado ao arquivo texto escrito de acordo com uma linguagem de programação.

¹⁹Do inglês *Message Passing Interface*.

²⁰Do inglês *Parallel Virtual Machine*.

²¹Subrotina é um segmento do programa que executa um procedimento específico de cálculo, normalmente utilizado diversas vezes ao longo da execução do programa.

²²Do inglês *High Performance Fortran*.

²³Do inglês *Parallel Fortran*.

²⁴Linguagem de programação amplamente utilizada em cálculos numéricos. Normalmente é acompanhada de um número, que indica a padronização da linguagem.

ferramentas em um pacote chamada *OSCAR* [88].

3.2.6 Aplicativos para ambiente de alto desempenho

Existem diversos aplicativos que fazem uso de processamento paralelo. Os mais importantes no contexto deste trabalho são:

Scilab : Clone do programa Matlab [89] desenvolvido e distribuído gratuitamente pela INRIA [90]. O Scilab utiliza a biblioteca *PVM* para a execução de tarefas em ambiente paralelo, possuindo uma interface gráfica com o usuário e diversos módulos de cálculo específicos, similares aos *Toolboxes*²⁵ do Matlab. O programa está disponível para uma ampla gama de sistemas operacionais, incluindo o Microsoft Windows [69], sendo uma excelente alternativa para fins educacionais. Os principais módulos de interesse para a Engenharia Química são os de simulação dinâmica, de redes neuronais, de tratamento de ruídos e de controle.

Dalton [91], GAMESS [92] e NWCHEM [93] Programas para cálculos quânticos normalmente encontrados em centros de processamento de alto desempenho. Estes programas são distribuídos gratuitamente, possuindo uma excelente aceitação pela comunidade acadêmica internacional.

Todos os *softwares* citados no item precedente já se encontram disponíveis no laboratório de pesquisa coordenado pelo Prof. Dr. Rubens Maciel Filho. Novos *softwares* podem ser incorporados, conforme a necessidade dos usuários.

3.3 O *cluster* proposto

O *cluster* proposto consiste em diversos computadores, divididos em mestre, escravos e usuários, e uma infra-estrutura de rede para interconexão destes computadores, conforme digrama esquemático apresentado na figura 3.1.

O computador mestre funciona como porta de entrada para os usuários do ambiente computacional, sendo o responsável pela compilação dos programas desenvolvidos e solicitação de execução dos programas nos computadores escravos.

²⁵São ferramentas para tratamento de problemas específicos, como tratamento de sinais e controle de processos.

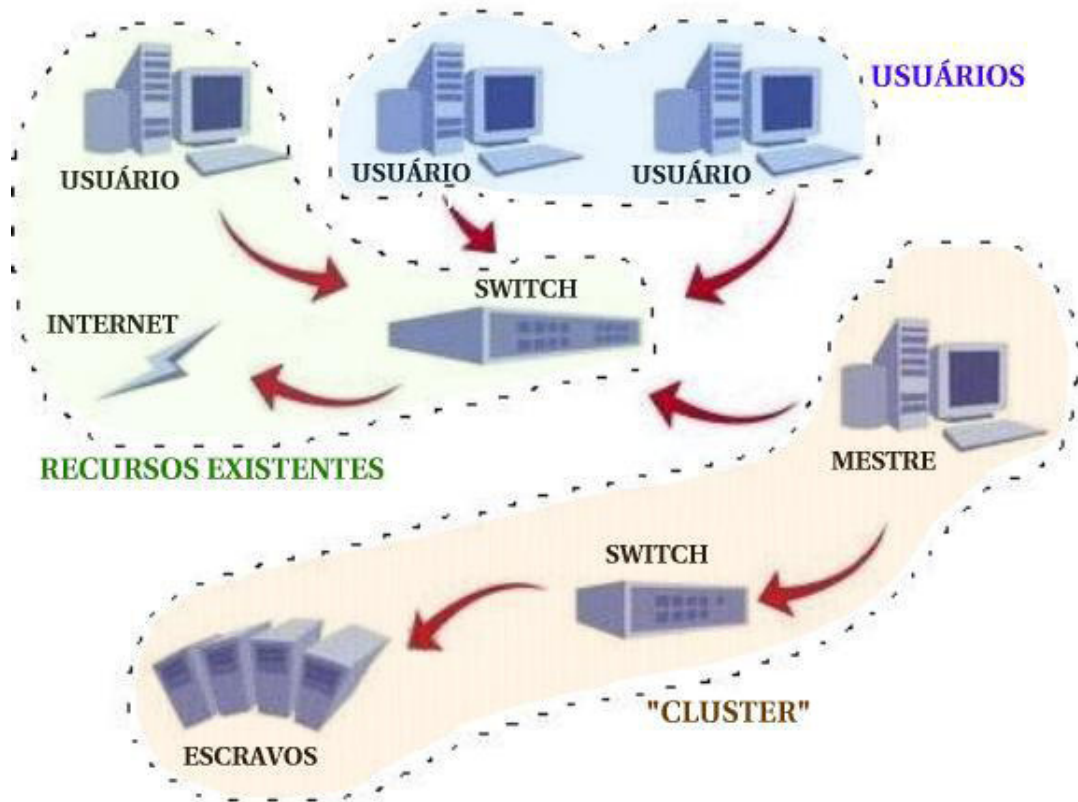


Figura 3.1: O *cluster* proposto.

O objetivo dos computadores escravos é a execução exclusiva de tarefas, paralelizadas ou não, seguindo critérios de execução do computador mestre. Não é necessário utilizar monitor, teclado e mouse nos computadores escravos, sendo todos eles gerenciados e controlados remotamente pelo computador mestre. Como os computadores escravos não são acessados diretamente pelos usuários utilizam-se os endereços IP^{26} reservados 10.0.0.0/8, ou 192.168.0.0/16.

O computador usuário será adicionado a infra-estrutura já existente e será utilizado para a preparação dos programas a serem executados no *cluster*. É importante salientar que o *cluster* pode ser disponibilizado para uso por parte de todos os interessados pertencentes ao grupo de pesquisa do LOPCA.

A arquitetura recomendada é a da família *x86*, por apresentar excelente relação custo-benefício, além de ser facilmente encontrada no mercado. O desempenho em cálculos com ponto flutuante, o mais comum em computação científica, é quase que equivalente aos obtidos por estações de trabalho de custo muito superior. Atualmente, a escolha mais lógica são os processadores da linha Opteron da *AMD* [67].

²⁶Do inglês *Internet Protocol*. É uma sequência numérica que identifica um computador na rede.

O ambiente paralelo sugerido leva à adoção de uma topologia de rede dinâmica, com a adoção de protocolo de transferência padrão Gigabit Ethernet apresentando bom desempenho de transferência e custo vantajoso quando comparada as outras topologias de rede. Além disso, é o padrão comumente adotado pelos fabricantes de *hardware*, com suporte para diversos sistemas operacionais. A conexão interna do *cluster* deve ser realizada com uma *switch*, apesar desta apresentar custo mais elevado que o *hub*, pelas características apresentadas anteriormente, podendo ser aproveitada a infra-estrutura de rede existente para a comunicação externa entre *cluster* e os computadores dos usuários.

O sistema deve ser de domínio público, tanto para redução do custo quanto para aproveitar diversas experiências já realizadas e amplamente divulgadas na literatura. O ambiente computacional sugerido permite a coexistência de mais de um sistema operacional, portanto, este não será definido *a priori*.

Os sistemas operacionais de domínio público possuem disponíveis um grande número de linguagens de programação, incluindo *FORTRAN* 77 e 90/95, *C*, *C++*, Java, entre outras, que permitem a utilização dos softwares citados no item 3.2.6.

3.4 Conclusões

É possível criar um ambiente computacional paralelo de alto desempenho utilizando apenas componentes facilmente disponíveis no mercado. Entre as inúmeras vantagens do ambiente computacional proposto podemos citar:

- o custo de aquisição corresponde a aproximadamente 30% do custo de um sistema comercial de desempenho computacional equivalente, ficando limitado a aquisição de *hardware*, uma vez que toda a infra-estrutura de *software*, incluindo sistema operacional, linguagens de programação e programas aplicativos, pode ser feita com programas de domínio público e/ou gratuito;
- diminuição de custos e facilidade de manutenção, uma vez que permite a utilização de componentes de fácil disponibilidade no mercado, tornando a solução independente de fornecedor;
- continuidade tecnológica, pois a atualização total ou parcial dos componentes utilizados pode ser realizada facilmente com componentes encontrados no mercado;

- flexibilidade de configuração, permitindo adequar a capacidade do ambiente computacional ao tipo de aplicação e ao orçamento. Exemplo desta flexibilidade pode ser observada na figura 3.2 que apresenta dois *clusters* desenvolvidos no laboratório de pesquisas Sandia [94];
- escalabilidade, permitindo anexar mais computadores ao *cluster* quando houver necessidade de aumento da carga de processamento;
- alta disponibilidade, uma vez que cada computador do *cluster*, excluindo-se o que atuará como servidor, é uma unidade independente, se algum deles falhar não ocorre comprometimento geral do ambiente;
- exclusividade, uma vez que geralmente são mantidos pelos próprios grupos de pesquisadores/usuários, não ficando vinculados aos fornecedores comerciais ou aos grandes centros de processamento de alto desempenho.

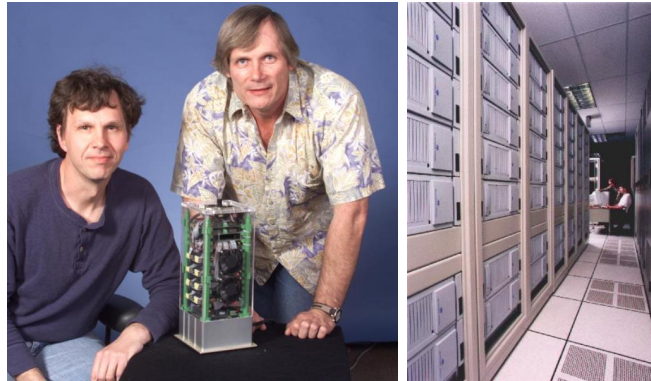


Figura 3.2: Demonstração da flexibilidade dos *clusters*: o da esquerda possui cinco nós, enquanto o da direita possui acima de 1500 nós, ambos do Laboratório Sandia.

Capítulo 4

Análise de desempenho de Sistemas Operacionais

4.1 Introdução

Ao longo dos últimos anos foram incorporados aos microcomputadores de uso pessoal e estações de trabalho uma série de técnicas e recursos derivados da supercomputação. Como resultado da redução dos preços e destes avanços, a diferença de desempenho entre estes equipamentos e os servidores específicos diminuiu ao ponto de permitir a popularização de métodos computacionais.

Ao mesmo tempo em que o desempenho computacional acessível ao pesquisador cresceu várias ordens de grandeza, a sua medida ou forma de expressão foi se tornando gradativamente mais complexa. Enquanto, no início da década de 90, o desempenho era expresso adequadamente através do número de operações de ponto flutuante¹ por segundo (*FLOPS*)², nos anos seguintes os fabricantes e organizações buscaram formas de tentar expressar o desempenho computacional do equipamento como um todo e não apenas do processador. Outros aspectos relevantes, como o desempenho do sistema de discos, a velocidade de acesso à memória, o desempenho em redes de alta velocidade e a capacidade dos equipamentos executarem diversas tarefas simultaneamente, passaram a ser medidos e os resultados obtidos ponderados em índices que expressavam o desempenho geral do equipamento.

A tabela 4.1 apresenta a frequência de processamento, a memória de acesso

¹Metodologia utilizada para o armazenamento de um número inteiro na forma binária.

²Do inglês *Floating Point Operations per Second*.

aleatório (*RAM*)³ e o desempenho em cálculos de ponto flutuante de diversos processadores da arquitetura *x86*, permitindo acompanhar sua evolução ao longo de diversas gerações de processadores.

Tabela 4.1: Evolução do desempenho dos processadores da arquitetura *x86*.

Processador	Frequência (<i>MHz</i>)	Memória típica (<i>Mb</i>)	<i>MFLOPS</i>
386SX	33	8	0,0052
386DX/387	40	8	0,7047
Pentium	60	32	16,3323
PentiumPro	233	64	80,9883
Pentium II	300	128	125,7116
Pentium III	700	256	312,9155
Athlon	800	256	562,5306

Testes de desempenho consagrados como o *Linpack* [95] deram lugar aos testes desenvolvidos pela *SPEC*⁴ [96]. Basta analisar os valores dos resultados dos equipamentos e optar por aquele que fornecer o melhor conjunto de índices para o orçamento disponível. Infelizmente, na maioria dos casos isto não ocorre, pois o acesso ao conjunto de testes que compõe o *SPEC* não é público, devendo ser adquirido a um custo relativamente alto, restringindo aos grandes fabricantes a medida e a publicação destes índices. Por outro lado, de meados da década de 90 em diante a relação entre custo e desempenho de computadores pessoais baseados na arquitetura *x86* tornou-os extremamente atrativo como alternativa às dispendiosas estações de trabalho. A natureza quase artesanal destes sistemas dificulta a análise de desempenho por parte do usuário final, uma vez que o computador não é constituído apenas pelo seu processador, e outros componentes como o sistema de cache, memória e discos são extremamente relevantes para o desempenho computacional.

O objetivo deste capítulo é apresentar a avaliação do desempenho de diferentes sistemas operacionais para a arquitetura *x86*. Esta avaliação foi realizada através do uso de programas específicos para este fim, sendo avaliado o desempenho de cálculos de números inteiros e números reais (ponto flutuante), desempenho de acesso à memória aleatória e o desempenho de acesso ao disco de armazenamento.

Todos os sistemas operacionais avaliados foram instalados em um microcomputador Pentium II de 350 *MHz*, com 64 *Mb* de memória *RAM* e disco de

³Do inglês *Random Access Memory*.

⁴*Systems Performance Evaluation Cooperative*.

armazenamento de 2,0 Gb. Foi seguida a instalação padrão de cada sistema operacional, sem adicionar recursos de som, rede e vídeo, além dos já presentes na instalação padrão.

4.2 Fatores relevantes para o desempenho computacional

O computador é formado pela união de diversos componentes, tais como placa-mãe, processadores, módulos de memória, placas de vídeo, discos de armazenamento de dados, entre outros.

Os fatores mais relevantes para um bom desempenho computacional em programas científicos são:

- desempenho de ponto flutuante do processador;
- desempenho do sistema de memória;
- desempenho do sistema de discos.

Outros fatores como o desempenho em aplicações de computação gráfica e visualização, redes de alta velocidade, multimídia, ou desempenho como servidor, embora relevantes, não foram abordados. Em grande parte, o desempenho destas últimas aplicações depende diretamente do balanço adequado do desempenho dos três itens estudados.

4.2.1 Análise de desempenho de ponto flutuante

A análise de desempenho de ponto flutuante de um computador pode ser realizada de diversas formas. Se o objetivo for determinar a sua capacidade máxima teórica de cálculos, basta recorrer às especificações do fabricante no que diz respeito ao número máximo de operações de ponto flutuante simultâneas e multiplicar esse valor pela velocidade do processador. Via de regra, este resultado é de pouca valia, uma vez que em processamentos reais o sistema dificilmente opera nas condições ótimas do fabricante.

Uma maneira alternativa para a análise de desempenho de ponto flutuante tem sido feita através do teste *Linpack* [95]. Neste teste, é medida a velocidade

do computador para realizar operações de álgebra linear com matrizes de diferentes dimensões e o resultado é expresso em *MFLOPS*⁵ para cada dimensão. Uma desvantagem do *Linpac* é que ele expressa simultaneamente o desempenho do sistema de ponto flutuante e do sistema de memória do computador, não fornecendo de forma clara o desempenho individual dos componentes.

Em 1992, Aburto [97] propôs um teste simples para avaliar a capacidade bruta de cálculos de ponto flutuante dos processadores. Neste teste, o programa *flops* calcula oito séries numéricas que exercitam as quatro operações aritméticas de ponto flutuante em precisão dupla de um processador com diferentes frequências. As séries foram construídas de forma a não solicitar desempenho do sistema de memória e apenas refletir o desempenho de cálculos de ponto flutuante a partir do tempo gasto pelo processador para calcular cada série.

A tabela 4.2 ilustra o balanço de operações de ponto flutuante em cada série executada pelo programa *flops*. Pode-se observar pelos dados desta tabela, que normalmente são analisados os desempenhos de operações de soma e multiplicação, sendo estas as operações mais otimizadas pelos fabricantes de processadores, sendo que muitos processadores podem realizar estas operações simultaneamente em expressões da álgebra linear do tipo $A = A * X + Y$, onde A , X e Y são números reais. Em operações de divisão, por outro lado, o desempenho é comparativamente sofrível e uma análise dos resultados da série 7 (com 25% de divisões) deve fornecer um bom indicativo para pesquisadores que realizam, por exemplo, simulações de dinâmica molecular.

Tabela 4.2: Número e tipo de operações de ponto flutuante para cada série do programa *flops*.

Série	Adição	Subtração	Multiplicação	Divisão	Total
1	7	0	6	1	14
2	3	2	1	1	7
3	6	2	9	0	17
4	7	0	8	0	15
5	13	0	15	1	29
6	13	0	16	0	29
7	3	3	3	3	12
8	13	0	17	0	30

A tabela 4.3 apresenta os resultados de desempenho obtidos (em *MFLOPS*)

⁵Mega *FLOPS*.

pelo programa *flops* para os diversos sistemas operacionais utilizados neste estudo. Esta tabela também apresenta os compiladores utilizados e as diretrizes de otimização de compilação, no caso *o* e *o3*, que correspondem ao nível de otimização mínima e máxima, respectivamente.

Tabela 4.3: Desempenho em *MFLOPS* de diversos sistemas operacionais medidos pelo programa *flops*.

Sistema Operac.	1	2	3	4	5	6	7	8
BeOS 5.0 ^a	35,84	17,27	41,40	38,84	39,27	46,09	16,27	48,19
BeOS 5.0 ^b	102,56	72,50	117,60	112,06	81,41	114,38	32,86	123,01
FreeBSD 3.3 ^c	52,40	39,90	63,74	109,73	83,63	111,70	28,27	111,13
FreeBSD 3.3 ^d	123,74	83,98	136,03	111,53	101,14	131,86	38,21	129,49
NetBSD 1.5.2 ^e	107,86	90,23	122,31	113,95	91,73	119,91	39,60	121,27
NetBSD 1.5.2 ^f	103,91	91,38	132,27	119,94	102,51	126,28	39,09	141,28
QNX ^g	46,34	23,70	61,67	71,98	44,62	57,89	24,05	61,40
QNX ^h	97,13	70,65	135,08	118,18	98,11	139,31	31,86	140,98
Suse 6.4 ⁱ	42,20	15,97	49,41	48,09	43,27	47,25	18,18	54,61
Suse 6.4 ^j	39,74	25,24	51,42	55,62	66,48	62,64	23,50	70,93
Windows 2000 ^k	36,80	24,57	46,26	57,11	66,24	61,97	23,30	73,25
Windows 2000 ^l	36,90	24,56	46,28	57,11	66,26	62,01	23,27	73,28

Diretriz de compilação:

^a gcc -DUNIX -o (gcc 2.95.2), ^b gcc -DUNIX -o3 (gcc 2.95.2),

^c cc -DUNIX -o (gcc 2.7.2.3), ^d cc -DUNIX -o3 (gcc 2.7.2.3)

^e cc -DUNIX -o (egcs 1.1.2), ^f cc -DUNIX -o3 (egcs 1.1.2)

^g gcc -DUNIX -o (gcc 2.95.2), ^h gcc -DUNIX -o3 (gcc 2.95.2)

ⁱ cc -DUNIX -o (gcc 2.95.2), ^j cc -DUNIX -o3 (gcc 2.95.2)

^k gcc -DUNIX -o (gcc 2.95.2), ^l gcc -DUNIX -o3 (gcc 2.95.2)

Analisando-se os dados da tabela 4.3 é possível observar que:

- a diretriz de otimização teve grande efeito no desempenho de cálculo nos sistemas operacionais BeOS, FreeBSD e QNX, enquanto o efeito foi praticamente nulo para o sistema operacional Windows;
- os sistemas operacionais FreeBSD e NetBSD apresentaram os melhores desempenhos, enquanto os sistemas Suse e Windows apresentaram os piores, considerando sempre os maiores valores para cada sistema operacional;

- existe uma grande diferença de desempenho em função do sistema operacional. A relação melhor/pior desempenho foi de 123,74/36,90 para o par FreeBSD/Windows na série 1 e de 39,60/23,30 para o par NetBSD/Windows na série 7.
- os resultados refletem o desempenho de cálculo obtido pelos compiladores utilizados, todos de distribuição gratuita e de grande aceitação no meio acadêmico. A utilização de outros compiladores pode levar a diferenças significativas nos indicadores de desempenho.

4.2.2 Análise de desempenho de acesso à memória

Tomando por base os resultados do programa *flops* pode-se estimar qual o nível de desempenho do sistema de memória necessário para sustentar a velocidade de cálculo. O melhor desempenho foi do NetBSD, versão 1.5.2, com um desempenho de 141,28 *MFLOPS* na série 8. Para que este sistema operacional consiga manter este desempenho em aplicações reais é necessário que o sistema de memória forneça sem restrições a quantidade de dados que o processador é capaz de lidar.

Uma variável de precisão dupla possui 8 bytes de tamanho, uma operação matricial do tipo $A(i) = A(i) * X + Y$, onde A é um número real e X e Y são constantes reais, necessita realizar um acesso a memória e ler e escrever 8 bytes a cada duas operações de ponto flutuante. Isto implica que o sistema de memória deve ser capaz de transferir $8 * 141,28/2 = 565,12$ MBytes/s. Este sistema de memória, por sua vez, varia em desempenho de acordo com a interação do sistema operacional com diversos dispositivos como: placa mãe utilizada (*chipset*⁶); processador (barramento e memória cache); memória *RAM*, entre outros.

O gargalo representado pelo desequilíbrio entre a velocidade do sistema de memória e a velocidade de processamento foi estudado por McCalpin EM 1995 através do programa *STREAM* [98]. Este programa realiza medidas de desempenho para quatro tipos básicos de operações da álgebra linear comuns em computação científica. Considerando $A(i)$, $B(i)$ e $C(i)$ elementos de vetores de precisão dupla de grandes dimensões e X uma constante de precisão dupla, os testes realizados por este programa são:

Atribuição ou cópia: $A(i) = B(i)$;

⁶Componente responsável pela comunicação do processador com os outros recursos do sistema.

Escalonamento e cópia: $A(i) = X * B(i)$;

Soma: $A(i) = A(i) + B(i)$; e

Soma com escalonamento: $A(i) = B(i) + X * C(i)$.

O programa *STREAM* foi compilado após definir um tamanho para os vetores A , B e C suficientemente grande para evitar interferência do sistema de memória cache do processador, o que super dimensionaria os resultados do teste. Os resultados obtidos estão apresentados na tabela 4.4 e a análise destes resultados permite diferenciar a interação dos sistemas operacionais com o sistema de memória de forma bastante clara.

Tabela 4.4: Velocidade de acesso à memória medida em *MBytes/s* com o programa *STREAM*.

Sistema Operacional	Cópia	Escal.	Soma	Cópia/escal.
FreeBSD 3.3 ^a	140,56	137,70	153,80	151,24
FreeBSD 3.3 ^b	140,70	95,88	117,57	118,96
NetBSD 1.5.2 ^c	140,69	95,88	117,56	118,95
NetBSD 1.5.2 ^d	142,15	147,62	170,41	142,69
Suse 6.4 ^e	142,47	149,02	168,06	144,32
Suse 6.4 ^f	144,78	150,99	169,10	158,84
Windows 2000 ^g	135,45	142,38	157,78	135,45
Windows 2000 ^h	139,53	135,45	151,43	129,13

Diretriz de compilação:

^a cc -o -lm (gcc 2.7.2.3), ^b cc -o3 -lm (gcc 2.7.2.3)

^c cc -o -lm (egcs 1.1.2), ^d cc -o3 -lm (egcs 1.1.2)

^e cc -o -lm (gcc 2.95.2), ^f cc -o3 -lm (gcc 2.95.2)

^g gcc -o -lm (gcc 2.95.2), ^h gcc -o3 -lm (gcc 2.95.2)

Pela análise da tabela 4.4 é possível perceber que:

- os sistemas FreeBSD e NetBSD apresentaram resultados máximos próximos, o que pode ser explicada pela origem comum destes dois sistemas operacionais;
- nos sistemas FreeBSD e Windows a diretriz de otimização do compilador prejudicou, em alguns casos, o desempenho de acesso a memória;

- os sistemas operacionais BeOS e QNX não puderam ser avaliados pois não foi possível executar o programa *STREAM*, devido a indisponibilidade de uma biblioteca dinâmica para estes sistemas;
- o sistema Suse apresentou o melhor desempenho na maioria das operações de acesso a memória *RAM*;
- a diferença entre os melhores resultados obtidos para cada sistema operacional ficou abaixo de 8% para a maioria dos casos, exceto para a operação de cópia com escalonamento, que atingiu 18% de diferença entre o melhor desempenho (Suse) e o pior (Windows).

4.2.3 Análise da velocidade de acesso aos discos

O desempenho dos computadores com relação à velocidade de acesso aos discos torna-se extremamente relevante naqueles métodos computacionais onde o volume de dados excede a capacidade de armazenamento da memória principal. Entre estes se destacam, por exemplo, as análises de trajetórias de dinâmica molecular e o armazenamento e leitura de integrais em alguns tipos de cálculos *ab-initio*. Vários fatores afetam o desempenho do sistema de discos de um computador, sendo que os principais são:

- tecnologia utilizada para armazenamento (*ATA*, *SCSI* ou *SATA*);
- velocidade dos dispositivos controladores e dos discos;
- sobrecarga sobre o processador associada ao controle de tarefas de leitura e escrita; e
- tecnologia do sistema operacional para lidar com as operações de leitura e escrita.

Para uma bom desempenho computacional é fundamental manter o processador ocupado a maior parte do tempo para cálculo, com a menor latência associada ao tempo que o processador aguarda o recebimento ou armazenagem dos dados em disco. Além disto, em ambientes computacionais onde vários usuários ou processos compartilham simultaneamente o sistema de discos, é essencial que este não sofra uma degradação sensível em função dos vários processos acessando diferentes regiões do disco para leitura e escrita. Várias características dos dados afetam o desempenho de um disco, entre eles

podemos citar a forma de acesso (em blocos de dados ou em bytes) e o tamanho dos blocos transferidos em cada operação de acesso.

A metodologia que o sistema operacional adota para o uso da memória principal disponível entre processos ou cache de disco de entrada e saída é extremamente variado, sendo que os sistemas mais modernos tentam minimizar ao extremo o acesso aos discos. Quando um programa solicita ao sistema operacional a leitura de uma parte de um arquivo, o sistema lerá uma quantidade maior de dados tentando antecipar a próxima solicitação. Em operações de escrita, o sistema procura reter na memória principal o máximo de dados possíveis antes de transferi-los para o disco. O desempenho deste sistema de cache em memória aumenta com o aumento da quantidade de memória disponível para o sistema operacional.

Vários programas foram desenvolvidos para tentar quantificar o desempenho de disco sob diferentes condições de uso. Neste trabalho nos ateremos ao programa *Bonnie* [99], que trabalha com bloco fixo e igual à 8192 bytes, realizando as seguintes medidas de velocidade:

- leitura e escrita de caracteres e a porcentagem de uso do processador nestas operações;
- leitura e escrita de blocos de 8 *Kbytes* e a porcentagem de uso do processador nestas operações;
- busca e acesso aleatório aos blocos de um arquivo e a porcentagem de uso do processador nestas operações;

Os resultados obtidos pelo *Bonnie* estão apresentados na tabela 4.5, e não indicam necessariamente o desempenho máximo do sistema de disco por trabalhar com tamanho de bloco fixo, mas reflete o desempenho médio que um computador terá com programas de uso rotineiro. A medida do custo computacional das operações, expressa pela porcentagem de uso do processador em cada etapa do teste, permite avaliar a relação real entre velocidade de acesso aos discos e perda de capacidade de processamento para os cálculos enquanto estas operações ocorrem.

Analisando-se os dados da tabela 4.5 é possível observar que:

- o sistema operacional QNX possui a melhor interface de acesso a disco, pois apresentou as maiores taxas de transferência com pequena utilização de CPU (máximo de 0,8%;

Tabela 4.5: Velocidade de acesso à disco medido pelo programa Bonnie.

Sistema Operacional	regravação ^a	CPU ^b	gravação ^a	CPU ^b	busca ^a	CPU ^b
FreeBSD 3.3	2030	8,6	3978	5,2	88,4	2,6
NetBSD 1.5.2	1755	3,1	6744	6,4	52,0	2,0
QNX	3138	0,8	6599	0,3	58,5	0,2
Suse 6.4	1875	17,8	4308	81,9	95,3	2,3

^a Velocidade medida em *kbytes* por segundo, e ^b utilização de CPU em %.

As diretrizes de otimização de compilação não tiveram efeito na taxa de transferência.

- o sistema operacional Suse apresentou alta utilização de CPU por taxa de transferência. O processo de gravação consumiu 81,9% de processamento para uma taxa de gravação de 4308 *Kbytes/s*, enquanto o NetBSD utilizou 6,4% para uma taxa de gravação de 6744 *Kbytes/s*;
- os sistemas operacionais BeOS e Windows não foram analisados por problemas de compilação e de execução, respectivamente, do programa *Bonnie* nestes sistemas.

4.3 Conclusões

Ao longo deste capítulo foram apresentadas as técnicas e as ferramentas básicas para análise de desempenho de computadores utilizando programas de domínio público.

A análise dos resultados obtidos através destes programas forneceu indicadores mais consistentes sobre o desempenho em função do sistema operacional instalado. Estes indicadores permitem uma melhor especificação de compra de computadores por parte dos interessados em pesquisas computacionais.

A diferença de desempenho, em função do sistema operacional, foi de até 341% para operações de ponto flutuante, de até 18% para o acesso a memória *RAM* e de até 67% para o acesso ao disco de armazenamento.

O desempenho de um computador não deve ser medido somente em função da velocidade do processador em *MHz* ou *GHz*, devido a grande diferença de desempenho em função do sistema operacional. Em projetos de grande porte, o ganho computacional obtido na utilização do sistema operacional de melhor desempenho pode ser determinante para o sucesso.

Capítulo 5

Métodos de química computacional

5.1 Introdução

Neste capítulo estão apresentados os conceitos mais importantes para a realização e compreensão de experimentos em química computacional, a saber:

- Estimativa dos recursos computacionais necessários para desenvolvimento da pesquisa;
- Preparação dos arquivos de coordenadas atômicas, que juntamente com as diretrizes de execução específicas de cada programa, representam o ambiente e as condições a serem simuladas;
- Metodologia quântica, clássica e correlacional dos programas de *QC*, e suas respectivas técnicas e variantes.

5.2 Recursos computacionais

Os parâmetros computacionais mais importantes a serem considerados em pesquisas envolvendo *QC* são o poder de processamento, o espaço em disco e o tempo de processamento disponíveis, uma vez que são utilizados intensamente os recursos computacionais.

Outra consideração importante é quanto a facilidade de uso do programas e sua versatilidade e eficiência. Normalmente os programas mais fáceis de serem

utilizados possuem opções limitadas de simulação e implementações numéricas não muito otimizadas [100].

Um projeto de pesquisa em química computacional deve considerar quais as propriedades a serem obtidas, quanto exato devem ser os valores e por que estes resultados devem ser obtidos. As respostas para estas perguntas definem o melhor método a ser utilizado.

5.3 Arquivos de entrada

A grande maioria dos programas de *QC* são executados por linha de comando. O procedimento de execução do programa requer um ou mais arquivos de entrada, que contenham a representação espacial dos átomos envolvidos e um conjunto de instruções de cálculo para obtenção dos resultados almejados. A representação espacial dos átomos podem ser realizada em coordenadas internas ou cartesianas, conforme detalhamento a seguir.

5.3.1 Coordenada interna

A representação por coordenada interna, também conhecida como matriz Z , é realizada através de parâmetros que correlacionam as posições atômicas no sistema estudado, tendo como referência um ou mais átomos definidos anteriormente.

A coordenada interna para qualquer átomo i consiste na distância interatômica, em Å, entre os átomos i e j , um ângulo, em graus, entre os átomos i , j e k e finalmente um ângulo torsional (diedro) formado entre os planos definidos pelo átomos i , j , k e l , sendo que os átomos j , k , l já tenham sido definidos.

As regras apresentadas no parágrafo anterior possuem as seguintes exceções:

1. O átomo 1 não possui coordenadas, pois é a origem;
2. O átomo 2 deve estar conectado ao átomo 1 apenas por uma distância interatômica;
3. O átomo 3 pode estar conectado ao átomo 1 ou 2, devendo definir um ângulo em relação ao alinhamento dos átomos 2 e 1 ($3 - 2 - 1$ ou $3 - 1 - 2$), não sendo possível a formação de um ângulo torsional (diedro). Por padrão, o átomo 3 está ligado ao átomo 2.

Existem, ainda, algumas restrições quanto a representação das posições atômicas por coordenadas internas. Estas restrições estão associadas ao procedimento de otimização geométrica molecular, no qual o algoritmo de otimização varia a posição atômica em busca de uma configuração que apresente a menor energia interna molecular. Estas restrições, quanto a representação das posições atômicas, são:

1. As distâncias interatômicas devem ser maiores que zero, sendo zero aceitável apenas para relatar simetria com outro átomo;
2. Os ângulos devem estar compreendidos entre 0 e 180 graus, inclusive, de forma a beneficiar o usuário. Ângulos negativos ou maiores que 180° indicam erro na construção da matriz representativa ou na especificação dos ângulos diedros, respectivamente;
3. Os ângulos diedros devem ser sempre definidos. Se durante o procedimento de minimização o átomo i formar um ângulo diedro com os átomos j , k e l , e estes formarem uma linha, o ângulo diedro fica indefinido, causando erro na execução do programa, erro este normalmente de difícil identificação. Esta restrição não se aplica quando o ângulo diedro for 0,0 ou 180,0 graus, ou quando os átomos j , k e l estiverem em linha reta devido simetria molecular, como no caso do acetileno, da acetonitrila, do butino-2, etc. No Mopac [101], os ângulos diedros são confirmados periodicamente, e caso se aproximem em 0,2 Å de uma linha reta o programa exibe uma mensagem de erro e interrompe sua execução.

O processo sugerido de seleção de átomos para representação interna das posições atômicas está apresentado na figura 5.1. Observa-se nesta figura que devem ser selecionados três átomos que fazem parte da cadeia principal, seguido da seleção dos átomos ligados aos átomos 1 e 2. O procedimento segue ao longo de toda a extensão da cadeia molecular.

5.3.2 Coordenada cartesiana

O arquivo de coordenada cartesiana consiste em uma matriz formada pelo símbolo químico ou número atômico, seguido pelas coordenadas cartesianas do átomo no espaço. O arquivo pode conter diretivas de otimização, mas não expressa a distância ou os ângulos formados entre os átomos, não sendo permitida a representação de átomos fictícios ou de simetria molecular.

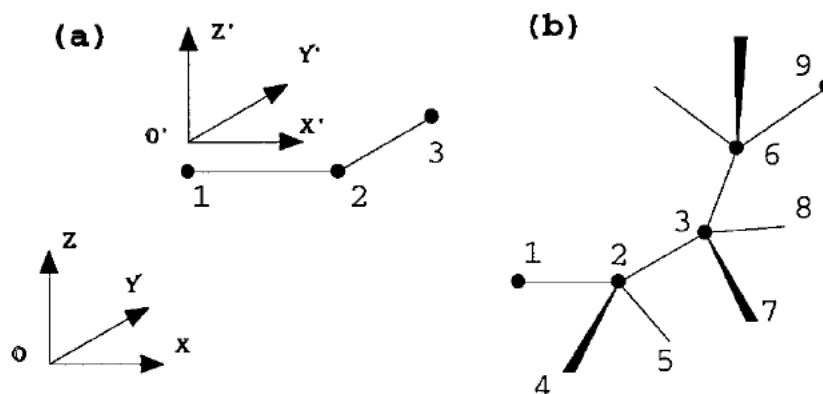


Figura 5.1: Ordem atômica sugerida para representação molecular por coordenadas internas: **a)** seleção inicial e **b)** sequência de seleções.

Por definição, o átomo 1 é a origem do espaço cartesiano, o átomo 2 está posicionado no eixo X , apresentando portanto coordenada $Y = 0$ e $Z = 0$. O átomo 3 está posicionado no plano $X - Y$, a menos que o ângulo $3 - 2 - 1$ seja $0,0$ ou $180,0$. Os demais átomos podem estar posicionados em qualquer lugar do espaço.

5.3.3 Exemplos

A seguir tem-se a representação da molécula de ácido fórmico (figura 5.2) através de coordenadas internas e coordenadas cartesianas.

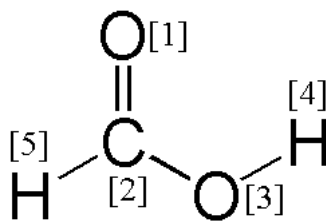


Figura 5.2: Molécula de ácido fórmico utilizada para representação das coordenadas. O número entre colchetes indica a ordem de citação.

Coordenada interna

A representação por coordenadas internas pode ser visualizada abaixo:

O	0.0000		0.000		0.000			
C	1.2000	1	0.000	0	0.000	0	1	0 0

O	1.3200	1	116.800	1	0.000	0	2	1	0
H	1.1100	1	127.300	1	180.000	1	2	1	3
H	0.9800	2	123.900	1	0.000	1	3	2	1

Nesta representação o primeiro átomo (oxigênio) é o ponto referencial, não necessitando de qualquer referência. O átomo 2 (carbono) está ligado ao átomo 1 por um comprimento de ligação de 1,200 Å. O átomo 3 (oxigênio) está ligado ao átomo de carbono com um comprimento de ligação de 1,320 Å, formando ainda um ângulo no conjunto $O - C - O$ de 116,800 graus. O primeiro hidrogênio está ligado ao oxigênio formando o grupo hidroxil, enquanto o segundo hidrogênio está ligado ao carbono. O ângulo diedro $H - C - O - O$ é de 180 graus.

Coordenada cartesiana

A seguir, temos a representação da mesma molécula, agora em função das coordenadas cartesianas atômicas:

O	8.00	0.0000	0.0000	0.0000
C	6.00	1.2000	0.0000	0.0000
O	8.00	1.7952	1.1782	0.0000
H	1.00	1.8726	-0.8830	0.0000
H	1.00	1.3156	2.0328	0.0000

Na representação cartesiana o primeiro átomo (oxigênio) marca a origem dos planos cartesianos. O segundo átomo (carbono) está posicionado sobre o eixo x , enquanto o terceiro átomo (oxigênio) está posicionado sobre o plano $x - y$. Os demais átomos estão distribuídos no espaço cartesiano.

5.4 Geometria molecular

A forma como o programa considera internamente a geometria molecular é fundamental para se poder avaliar o desempenho da otimização geométrica. Muitos programas quânticos otimizam a geometria da molécula através de mudanças na matriz Z , o que é uma boa estratégia por corresponder a movimentos moleculares semelhantes aos vistos nos modos vibracionais. Contudo, uma pequena mudança das coordenadas internas pode corresponder a uma grande distorção em algumas partes da molécula.

Normalmente, a representação por coordenadas cartesianas é a única forma possível em simulações de mecânica clássica, uma vez que pode envolver uma grande quantidade de átomos, ou moléculas, sem ligações químicas.

A figura 5.3 apresenta um comparativo das etapas de otimização geométrica quando são utilizadas coordenadas internas ou coordenadas cartesianas na representação molecular. A parte **a)** deste figura representa a etapa de otimização considerando coordenadas internas, com alteração do ângulo de ligação mas mantendo a distância entre os átomos, enquanto a parte **b)** representa a etapa de otimização com coordenadas cartesianas, com alteração de duas coordenadas atômicas.

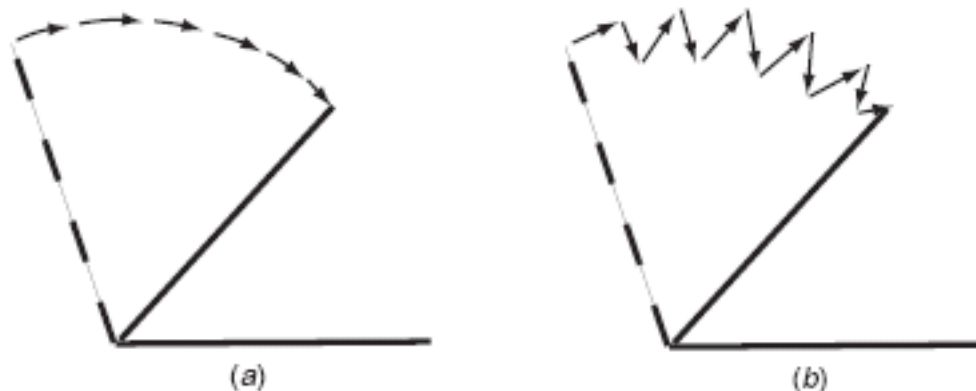


Figura 5.3: Etapas de otimização geométrica molecular quando são utilizadas: **a)** coordenadas internas; e **b)** coordenadas cartesianas.

Alguns programas de *QC*, durante as etapas de cálculo, convertem uma coordenada para outra, de forma a melhorar a eficiência de otimização. Quando é considerada a redundância de coordenadas, inicialmente o sistema converte a matriz interna em coordenadas cartesianas e confere se todas as distâncias interatômicas estão dentro de uma margem razoável para ligações químicas. O programa gera, então, uma lista de distâncias atômicas e ângulos de ligação, e a partir desta lista constrói uma matriz interna que possui mais informações do que as necessárias para especificação da geometria do sistema, evitando as restrições já apresentadas no item 5.3.1. Este procedimento costuma ser o mais eficiente para otimização da geometria, com exceção de quando estiverem envolvidas ligações extremamente longas, como por exemplo na formação ou quebra de ligação no estado de transição ou interações intermoleculares.

5.5 Métodos quânticos

5.5.1 Conjunto de base

Inicialmente os cálculos quânticos moleculares, também conhecidos como métodos de estrutura eletrônica, foram realizados como uma Combinação Linear de Orbitais Atômicos - Orbitais Moleculares, conforme apresentado na equação 5.1:

$$\psi_i = \sum_{\mu=1}^n c_{\mu i} \phi_{\mu} \quad (5.1)$$

onde ψ_i é a função de onda para o i_{simo} orbital molecular, $c_{\mu i}$ são coeficientes de combinação linear, ϕ_{μ} é o μ_{simo} orbital molecular e n é o número de orbitais. Este procedimento permitia a solução de equações de Hartree-Fock para átomos¹. Posteriormente o conceito de orbital molecular foi substituído pelo conceito de funções de base.

A expressão exata para a função de base é do tipo STO ², escrita em coordenadas esféricas segundo a equação 5.2:

$$\phi_i(\zeta, n, l, m, r, \theta, \phi) = N r^{n-1} e^{-\zeta r} Y_{lm}(\theta, \phi) \quad (5.2)$$

onde N é a constante de normalização, ζ é o chamado expoente, θ e ϕ são coordenadas esféricas e Y_{lm} é uma função que descreve o contorno (parte do momento angular). n , l e m são os números quânticos principal, momento angular e magnético, respectivamente.

A representação dos orbitais do tipo STO implica na solução numérica dos orbitais. Com o objetivo de reduzir o esforço computacional foi introduzida a representação GTO ³, que permite solução analítica de acordo com a equação 5.3:

$$g(\alpha, l, m, n, x, y, z) = N e^{-\alpha r^2} x^l y^m z^n \quad (5.3)$$

onde g é a função gaussiana, N é constante de normalização, α é o expoente da função gaussiana, x , y , e z são coordenadas cartesianas e l , m e n não são números quânticos, mas sim expoentes integrais em coordenadas cartesianas. Os valores de x , y e z estão relacionados de acordo com a equação 5.4:

¹Na verdade a função de onda é válida apenas para um único elétron atômico.

²Do inglês *Slater Type Orbital*

³Do inglês *Gaussian Type Orbital*.

$$R^2 = x^2 + y^2 + z^2 \quad (5.4)$$

As principais vantagens matemáticas da função *GTO* em relação a função *STO* são a eliminação do fator pré-exponencial r^{n-1} , o enquadramento do r como função quadrático e a descrição do contorno em função de coordenadas cartesianas. A ausência do termo r^{n-1} restringe a utilização de uma primitiva gaussiana simples apenas aos orbitais $1s$, $2p$, $3d$ e $4f$, sendo necessária a utilização de múltiplas primitivas gaussianas para a representação dos outros orbitais.

Um conjunto de base do tipo *GTO* requer mais primitivas do que um conjunto de bases *STO*, porém no primeiro caso as integrais podem ser computadas analiticamente, o que é muito mais rápido do que a integração numérica exigido pelo conjunto de base *STO*. A figura 5.4 apresenta a utilização de três funções gaussianas para ajuste da curva à função de Slater, mas mesmo quando consideramos 4 ou 5 funções gaussianas para descrição do orbital a solução analítica dos cálculos das integrais é muito mais rápida que a solução numérica exigida pela função de Slater [102].

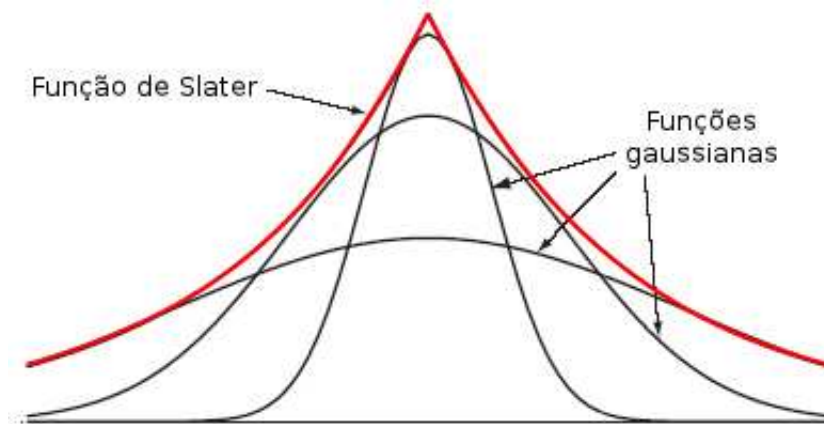


Figura 5.4: Aproximação da função de Slater através de três funções gaussianas.

As primitivas gaussianas são determinadas para átomos isolados e não consideram a deformação do orbital causada pela existência de um átomo próximo. Para resolver este problema foram introduzidas combinações lineares de primitivas gaussianas, conhecidas como contrações. Estas contrações podem ser utilizadas como conjunto de base, possuindo seus coeficientes e expoentes fixos específicos.

As contrações originais derivadas de cálculos atômicos de Hartree-Fock são freqüentemente expandidas com outras funções, sendo as funções de polarização e difusas as mais populares. As funções de polarização são funções simples que

possuem valores de l maiores que os presentes nos orbitais atômicos ocupados para o átomo correspondente. Deve-se observar, por exemplo, que para o elemento lítio, as funções do tipo p não são consideradas funções de polarização, enquanto que para o elemento enxofre as funções do tipo d são consideradas de polarização, o que pode parecer ambíguo, uma vez que para ambos os casos estes orbitais não estão ocupados. Polarizações do tipo d acrescentam de 5 a 6 funções de base e as do tipo f acrescentam de 7 a 10 funções de base, isto para cada elétron.

As funções gaussianas difusas possuem expoentes pequenos que decaem lentamente com o aumento da distância do elétron até o núcleo atômico, sendo indispensáveis para a correta descrição de ânions e ligações fracas, na obtenção de propriedades como momento dipolo e polarizabilidade, entre outras.

Alguns conjuntos de bases são otimizados para cálculos geométricos e energéticos, enquanto outros para obtenção de propriedades, como polarizabilidade. Alguns foram idealizados apenas para métodos Hartree-Fock, outros ajustados para exatidão dos cálculos. Existem também conjuntos indicados para ânions e outros indicados para cátions e moléculas neutras. Em alguns casos, é necessário a descrição dos orbitais internos, como para análise de espectro de NMR^4 , enquanto a representação dos orbitais de valência podem ser mais importantes em estudos de estado de transição.

A descrição de funções de polarização e funções difusas nos conjuntos de base não possuem um padrão oficial, dependendo do autor destes conjuntos. Por exemplo, o grupo do Pople descreve da seguinte forma os seguintes conjuntos de base:

- 6-31G** como sinônimo do grupo 6-31G(dp), representando o conjunto 6-31G com uma função de polarização em cada um dos orbitais d e p ;
- 6-311G(3d3f,2p) representa o conjunto 6-311G com três funções de polarização do tipo d e duas funções de polarização do tipo f nos átomos pesados e duas funções de polarização do tipo p nos átomos de hidrogênio;
- 6-31+G representa a inclusão de uma função difusa dos tipos s e p , com o mesmo expoente, no conjunto de base original;
- 6-311++G representa a inclusão de uma função difusa dos tipos s e p nos átomos pesados e uma função difusa do tipo s nos átomos de hidrogênio.

⁴Do inglês *Nuclear Magnetic Resonance*.

Número total de funções de base e de primitivas

O número total de primitivas e funções de base dependem do conjunto de base utilizado. Por exemplo, considerando um cálculo da molécula de ácido sulfônico, H_2SO_3 , pelo conjunto de base 6-311++G(3df,2p) teremos os números de funções de base e o número de primitivas conforme apresentado abaixo:

Número de funções de base:

S 7 funções do tipo s , 6×3 funções do tipo p , 3×5 funções do tipo d e 1×7 funções do tipo f ;

O 5 funções do tipo s , 4×3 funções do tipo p , 3×5 funções do tipo d e 1×7 funções do tipo f ;

H 4 funções do tipo s e 2×3 funções do tipo p (3 funções do tipo p para cada contração do tipo p , ou seja, p_x , p_y e p_z);

$$H_2SO_3 \quad (4 + 2 \times 3) \times 2 + (7 + 6 \times 3 + 3 \times 5 + 1 \times 7) + (5 + 4 \times 3 + 3 \times 5 + 7) \times 3 = 184$$

Número de primitivas gaussianas:

$$\mathbf{S} \quad 1 \times (6 + 3 + 1 + 1 + 1 + 1 + 1) + 3 \times (4 + 2 + 1 + 1 + 1 + 1) + 5 \times (1 + 1 + 1) + 7 \times 1 = 66;$$

$$\mathbf{O} \quad 1 \times (6 + 3 + 1 + 1 + 1) + 3 \times (3 + 1 + 1 + 1) + 5 \times (1 + 1 + 1) + 7 \times 1 = 52$$

$$\mathbf{H} \quad 1 \times (3 + 1 + 1 + 1) + 3 \times (1 + 1) = 12;$$

$$H_2SO_3 \quad 2 \times 12 + 66 + 3 \times 42 = 246$$

Existe ainda a contração geral, que corresponde a um conjunto de base onde as mesmas primitivas gaussianas podem aparecer em diversas funções de base. Em um esquema geral de contração, as funções de base são formadas como diferentes combinações lineares das mesmas primitivas, o que contrasta com o esquema segmentado apresentado anteriormente. A principal vantagem associada ao uso de contrações gerais é que a mesma pode ser selecionada para uma aproximação mais verdadeira dos orbitais atômicos, o que torna a interpretação dos coeficientes dos orbitais moleculares significativa. Uma segunda vantagem é que possui uma forma mais padronizável que a contração segmentada. O grande problema é que existem poucos programas de QC com recurso de representação por contrações gerais, devido a grande

difículdade de implementação computacional, pois o programa terá que considerar um bloco de integrais simultaneamente, para calcular a contribuição de um dado conjunto de primitivas uma única vez. Uma maneira de contornar este problema é considerar as contrações gerais como um conjunto de base customizado, repetindo as mesmas primitivas gaussianas em diferentes contrações. Entretanto, esta solução apresenta um enorme custo computacional.

Os orbitais atômicos internos, na maioria dos casos, não são afetados significativamente pelas alterações dos orbitais mais externos, alterações estas causadas pelas ligações químicas. As técnicas do potencial efetivo interno (*ECP*)⁵ e Potenciais Efetivos (*EP*)⁶ levam em conta esta característica, tratando os elétrons das camadas mais internas como um potencial médio, ao invés de partículas independentes. Ou seja, os potenciais internos não são orbitais, mas sim modificações do hamiltoniano, com grande redução do custo computacional, permitindo a inclusão de efeitos relativistas em cálculos envolvendo átomos pesados, o que é praticamente inviável quando são considerados todos os elétrons, devido ao grande custo computacional.

Os próximos itens apresentam as funções de base mais comuns em cálculos quânticos:

STO-3G

Conjunto de base mínimo onde a função de Slater é aproximada com três funções gaussianas primitivas. Normalmente, é utilizado em moléculas grandes, fornecendo valores qualitativos.

STO-*n*G

Conjunto de base onde *n* pode ser variar de 2 até 6, indicando o número de primitivas por camada eletrônica com momento angular ocupado. Estes conjuntos de base são utilizados para obtenção de resultados qualitativos e apresentam baixa flexibilidade. Disponível do *H* até o *Xe*.

⁵Do inglês *Effective Core Potential*.

⁶Do inglês *Effective Potentials*.

3-21G

Conjunto de base com três primitivas, mas com flexibilidade nos orbitais de valência. Disponível para H até Cs .

6-31G

Conjunto de base bastante utilizada por fornecer resultados quantitativos para moléculas orgânicas, estando disponível do H até Ar .

6-311G

Outro conjunto de base bastante utilizada, estando disponível do H até o Kr .

pVTZ

Conjunto de base otimizada para obtenção de resultados de polarizabilidade, estando disponível do H até Ca , Br , Rb até Sr e I

5.5.2 Métodos *ab-initio*

No método quântico *ab-initio* de Hartree-Fock (HF) a aproximação primária é do tipo central, o que significa que a repulsão coulombiana elétron-elétron é levada em consideração pela integração do termo de repulsão. Com isto é obtido o efeito médio da repulsão, mas não a interação explícita da repulsão entre os elétrons. Isto é um cálculo variacional, de forma que a energia calculada é igual ou superior a energia experimental. A energia resultante deste cálculo é expressa em Hartrees, onde $1 \text{ hartree} = 27,2116 \text{ eV}$. A figura 5.5 apresenta duas condições geométricas distintas que apresentam a mesma energia eletrônica pelo método de HF [30].

A grande vantagem do método de HF é quebrar a equação de Schrödinger multi-elétrons em muitas equações de um elétron. Cada equação de um elétron é solucionada para produzir uma função de onda de um elétron (orbital) e uma energia (energia do orbital).

A segunda aproximação do método é que a função de onda deve ser descrita na forma de funções matemáticas, sendo exata apenas para sistemas de um elétron. As funções mais utilizadas são combinações lineares de orbitais do tipo gaussiana. A função

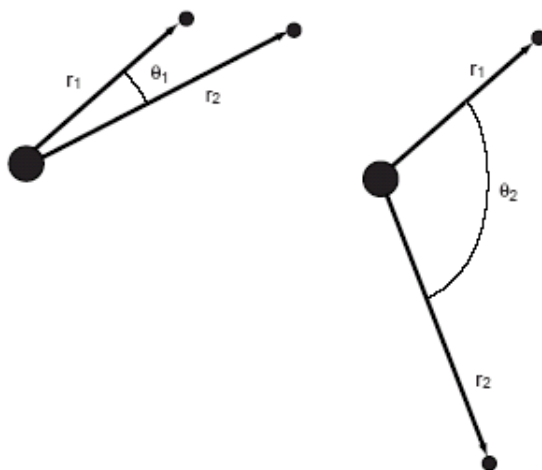


Figura 5.5: Efeito variacional do método de Hartree-Fock: a energia do elétron depende apenas da distância deste para o núcleo, mas independe da distância entre os elétrons, o que leva à mesma energia para estas duas geometrias distintas.

de onda pode ser representada como uma combinação linear de orbitais atômicos, ou de funções de base. Esta aproximação também resulta em valores de energia maiores que os experimentais, mas como este desvio é conhecido os valores obtidos podem ser corrigidos.

As funções gaussianas são multiplicadas por uma função angular de maneira a representar o orbital de simetria s , p , d , e assim por diante. Um termo angular constante fornece o orbital s . Termos angulares x , y e z fornecem o orbital p .

Uma grande restrição do método de HF é que os elétrons estão sempre emparelhados. Isto significa que os cálculos falham em representar casos onde os elétrons podem sofrer desemparelhamento. Por exemplo, ao aumentar a distância entre os átomos da molécula H_2 os cálculos pelo método de HF levam à dissociação na forma de H^+ e H^- .

Para contornar a natureza variacional do método HF foram criados métodos que introduzem uma correção por correlação. Alguns destes métodos são a teoria da perturbação de Møller-Plesset MPn , onde n é a ordem de correlação, a ligação de valência generalizada (GVB)⁷, campo alto consistente multi-configuracional ($MCSCF$)⁸, interação de configuração (CI)⁹ e teoria de agrupamentos acoplados

⁷Do inglês *Generalized Valence Bond*.

⁸Do inglês *Multi-Configurational Self-Consistent Field*.

⁹Do inglês *Configuration Interaction*.

(*CC*)¹⁰.

Perturbação de Møller-Plesset

Ao mapear a função de onda pela formulação da teoria da perturbação, o método de *HF* pode ser considerado como perturbação de primeira ordem. Desta forma pode ser introduzida uma pequena correlação utilizando o método de Møller-Plesset, gerando o método *MP2*. As correlações de terceira ordem (*MP3*) e de quarta ordem (*MP4*) também são comuns, sendo esta última equivalente ao método de interação de configuração de excitação dupla (a ser apresentado no próximo item). A correlação de quinta ordem (*MP5*) apresenta um alto custo computacional, na ordem de N^{10} , onde N corresponde ao número de orbitais moleculares considerados no cálculo.

O método da teoria de perturbação de Møller-Plesset, por quebrar o caráter variacional, pode levar a resultados de energia menores que os valores experimentais, conforme apresentado na figura 5.6.

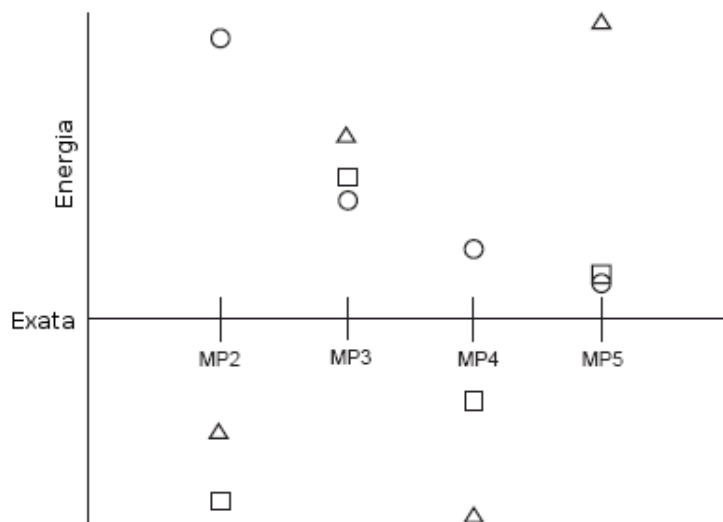


Figura 5.6: Possíveis resultados para várias correlações de Møller-Plesset, onde o círculo representa convergência monoatômica, o quadrado convergência oscilante e séries divergentes são mostradas pelos triângulos.

Interação de configuração

A função de onda de interação de configuração possui múltiplos determinantes. Esta função de onda é construída a partir de funções de onda de *HF*, com a criação de

¹⁰Do inglês *Coupled Cluster*.

novos determinantes pelo deslocamento de elétrons de orbitais ocupados para orbitais vazios. O resultado pode ser bastante exato, mas o custo computacional é da ordem de N^8 .

A interação de configuração de excitação única recebe o nome de *CIS*, a de excitação dupla recebe o nome de *CISD*, e assim por diante. Quando são consideradas todas as excitações possíveis temos a interação de configuração cheia. Esta abordagem apresenta um imenso custo computacional, e se for considerada uma função de base infinita será obtido o resultado exato para a mecânica quântica.

Campo auto-consistente multi-configuracional

Este é um método multi-determinantes onde os orbitais atômicos são otimizados, resultando em um método que permite obter resultado com a mesma exatidão que outros métodos de custo computacional mais elevado. Este método é particularmente recomendado quando a função de onda obtida pelo método de *HF* apresenta um determinante de referência com peso menor que 0,9.

O grande inconveniente deste método é que ele exige um maior conhecimento técnico do usuário, uma vez que não existe uma maneira de se automatizar quais as configurações que devem ser exploradas pelo método, devendo o usuário especificar quais os orbitais moleculares que serão utilizados.

Interação de configuração multi-referência

A função de onda inicial usada neste método, chamada de estado referencial, é obtida a partir do método *MCSCF* ao invés do método *HF*. Existem mais determinantes neste método do que no método *CI*, o que implica em elevado custo computacional, mas pode fornecer uma ótima correção energética por efeitos de correlação.

Agrupamentos acoplados

Este método é similar ao *CI* no qual a função de onda é uma combinação linear de muitos determinantes, mas, neste caso, a escolha dos determinantes é mais complexa. Como no método *CI*, existem várias ordens, como *CCS*, *CCSD*, *CCSDT*, e assim por diante. Estes métodos podem apresentar custo computacional na ordem de N^7 , o que significa que dobrar o tamanho da molécula implica em um esforço computacional 128 vezes maior, mas os resultados de constantes rotacionais podem apresentar um desvio

de apenas 0,1% em relação ao valor experimental, conforme apresentado na figura 5.7 para a molécula de N_2H_2 otimizada pelo método $CCSD(T)$ no nível cc-pVQZ, mais interação de valência-central [30].

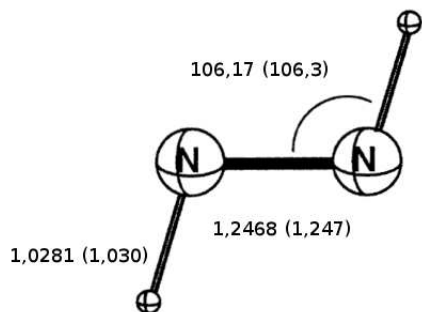


Figura 5.7: Geometria molecular do N_2H_2 obtida através do método $CCSD(T)$ no nível cc-pVQZ, mais interação de valência-central. As distâncias atômicas estão em ângstrons (Å) e os ângulos em graus ($^{\circ}$) (valores experimentais entre parênteses).

Método quântico de Monte-Carlo

Este método trabalha com uma função de onda correlacionada explicitamente em função da distância elétron-elétron, eliminando os desvios implícitos do método de HF .

A função de onda deste método possui a seguinte forma funcional:

$$\Psi = D_{\alpha} D_{\beta} \prod_{i < j} f(r_{ij}) \quad (5.5)$$

onde D_{α} e D_{β} são determinantes dos spins eletrônicos α e β . O uso destes dois determinantes não introduz erro pois não existem muitos spins eletrônicos nos hamiltonianos. O termo $f(r_{ij})$ é o termo que corresponde ao efeito de correlação eletrônica, que pode incluir termos de três corpos, denominados $f(R - i, r_j, r_{ij})$. A forma geral do contorno da função de correlação é conhecida, mas ainda não existe um consenso sobre qual a melhor função matemática a ser usada, devendo o usuário avaliar diversas funções e selecionar a que mais se adequa ao sistema estudado.

Orbitais naturais

Uma vez encontrada a energia e a função de onda, normalmente é necessário calcular outras propriedades moleculares de interesse. Para os métodos

multi-determinantes (*MCSCF*, *CI*, *CC* e *MRCI*), a melhor maneira é utilizar os Orbitais Naturais (*NO*)¹¹. Os *NO* permitem a compressão da matriz de funções de muitos determinantes para um conjunto de orbitais e de elétrons em cada orbital natural.

5.5.3 Termoquímica através de cálculos *ab-initio*

Os métodos *ab-initio* calculam a energia total da molécula como sendo o resultado da soma da energia eletrônica e da energia de repulsão entre núcleos atômicos, estando a molécula isolada no vácuo e sem vibração ($0K$), conforme apresentado na equação 5.6:

$$E_{eq} = E_{el} + E_{núcleo-núcleo} \quad (5.6)$$

A partir da superfície de energia potencial em $0K$ e da aproximação do oscilador harmônico podem ser calculadas as frequências vibracionais, ν_i , do modo vibracional normal. Usando estas frequências podem ser calculadas as contribuições vibracionais, rotacionais e translacionais para as propriedades termodinâmicas, como a função de ou a capacidade calorífica, que aparecem quando se eleva a temperatura de 0 para TK . Considerando Q a função de partição, E a energia, S a entropia e C o calor específico tem-se as contribuições energéticas para o cálculo da energia total.

Contribuição vibracional

O detalhamento para o cálculo da contribuição vibracional nas propriedades termodinâmicas está descrito da equação 5.7 até a equação 5.14.

$$Q_{\text{vib}} = \sum_i \frac{1}{[1 - \exp(-h\nu_i/kT)]} \quad (5.7)$$

onde E_{vib} para uma molécula a temperatura T é:

$$E_{\text{vib}} = \sum_i \left\{ \frac{h\nu_i}{2} + \frac{h\nu_i \exp(-h\nu_i/kT)}{[1 - \exp(-h\nu_i/kT)]} \right\} \quad (5.8)$$

onde h é a constante de Planck, ν_i é a i ésima frequência vibracional normal, e k a constante de Boltzmann. Para 1 mol de moléculas a energia E_{vib} deve ser multiplicada pelo número de Avogadro $N_a = R/k$. Então:

¹¹Do inglês *Natural Orbitals*.

$$E_{\text{vib}} = N_a \sum_i \left\{ \frac{h\nu_i}{2} + \frac{h\nu_i \exp(-h\nu_i/kT)}{[1 - \exp(-h\nu_i/kT)]} \right\} \quad (5.9)$$

Notar que na equação 5.9 o primeiro termo é a energia vibracional no ponto zero, e o segundo termo é a contribuição vibracional devido ao aumento de temperatura de $0K$ para T . Desta forma, a equação 5.9 pode ser decomposta em:

$$E_{\text{vib}} = E_{\text{zero}} + E_{\text{vib}}(0 \rightarrow T) \quad (5.10)$$

$$E_{\text{zero}} = N_a \sum_i \frac{h\nu_i}{2} \quad (5.11)$$

$$E_{\text{vib}}(0 \rightarrow T) = N_a \sum_i \frac{h\nu_i \exp(-h\nu_i/kT)}{[1 - \exp(-h\nu_i/kT)]} \quad (5.12)$$

$$S_{\text{vib}} = R \sum_i \left\{ \frac{(h\nu_i/kT) \exp(-h\nu_i/kT)}{[1 - \exp(-h\nu_i/kT)]} - \ln[1 - \exp(-h\nu_i/kT)] \right\} \quad (5.13)$$

$$C_{\text{vib}} = R \sum_i \left\{ \frac{(h\nu_i/kT)^2 \exp(-h\nu_i/kT)}{[1 - \exp(-h\nu_i/kT)]^2} \right\} \quad (5.14)$$

Em qualquer temperatura acima de $0K$ a molécula rotaciona nos eixos x , y e z , e translaciona nas direções x , y e z . Assumindo a equípartição de energia é possível calcular as energias de rotação (E_{rot}) e de translação (E_{tr}).

Contribuição rotacional

Considerando o número de simetria σ e o momento de inércia I_w , onde w corresponde ao eixo de rotação, as contribuições energéticas devido a rotação molecular podem ser obtidas a partir das equações 5.15 até 5.18 para moléculas lineares e das equações 5.19 até 5.23 para moléculas não lineares.

Molécula linear

$$Q_{\text{rot}} = \frac{8\pi^2 I kT}{\sigma h^2} \quad (5.15)$$

$$E_{\text{rot}} = (2/2)RT \quad (5.16)$$

$$\begin{aligned} S_{\text{rot}} &= R \ln \left[\frac{8\pi^2 I k T}{\sigma h^2} \right] + R \\ &= R \ln I + R \ln T - R \ln \sigma - 4.349203 \end{aligned} \quad (5.17)$$

$$C_{\text{rot}} = (2/2)R \quad (5.18)$$

onde $-4,349203 = R \ln [8 \times 10^{-16} \pi^2 k / (N_a h^2)] + R$.

Molécula não-linear

$$\begin{aligned} Q_{\text{rot}} &= \left(\frac{\sqrt{\pi}}{\sigma} \right) \left[\frac{8\pi^2 k T}{h^2} \right]^{3/2} \sqrt{I_A I_B I_C} \\ &= \left(\frac{\sqrt{\pi}}{\sigma} \right) \left[\left(\frac{8\pi^2 c I_A}{h} \right) \left(\frac{8\pi^2 c I_B}{h} \right) \left(\frac{8\pi^2 c I_C}{h} \right) \right]^{1/2} \left(\frac{k T}{hc} \right)^{3/2} \end{aligned} \quad (5.19)$$

$$E_{\text{rot}} = (3/2)RT \quad (5.20)$$

$$\begin{aligned} S_{\text{rot}} &= \frac{R}{2} \ln \left\{ \left(\frac{\pi}{\sqrt{\sigma}} \right) \left(\frac{8\pi^2 c I_A}{h} \right) \left(\frac{8\pi^2 c I_B}{h} \right) \left(\frac{8\pi^2 c I_C}{h} \right) \left(\frac{k T}{hc} \right)^3 \right\} \\ &\quad + (3/2)R \\ &= (R/2) \ln (I_A I_B I_C) + (3/2)R \ln T - R \ln \sigma - 5.3863921 \end{aligned} \quad (5.21)$$

$$C_{\text{rot}} = (3/2)R \quad (5.22)$$

onde -5.3863921 é calculado como:

$$R \ln \left\{ \frac{1}{h^3} \left(\frac{10^{-16}}{N_a} \right)^{3/2} \sqrt{(3 \times 2^9 \times \pi^7 \times k)} \right\} + (3/2)R \quad (5.23)$$

Contribuição translacional

Considerando o peso molecular M temos as contribuições energéticas translacionais calculadas de acordo com as equações 5.24 até 5.33.

$$Q_{\text{tra}} = \left(\frac{\sqrt{2\pi M k T / N_a}}{h} \right)^3 \quad (5.24)$$

$$E_{\text{tra}} = (3/2)RT \quad (5.25)$$

$$S_{\text{tra}} = R \left\{ \frac{5}{2} + \frac{3}{2} \ln \left(\frac{2\pi k}{h^2} \right) + \ln k + \frac{3}{2} \ln \left(\frac{M}{N_a} \right) + \frac{5}{2} \ln T - \ln p \right\} \quad (5.26)$$

$$= (5/2)R \ln T + (3/2)R \ln M - R \ln p - 2.31482 \quad (5.27)$$

$$C_{\text{tra}} = (5/2)R \quad (5.28)$$

ou $H_{\text{tra}} = (5/2)RT$ devido ao termo pV (conforme equação 5.31). A energia interna U na temperatura T é dada por:

$$U = E_{\text{eq}} + [E_{\text{vib}} + E_{\text{rot}} + E_{\text{tra}}] \quad (5.29)$$

$$= E_{\text{eq}} + [(E_{\text{zero}} + E_{\text{vib}}(0 \rightarrow T)) + E_{\text{rot}} + E_{\text{tra}}] \quad (5.30)$$

A entalpia H para um mol de gás é definida como:

$$H = U + pV \quad (5.31)$$

Assumindo a relação dos gases ideais ($pV = RT$) obtemos a equação:

$$H = U + pV = U + RT \quad (5.32)$$

A energia livre de Gibbs pode ser calculada pela equação:

$$G = H - TS(0 \rightarrow T) \quad (5.33)$$

5.5.4 Métodos da Teoria do Funcional de Densidade

Uma alternativa simples para a equação de Schrödinger foi proposta por H. L. Thomas e Enrico Fermi em 1927, conhecida como a aproximação de Thomas-Fermi, onde ao invés da função de onda do sistema, considera-se a densidade funcional de $\rho(r)$. A energia total E do sistema é escrita como uma funcional de $\rho(r)$ ¹², ou seja, $E[\rho(r)]^E$. Uma simplificação drástica e admirável: a equação para um sistema de N elétrons, e portanto, com $3N$ coordenadas independentes, é reduzida pela aproximação de Thomas-Fermi a um problema em 3 dimensões apenas. Cálculos com a equação de Schrödinger para sistemas multieletrônicos requerem a solução de integrais envolvendo, simultaneamente, coordenadas de um elétron ou de dois elétrons, com recorrências a espaços de três ou seis dimensões, respectivamente. Sendo os elétrons partículas indistinguíveis, percebe-se que bastaria conhecer as funções de distribuição de um ou dois elétrons para se calcular estas integrais. Logo, pode-se argumentar que a equação de Schrödinger, ao considerar as coordenadas de todos os elétrons simultaneamente, contém mais informações do que o necessário.

Inicialmente as metodologias originadas da aproximação de Thomas-Fermi, entre as quais se pode citar o Método $X\alpha$ desenvolvida por Slater *et al.*, não produziram impactos importantes que justificassem mudanças no caminho trilhado para a resolução direta da equação de Schrödinger. Esta preferência pode ter sido induzida também por razões históricas, pois deste o átomo de Dalton os modelos químicos fazem referência direta aos átomos como blocos constituintes da matéria. A utilização da tabela periódica, a recorrência a propriedades de grupos funcionais, etc., contribuíram para uma cultura onde a identificação de átomos em moléculas é a parte central do pensamento químico. A expansão da função de onda de Schrödinger em bases construídas para átomos preservou esta perspectiva histórica, uma condição favorável para a utilização da nova teoria na investigação de problemas clássicos e fundamentais para a química. Este panorama começou a mudar a partir dos trabalhos de Walter Kohn, laureado com o prêmio Nobel de química de 1998 juntamente com John A. Pople [31].

Em 1964 Hohenberg e Kohn publicaram o trabalho “Inhomogeneous electron gas” no qual demonstram um teorema importante: a densidade eletrônica exata do estado fundamental de uma molécula determina, de forma unívoca, todas as propriedades do estado fundamental desta molécula, ou seja, provaram a existência de

¹²Um funcional é uma expressão matemática onde a variável independente é uma função.

uma teoria baseada na aproximação de Thomas-Fermi exata. Em um trabalho posterior, intitulado “Self-consistent equations including exchange and correlation effects”, Kohn e Sham deduzem as equações necessárias para calcular a densidade eletrônica $\rho(r)$ utilizando procedimentos baseados no método variacional, recebendo esta metodologia o nome de Teoria do Funcional de Densidade (*DFT*). Cabe ressaltar que a viabilidade desta metodologia depende do conhecimento do funcional $E[\rho(r)]$ que expressa a energia total E do sistema em função de $\rho(r)$. Esta tarefa não é simples, pois não existe um formalismo exato que auxilie a construção deste funcional.

A *DFT* tem sido muito utilizada para estudar a estrutura eletrônica de sólidos, mas avanços consideráveis foram atingidos na proposição de funcionais de densidade adequados para a utilização da *DFT* no cálculo de propriedades moleculares, com destaque para o trabalho de Becke, Lee e Parr, cuja funcional conhecida como B3LYP tem proporcionado bons resultados para cálculos de química quântica.

O grande impacto da *DFT* na química quântica está relacionado com o fato desta teoria produzir resultados superiores aos obtidos com o método *HF*, a um custo computacional muito menor. Destaca-se também a evolução de conceitos e metodologias para extrair da *DFT* grandezas úteis para a interpretação de propriedades moleculares. Estas novas facilidades, apoiadas por uma fundamentação teórica consistente, tem incentivado a utilização da *DFT* na investigação de propriedades eletrônicas de sistemas químicos poliatômicos, como por exemplo o estudo de sítios ativos de proteínas, reações em superfície, reações em solução, polímeros, etc.

A grande vantagem da *DFT* sobre os métodos *ab-initio* consiste no ganho de tempo computacional e na redução da memória alocada. Considerando-se um sistema arbitrário qualquer modelado com n funções de base, o esforço computacional no estudo deste sistema utilizando-se a *DFT* aumenta na ordem de n^3 , enquanto para os métodos *ab-initio* aumenta em n^4 ou n^5 para técnicas de interação de configuração. Em contraste aos métodos semi-empíricos, o hamiltoniano na *DFT* é bem definido e suas características não são obscurecidas por freqüentes e injustificáveis aproximações no procedimento computacional. Nas derivações das equações do funcional de densidade, nenhum parâmetro, em princípio, necessita ser ajustado ou determinado empiricamente, sob este aspecto a *DFT* pode ser considerada de natureza *ab-initio*.

A tabela 5.1 apresenta as equações de definem os as teorias de Hartree-Fock e do Funcional de Densidade, sendo possível observar a grande semelhança entre os cálculos. Nesta tabela o $\hat{h}_i$ é o operador mono-eletrônico definido pela equação 5.34, contendo as contribuições cinética e de atração nuclear, e o termo $\frac{1}{r_{ij}}$ é o operador de

dois elétrons [103].

$$\hat{h}_i = \hat{t}_i + \sum_A \frac{Z_A}{r_{Ai}} \quad (5.34)$$

Tabela 5.1: Equações das teorias de Hartree-Fock e do Funcional de Densidade

Hartree-Fock	Funcional de Densidade
$E = E[\Psi, R_\alpha]$ $E = \int \Psi * \left[\sum_i \hat{h}_i + \sum_{i>j} \frac{1}{r_{ij}} \right] \Psi \tau$ $\Psi = \ \psi_1(1)\psi_2(2) \dots \psi_3(3)\ $ $\frac{\partial E}{\partial \Psi} = 0$ $\left[-\frac{1}{2}\nabla^2 + V_c(r) + V_x^i(r) \right] \psi_i = \varepsilon_i \psi_i$	$E = E[\rho, R_\alpha]$ $E = T[\rho] = U_{ext+coul}[\rho] + E_{xc}[\rho]$ $\rho = \sum \psi_i(r) ^2$ $\frac{\partial E}{\partial \rho} = 0$ $\left[-\frac{1}{2}\nabla^2 + V_c(r) + V_{xc}(r) \right] \psi_i = \varepsilon_i \psi_i$

A *DFT* tem sido amplamente utilizada em estudos farmacêuticos, agroquímicos e biotecnológicos; na área de materiais e polímeros; em pesquisas com complexos organometálicos e aglomerados típicos de catálise; em estudos de superfície e estado sólido; em eletroquímica e microeletrônica.

Independentemente da forma como a *DFT* está sendo apresentada, em termos genéricos temos dois postulados básicos:

A função de onda do estado fundamental, e daí todas as propriedades deste estado, é funcional da densidade eletrônica - $\rho(r)$;

A energia do estado fundamental de um sistema multieletrônico sob um dado potencial externo $v(r)$ pode ser escrito como:

$$E_v[\rho(r)] = \int v(r)\rho(r)d(r) + F[\rho]$$

onde F é denominado de funcional universal de ρ , que independe do potencial externo $v(r)$. Substituindo-se a função densidade exata deste estado por uma função aproximada $\tilde{\rho}$, E_v aumenta como demonstrado em:

$$\tilde{E} = \tilde{E}[\tilde{\rho}] = \int v(r)\tilde{\rho}(r)d(r) + F[\tilde{\rho}] \geq E_0 = E[\rho]$$

Não se conhece a expressão analítica desta função densidade, sendo esta determinação geralmente realizada partindo-se das equações de *HF* e através de um processo de autoconsistência determinar qual a melhor função ρ que minimizaria a energia total do sistema.

A qualidade dos resultados obtidos é função principalmente do tipo do potencial $V_{XC}(r)$, definido como a derivada funcional da energia total de troca-correlação:

$$V_{XC}(r) = \frac{\partial E_{XC}[\rho(r)]}{\partial \rho(r)} \quad (5.35)$$

A grande vantagem em utilizar a densidade eletrônica é que as integrais de repulsão pelo efeito Coulomb são calculadas apenas sobre a densidade eletrônica, que é uma função tridimensional, com escalonamento N^3 . Além disto, podem ser incluídos efeitos de correlações eletrônicas, o que contribui para redução do custo computacional sem afetar a exatidão dos resultados.

Os principais modelos propostos nos últimos 60 anos para diferentes representações e correções para o funcional troca-correlação são apresentados a seguir.

Funcional de troca-correlação

O funcional $F[\rho]$ pode ser representado pela soma dos funcionais de energia cinética $T[\rho]$ e de repulsão eletrônica $V_{ee}[\rho]$, assim a energia eletrônica total torna-se:

$$E_{el}[\rho] = T[\rho] + \int \rho(r)v(r)d(r) + V_{ee}[\rho] \quad (5.36)$$

Dentre os modelos mais importantes empregados para representar os componentes da equação 5.36 encontram-se:

Modelo de Thomas-Fermi

No modelo de Thomas-Fermi (TF) a representação para energia de sistemas moleculares é descrita como:

$$E_{el}^{TF}[\rho] = \frac{3}{10}(3\pi^2)^{2/3} \int \rho(r)^{5/3}d + \sum_A \int \rho(r)v(r)dr + \frac{1}{2} \int \int \frac{\rho(r_1)\rho(r_2)}{|r_2 - r_1|}dr_1dr_2 \quad (5.37)$$

onde o primeiro termo à direita representa $T[\rho]$, o potencial $v(r)$ na ausência de campo externo do segundo termo é exatamente o potencial nuclear $-Z_A/|R_A - r|$ e o último termo à direita representa a repulsão inter-eletrônica.

Modelo de Thomas-Fermi-Dirac

O modelo de Thomas-Fermi-Dirac (*M-TFD*) utiliza a representação de um gás uniforme de elétrons, onde adiciona-se a energia de troca $E_X[\rho]$ (parte quântica) à energia potencial de Coulomb, $U[\rho]$ (terceira parte à direita da equação 5.37, parte clássica de $V_{ee}[\rho]$):

$$V_{ee}[\rho] = U_{coul}[\rho] + E_X^{DIRAC}[\rho] \quad (5.38)$$

sendo

$$E_X^{DIRAC}[\rho] = \frac{3}{4} \left(\frac{3}{\pi} \right)^{1/3} \int \rho(r)^{4/3} dr \quad (5.39)$$

Modelo χ_α

O modelo χ_α substitui o operador não local exato de Fock por uma representação aproximada do potencial de troca local proporcional a $\rho^{1/3}$. Esta representação é dada em uma função de um parâmetro ajustável α , fortemente dependente da carga nuclear. Para sistemas atômicos este parâmetro varia entre 2/3 e 1:

$$V_{\chi\alpha} = \frac{3}{2}\alpha \left(\frac{3}{\pi} \right) \rho(r)^{1/3} \quad (5.40)$$

Desta forma, a energia de troca total pode ser escrita como:

$$E_{\chi\alpha} = \frac{3}{4}\alpha \left(\frac{3}{\pi} \right)^{1/3} \int \rho(r)^{4/3} dr \quad (5.41)$$

Apesar de desprezar a contribuição da correlação eletrônica, este modelo é o mais simples para introduzir a energia de troca dentro dos métodos de funcional de densidade, sendo útil em sistemas atômicos, moleculares e do estado sólido.

Multi Scattering χ_α

O modelo Multi Scattering χ_α (*MS- χ_α*) divide o espaço em regiões onde o potencial é esfericamente simétrico ou onde o potencial é constante¹³. Fornece bons

¹³Esta abordagem também é conhecida como *muffin-tin*.

resultados para sistemas que apresentam simetria similar a uma esfera, mas falha em sistemas moleculares que exigem uma maior flexibilidade do potencial.

Modelo Hartree-Fock Slater

O modelo Hartree-Fock Slater (*HFS*) é um procedimento numérico variacional que utiliza a função local como funcional de troca, mas evita regiões onde o potencial é esfericamente simétrico ou constante, sendo válido para sistemas onde a densidade de carga varia lentamente em uma escala atômica, com tratamento exato para a energia cinética. A energia total de troca-correlação é obtida integrando o espaço total desta molécula, conforme equação 5.42.

$$E_{XC}^{\chi\alpha} = \frac{3}{4}\alpha \left(\frac{3}{\pi}\right)^{1/3} \int \rho(r)^{4/3} dr \quad (5.42)$$

A introdução do efeito de correlação eletrônica no modelo de gás uniforme de elétrons, leva a uma aproximação mais geral, conhecida por Aproximação da Densidade Local (*LDA*)¹⁴, que representa o potencial de troca-correlação com caráter local.

BLYP

Método do funcional de correlação de Becke, com troca de Lee, Yang e Parr.

B3LYP

Método onde o funcional de correlação de Becke possui três termos.

5.5.5 Método semi-empírico

Os métodos semi-empíricos possuem a mesma estrutura que os métodos *ab-initio*, ou seja também apresentam um hamiltoniano e uma função de onda. No entanto, nos métodos semi-empíricos algumas informações são aproximadas ou completamente omitidas, como algumas integrais de elétrons duplos e considerações sobre os elétrons mais internos, o que simplifica o conjunto de base.

A correção destas simplificações e/ou omissões são realizadas através de parametrização com resultados obtidos experimentalmente ou através de métodos

¹⁴Do inglês *Local Density Approximation*.

ab-initio, porém esta parametrização não é tão importante quanto nos métodos clássicos.

A parametrização pode ser realizada para otimização de várias características, sendo as mais comuns a geometria molecular e a energia de formação. Alguns métodos foram ampliados para obtenção de outros parâmetros, como momento dipolo, calor de reação, potencial de ionização, e algumas propriedades mais específicas, como espectro eletrônico ou bandas de ressonância magnética nuclear.

A seguir serão apresentados os principais métodos semi-empíricos.

Método de Hückel

Um dos métodos mais antigos e simples, onde apenas são modelados os elétrons de valência π nos carbonos planares conjugados. Um parâmetro é utilizado para descrever a interação entre átomos ligados. Atualmente o método de Hückel é normalmente utilizado como exercício em sala de aula, pois os cálculos envolvidos podem ser feitos à mão.

Método extendido de Hückel

A grande vantagem deste método é que todos os orbitais de valência são modelados. Este método pode ser aplicado a praticamente todos os elementos da tabela periódica, e apesar de não ser muito preciso é uma das poucas alternativas para utilização em átomos inorgânicos. A grande vantagem deste método é para determinação de estruturas de banda e para simetria de orbitais, com um baixo custo computacional, porém não deve ser considerado para predição de geometria molecular devido a baixa qualidade.

CNDO

O método *CNDO*¹⁵ é o mais simples de uma família de métodos que modelam os orbitais de valência utilizando apenas um conjunto de base mínimo de orbitais do tipo Slater. Este método normalmente é o utilizado em pacotes *ab-initio* para gerar o passo inicial de otimização.

A melhoria da parametrização inicial levou ao desenvolvimento do método

¹⁵Do inglês *Complete Neglect of Differential Overlap*.

CNDO/2 e também do método *CNDO/S*, que reproduzem o espectro eletrônico molecular com boa predição de energias de excitação, mas com baixa qualidade de predição geométrica.

MINDO

O método *MINDO*¹⁶ pode ser subdividido em *MINDO/1*, *MINDO/2* e *MINDO/3*, sendo utilizado normalmente em pacotes *ab-initio* para gerar o passo inicial de otimização.

MNDO

O método *MNDO*¹⁷ é conhecido por fornecer resultados qualitativos razoáveis para diversos compostos orgânicos. Os principais desvios apresentados por este método são a energia de excitação subestimada, a barreira de ativação superestimada, a geometria normalmente não corresponde ao mínimo energético e as barreiras para rotação angular são muito pequenas.

ZINDO

O método *ZINDO*¹⁸ possui uma parametrização específica para reprodução de espectros eletrônicos, sendo de grande utilização para modelar sistemas contendo metais de transição, por ser um dos poucos métodos parametrizados para metais.

PRDDO

O método *PRDDO*¹⁹ foi desenvolvido para melhor aproveitamento do tempo computacional. Foi parametrizado para representar elementos a partir do Bromo, incluindo os metais de transição da terceira linha da tabela periódica, sendo utilizado normalmente em compostos inorgânicos, compostos organometálicos, cálculos do estado sólido e modelagem de polímeros.

Este método, apesar de gerar resultados com confiabilidade semelhante a outros métodos, não é normalmente encontrado nos programas semi-empíricos mais difundidos.

¹⁶Do inglês *Modified Intermediate Neglect of Differential Overlap*.

¹⁷Do inglês *Modified Neglect of Diatomic Overlap*.

¹⁸Do inglês *Zerner Intermediate Neglect of Differential Overlap*.

¹⁹Do inglês *Partial Redention of Diatomic Differential Overlap*.

AM1

O Modelo 1 de Austin (*AM1*)²⁰ é um dos modelos mais empregados em cálculos semi-empíricos, por apresentar um dos resultados mais exatos para moléculas orgânicas, permitindo incluir efeitos de solvatação.

Devido a grande utilização, as maiores qualidades e deficiências deste método são bem conhecidas, como a grande confiabilidade para moléculas que contenham o elemento alumínio e a baixa qualidade dos resultados para predição de barreiras rotacionais (aproximadamente 1/3 do valor experimental).

PM3

O método *PM3*²¹ utiliza praticamente as mesmas equações utilizadas pelo método *AM1*, porém com uma nova parametrização.

As grandes qualidades e limitações deste método também são bem conhecidas, como a excelente qualidade dos resultados de calor de formação e moléculas hipervalentes. Entretanto, existem problemas na predição de barreira de rotação em torno dos átomos *C – N*, nos estados eletrônicos dos compostos contendo o elemento germânio, entre outros.

PM3/TM

Este método é uma extensão do método *PM3*, onde foram incluídos os orbitais *d*, para aplicação em moléculas que contenham os metais de transição, tendo sido parametrizado quase que totalmente de forma a reproduzir geometrias caracterizadas pela técnica de difração de raios X.

SAM1

Este método pode ser considerado um intermediário entre os demais métodos semi-empíricos e os métodos *ab-initio*. Ele continua negligenciando algumas integrais incluídas nos métodos *ab-initio*, mas inclui mais orbitais que os demais métodos semi-empíricos, notadamente os orbitais do subnível *d*. Esta característica faz com que este método também possua tempo computacional intermediário, sendo maior que os

²⁰Do inglês *Austin Model 1*.

²¹Do inglês *Parameterization Method 3*.

demais métodos semi-empíricos mas bem menor que os métodos *ab-initio* com o menor conjunto de base. Os resultados são os melhores entre os métodos semi-empíricos, incluindo resultados de frequências vibracionais.

5.5.6 Termoquímica através de cálculos semi-empíricos

Devemos observar que os parâmetros dos orbitais moleculares pelos métodos *MINDO/3*, *MNDO*, *AM1* e *PM3* são otimizados para reprodução do calor de formação experimental²² e da geometria molecular à 25°C , e não para reproduzir a energia total ou geometria à 0K , como acontece com os métodos *ab-initio*.

Portanto, os valores calculados de calor de formação, constantes de força, frequências vibracionais são relacionados a temperatura de 25°C . Da mesma forma, o E_{zero} calculado também não corresponde a uma temperatura de 0K , devendo ser utilizado com critério.

O calor de formação, definido como E_{scf} , à 25°C pode ser obtida através da equação 5.43.

$$\begin{aligned} E_{\text{scf}} &= E_{\text{eq}} + E_{\text{zero}} + E_{\text{vib}}(0 \rightarrow 298.15) + E_{\text{rot}} + E_{\text{tra}} + pV \\ &= + \sum [-E_{\text{elec}}(\text{átomo}) + \Delta H_f(\text{átomo})] \end{aligned} \quad (5.43)$$

A equação 5.43 pode ser expressa de uma forma mais compacta através de outros conceitos termodinâmicos. Por exemplo, o calor padrão de formação pode ser calculado pela equação:

$$\Delta H = E_{\text{scf}} + (H_T - H_{298}) \quad (5.44)$$

onde H_T e H_{298} são as contribuições para o aumento de temperatura de 0K para a temperatura T e $298,15^\circ\text{C}$, respectivamente. Em outras palavras, a entalpia de formação é corrigida pela diferença de temperatura entre $298,15^\circ\text{C}$ e T .

No programa Mopac são utilizadas as seguintes variáveis [101]:

$$C_1 = \frac{hc}{kT} \quad (5.45)$$

²²O calor padrão de formação corresponde a energia de formação de um mol de produto à 25°C a partir dos seus elementos no estado padrão.

Considerando o número de onda ω_i em cm^{-1} temos:

$$\nu_i = \omega_i c \quad (5.46)$$

$$E_{\text{WJ}} = \exp(-h\nu_i/kT) = \exp(-\omega_i hc/kT) = \exp(-\omega_i C_1) \quad (5.47)$$

As constantes rotacionais A , B , and C , dadas em cm^{-1} , levam a seguinte equação:

$$A = \frac{h}{(8\pi^2 c I_A)} \quad (5.48)$$

Considerando a entalpia em cal/mol e a entropia em cal/mol/K as equações de 5.49 até 5.55 correspondem as contribuições energéticas vibracionais, as equações de 5.56 até 5.66 correspondem as contribuições rotacionais e as equações de 5.67 até 5.76 correspondem as contribuições translacionais [101].

Contribuição vibracional

$$Q_{\text{vib}} = \pi \sum_i \frac{1}{(1 - E_{\text{WJ}})} \quad (5.49)$$

$$E_0 = \frac{0.5 N_a h c}{4.184 \times 10^7} \sum_i \omega_i \quad (5.50)$$

$$= 1.429572 \sum_i \omega_i \quad (5.51)$$

$$E_{\text{vib}}(0 \rightarrow T) = N_a h c \sum_i \frac{\omega_i E_{\text{WJ}}}{1 - E_{\text{WJ}}} \quad (5.52)$$

$$= (R/k) h c \sum_i \frac{W_i E_{\text{WJ}}}{1 - E_{\text{WJ}}} \quad (5.53)$$

$$\begin{aligned} S_{\text{vib}} &= R(hc/kT) \sum_i \left\{ \frac{\omega_i E_{\text{WJ}}}{(1 - E_{\text{WJ}})} \right\} - R \sum_i \ln(1 - E_{\text{WJ}}) \\ &= R C_1 \sum_i \left\{ \frac{\omega_i E_{\text{WJ}}}{(1 - E_{\text{WJ}})} \right\} - R \sum_i \ln(1 - E_{\text{WJ}}) \end{aligned} \quad (5.54)$$

$$\begin{aligned}
 C_{\text{vib}} &= R(hc/kT)^2 \sum_i \left\{ \frac{\omega_i^2 E_{\text{WJ}}}{(1 - E_{\text{WJ}})^2} \right\} \\
 &= RC_1^2 \sum_i \left\{ \frac{\omega_i^2 E_{\text{WJ}}}{(1 - E_{\text{WJ}})^2} \right\}
 \end{aligned} \tag{5.55}$$

Contribuição rotação

Molécula Linear

$$Q_{\text{rot}} = (1/\sigma)(1/A)(kT/hc) = \frac{1}{\sigma AC_1} \tag{5.56}$$

$$E_{\text{rot}} = (2/2)RT \tag{5.57}$$

$$S_{\text{rot}} = R \ln \left(\frac{kT}{\sigma hc A} \right) + R \tag{5.58}$$

$$= R \ln \left(\frac{1}{\sigma AC_1} \right) + R \tag{5.59}$$

$$= R \ln \left(\frac{kT}{\sigma hc A} \right) + R \tag{5.60}$$

$$C_{\text{rot}} = (2/2)R \tag{5.61}$$

Molécula não linear

$$Q_{\text{rot}} = \frac{1}{\sigma} \left[\frac{\pi}{(ABC C_1^3)} \right]^{1/2} \tag{5.62}$$

$$E_{\text{rot}} = (3/2)RT \tag{5.63}$$

$$\begin{aligned}
 S_{\text{rot}} &= \frac{R}{2} \ln \left\{ \frac{\pi}{\sigma^2 ABC} \left(\frac{kT}{hc} \right)^3 \right\} + (3/2)R \\
 &= 0.5R3 \ln(kT/hc) - 2 \ln \sigma + \ln \left(\frac{\pi}{ABC} \right) + 3
 \end{aligned} \tag{5.64}$$

$$= 0.5R - 3 \ln C_1 - 2 \ln \sigma + \ln \left(\frac{\pi}{ABC} \right) + 3 \tag{5.65}$$

$$C_{\text{rot}} = (3/2)R \quad (5.66)$$

Contribuição translacional

$$Q_{\text{tra}} = \left(\frac{\sqrt{2\pi M k T / N_a}}{h} \right)^3 = \left(\frac{\sqrt{1.660540 \times 10^{-24} \times 2\pi M k T}}{h} \right)^3 \quad (5.67)$$

$$E_{\text{tra}} = (3/2)RT \quad (5.68)$$

$$H_{\text{tra}} = (3/2)RT + pV = (5/2)RT \quad (5.69)$$

$$\begin{aligned} S_{\text{tra}} &= (R/2)[5 \ln T + 3 \ln M] - 2.31482 \text{ cf. } p = 1 \text{ atm} \\ &= 0.993608[5 \ln T + 3 \ln M] - 2.31482 \end{aligned} \quad (5.70)$$

$$H_{\text{vib}} = E_{\text{vib}}(0 \rightarrow T) \quad (5.71)$$

Deve-se notar que o valor de E_{zero} não está incluso no cálculo do valor sw H_{vib} e que ω_i não é derivado das constantes de força à $0K$. Desta forma, para uma temperatura TK :

$$H_T = [H_{\text{vib}} + H_{\text{rot}} + H_{\text{tra}}] \quad (5.72)$$

enquanto para uma temperatura $T = 298.15K$:

$$H_{298} = [H_{\text{vib}} + H_{\text{rot}} + H_{\text{tra}}] \quad (5.73)$$

Pode-se notar também que H_T e H_{298} são equivalentes a seguinte equação:

$$(E_{\text{vib}} - E_{\text{zero}}) + E_{\text{rot}} + (E_{\text{tra}} + pV) \quad (5.74)$$

Pode-se observar ainda que na equação 5.74 as frequências normais são aquelas obtidas normalmente à $25^\circ C$, mas nunca aquelas obtidas à $0K$.

A entalpia padrão de formação, ΔH , pode ser calculada pela equação 5.44.

A energia interna padrão de formação, ΔU , pode ser calculada pela equação 5.75:

$$\Delta U = \Delta H - R(T - 298.15) \quad (5.75)$$

A energia livre de Gibbs, ΔG , pode ser calculada levando-se em conta a diferença do valor padrão com o valor calculado para um isômero ou para uma temperatura diferente, conforme equação 5.76, com o objetivo de cancelar os valores desconhecidos da entropia de formação dos elementos que constituem a molécula.

$$\Delta G = [\Delta H - TS] \text{ (estado considerado)} - [\Delta H - TS] \text{ (estado padrão)} \quad (5.76)$$

5.6 Métodos clássicos

Os métodos clássicos são todos aqueles onde não são considerados explicitamente os elétrons dos átomos ou moléculas do sistema, ou seja, não são realizados os cálculos da distribuição eletrônica. Nestes métodos são realizados cálculos de atração e repulsão entre corpos (átomos) de acordo com a lei de Newton, onde as forças envolvidas são determinadas de acordo com o modelo teórico ou empírico utilizado [104].

A história da utilização dos métodos clássicos em *QC* se confunde com a própria história dos computadores. A primeira simulação molecular relatada de um líquido foi realizada por Metropolis e colaboradores (1953) no histórico computador *MANIAC* no Centro de Pesquisas de Los Alamos, com utilização do método de Monte-Carlo (*MC*)²³ para simulação de líquidos. Posteriormente, Alder e Wainwright (1957) introduziram o método de Dinâmica Molecular (*MD*)²⁴ [105]. Estes métodos passaram a receber maior atenção em meados da década de 70, quando os computadores se tornaram poderosos o bastante e relativamente comuns para a comunidade científica [106].

De maneira geral o método *MD* é empregado para a obtenção de propriedades dinâmicas, em função do tempo, através das forças intermoleculares do sistema, enquanto que o método *MC* é empregado para situações onde o sistema não

²³Termo relacionado ao caráter aleatório do método para geração de novas configurações.

²⁴Do inglês *Molecular Dynamics*. Este método reproduz o comportamento espacial dos átomos ao longo do espaço de tempo simulado.

possua características dinâmicas, envolvendo o cálculo de alterações de forças intermoleculares [105, 107].

5.6.1 Equações de movimento

Equações de Newton

Para uma partícula isolada a equação do movimento é dada por:

$$\mathbf{F} = m\ddot{\mathbf{r}} \quad (5.77)$$

onde $\mathbf{F}$ é a energia total, m a massa e $\ddot{\mathbf{r}}$ é a segunda derivada do vetor de posição $\mathbf{r}$.

Em QC clássica estamos interessados em descrever a energia de um sistema com muitas partículas. Estas partículas podem sofrer influência de forças externas, como as paredes de um recipiente, e forças internas de atração e repulsão. Para um sistema multiparticulado a equação do movimento para cada partícula e para o conjunto de partículas podem ser vistas nas equações 5.78 e 5.79, respectivamente.

$$\sum_j \mathbf{F}_{ji} + \mathbf{F}_i^e = m_i \ddot{\mathbf{r}}_i \quad (5.78)$$

$$\sum_{j>i} \sum_i \mathbf{F}_{ji} + \sum_i \mathbf{F}_i^e = \sum_i m_i \ddot{\mathbf{r}}_i \quad (5.79)$$

O primeiro termo da equação 5.79 é consequência da segunda lei de Newton, onde toda ação causa uma reação igual e contrária, ou seja, $\mathbf{F}_{ij} + \mathbf{F}_{ji}$ é nulo. Portanto:

$$\mathbf{F}_i^e = \sum_i m_i \ddot{\mathbf{r}}_i \quad (5.80)$$

O tratamento acima não considera as restrições do sistema, restrições estas comuns em sistemas reais. Uma das maneiras de expressar as restrições de um sistema é expressar as equações conectando as coordenadas das partículas do tipo:

$$f(\mathbf{r}_1 + \mathbf{r}_2 + \mathbf{r}_3 \cdots t) = 0 \quad (5.81)$$

onde as restrições podem ser formuladas como:

$$(\mathbf{r}_i - \mathbf{r}_j)^2 - c_{ij}^2 = 0 \quad (5.82)$$

A equação 5.82 é o que garante a preservação das ligações entre os átomos de uma molécula. As restrições impõem duas dificuldades no cálculo de sistemas multiparticulados: **a)** as coordenadas e as equações de movimento não são totalmente independentes; e **b)** as forças de restrição não são conhecidas inicialmente.

A primeira dificuldade é solucionada através do emprego de coordenadas generalizadas. Em um sistema irrestrito, um sistema de N partículas possui $3N$ coordenadas independentes, ou graus de liberdade. Ao serem introduzidas k restrições, o número de graus de liberdade é reduzido para $3N - k$. Conseqüentemente, são introduzidas $3N - k$ variáveis independentes ($q_1, q_2 \cdots q_{3N-k}$) em cada uma das velhas coordenadas ($\mathbf{r}_1 \cdots \mathbf{r}_N$) (equação 5.83).

$$\mathbf{r}_N = \mathbf{r}_N(q_1, q_2 \cdots q_{3N-k}, t) \quad (5.83)$$

A segunda dificuldade exige a reformulação das equações com o objetivo de excluir as forças de restrição do conjunto de equações, resultando nas equações de movimento de Lagrange e/ou Hamilton.

Equações de Lagrange

Ao considerarmos um deslocamento infinitesimal da partícula, consistente com as forças e restrições impostas ao sistema naquele momento, não existe a possibilidade de alteração da força de restrição. A somatório do trabalho sobre todas as partículas é zero:

$$\sum_i \mathbf{F}_i \delta \mathbf{r}_i = 0 \quad (5.84)$$

A força $\mathbf{F}_i$ será resultado da soma da força aplicada $\mathbf{F}_i^a$ e da força da restrição $\mathbf{f}_i$. Substituindo esta relação na equação 5.84 teremos:

$$\sum_i \mathbf{F}_i^a \delta \mathbf{r}_i + \sum_i \mathbf{f}_i \delta \mathbf{r}_i = 0 \quad (5.85)$$

Normalmente o trabalho virtual das restrições é nula, e a condição de equilíbrio envolve apenas a força aplicada:

$$\sum_i \mathbf{F}_i^a \delta \mathbf{r}_i = 0 \quad (5.86)$$

A equação 5.86 é o princípio do trabalho virtual. Geralmente $\mathbf{F}_i^a \neq 0$ pois $\delta \mathbf{r}_i$ não é completamente independente, mas sim conectado as restrições. Torna-se necessário uma transformação de forma que $\delta \mathbf{r}$ seja zero. A equação de movimento de Newton (equação 5.77) pode ser reescrita como:

$$\mathbf{F}_i - \dot{\mathbf{p}}_i = 0 \quad (5.87)$$

onde $\mathbf{p}_i = m_i \dot{\mathbf{r}}_i$.

A equação 5.84 pode ser então reescrita como:

$$\sum_i (\mathbf{F}_i - \dot{\mathbf{p}}_i) \delta \mathbf{r}_i = 0 \quad (5.88)$$

Ao decompor a força em componentes aplicados e restrições se obtém:

$$\sum_i (\mathbf{F}_i^a - \dot{\mathbf{p}}_i) \delta \mathbf{r}_i + \sum_i \mathbf{F}_i \delta \mathbf{r}_i = 0 \quad (5.89)$$

onde o segundo termo pode ser desprezado, uma vez que a força das restrições é nula. Utilizando a relação $\mathbf{r}_i = \mathbf{r}_i(q_1, q_2 \dots q_n, t)$ e assumindo n coordenadas independentes a equação 5.89 pode ser reescrita como:

$$\sum_j \left\{ \left[\frac{d}{dt} \left(\frac{\partial E_{cin}}{\partial \dot{q}_j} \right) - \left(\frac{\partial E_{cin}}{\partial q_j} \right) \right] - Q_j \right\} \delta q_j = 0 \quad (5.90)$$

onde E_{cin} é a energia cinética e Q_j são os componentes de força generalizada definidos como:

$$Q_j = \sum_i \mathbf{F}_i \frac{\partial \mathbf{r}_i}{\partial q_j} \quad (5.91)$$

onde foi omitido o subscrito a do termo de força. Quando consideramos as restrições os termos q_j são mutuamente independentes. Como os termos ∂q_j são independentes de ∂q_k a equação 5.90 só será satisfeita quando existirem n equações:

$$\frac{d}{dt} \left(\frac{\partial E_{cin}}{\partial \dot{q}_j} \right) - \frac{\partial E_{cin}}{\partial q_j} = Q_j \quad (5.92)$$

A equação 5.92 é a equação de Lagrange para o caso geral onde as forças não são derivadas do potencial. Ao derivar a força do potencial teremos:

$$\mathbf{F}_i = -\nabla_i V \quad (5.93)$$

e as forças generalizadas podem ser obtidas de:

$$Q_j = -\sum_i \nabla_i V \frac{\partial \mathbf{r}_i}{\partial q_j} \quad (5.94)$$

que é idêntica à derivada parcial da função $-V(\mathbf{r}_1, \mathbf{r}_2 \cdots \mathbf{r}_N)$ em respeito à q_j :

$$Q_j = -\frac{\partial V}{\partial q_j} \quad (5.95)$$

que ao ser substituída na equação 5.92 gera:

$$\frac{d}{dt} \left(\frac{\partial E_{cin}}{\partial \dot{q}_j} \right) - \frac{\partial (E_{cin} - V)}{\partial q_j} = 0 \quad (5.96)$$

Como V é função apenas da posição e completamente independente da velocidade generalizada, pode ser incluída na derivada parcial em respeito a $\dot{q}$:

$$\frac{d}{dt} \left[\frac{\partial (E_{cin} - V)}{\partial \dot{q}_j} \right] - \frac{\partial (E_{cin} - V)}{\partial q_j} = 0 \quad (5.97)$$

Se definirmos um novo lagrangeano L como:

$$L = E_{cin} - V \quad (5.98)$$

a equação 5.97 pode ser reescrita como:

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_j} \right) - \frac{\partial L}{\partial q_j} = 0 \quad (5.99)$$

que corresponde a apresentação familiar das equações de Lagrange.

A grande vantagem das equações de Lagrange é que a necessidade dos vetores de força das equações de Newton são substituídas pelas funções escalares de E_{cin} e V . Esta propriedade permite que qualquer problema de mecânica seja transferido para coordenadas lagrangeanas pela representação de E_{cin} e V em coordenadas generalizadas, permitindo a obtenção de L que deve ser substituído na equação 5.99. Por exemplo, para o movimento de uma partícula em coordenadas cartesianas x , y e z obtemos $L = 1/2(x^2 + y^2 + z^2) - V$. Substituindo esta equação na 5.99 para cada uma das coordenadas obtemos as equações de movimento de Newton:

$$F_x = \frac{\partial V}{\partial x} = m\ddot{x} \quad (5.100)$$

$$F_y = \frac{\partial V}{\partial y} = m\ddot{y} \quad (5.101)$$

$$F_z = \frac{\partial V}{\partial z} = m\ddot{z} \quad (5.102)$$

Equações de Hamilton

Enquanto as equações de movimento Lagrange utilizam o tempo como parâmetro para as coordenadas generalizadas e para a velocidade, as equações de movimento de Hamilton utilizam as coordenadas generalizadas e momento generalizado como variáveis independentes:

$$p_i = \frac{\partial L(q_j, \dot{q}_i, t)}{\partial \dot{q}_i} \quad (5.103)$$

A mudança nas variáveis independentes pode ser obtida através de transformadas de Lagrange, resultando na definição da função hamiltoniana (H):

$$H(p, q, t) = \sum_i \dot{q}_i p_i - L(q, \dot{q}_i, t) \quad (5.104)$$

A diferencial da equação 5.104 em função de p , q e t é:

$$dH = \sum_i \frac{\partial H}{\partial q_i} dq_i + \sum_i \frac{\partial H}{\partial p_i} dp_i + \frac{\partial H}{\partial t} dt \quad (5.105)$$

A partir da equação 5.104 podemos obter:

$$dH = \sum_i \dot{q}_i dp_i + \sum_i p_i d\dot{q}_i - \sum_i \frac{\partial L}{\partial \dot{q}_i} d\dot{q}_i - \sum_i \frac{\partial L}{\partial q_i} dq_i - \frac{\partial L}{\partial t} dt \quad (5.106)$$

Os termos envolvendo $d\dot{q}_i$ da equação 5.106 são cancelados em consequência da seguinte definição de momento generalizado:

$$\sum_i p_i d\dot{q}_i - \sum_i \frac{\partial L}{\partial \dot{q}_i} d\dot{q}_i = 0 \quad (5.107)$$

A partir da equação de Lagrange é possível obter:

$$\frac{\partial L}{\partial q_i} = \dot{p}_i \quad (5.108)$$

A equação 5.106 pode ser reescrita como:

$$dH = \sum_i \dot{q}_i dp_i - \sum_i \dot{p}_i dq_i - \frac{\partial L}{\partial t} dt \quad (5.109)$$

Comparando a equação 5.105 com a equação 5.109 é possível obter as seguintes $2n + 1$ relações:

$$\dot{q}_i = \frac{\partial H}{\partial p_i} \quad (5.110)$$

$$\dot{p}_i = -\frac{\partial H}{\partial q_i} \quad (5.111)$$

$$\frac{\partial L}{\partial t} = -\frac{\partial H}{\partial t} \quad (5.112)$$

As equações 5.110 à 5.112 são conhecidas como equações de movimento canônicas de Hamilton.

Os problemas de mecânica podem ser resolvidos utilizando a equação 5.103 para determinar o lagrangeano do movimento, e então formulando o hamiltoniano através da equação 5.104. Substituindo o hamiltoniano obtido nas equações 5.110 até 5.112 teremos as equações de movimento de primeira ordem.

O hamiltoniano é constante em respeito ao movimento. Se o potencial for independente da velocidade e as equações que definem as coordenadas generalizadas (equação 5.79) sejam independentes do tempo, o hamiltoniano passa a corresponder a energia total do sistema:

$$H = E_{cin} + V \quad (5.113)$$

e as equações de movimento podem ser obtidas diretamente pela substituição do lado direito da equação 5.113 nas equações 5.110 até 5.112.

Enquanto as equações de Newton são suficientes para a simulação de sistemas moleculares simples, Dinâmicas Lagrangeanas e Hamiltonianas apresentam vantagens para sistemas mais complexos [105].

5.6.2 Agrupamentos (*Ensembles*)

Do ponto de vista puramente estatístico as propriedades macroscópicas do sistema, como pressão, temperatura ou densidade, representam a média de todos os estados quânticos possíveis. Para calcular as propriedades de um sistema real é conveniente definir um agrupamento, ou *ensemble*, que consiste em uma coleção imaginária de um número extremamente grande de sistemas em diferentes estados quânticos, mas com propriedades macroscópicas comuns. Para economia computacional é definido uma caixa imaginária contendo moléculas, e esta caixa é repetida ao longo dos eixos, minimizando a interferência causada pelos vértices da caixa de simulação (figura 5.8).

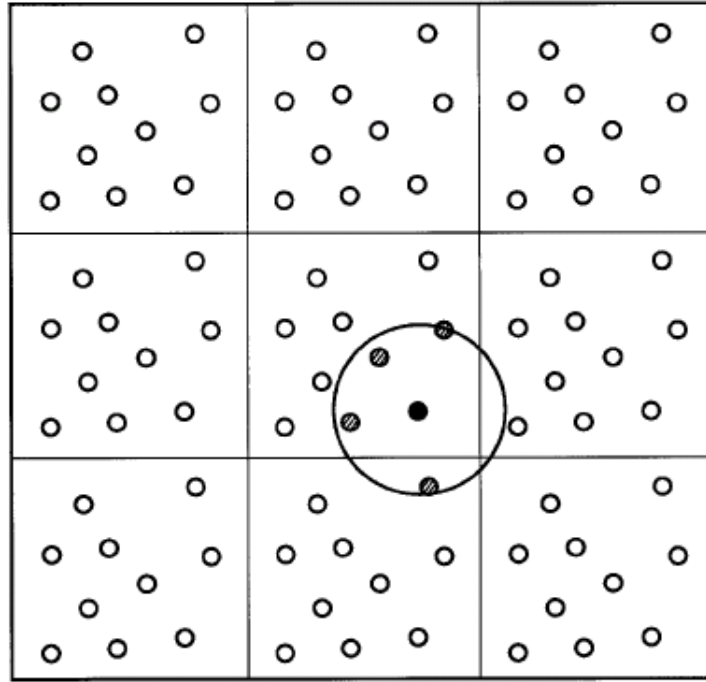


Figura 5.8: Duplicação da caixa de simulação

Por exemplo, cada sistema de um *ensemble* deve ter a mesma temperatura, pressão e número de moléculas que o sistema real que ele representa. A média de qualquer propriedade dinâmica do *ensemble* pode ser obtida pela relação:

$$\langle A \rangle = \sum_i A_i p_i \quad (5.114)$$

onde A_i é o valor da propriedade A no estado quântico i , p_i é a probabilidade de observar o i -ésimo estado e o termo da esquerda é o valor médio observado para o *ensemble*.

Desta forma, as propriedades do sistema no tempo estão relacionadas com o valor médio dos *ensembles*, através da relação conhecida como hipótese ergódica:

$$A_{1=\infty} = \langle A \rangle \quad (5.115)$$

A forma como p_i é determinada depende da escolha de quais propriedades macroscópicas são comuns a todos os elementos do *ensemble*. Os tipos mais comuns de *ensembles*, e suas características, estão apresentados na tabela 5.2.

Tabela 5.2: Principais *ensembles* termodinâmicos

<i>Ensemble</i>	Constantes	Z	p_i
Microcanônico	N, V e E	$\sum_i \delta(E_i - E)$	$\frac{\delta(E_i - E)}{Z_{NVE}}$
Canônico	N, V e T	$\sum_i e^{-\beta E_i(N,V)}$	$\frac{e^{-\beta E_i(N,V)}}{Z_{NVT}}$
Grão-Canônico	μ, V e T	$\sum_i e^{\beta N_i \mu} Z_{NVT}$	$\frac{e^{-\beta(E_i - \mu N_i)}}{Z_{\mu VT}}$
Isotérmico-Isobárico	N, P e T	$\sum_i e^{\beta P V_i} Z_{NVT}$	$\frac{e^{-\beta(E_i + P V_i)}}{Z_{NPT}}$

Os *ensembles* microcanônicos, canônicos e isotérmico-isobárico descrevem sistemas fechados para os quais não há mudança de número de partícula, enquanto o *ensemble* grão-canônico é apropriado para sistemas abertos, onde é possível troca de elementos (átomos ou moléculas). No limite termodinâmico todos os *ensembles* são equivalentes, sendo possível trocar suas características, e a escolha do melhor *ensemble* é uma questão de conveniência [6].

Muitos processos físicos não envolvem troca de partículas e ocorrem tanto em temperatura e volume quanto em temperatura e pressão constantes, sendo que nestes casos a seleção natural são os *ensembles* canônicos e isotérmico-isobárico, respectivamente [105].

Cada *ensemble* está associado a uma função termodinâmica característica. A entropia (S) pode ser obtida diretamente a partir da função de partição microcanônica:

$$S = K \ln Z_{NVE} \quad (5.116)$$

A função de Helmholtz (A) está associada ao *ensemble* canônico:

$$A = -kT \ln Z_{NVT} \quad (5.117)$$

O *ensemble* isotérmico-isobárico fornece a definição da função de Gibbs (G):

$$G = -kT \ln Z_{NPT} \quad (5.118)$$

A função termodinâmica de pressão está associada ao *ensemble* grão-canônico:

$$PV = -kT \ln Z_{\mu VT} \quad (5.119)$$

Propriedades termodinâmicas

Energia

A energia total (E) ou o Hamiltoniano (H) do agrupamento pode ser obtido a partir das médias entre a energia cinética (E_{cin}) e potencial (E_{pot}) do sistema:

$$E = \langle E_{cin} \rangle + \langle E_{pot} \rangle \quad (5.120)$$

A energia cinética pode ser estimada a partir dos momentos individuais das partículas. A energia potencial é a somatória das interações interpartículas, calculada segundo um determinado potencial de força.

Temperatura

A temperatura pode ser obtida a partir do teorema do virial. O teorema pode ser expresso em termos de momentos generalizados como:

$$\left\langle p_k \frac{\partial H}{\partial p_k} \right\rangle = kT \quad (5.121)$$

Para N átomos, cada qual com três graus de liberdade, obtemos a seguinte equação:

$$T = \frac{2\langle E_{cin} \rangle}{3Nk} \quad (5.122)$$

Podemos ainda considerar a T como sendo a média dos valores das contribuições instantâneas de temperatura (t):

$$t = \frac{2E_{cin}}{3Nk} \quad (5.123)$$

Pressão

A pressão também pode ser obtida a partir do teorema do virial, que pode ser expresso em termos de coordenadas generalizadas como:

$$\left\langle q_k \frac{\partial H}{\partial q_k} \right\rangle = kT \quad (5.124)$$

que pode ser reescrita na forma de:

$$PV = Nkt + \langle W \rangle \quad (5.125)$$

onde W é o virial interno definido como:

$$W = \frac{1}{3} \sum_i \sum_{j>i} \mathbf{r}_{ij} \mathbf{f}_{ij} = -\frac{1}{3} \sum_i \sum_{j>i} w(r_{ij}) \quad (5.126)$$

onde $w(r)$ função do virial para o par intermolecular:

$$w(r) = r \frac{du(r)}{dr} \quad (5.127)$$

Finalmente podemos definir uma função de pressão instantânea análoga a função de temperatura instantânea:

$$p = \rho kt + \frac{W}{V} \quad (5.128)$$

Excesso de potencial químico

O excesso de potencial químico μ^{ex} pode ser obtido através da inserção randômica de uma partícula em um sistema do qual já conhecemos a energia potencial. A diferença na energia potencial entre o novo valor e o anterior a inserção da partícula é chamada de E_{test} , e será utilizada nos cálculos a seguir.

Para os *ensembles* canônicos e grão-canônicos deve ser utilizada a equação 5.129.

$$\mu^{ex} = -kT \ln \langle \exp(-\beta E_{test}) \rangle \quad (5.129)$$

Devido as flutuações da temperatura cinética existe a seguinte equação para o *ensemble* microcanônico:

$$\mu_{ex} = -kT \ln \left[\frac{\langle t_{3/2} \exp(-\frac{E_{test}}{kt}) \rangle}{\langle t \rangle^{3/2}} \right] \quad (5.130)$$

Para o *ensemble* isotérmico-isobárico o cálculo é realizado considerando as flutuações no volume do sistema, conforme a seguinte equação:

$$\mu_{ex} = -kT \ln \left[\frac{\langle V \exp(-\beta E_{test}) \rangle}{\langle V \rangle} \right] \quad (5.131)$$

Capacidade calorífera em volume constante

Considerando U a energia interna do sistema, podemos expressar a capacidade calorífera em volume constante (C_V) como:

$$C_V = \left(\frac{\partial U}{\partial T} \right)_V \quad (5.132)$$

Para o *ensemble* microcanônico a C_V pode ser obtida através das equações 5.133 ou 5.134.

$$C_V = \frac{1}{\frac{2}{3Nk} \left(1 - \frac{2 \langle \delta E_{pot}^2 \rangle_{NVE}}{3N(kT)^2} \right)} \quad (5.133)$$

$$C_V = \frac{1}{\frac{2}{3Nk} \left(1 - \frac{2 \langle \delta E_{cin}^2 \rangle_{NVE}}{3N(kT)^2} \right)} \quad (5.134)$$

Para o *ensemble* canônico temos a equação 5.135:

$$C_V = \frac{\langle \delta E_{pot}^2 \rangle_{NVT}}{kT^2} + \frac{3Nk}{2} \quad (5.135)$$

E a equação 5.136 para o *ensemble* grão-canônico:

$$C_V = \frac{3Nk}{2} + \frac{1}{kT^2} \left(\langle \delta E_{pot}^2 \rangle_{\mu VT} - \frac{\langle \delta E_{pot} \delta N \rangle_{\mu VT}^2}{\langle \delta N^2 \rangle_{\mu VT}} \right) \quad (5.136)$$

Capacidade calorífica em pressão constante

A capacidade calorífica em pressão constante (C_P) pode ser definida como:

$$C_P = \left(\frac{\partial H}{\partial T} \right)_P \quad (5.137)$$

onde H é a entalpia do sistema.

Para o *ensemble* isobárico-ísotérmico:

$$C_P = \frac{\langle \delta (E_{cin} + E_{pot} + PV)^2 \rangle_{NPT}}{kT^2} \quad (5.138)$$

Coefficiente de pressão térmica

O coeficiente de pressão térmica (γ_V) pode ser definido como:

$$\gamma_V = \left(\frac{\partial P}{\partial T} \right)_V \quad (5.139)$$

Para o *ensemble* microcanônico:

$$\gamma_V = \frac{2C_V}{3} \left(\frac{1}{V} \frac{\langle \delta P \delta E_{pot} \rangle_{NVE}}{N (kT)^2} \right) \quad (5.140)$$

ou:

$$\gamma_V = \frac{2C_V}{3} \left(\frac{1}{V} \frac{\langle \delta P \delta E_{cin} \rangle_{NVE}}{N (kT)^2} \right) \quad (5.141)$$

onde C_V pode ser obtida pela equação 5.133 ou 5.134.

Para o *ensemble* canônico:

$$\gamma_V = \frac{1}{V} \left(\frac{\langle \delta E_{pot} \delta W \rangle_{NVT}}{kt^2} + Nk \right) \quad (5.142)$$

ou:

$$\gamma_V = \frac{\langle \delta E_{pot} \delta W \rangle_{NVT}}{kt^2} + Nk + \frac{NK}{V} \quad (5.143)$$

Para o *ensemble* grão-canônico:

$$\gamma_V = \frac{NK}{V} + \frac{\langle \delta E_{pot} \delta N \rangle_{\mu VT}}{VT} \left(1 - \frac{N}{\langle \delta N^2 \rangle_{\mu VT}} \right) + \frac{\langle \delta E_{pot} \delta W \rangle_{\mu VT}}{VkT^2} \quad (5.144)$$

Compressibilidade isotérmica

A compressibilidade isotérmica (β_T) pode ser obtida a partir da equação:

$$\beta_T = -\frac{1}{V} \left(\frac{\partial V}{\partial P} \right)_T \quad (5.145)$$

Para o *ensemble* microcanônico:

$$\beta_T^{-1} = \beta_s^{-1} - \frac{TV\gamma_V^2}{C_V} \quad (5.146)$$

onde C_v e γ_V são obtidos a partir das equações 5.133 ou 5.134 e equações 5.140 ou 5.141, respectivamente. O termo β_s , correspondente a compressibilidade adiabática, e pode ser obtido através da equação 5.147.

$$\beta_s^{-1} = \frac{2Nkt}{3V} + \frac{\langle F \rangle_{NVE}}{V} + \langle P \rangle_{NVE} - \frac{V\langle \delta P^2 \rangle_{NVE}}{kt} \quad (5.147)$$

O termo F da equação 5.147 pode ser obtido por:

$$F = \frac{1}{9} \sum_i \sum_{j>i} x(r_{ij}) \quad (5.148)$$

onde:

$$x(r) = r \frac{dw(r)}{dr} \quad (5.149)$$

Para o *ensemble* canônico:

$$\beta_T^{-1} = \frac{1}{V} \left(NkT + \langle W \rangle_{NVT} + \langle F \rangle_{NVT} - \frac{\langle \delta W^2 \rangle_{NVT}}{kT} \right) \quad (5.150)$$

ou:

$$\beta_T^{-1} = \frac{2NkT}{3V} + \langle P \rangle_{NVT} + \frac{\langle F \rangle_{NVT}}{V} - \frac{V \langle \delta P^2 \rangle_{NVT}}{kT} \quad (5.151)$$

Para o *ensemble* isotérmico-isobárico o valor da β_T está relacionada diretamente as flutuações do volume:

$$\beta_T = \frac{\langle \delta V^2 \rangle_{NPT}}{kTV} \quad (5.152)$$

Para o *ensemble* grão-canônico o valor da β_T está relacionada diretamente as flutuações do número de partículas:

$$\beta_T = \frac{V \langle \delta N^2 \rangle_{\mu VT}}{N^2 kT} \quad (5.153)$$

Coefficiente de expansão térmica

O coeficiente de expansão térmica (α_P) pode ser definido como:

$$\alpha_P = \frac{1}{V} \left(\frac{\partial V}{\partial T} \right)_P \quad (5.154)$$

Para o *ensemble* isotérmico-isobárico o valor pode ser obtido pela equação:

$$\alpha_P = \frac{\langle \delta V \delta (E_{cin} + E_{pot} + PV) \rangle_{NPT}}{NkT^2} \quad (5.155)$$

E para o *ensemble* grão-canônico:

$$\alpha_P = \frac{P\beta_T}{T} - \frac{\langle \delta E_{pot} \delta N \rangle_{\mu VT}}{NkT^2} + \frac{\langle E_{pot} \rangle_{\mu VT} \langle \delta N^2 \rangle_{\mu VT}}{N^2 kT^2} \quad (5.156)$$

Coefficiente de difusão

O coeficiente de difusão (D) pode ser obtido a partir da seguinte equação:

$$D = \lim_{t \rightarrow \infty} \frac{1}{6Nt} \left\langle \sum_{j=1}^N |r_j(t) - r_j(0)|^2 \right\rangle \quad (5.157)$$

onde t significa tempo.

Para resolver corretamente o lado direito da equação 5.157 deve ser utilizado condições periódicas [108, 109]. Alternativamente, se for assumido que o deslocamento atômico por unidade de tempo seja pequeno em relação ao tamanho do sistema, os átomos deslocados r' podem ser obtidos através da equação 5.158.

$$r'_{xj}(t) = r_{xj}(t) + \text{nint} \left(\frac{r'_{xj}(t - \Delta t) - r_{xj}(t)}{L_x} \right) L_x \quad (5.158)$$

onde $\text{nint}(x)$ é o inteiro mais próximo de x .

5.6.3 Potenciais de força intermoleculares

O cálculo de qualquer propriedade termodinâmica exige a solução das equações de movimento dos átomos do sistema, para obtenção dos valores das energias cinética e potencial [46].

O cálculo destas energias envolve um detalhamento a respeito da natureza de atração e repulsão entre os átomos. Estas forças, doravante denominada potencial de força intermolecular, podem ser resultado de interações de pequena e de longa distância.

Efeitos eletrostáticos, de indução e de dispersão são exemplos de interações de longa distância, onde a energia de interação é inversamente proporcional a distância interatômica. As interações de pequena distância são caracterizadas pelo decaimento exponencial na energia de interação em relação a distância intermolecular.

Em teoria, é possível calcular as interações intermoleculares a partir de cálculos de estrutura eletrônica, mas na prática este procedimento tem se mostrado válido apenas para sistemas compostos por moléculas relativamente simples [105].

A aproximação mais simples para o potencial de força intermolecular é tratar os átomos como esferas rígidas:

$$X = \begin{cases} \infty & r \leq \sigma \\ 0 & r > \sigma \end{cases} \quad (5.159)$$

onde σ é o diâmetro da esfera rígida.

O potencial de força apresentado na equação 5.159 foi utilizado nos primeiros relatos de simulação molecular, nos anos 50. Contudo, a ausência de contribuições de interações de atração impede predição de propriedades de líquidos reais, mesmo qualitativas [105].

Uma representação mais realista das interações intermoleculares é dada pelo potencial de Lennard-Jones, sendo este um dos potenciais mais utilizados em simulação molecular:

$$u(r) = \epsilon \left[\left(\frac{m}{n-m} \right) x_{-n} - \left(\frac{n}{n-m} \right) x_{-m} \right] \quad (5.160)$$

onde n e m são constantes, $x = r/r_m$ e r_m é a separação correspondente ao mínimo energético. O diâmetro da esfera rígida é relacionada separação de energia mínima:

$$\sigma = r_m \left(\frac{m}{n} \right)^{\frac{1}{n-m}} \quad (5.161)$$

A apresentação mais comum do potencial de Lennard-Jones é quando $n = 12$ e $m = 6$, ou seja:

$$u(r) = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] \quad (5.162)$$

A repulsão intermolecular e a interação de dispersão são representadas por diferentes dependências inversas das separações intermoleculares. O dependência do termo r^{-6} em relação ao termo de atração é consistente com a expansão da energia de dispersão em $1/r$. Contudo, o coeficiente de dispersão tipicamente é superestimado em um fator de 2 em relação aos valores experimentais, o que pode ser parcialmente explicado pela ausência de termos r^{-8} e r^{-10} . Não existe definição teórica para o termo de repulsão r^{-12} .

Diferente do que acontece com as moléculas apolares, a interação entre moléculas polares ou íons são de longa distância, que pode ser definida como aquele potencial que diminui mais lentamente que r^{-d} , onde r é a separação intermolecular e d a dimensionalidade do sistema. Tipicamente potenciais íon-íon e dipolo-dipolo são proporcionais a r^{-1} e r^{-3} , respectivamente. Consequentemente, devem ser incorporadas técnicas para o cálculo destas forças além dos limites do sistema que está sendo simulado.

De uma forma mais ampla, a energia de um sistema molecular resultante das interações atômicas pode ser expressa como:

$$\begin{aligned} E(r, l, \theta, \tau, \chi) = & \sum_{\text{não ligantes}} u_{nb}(r) + \sum_{\text{ligações}} u_b(l) + \sum_{\text{ângulo}} u_\theta(\theta) + \sum_{\text{dihedro}} u_\tau(\tau) + \\ & \sum_{\text{fora do plano}} u_\chi(\chi) + \sum u_{el} + \dots \end{aligned} \quad (5.163)$$

onde U_{nb} é o potencial dos elementos não ligados, u_b é o potencial das ligações, u_θ é o potencial dos ângulos lineares, u_ϕ é o potencial dos ângulos diedros, u_χ é o potencial dos ângulos fora do plano (torções impróprias) e u_{el} é o potencial de Coulumb.

A equação 5.163 inclui os termos diagonais, podendo ser expandida para inclusão dos termos cruzados.

A exatidão do potencia de força depende da exatidão das informações utilizadas em sua parametrização. Os comprimentos e ângulos de ligação podem ser obtidos por técnicas de raios-X, infravermelho, ressonância magnética-nuclear e microondas, enquanto os demais termos podem ser obtidos por detalhamento de resultados experimentais. Os valores também podem ser estimados por métodos *ab-initio* e semi-empíricos [105].

A seguir estão apresentados diversos campos de força comumente empregados em simulação molecular.

AMBER

O campo de força AMBER²⁵ foi parametrizado especificamente para proteínas e ácidos nucleicos. Este campo de força utiliza cinco termos ligantes e não ligantes com um refinado tratamento eletrostático, mas não inclui termos cruzados. Existe um programa para simulação de mecânica molecular com este mesmo nome.

CHARMM

CHARMM²⁶ é o nome de um campo de força e também de um programa de simulação de mecânica molecular. Inicialmente foi parametrizado para proteínas e ácidos nucleicos, mas atualmente possui nova parametrização, podendo ser utilizado em estudos de biomoléculas, efeitos de solvatação, empacotamento de cristais, análise vibracional. Este campo de força utiliza cinco termos de valência, sendo um eletrostático.

²⁵Do inglês *Assisted Model Building with Energy Refinement*.

²⁶Do inglês *Chemistry at Harvard Macromolecular Mechanics*.

CFF

O CFF²⁷ foi desenvolvido com o objetivo de fornecer bons resultados de geometria conformacional, espectros vibracionais, energia de cadeia e valores para entalpia de proeínas. Existem diversas extensões deste campo de força, como UBCFF²⁸, CVFF²⁹ e LCFF³⁰. Esta família de campos de força utiliza cinco ou seis termos de valência, sendo um deles eletrostático, e de quatro a seis termos cruzados.

CHEAT

O campo de força CHEAT³¹, foi parametrizado especificamente para carboidratos.

DREIDING

Campo de força multipropósito para moléculas orgânicas, que utiliza cinco termos, sendo um eletrostático. Este campo de força já foi muito utilizado para grandes sistemas biomoleculares, mas atualmente seu uso está sendo substituído por outros campos de força mais completos.

ECEPP

O ECEPP³² foi um dos primeiros campos de força específicos para peptídeos a ser substituído por outros mais expandidos. Utiliza três termos fixos, um termo de interação de Van der Walls e um termo eletrostático.

GROMOS

GROMOS³³ é o nome tanto de um campo de força quanto de uma pacote computacional. Este campo de força é comumente empregado para predição de

²⁷Do inglês *Consistent Force Field*.

²⁸Do Inglês *Ure-Bradley Consistent Force Field*.

²⁹Do inglês *Valence Version Consistent Force Field*.

³⁰Do inglês *Lynghy Consistent Force Field*.

³¹Do inglês *Carbohydrate Hydroxys Represented by External Atoms*.

³²Do inglês *Empirical Conformacional Energy Program for Peptides*.

³³Do inglês *Gronigen Molecular Simulation*.

movimento dinâmico de moléculas e misturas líquidas. Utiliza cinco termos, sendo um eletrostático.

MM1, MM2, MM3 e MM4

Campos de força de uso geral. Esta família de campos de força tem apresentado melhorias significativas na qualidade dos resultados em cada nova geração desenvolvida, sendo um dos campos de força mais utilizados atualmente. Esta família de campos de força utiliza de cinco a seis termos de valência, sendo um eletrostático, e de um à nove termos cruzados.

MMFF

O MMFF³⁴ é um dos mais novos campos de força publicados, possuindo aplicação genérica. Foi concebido inicialmente para utilização em simulações de dinâmica molecular, mas também está sendo utilizado para otimização geométrica. Utiliza cinco termos de valência, sendo um eletrostático, e um termo cruzado.

MOMEC

Campo de força para aplicação em coordenação de compostos contendo metais de transição. Utiliza quatro termos de valência e a interação metal ligante consiste em um termo de estiramento de ligação. A coordenação esférica é mantida por interações não-ligantes entre elementos ligantes, possuindo boa resposta para compostos de coordenação octaédrica.

OPLS

O OPLS³⁵ foi desenvolvido para simulação de misturas líquidas, sendo também utilizado para modelagem de dinâmica molecular de biomoléculas. Este campo de força utiliza cinco termos de valência, onde um dos termos é eletrostático.

³⁴Do inglês *Merck Molecular Force Field*.

³⁵Do inglês *Optimized Potentials for Liquid Simulation*.

Tripes

Também conhecido como campo de força Sybyl, pois foi concebido pela empresa Tripos Inc. para inclusão em seus programas Alchemy e Sybyl. Este campo de força foi desenvolvido para modelagem de moléculas orgânicas e bioquímicas, sendo também utilizado para análise *ComFa*, uma técnica de correlação estrutura-atividade. Possui cinco termos de valência, sendo um eletrostático.

UFF

Muitos campos de força foram criados para aplicação universal, mas o UFF³⁶ é um dos campos de força mais promissores para contemplação de todos os elementos da tabela periódica. Este campo de força atualmente é o mais utilizado para simulação de sistemas contendo elementos inorgânicos. Não utiliza termo eletrostático, possuindo quatro termos de valência. A literatura original recomenda utilizar cargas obtidas pelo método *Q-equilibrate*, mas estudos independentes revelam melhores resultados sem a utilização de cargas.

YETI

Este campo de força fornece bons resultados para interações entre não-ligantes, como a interação entre biomoléculas e substrato formado por pequenas moléculas. Como não recomendado para otimização geométrica, devido a baixa qualidade dos resultados, esta deve ser realizada através de outros campos de força, e então aplicar o YETI na geometria já otimizada para obtenção dos resultados. Este campo de força tem sido expandido para inclusão de metais e efeitos de interação soluto-solvente.

5.6.4 Métodos de Monte-Carlo

O método de *MC* consiste basicamente em uma estratégia estocástica, onde os objetivos são atingidos pelos seguintes passos:

1. gerar uma configuração de teste randomicamente, através do deslocamento, troca, remoção ou adição de moléculas;

³⁶Do inglês *Universal Force Field*.

2. calcular a mudança na energia ou outra propriedade da configuração teste, e avaliar o resultado de acordo com um critério específico;
3. comparar o critério com o número randômico e aceitar ou rejeitar a configuração teste correspondente.

É importante salientar que nem todos os estados farão uma contribuição significativa para as propriedades do sistema. Para determinar com acurácia as propriedades do sistema no tempo disponível para simulação, é importante amostrar os estados que fazem as contribuições mais significativas. Este objetivo pode ser atingido gerando-se as Cadeias de Markov, que consistem em uma seqüência de configurações testes, na qual a seqüência de sucessivos testes dependem apenas do antecessor imediato, de forma que um estado novo será aceito apenas se for mais favorável que o estado existente.

E relativamente fácil implementar o método de *MC* para outros tipos de agrupamentos porque a técnica de amostragem pode ser aplicada para qualquer problema para o qual seja possível construir a cadeia de Markov. A transformação para outros *ensembles* é obtida pela amostragem da variável, que é conjugada a um novo parâmetro, em adição a amostragem das coordenadas das partículas.

5.6.5 Métodos de Dinâmica Molecular

No método de *MD* as equações de movimento das moléculas são calculadas para que então seja gerado uma nova configuração, permitindo-se desta forma o cálculo de propriedades que dependam do tempo. De maneira geral o método *MD* é mais complexo para programar do que o método *MC*, apresentando também um maior custo computacional.

O método *MD* deve considerar os graus de liberdade vibracionais e rotacionais de moléculas poliatômicas na integração das equações de movimento, consideração esta que pode ser ignorada pelo método *MC*.

A generalização de *ensembles* na técnica de *MD* é mais complicada pelo fato de que as equações de movimento levam naturalmente ao *ensemble* microcanônico. Consequentemente, não existe um método bem definido para modificar as equações de movimento

5.6.6 Aplicação em equilíbrio de fase

As técnicas tradicionais de *MC* e *MD*, baseadas nos *ensembles* canônico e isobárico-isotérmico, respectivamente, não permitem a obtenção de potencial químico com a precisão necessária para cálculos de equilíbrio de fase, o que fazia com que as aplicações nesta área fossem bastante limitadas.

O cenário mudou com a introdução do *ensemble* de Gibbs associado ao método *MC* por Panagiotopoulos em 1987. No *ensemble* de Gibbs, o problema de obtenção de valores de potencial químico confiáveis e as incertezas causadas pela interface são eliminados pela representação das fases que coexistem no sistema em caixas de simulação separadas. O equilíbrio entre as fases, representadas pelas caixas de simulação (figura 5.9, é obtido através de deslocamentos moleculares e mudanças de volume das caixas de simulação, assim como trocas moleculares entre as caixas de simulação. O *ensemble* de Gibbs pode ser aplicado a fluídos puros, misturas binárias e ternárias de fluídos, fluídos iônicos e materiais poliméricos [105].

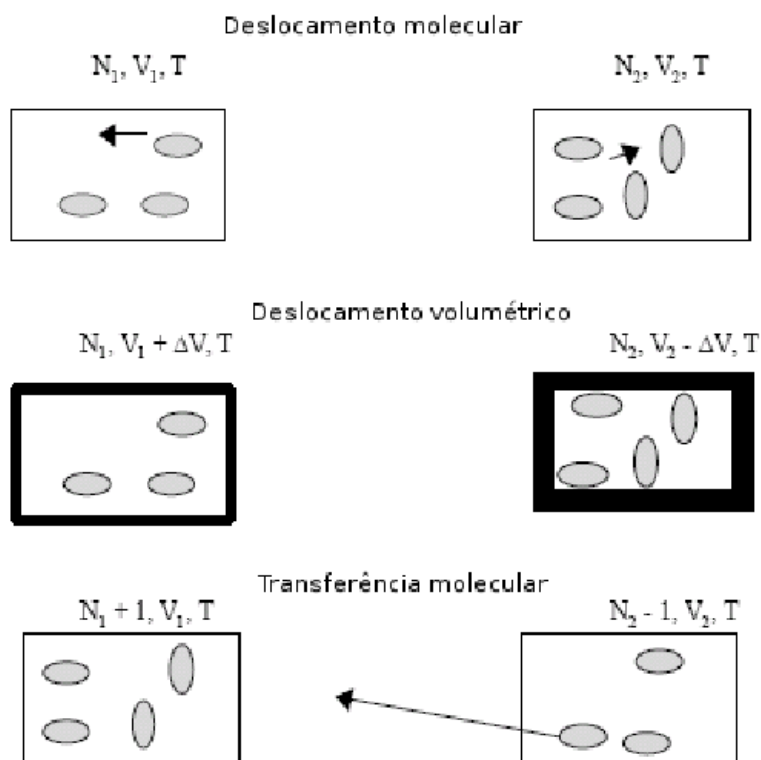


Figura 5.9: *Ensemble* de Gibbs.

5.7 Metodologia correlacional

Os métodos correlacionais, como o próprio nome leva a crer, correlacionam características dos materiais de referência, denominadas descritores³⁷ moleculares. Estes métodos podem ser divididos em *QSPR*³⁸, *QSAR*³⁹ e *QSAR-3D*⁴⁰.

5.8 Descrição do processo

O processo de desenvolvimento de pesquisa correlacionais segue o fluxograma apresentado na figura 5.10.

Pode ser observado na figura 5.10 que existem dois caminhos a serem seguidos: o desenvolvimento de um modelo que correlaciona descritores/propriedades e o emprego do modelo desenvolvido para predição de propriedades de interesse.

O desenvolvimento do modelo de correlação é um processo bem conhecido pela comunidade de química computacional. Na prática, é um trabalho bastante complicado que requer um sólido conhecimento de várias disciplinas científicas. A primeira etapa é a preparação de um conjunto de dados que contenha a lista de estruturas químicas e suas propriedades ou atividades associadas. Esta etapa é realizada com ajuda de banco de dados de propriedades moleculares, sendo que as informações podem ter origem experimental ou computacional. O ideal é utilizar mais de uma fonte de informação, para estimativa do desvio dos valores.

A etapa seguinte consiste na aplicação de métodos estatísticos para obtenção de modelos matemáticos que representem a relação entre os descritores utilizados e as propriedades/atividades investigadas. Os dados utilizados para desenvolvimento do modelo formam uma tabela, com os valores das variáveis dependentes (propriedades e/ou atividades), variáveis independentes (descritores moleculares), e um conjunto de parâmetros de controle para construção do modelo. O processo de desenvolvimento é iterativo, e os diversos modelos obtidos são comparados e analisados, mediante comparação com dados conhecidos, para seleção do modelo mais preditivo.

³⁷Descritores são características ligadas as moléculas. Podem ser características estruturais, físico-químicas, quânticas, bioquímicas, ou ainda derivadas destas.

³⁸Do inglês *Quantitative Structure/Properties Relationship*.

³⁹Do inglês *Quantitative Structure/Activity Relationship*.

⁴⁰Utilização de modelos tridimensionais para QSAR.

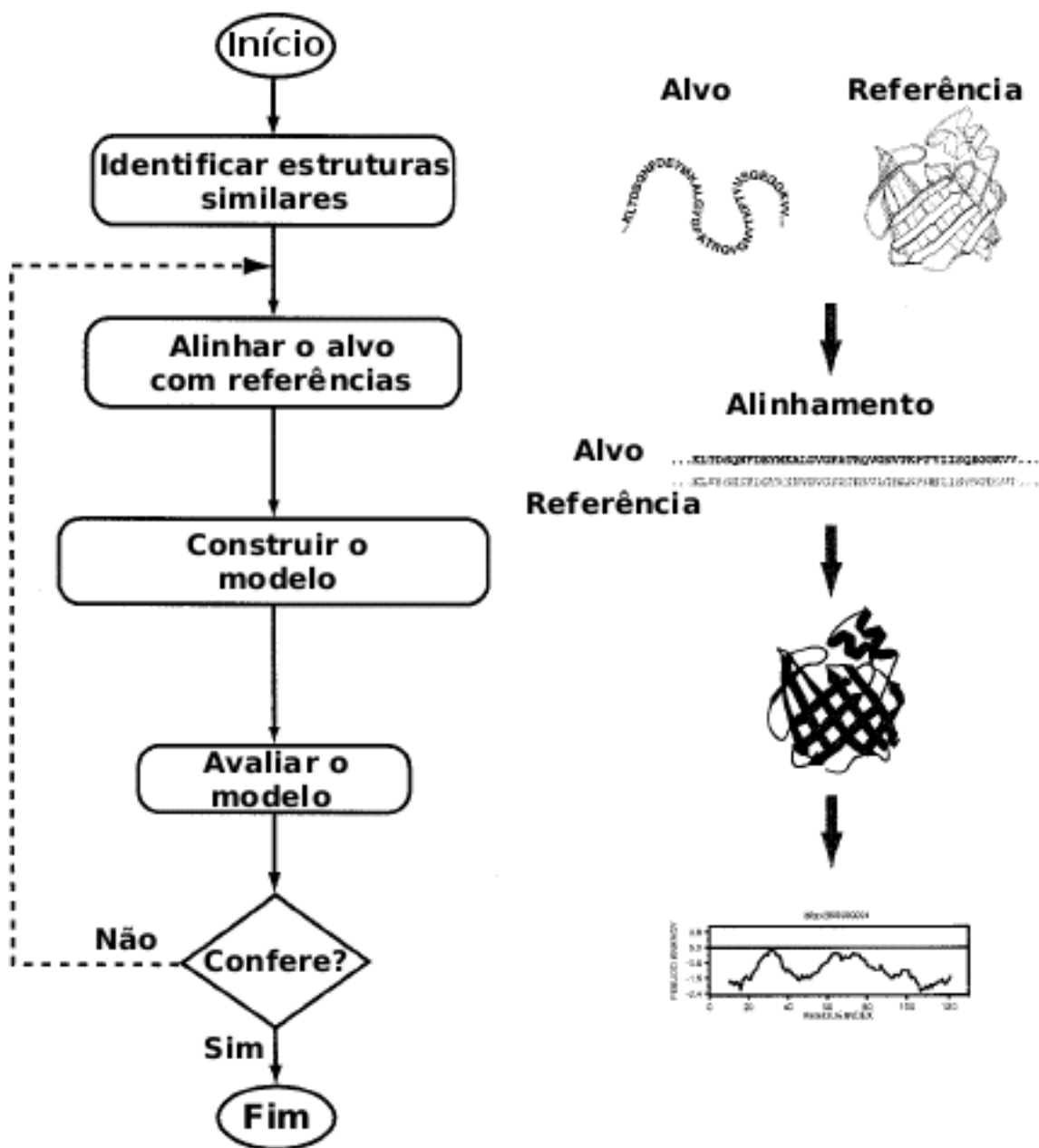


Figura 5.10: Modelagem QSAR

5.8.1 Etapas do processo

Pré processamento

Os dados iniciais a serem utilizados devem ser analisados detalhadamente. O banco de dados de informações moleculares deve possuir uma representação estrutural da molécula, em duas ou três dimensões, como forma de evitar entradas repetitivas, o que fatalmente ocorreria se fosse considerado apenas o nome do composto para sua

identificação.

Conversão de estruturas 2D em 3D

A conversão de estruturas 2D em 3D é uma etapa comum em estudos correlacionais, pois a maioria dos bancos de dados de informações moleculares armazenam a representação 2D do composto, mas a representação 3D é fundamental para o entendimento dos descritores quânticos e moleculares, além de serem indispensáveis, como arquivo de entrada, para execução de programas de química computacional. A conversão de duas para três dimensões pode ser realizada pela utilização de comprimentos padrão de ligação e estados de hibridização dos átomos.

Otimização conformacional

A representação 3D obtida na etapa anterior é submetida a um processo de minimização energética, para obtenção da representação conformacional ótima. O nível da análise conformacional depende da flexibilidade dos compostos estudados, mas de forma geral o menor estado energético é utilizado para reprodução dos resultados.

Cálculos de estrutura eletrônica

Os cálculos de estrutura eletrônica, sejam eles *ab-initio* ou semi-empíricos, são utilizados para obtenção da configuração eletrônica molecular, e através desta os descritores quânticos moleculares. Esta etapa envolve cálculos rigorosos para otimização geométrica da molécula e determinação de seu estado eletrônico. Alguns resultados obtidos serão utilizados diretamente como descritores moleculares, como calor de formação e momento dipolo, enquanto outros serão utilizados para calcular descritores adicionais, como parâmetros de polaridade, área parcialmente carregada positivamente, etc.

Cálculos dos descritores moleculares

As estruturas 2D e 3D e a distribuição eletrônica são utilizadas para o cálculos dos descritores moleculares que serão utilizados no desenvolvimento do modelo.

Desenvolvimento do modelo correlacional

Diversas técnicas podem ser utilizadas para o desenvolvimento do modelo matemático que correlaciona propriedade/atividade com a estrutura molecular. AS técnicas mais empregadas são:

- Modelos de regressão lineares;
- Mínimos quadrados parciais;
- Regressão por componentes principais; e
- Redes neurais

Modelos de regressão multilineares podem ser obtidos no espaço escalar dos descritores originais. A regressão por componentes principais e por mínimos quadrados parciais são aplicados no espaço dos principais descritores ortogonalizados e no espaço dos descritores alvo, respectivamente. As redes neurais são utilizadas para o desenvolvimento de modelos não-lineares. Existem ainda muitas estratégias heurísticas para seleção dos descritores na procura por uma correlação multivariada que modele a propriedade investigada. O modelo final obtido é analisado através de teste estatísticos, técnicas de validação cruzada e aplicação de conjunto de testes internos e externos.

5.9 Conclusões

Os diversos métodos de cálculo de QC são bem definidos e fundamentados em conceitos teóricos consolidados, de forma que mesmo suas deficiências são bem relatadas na literatura.

O resultado obtido para um determinado cálculo quântico (método e conjunto de base) deve ser sempre o mesmo, independente do programa utilizado. Os eventuais desvios devem ser consequência de arredondamento de cálculo, aparecendo apenas após o quinto ou sexto algarismo significativo.

Os métodos de mecânica molecular possuem um caráter mais dinâmico. É recomendável que sejam realizadas diversas simulações para que o usuário consiga acompanhar a dinâmica do sistema, ou seja, para que o efeito observado seja realmente consequência das condições consideradas. No caso da MD o sistema deve ser simulado

para diferentes intervalos de tempo e para a MC o sistema deve ser simulado para diferentes configurações iniciais.

Não existe, *a priori*, uma metodologia recomendada para seleção dos descritores moleculares a serem utilizados nos métodos correlacionais. O usuário deve considerar os aspectos teóricos que influenciam as propriedades investigadas para seleção destes descritores.

Capítulo 6

Ambiente computacional criado

6.1 Introdução

Neste capítulo são apresentados os *softwares* adquiridos e disponibilizados para uso ao longo deste trabalho.

Cabe destacar que os resultados obtidos para uma determinada molécula por um modelo teórico bem definido é o mesmo independente do *software* empregado. Entretanto a escolha do *software* é uma etapa importante, onde o usuário deve considerar fatores como custo, funcionalidade, eficiência computacional, facilidade de uso, automação e robustez.

6.2 Programas disponibilizados

Os programas estão agrupados em função da metodologia de cálculo, estando citados em ordem alfabética, com a descrição das principais funcionalidades e algumas observações importantes. O anexo A contem o detalhamento destes *softwares*.

A tabela 6.1 apresenta os programas de cálculo de estrutura eletrônica, a tabela 6.2 os programas de simulação de mecânica molecular e a tabela 6.3 os programas correlacionais. Alguns aplicativos de apoio estão apresentados na tabela 6.4.

Tabela 6.1: Programas de estrutura eletrônica.

Programa	SE ^a	HF ^b	DFT ^c	Observação
Aimpac [110]	X	X		
Ampac [111]	X			Deu origem ao Mopac
Amsol [112]	X			Ampac para solvatação
Dalton [91]		X	X	Propriedades magnéticas
Deft [113]			X	
Eceppak [114]	X		X	Voltado para peptídeos
Gamess [92]	X	X	X	
Mobosol [115]	X			Mopac para solvatação
Mopac [101]	X			
NWChem [93]	X	X	X	
Polyrate [116]	X			Análise de reações

^a Semi-empírico, ^b Hartree-Fock e ^c Teoria do Funcional de Densidade.

6.3 Banco de dados

Os programas citados anteriormente podem ser utilizados para obtenção de diversos parâmetros físico-químicos. Contudo, a obtenção destes parâmetros pode demandar um longo tempo computacional, requerendo ainda uma análise criteriosa para avaliação da exatidão dos resultados.

Uma alternativa é a utilização de banco de dados de propriedades físico-químicas para obtenção dos parâmetros de interesse. A tabela 6.5 apresenta alguns dos fornecedores de banco de dados de informações físico-químicas, o endereço eletrônico e o número aproximado de compostos disponíveis.

6.4 Conclusões

As ferramentas computacionais disponibilizadas permitem a realização de qualquer trabalho na área de química computacional, independentemente da metodologia desejada. O usuário deve, entretanto, considerar o esforço computacional requerido para realização do trabalho.

Tabela 6.2: Programas de mecânica molecular.

Programa	MD ^a	MC ^b	Observação
Alcmd [117]	X		
Allen e Tildesley [6]	X	X	
Amber [118]	X	X	Campo de força
B (ex-Biomer) [119]	X		
CCP5 [120]	X	X	
Charmm [121, 122]	X	X	Campo de força
Frenkel e Smit [123]	X	X	
Gromacs [124]	X	X	Moléculas bioquímicas e polímeros
Mindy [125]	X		
Moil [126]	X		Biomoléculas
Moldy [127]	X		Matéria condensada
Namd [128]	X	X	
Sadus [105, 129]	X	X	
Solvate [130]	X		Solvatação
Tinker [131]	X	X	
Towhee [132]		X	Equilíbrio de fases

^a Dinâmica Molecular e ^b Método de Monte-Carlo.

Tabela 6.3: Programas correlacionais.

Programa	Principal funcionalidade
Absolv [133]	Parâmetros de solvatação
Asad [134]	Reações atmosféricas
Gepol [135]	Superfície e volume molecular
Kintecus [136]	Reações, com interface com o MS Excel
Loop [137, 138]	Geração de coordenadas atômicas
Protarch [139]	Genérico

Tabela 6.4: Ferramentas de apoio.

Programa	Principal funcionalidade
Acclabs [140]	Desenho e manipulação com descritores
Babel ou OpenBabel [141]	Manipulação de formatos de arquivos
Biocore [142]	integração via rede
Chemoffice [143]	Desenho e manipulação com descritores
MMFF [144]	Campo de força
Molda [145]	Desenho e manipulação com descritores
Molden [146]	Visualização de resultados
Molmol [147]	Visualização de macromoléculas
Rasmol [148]	Visualização de biomoléculas

Tabela 6.5: Fornecedores de banco de dados de informações químicas

Empresa	Fonte	Nº de comp.
ACB Blocks	http://www.acbblocks.com	61237
Akos	http://www.akosgmbh.com	161316
AnalytiCon Disc.	http://www.ac-discovery.com	5438
Arkive	http://ark.chem.ufl.edu/pages/arkive.htm	28504
Array BioPharma	http://www.arraybiopharma.com	517
Asinex	http://www.asinex.com	348203
Aurora	http://www.aurora-feinchemie.com	25295
BioFocus	http://www.biofocus.com	23836
Chembridge	http://www.chembridge.com	387859
ChemDiv	http://www.chemdiv.com	451205
Chemical Block	http://www.chemical-block.com	1993
ChemStar	http://www.chemstar.ru	60066
Chem. Nat.	http://chimiotheque-nationale.ujf-grenoble.fr/GDS/	14946
Enamine	http://www.enamine.relc.com	385175
ICOA	http://www.univ-orleans.fr/icoa/	2811
IFLab	http://www.iflab.kiev.ua	100392
InterBioScreen	http://www.ibscreen.com	345060
Key Organics Ltd	http://www.keyorganics.ltd.uk	42414
Maybridge	http://www.maybridge.com	71041
MDPI	http://www.mdpi.org	8853
MSDiscovery	http://www.msdiscovery.com	1982
Nanosyn	http://www.nanosyn.com	65184
NCI	http://dtp.nci.nih.gov	244321
Pharmeks	http://www.pharmeks.com	83992
Prestwick	http://www.prestwickchemical.com	876
Sigma-Aldrich	http://www.sigma-aldrich.com	162171
Specs	http://www.specs.net	178492
TimTec	http://www.timtec.net	122238
TOSLab	http://www.toslab.com	21004
Tripos	http://leadquest.tripos.com	82370
VitasMLab	http://www.vitasmlab.com	193993
Worldmolecules	http://www.worldmolecules.com	33259

Capítulo 7

Avaliação das potencialidades e aplicabilidade

7.1 Introdução

Este capítulo apresenta diversas informações fundamentais para o sucesso de todo projeto que envolva a utilização de QC para obtenção e/ou análise de informações, tais como a estimativa dos recursos computacionais necessários, os fatores de correção a serem aplicados em cálculos de frequências vibracionais, os desvios no cálculo de diversos parâmetros físico-químicos, entre outras.

Também são apresentados neste capítulo os procedimentos e os resultados obtidos para a determinação da constante de equilíbrio e do coeficiente de auto-difusão, como exemplo de utilização das técnicas abordadas ao longo deste trabalho para obtenção de propriedades voltadas ao desenvolvimento de processos químicos.

7.2 Recursos computacionais

Como já destacado no início do capítulo 5 o pesquisador que utiliza QC deve considerar atentamente o poder de processamento, o espaço em disco e o tempo de processamento disponíveis. O ideal é que o usuário saiba prever quais os recursos computacionais necessários para a realização dos cálculos.

A utilização dos recursos computacionais também pode ser definida pelo usuário. Por exemplo, muitos dos programas *ab-initio* podem utilizar o disco rígido para

armazenamento de constantes que são calculadas uma vez e utilizadas constantemente, como as integrais que descrevem a sobreposição entre diversas funções de base. Existe a opção para que o usuário especifique que estas constantes sejam calculadas todas as vezes que forem necessárias, reduzindo os requerimentos de espaço em disco, mas aumentando o tempo de processamento.

7.2.1 Tempo de processamento

Os métodos *ab-initio* mais simples possuem um escalonamento na ordem de $N^{x>4}$, onde N é o número de orbitais moleculares. Vamos considerar, como exemplo, que o cálculo de otimização geométrica de um determinado monômero leve 20 minutos. O cálculo da otimização geométrica de um trímero, que possui três vezes o número de átomos do monômero, será $3^4 * 20$ minutos, ou aproximadamente 27 horas. A otimização geométrica de uma cadeia polimérica com quinze unidades levará $15^4 * 20$ minutos, ou aproximadamente 2 anos!

Existem muitos métodos que podem ser empregados para obtenção dos mesmos parâmetros de interesse, e o ideal é que o usuário saiba avaliar qual o melhor método a ser utilizado, considerando a melhor relação entre o tempo computacional e a exatidão dos resultados obtidos. A tabela 7.1 apresenta algumas estimativas de tempo computacional necessário para diversos métodos de cálculo, onde M é o número de átomos, L é o comprimento de um lado da caixa de simulação molecular considerando condições periódicas, A é número de orbitais espaciais ativos, e N o número de orbitais considerados no cálculo.

7.2.2 Interação do usuário com o programa

Uma consideração importante é quanto a facilidade de uso do programas e sua versatilidade e eficiência. Normalmente os programas mais fáceis de serem utilizados possuem opções limitadas de simulação e implementações numéricas não muito otimizadas.

É comum o usuário iniciar seus trabalhos nesta área utilizando programas com interface gráfica fácil e intuitiva, mas conforme vão surgindo problemas mais complexos este usuário tende a migrar para programas mais eficientes, porém menos amigáveis. Esta migração faz mais sentido do que adquirir um supercomputador com um programa com interface gráfica para resolver um problema que pode ser solucionado em uma

Tabela 7.1: Estimativa do tempo computacional em função do método.

Método	Escalonamento	Método	Escalonamento
DFT ^a	N	CISD	N^6
MM	M^2	MP4	N^7
MD	M^2 ou L^6	CC3	N^7
Semi-empírico ^b	N^2	MP5	N^8
HF ^c	$N^2 - N^6$	CISDT	N^8
Semi-empíricos ^d	N^3	CCSDT	N^8
HF	N^3	MP6	N^9
DFT	N^3	MP7	N^{10}
QMC ^e	N^3	CISDTQ	N^{10}
MP2	N^5	CASSCF	$A!$
CC2	N^5	CI	$N!$
MP3	N^6	QMC ^f	$N!$
CCSD	N^6		

^a Escalonamento linear para moléculas muito grandes, ^b Para moléculas pequenas e médias limitadas pelas integrais, ^c Dependendo da simetria e exatidão das integrais, ^d Para moléculas muito grandes limitadas pela inversão matricial, ^e Aplicando a matriz inversa de Slater e ^f Sem inversão da matriz de Slater.

estação de trabalho com outro programa executado por linha de comando.

7.3 Correção da vibração molecular

A determinação do espectro vibracional molecular e das intensidades de seus picos são importantes para determinar a estrutura geométrica molecular e o ambiente eletrônico desta molécula. Estes valores podem ser obtidos experimentalmente através das técnicas espectroscópicas no infravermelho e de Raman, ou então computacionalmente através de métodos de estrutura eletrônica *ab-initio* ou *DFT*.

As vibrações moleculares de maior intensidade correspondem ao estiramento de ligações, as vibrações de intensidade mediana correspondem as torções de ligações, e as vibrações de baixa intensidade estão relacionadas aos movimentos torcionais moleculares. A tabela 7.2 apresenta os fatores de correção para as frequências vibracionais obtidas através de diferentes métodos de *QC*. Estes valores devem ser

aplicados em função do método de cálculo, não sofrendo influência do programa utilizado para obtenção dos resultados, pois estes desvios são inerentes ao detalhamento teórico de cada método.

Tabela 7.2: Fator de correção de frequência vibracional

Método/conjunto de base	Fator de correção
HF/3-12G	0,9085
HF/6-31G*	0,8953
ROHF/6-31G*	0,893
HF/6-31+G*	0,8970
HF/6-31G**	0,8992
HF/6-311G**	0,9051
HF/aug-cc-pVTZ	0,9
MNDO	0,9
AM1	0,95
SAM1	0,993
MP2/6-31G*	0,9434 ate 0,95
MP2/6311G**	0,9496
DFT	0,96
BLYP/6-31G*	0,9945
BP86/6-31G	0,9914
B3PW91/6-31G*	0,9573
B3LYP/6-31G	0,9614

7.4 Exatidão dos resultados computacionais

Em química analítica pode ser realizado um determinado número de análises independentes, sendo possível desta forma estimar o desvio padrão experimental. Em química computacional, repetir o mesmo procedimento leva a obtenção do mesmo resultado, com exceção ao método de Monte-Carlo¹.

Existem diversos artigos e referências citando o erro das técnicas de química computacional[149, 150]. Se o usuário não possuir nenhuma referência disponível, deve realizar alguns testes preliminares para estimar qual o melhor método para seus

¹Se o usuário especificar a mesma semente para geração dos números aleatórios os resultados também serão os mesmos

objetivos específicos, preferencialmente empregando diversos métodos em compostos similares ao sistema estudado.

A tabela 7.3 apresenta o erro esperado para diversas propriedades obtidas por técnicas de *QC* em relação aos resultados obtidos experimentalmente. A análise desta tabela auxilia na seleção do melhor método para obtenção da propriedade de interesse, e sua elaboração foi realizada através da análise do banco de dados da literatura de química quântica (*QCLDB*²) [151].

7.5 Estruturas de Transição

A estrutura de transição de uma molécula corresponde ao estado mais energético entre duas estruturas distintas desta molécula, dentro do mecanismo reacional.

De uma maneira geral, uma reação ocorre em cinco etapas distintas:

1. Reagentes isolados, sem interações intermoleculares;
2. Complexo van der Waals entre os reagentes;
3. Estruturas de transição entre reagentes/produtos;
4. Complexo van der Waals entre produtos;
5. Produtos isolados, sem interações intermoleculares.

Matematicamente, o estado de transição corresponde ao ponto onde a derivada da energia, em função da geometria molecular, é zero, e a segunda derivada da energia, em função de mudança de geometria de um dos elétrons, é positiva com inclinação negativa.

O ponto intermediário na inversão de uma estrutura cíclica, a estrutura obtida quando se considerada uma simetria superior a esperada para o composto, entre outras situações, também corresponde a um máximo energético que atende a definição matemática, mas não são estruturas de transição de um mecanismo reacional. Portanto a identificação da estrutura de transição deve ser criteriosa e incluir outros parâmetros, quando disponíveis.

²Do inglês *Quantum Chemistry Literature DataBase*.

Tabela 7.3: Resultados computacionais *versus* resultados experimentais

Método	Propriedade	Exatidão
MM2	ΔH_f^0	$0,5kcal/mol$
	Comprimento de ligação	$0,01 \text{ \AA}$
	Ângulo de ligação	$1,0^\circ$
	Ângulo diedro	$8,0^\circ$
	Dipolo	$0,1D$
MM3	ΔH_f^0	$0,6kcal/mol$
	Comprimento de ligação	$0,01 \text{ \AA}$
	Ângulo de ligação	$1,0^\circ$
	Ângulo diedro	$5,0^\circ$
	Dipolo	$0,07D$
CFF	ΔH_f^0	$2,0kcal/mol$
	Comprimento de ligação	$0,01 \text{ \AA}$
	Ângulo de ligação	$1,0^\circ$
	Energia de sorção	$5,0kcal/mol$
CNDO	ΔH_f^0	$200,0kcal/mol$
INDO/1	ΔH_f^0	$100,0kcal/mol$
MINDO/3	ΔH_f^0	$5,0kcal/mol$
MNDO	ΔH_f^0	$11,0kcal/mol$
	Comprimento de ligação	$0,48 \text{ \AA}$
	Ângulo de ligação	$4,3^\circ$
	Dipolo	$0,3D$
	Pot. de ionização	$0,8eV$
MNDO/d	ΔH_f^0	$5,0kcal/mol$
	Dipolo	$0,4D$
	Pot. de ionização	$0,6eV$
AM1	ΔH_f^0	$8,0kcal/mol$
	Energia total	$18,8kcal/mol$

Tabela 7.4: Resultados computacionais *versus* resultados experimentais (continuação).

Método	Propriedade	Exatidão
AM1	Comprimento de ligação	0,048 Å
	Ângulo de ligação	3,3°
	Dipolo	0,5D
	Pot. de ionização	0,6eV
PM3	ΔH_f^0	8,0kcal/mol
	Energia total	17,2kcal/mol
	Comprimento de ligação	0,037 Å
	Ângulo de ligação	3,9°
SAM	Dipolo	0,6D
	Pot. de ionização	0,7eV
	ΔH_f^0	8,0kcal/mol
	Pot. de ionização	0,4eV
BLYP/6-31G**	Energia de reação	9,95kcal/mol
BLYP/6-31+G	Energia total	3,9kcal/mol
BLYP/6-311+G	Energia total	3,9kcal/mol
	Pot. de ionização	0,27eV
BLYP/DZVP	Energia de reação	7,73kcal/mol
B3LYP	ΔH_f^0	2,0kcal/mol
B3LYP/6-31G	Energia total	7,9kcal/mol
B3LYP/6-311+G	Comprimento de ligação	0,020 Å
	Ângulo de ligação	1,4°
	Energia total	3,1kcal/mol
	Comprimento de ligação	0,017 Å
B3PW91/6-311+G	Ângulo de ligação	1,4°
	Pot. de ionização	0,191eV
	Dipolo	0,5D
HF/STO-3G	Energia total	93,3kcal/mol

Tabela 7.5: Resultados computacionais *versus* resultados experimentais (continuação).

Método	Propriedade	Exatidão
HF/STO-3G	Ângulo de ligação	1,7°
	Comprimento de ligação	0,055 Å
HF/3-21G	ΔH_f^0	7,0kcal/mol
	Dipolo	0,4D
HF/3-21G*	ΔH_f^0	4,0kcal/mol
	Energia total	51,0kcal/mol
	Dipolo	0,2D
	Comprimento de ligação	0,032 Å
HF/6-311+G	Ângulo de ligação	1,4°
	Ângulo de ligação	1,3°
	Comprimento de ligação	0,035 Å
	Ângulo de ligação	2,2°
MP2/3-21G	Comprimento de ligação	0,044 Å
	Ângulo de ligação	1,5°
MP2/6-31G	Comprimento de ligação	0,048 Å
	Energia total	11,4kcal/mol
MP2/6-331G	Energia total	8,9kcal/mol
	Energia total	1,6kcal/mol
G1	Energia total	1,2kcal/mol
G2	Energia total	1,5kcal/mol
G2(MP2)	Energia total	

O detalhamento do deslocamento do núcleo atômico do estado de transição até o estado de mínima energia é conhecido como Coordenada de Reação, e permite o mapeamento energético ao longo do mecanismo reacional.

7.6 Constante de equilíbrio

O conhecimento da constante de equilíbrio é muito importante no estudo de mecanismos reacionais [152]. Além desta constante permitir prever a concentração das espécies presentes na mistura reacional, ela está conectada a constante de reação de duas maneiras distintas: **a)**- a existência de equilíbrio entre espécies estáveis em mecanismos reacionais de múltiplas etapas; e **b)**- o equilíbrio entre reagentes e complexos ativados de acordo com a teoria do estado de transição [32]. Desta forma, a constante de equilíbrio

total corresponde ao produto das constantes de reação para cada uma das etapas do mecanismo reacional.

A exatidão da constante de equilíbrio obtida computacionalmente depende dos valores energéticos, dos parâmetros geométricos e da estrutura eletrônica de cada um dos compostos presentes no equilíbrio. A constante de equilíbrio k_{eq} (ou $\Delta_r G^0$) pode ser obtida a partir do tratamento termodinâmico apresentado nas equações 7.1 e 7.2:

$$K_{eq} = \exp -(\Delta_r G^{IG})/(RT) \quad (7.1)$$

$$G = E_{\text{elettronica}} + ZPE + (H_T - H_0) - TS \quad (7.2)$$

onde $\Delta_r G^{IG} = \Delta_r G^0$ para o gás ideal (relativo a uma reação com coeficientes estequiométricos dados e espécies gasosas no estado padrão), H e S são obtidas através de função de partição molecular, R é a constante dos gases e T é a temperatura. A energia de reação de Gibbs deve corresponder ao valor calculado para a reação na temperatura de interesse.

Os coeficientes de atividade podem ser expressos como concentração molar ou fugacidade, resultando em uma constante de equilíbrio adimensional. Esta constante de equilíbrio é apenas uma aproximação, e seu valor diverge do valor exato quanto mais o coeficiente de atividade difere do valor unitário.

A principal dificuldade na predição de taxas de reação por métodos computacionais pode ser exemplificada através da análise da figura 7.1, que apresenta a coordenada de reação de hidrogenação do CO em catalisadores de Pd e Li . Nesta figura pode-se observar que a dissociação do $C - O$ é menos favorecida quando se utiliza catalisador de Pd , tornando a síntese de metanol mais favorável neste mecanismo reacional. Por outro lado, quando se utiliza catalisador de Ni o mecanismo de dissociação é mais favorecido, levando preferencialmente a formação de metano. Os métodos computacionais podem predizer os estados de transição dos mecanismos reacionais considerando apenas as espécies químicas envolvidas na reação, mas o valor energético destes estados depende da interação com o catalisador.

A tabela 7.6 apresenta os resultados obtidos para o $\ln$ da constante de equilíbrio calculada computacionalmente através de diversos métodos implementados no programa Gamess (versão de 20/11/2004) e o valor obtido experimentalmente para a reação $NO_2 \rightleftharpoons NO + 1/2O_2$, para diferentes temperaturas.

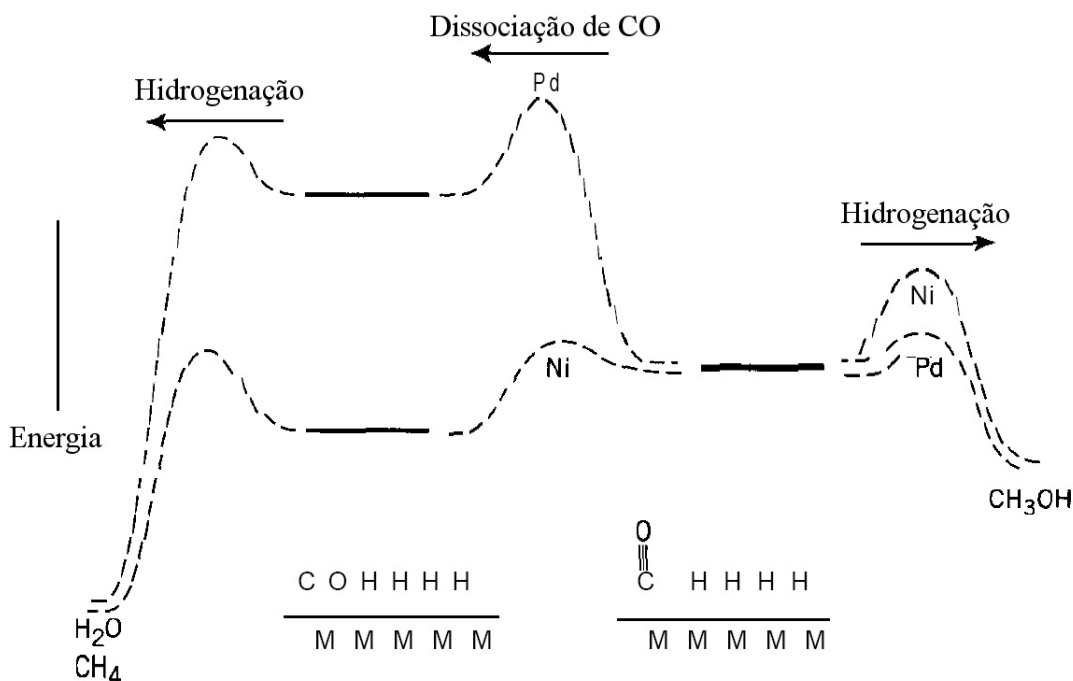


Figura 7.1: Coordenada de reação de hidrogenação do CO em catalisadores de Pd e de Ni .

Tabela 7.6: Valores teóricos e experimentais do $\ln K_{eq}$ para a reação $NO_2 \rightleftharpoons NO + 1/2O_2$ em diferentes temperaturas.

Método	499 K	577 K	700 K	800 K	825,4K
HF/6-311G*	25,93	23,65	21,11	19,62	19,30
MP2	-7,74	-5,46	-2,90	-1,38	-1,06
MP4	-6,86	-4,72	-2,28	-0,83	-0,53
Exper. ^a	-4,86	-2,97	-0,85	0,38	0,66

^a Referência [32].

A análise dos dados da tabela 7.6 permite observar a baixa exatidão do método de HF , e a melhoria na exatidão dos resultados proporcionada pela inclusão dos efeitos de correlação dos métodos $MP2$ e $MP4$.

A tabela 7.7 apresenta os valores obtidos computacionalmente através de métodos implementados no programa Gamess (versão de 20/11/2004) e os valores experimentais do $\ln$ da constante de equilíbrio da reação $1/2N_2 + 3/2H_2 \rightleftharpoons NH_3$, para diferentes temperaturas.

Pode-se observar nos dados da tabela 7.7 que os resultados experimentais

Tabela 7.7: Valores teóricos e experimentais do $\ln K_{eq}$ para a reação $1/2N_2 + 3/2H_2 \rightleftharpoons NH_3$ em diferentes temperaturas.

Método	200 K	300 K	400 K	500 K	600 K
HF/6-311G*	7,90	1,49	-1,96	-4,14	-5,97
MP2	-2,34	-5,35	-7,10	-8,27	-9,13
MP4	-5,84	-7,68	-8,84	-9,67	-10,29
Exper. ^a	15,5	6,50	1,80	-1,10	3,20

^a Referência [32].

apresentam um $\ln$ da constante de equilíbrio mínimo para uma temperatura próxima de $500K$, mas os resultados obtidos computacionalmente não acompanham esta tendência. O desvio entre os resultados experimentais e os obtidos computacionalmente pode ser atribuído ao caráter iônico da molécula formada, que por sua vez influencia no coeficiente de atividade das espécies envolvidas no mecanismo reacional.

7.7 Auto-difusão

O coeficiente de auto-difusão D_i^* pode ser definido como o coeficiente de difusão D_i da espécie i na ausência de gradiente de potencial químico, de acordo com a equação 7.3:

$$D_i^* = D_i \frac{\partial \ln c_i}{\partial \ln a_i} \quad (7.3)$$

onde a_i é a atividade da espécie i em solução e c_i é a concentração de i . O coeficiente de difusão, por sua vez, pode ser definido como o fluxo mássico da espécie i em função do gradiente de concentração desta espécie.

A determinação computacional do coeficiente de auto-difusão pode ser obtido a partir das velocidades atômicas calculadas por simulação de dinâmica molecular.

A tabela 7.8 apresenta os resultados de auto-difusão obtidos pela simulação de dinâmica molecular de um *ensemble* contendo 851 moléculas formadas por 2741 átomos, sendo 1 molécula da enzima α Ribonuclease e 850 moléculas de água, considerando uma temperatura de $300K$ e pressão de 1 atmosfera. A simulação foi realizada para um intervalo de 0 à 10 ps e de 0 à 50 ps, com propagação periódica ao longo dos três eixos cartesianos.

A α Ribonuclease é uma enzima digestiva, secretada pelo pâncreas, e tem sido estudada experimentalmente e computacionalmente, devido sua configuração α hélice em solução aquosa, o que é notável para uma molécula enzimática tão pequena, conforme relatado por Stryer (1988) [153].

Tabela 7.8: Coeficiente D_i^*) para a α Ribonuclease em água, em diferentes intervalos de tempo.

Intervalo (ps)	$D_i^* \alpha$ Ribonuclease (cm^2/s)	D_i^* água (cm^2/s)
0 até 10	$1,041 \times 10^{-5} (+/- 0,020)$	$4,243 \times 10^{-5} (+/- 0,227)$
0 até 50	$0,566 \times 10^{-5} (+/- 0,244)$	$4,683 \times 10^{-5} (+/- 0,313)$

Pode-se notar na tabela 7.8 que o valor de auto-difusão para as moléculas de água possuem valores estatisticamente concordantes para os dois intervalos de tempo estudados. Isto pode ser explicado pelo fato de que as moléculas de água são pequenas e se deslocam livremente pelo meio simulado, e este deslocamento representa ainda uma pequena influência energética ao equilíbrio do sistema.

Já para a α Ribonuclease existe uma diferença significativa entre os resultados para os dois intervalos de tempo simulados. No intervalo de 10 ps o valor obtido corresponde à acomodação geométrica no meio estudado, estando associada a um grande deslocamento atômico, o que leva a uma estimativa alta para a auto-difusão. Já no intervalo de tempo de 50 ps o deslocamento atômico é menor, tendo sido atingida a configuração estável no meio simulado, e realmente corresponde a mobilidade na molécula de α Ribonuclease. Além disso, a dimensão da enzima estudada é bem maior que a das outras moléculas do meio, e seu deslocamento implica em uma grande movimentação das outras moléculas para obtenção de uma configuração estável, o que explica o alto valor do desvio padrão para os valores calculados de auto-difusão.

7.8 Avaliação dos métodos

7.8.1 Métodos *ab-initio*

De forma geral a exatidão dos resultados obtidos por estes métodos é muito boa, convergindo para a solução exata com a inclusão de efeitos de correlação eletrônica. O esforço computacional destes métodos já foi apresentado na tabela 7.1, enquanto a exatidão destes métodos normalmente aumenta de acordo com a seguinte ordem:

$$HF \ll MP2 < CISD \cong MP4 \cong CCSD < CCSD(T) < CCSDT < fullC$$

As principais fontes de erro dos métodos *ab-initio* são:

- aproximações de Born-Oppenheimer;
- uso de um conjunto de base incompleto;
- correlação eletrônica incompleta; e
- omissão de efeitos relativistas.

O uso de recursos computacionais (tempo de processamento, memória e espaço em disco) destes métodos é alto, fazendo com que os métodos mais rigorosos sejam aplicados no estudo de moléculas com algumas dezenas de elétrons, enquanto os métodos *HF* e *MP2* são aplicados no estudo de moléculas de médio porte, considerando sempre o poder computacional disponível em centros de processamento de alto desempenho.

7.8.2 Métodos semi-empíricos

Os métodos semi-empíricos fornecem bons resultados para moléculas orgânicas, principalmente se o objetivo é simular moléculas semelhantes as utilizadas na parametrização do método empregado. Propriedades como geometria molecular e parâmetros energéticos apresentam uma boa qualidade de predição, enquanto estados vibracionais e estruturas de transição são preditos de forma mediana. Forças de Van der Walls e forças de dispersão intermoleculares apresentam predições de baixa qualidade, o que pode ser atribuída a falta de funções de base difusas nos campos de força semi-empíricos.

É recomendável a utilização de mais de um método para a predição confiável de propriedades moleculares. Mesmo os métodos mais empregados, como *AM1* e *PM3*, apresentam suas respectivas qualidades e deficiências. Pode-se citar, por exemplo, que o método *AM1* prediz ligações de hidrogênio com comprimentos reduzidos em 0,1 Å em relação ao valor experimental, mas a orientação predita é correta.

A escalabilidade típica destes métodos permite que o estudo de moléculas de grande porte sejam realizadas em estações de trabalho com alto poder de processamento.

7.8.3 Métodos de simulação molecular

É praticamente impossível avaliar os resultados obtidos por simulação molecular, dada a grande possibilidade de configurações do sistema e ao grande número de campos de força disponíveis. A escolha do número de moléculas no sistema, da propagação das condições periódicas, do campo de força empregado, e das condições físicas do sistema deve incluir o levantamento do comportamento do sistema em estudo, através de um procedimento iterativo de alteração e análise das condições do sistema estudado e obtenção das propriedades de interesse.

7.9 Conclusão

Ao longo deste capítulo foram apresentadas informações importantes para o desenvolvimento de um projeto envolvendo química computacional.

Os exemplos de utilização apresentados possuem caráter meramente demonstrativo, possuindo complexidade baixa quando comparada aos estudos realizados com objetivos reais. Contudo, servem para ilustrar a utilização de técnicas de *QC* para obtenção de propriedades que podem ser utilizadas no desenvolvimento e otimização de processos químicos.

Capítulo 8

Conclusões

A química computacional é freqüentemente citada como umas das áreas fundamentais para a obtenção da resposta para diversos problemas da química, biologia, física, e mais recentemente, da engenharia química. Os constantes aperfeiçoamentos na implementação de modelos teóricos, no desenvolvimento de algoritmos numéricos e no poder de processamento de cálculo permitem que hoje sejam obtidos resultados computacionais para problemas antes considerados insolúveis, e caso seja mantida esta tendência, teremos um futuro empolgante para esta área de conhecimento.

A *QC* vem ganhando o merecido respeito acadêmico e industrial, uma vez que os resultados obtidos possuem, em muitos casos, exatidão semelhante aos resultados obtidos experimentalmente. A grande vantagem é que podem ser simuladas condições e moléculas que não estejam disponíveis para a realização de experimentos tradicionais.

Neste trabalho, além de apresentar as diversas metodologias de química computacional, também foram incluídos estudos relacionados ao ambiente computacional. Estes estudos e seus resultados, apesar de não serem exigidos para a realização de cálculos teóricos, permitem que o usuário tenha uma visão mais ampla do projeto envolvendo a *QC*, auxiliando assim na seleção da melhor metodologia e estimativa do tempo para execução do projeto.

Foi projetado um ambiente computacional de alto desempenho e baixo custo, utilizando *hardware* facilmente disponível no mercado e *softwares* gratuitos, desde o sistema operacional até programas aplicativos. O ambiente proposto possui grande flexibilidade operacional, podendo ser facilmente expandido para atender demandas futuras, permitindo ainda sua atualização tecnológica ao longo do tempo.

O comparativo realizado entre diversos sistemas operacionais, instalados em um

mesmo microcomputador, permitiu observar diferenças significativas nos indicadores de desempenho. Os resultados obtidos para cálculo de ponto flutuante, acesso à memória randômica e acesso ao disco de armazenamento permitiu selecionar qual o melhor sistema operacional para aplicação em computação científica.

Foram apresentadas as metodologias de cálculo dos programas quânticos e de simulação molecular, assim como quais são as propriedades obtidas através destes cálculos. As propriedades obtidas de forma direta podem ser utilizadas para a obtenção de muitas outras. As principais potencialidades e aplicabilidades também foram apresentadas, juntamente com os desvios mais comuns esperados para os diferentes métodos.

Os cálculos da constante de equilíbrio e do coeficiente de auto-difusão realizados tiveram como objetivo apresentar a obtenção de propriedades de interesse para o desenvolvimento de processos. Um estudo mais aprofundado, considerando todas as espécies envolvidas em todas as etapas do mecanismo reacional no primeiro caso e diferentes agrupamentos e condições periódicas do sistema dinâmico no segundo caso, permitiria a obtenção de resultados mais confiáveis para as propriedades preditas.

O ambiente computacional criado, voltado para a realização de pesquisas envolvendo *QC*, pode ser utilizado para nos seguintes estudos:

- estruturas de transição;
- constantes e taxas de reação;
- interação solvente-catalisador;
- equilíbrio de fases;
- energia de solvatação;
- processamento paralelo.

A execução de programas de *QC* para obtenção de resultados quantitativos somente deve começar após serem formuladas respostas para as seguintes questões:

- qual a propriedade a ser predita;
- qual a exatidão esperada para o resultado; e
- qual o tempo computacional disponível.

A propriedade a ser predita é importante para definir qual o melhor modelo teórico a ser utilizado, ocorrendo nesta etapa o direcionamento para métodos quânticos, clássicos, correlacionais, ou ainda uma combinação entre eles.

A exatidão esperada para a resposta é importante para definição da técnica a ser utilizada, pois uma maior exatidão normalmente exige maior esforço computacional. Na verdade, a escolha da melhor técnica resulta em mais de uma resposta, sendo recomendado que o pesquisador utilize diversas técnicas para obtenção de um valor médio e também aplique estas técnicas em sistemas conhecidas para validação dos resultados obtidos.

O tempo computacional necessário pode ser reduzido pela escolha do método e da técnica utilizada ou ainda pela utilização de técnicas de processamento paralelo.

Trabalhos futuros

Neste trabalho foi criado o ambiente para pesquisas em QC , porém a utilização deste ambiente ficou limitada a testes de compilação e execução de simulações para avaliação do correto funcionamento dos programas disponibilizados e das bibliotecas numéricas utilizadas. O ambiente criado foi utilizado para obtenção da constante de equilíbrio e do coeficiente de auto-difusão em meio aquoso, apenas de forma ilustrativa.

As sugestões para continuação deste trabalho são:

- transferência do ambiente computacional criado para estações de trabalho com maior poder de processamento;
- aplicação das técnicas de paralelismo para execução simultânea destes programas em mais de uma estação de trabalho;
- desenvolvimento de interface gráfica para auxiliar na utilização dos programas executados por linha de comando, sem alterar as funcionalidades destes;
- aplicação do ambiente criado para obtenção da solução para um problema real, de interesse acadêmico ou industrial.

Referências Bibliográficas

- [1] EDGAR, T. F.; HIMMELBLAU, D. M.; LASDON, L. S. *Optimization of Chemical Processes*. 2. ed. New York: McGraw Hill, 2001.
- [2] LUYBEN, W. L. *Process Modelling, Simulation and Control for Chemical Engineers*. 2. ed. Columbus: McGraw Hill, 1999.
- [3] BIRD, R. B.; STEWART, W. E.; LIGHTFOOT, E. N. *Transport Phenomena*. 2. ed. New York: John Willey & Sons, Inc., 2002.
- [4] CHUNG, C. A. *Simulation Modeling Handbook*. Boca Raton: CRC Press, 2004.
- [5] TUNDO, P. et al. Synthetic pathways and processes in green chemistry. introductory overview. *Pure and Applied Chemistry*, v. 72, p. 1207–1228, 2000.
- [6] ALLEN, M.; TILDESLEY, D. *Computer Simulation of Liquids*. Oxford: Clarendon Pres, 1987.
- [7] DIXON, D. A.; FELLER, D. Computational chemistry and process design. *Chemical Engineering Science*, v. 54, p. 1929–1939, 1999.
- [8] SOCIETY, A. C. Technology vision 2020: The us chemical industry, a joint report by the american chemical society, american institute of chemical engineers, chemical manufactures association, council for chemical research and synthetic orgânica chemical manufacturers association. *American Chemical Society*, 1996.
- [9] ALDA, W. et al. Exothermic and endothermic chemical reactions involving very many particles modeled with molecular dynamics. *Physics D*, p. 261–274, 2000.
- [10] LIPKOWITZ, K. B. Applications of computational chemistry to the study of cyclodextrins. *Chemical Reviews*, v. 98, p. 1829–1873, 1998.
- [11] BROWN, R. D. et al. Acronyms used in theoretical chemistry. *Pure and Applied Chemistry*, v. 68, p. 387–456, 1996.

- [12] BOGGS, J. E. Guidelines for presentation of methodological choices in the publication of computational results - a: *Ab-initio* electronic structure calculations. *Pure and Applied Chemistry*, v. 72, p. 1–4, 1998.
- [13] BALA, P. et al. Quantum-classical molecular dynamics simulations of proton transfer processes in molecular complexes and in enzymes. *Journal of Physical Chemistry*, v. 100, p. 2535–2545, 1996.
- [14] MONARD, G.; JR, K. M. M. Combined quantum mechanical/molecular mechanical methodologies applied to biomolecular systems. *Accounts of Chemical Research*, v. 32, p. 904–911, 1999.
- [15] BALUCANI, N. et al. Dynamics of the $\text{cl} - \text{d}_2$ reaction: A comparison of crossed molecular beam experiments with quasi-classical trajectory calculations on a new *ab initio* potential energy surface. *Chemical Physics Letters*, v. 328, p. 500–508, 2000.
- [16] NGUYEN, M. T.; SENGUPTA, D.; VANQUICKENBORNE, L. G. Kinetic analyses combining quantum chemical and quantum statistical methods: Some case studies. *Journal of Physical Chemistry*, v. 100, p. 10956–10966, 1996.
- [17] ENGVIST, O.; ÅSTRAND, P.; KARLSTRÖM, G. Accurate intermolecular potentials obtained from molecular wave functions: Bridging the gap between quantum and molecular simulations. *Chemical Reviews*, v. 100, p. 4087–4108, 2000.
- [18] GURKA, D. F. et al. Environmental analysis by *Ab-initio* quantum mechanical computation and gas chromatography/fourier transform infrared spectrometry. *Analytical Chemistry*, v. 68, p. 4221–4227, 1996.
- [19] MEIER, R. J.; AAGAARD, O. M.; BUDA, F. First principles molecular dynamics applied to homogeneous catalysis: on ethylene insertion mechanisms and metathesis. *Journal of Molecular Catalysis A*, v. 160, p. 189–197, 2000.
- [20] BOUYER, F.; PICARD, G.; LEGENDRE, J. J. Computational chemistry: A way to reach spectroscopic and thermodynamic data for exotic compounds. *Journal of Chemical Information and Computer Sciences*, v. 36, p. 684–693, 1996.
- [21] ARMSTRONG, B. M.; MCKEE, M. L.; SHEVLIN, P. Experimental and computational study of the intramolecular reactivity of free *tert*-butylphenylmethyleno: Modification of the chemistry of *tert*-butylmethylene by the introduction of a phenyl group. *Journal of Organic Chemistry*, v. 63, p. 7408–7412, 1998.

- [22] BANKS, H. D.; WHITE, W. E. A computational study of the reactions of thiranes with ammonia and amines. *Journal of Organic Chemistry*, v. 66, p. 5981–5986, 2001.
- [23] ROGERS, D. W.; MCLAFFERTY, F. J. G3(mp2) calculations of enthalpies of hydrogenation, isomerization, and formation of [3]-radialene and related compounds. *Journal of Physical Chemistry A*, v. 106, p. 1054–1059, 2002.
- [24] CUMMINGS, P. T.; WESTMORELAND, P. R. Proceedings of the first international conference on the foundations of molecular modeling and simulation. *AIChE Symposium Series*, v. 93, p. 1–12, 2000.
- [25] LENARDAO, E. J. et al. Green chemistry - os 12 conceitos da química verde e sua inserção nas atividades de ensino e pesquisa. *Química Nova*, v. 26, p. 123–129, 2003.
- [26] GERMASHEV, I. V. et al. Computer-aided design of chemical compounds with controlled properties. *Theoretical Foundations of Chemical Engineering*, v. 38, p. 86–91, 2004.
- [27] KUHN, C.; BERATAN, D. N. Inverse strategies for molecular design. *Journal of Physical Chemistry*, v. 100, p. 10595–10599, 1996.
- [28] DUPONT, J. Economia de átomos, engenharia molecular e catálise organometálica bifásica: conceitos moleculares para tecnologias limpas. *Química Nova*, v. 23, p. 825–831, 2000.
- [29] ATKINS, P. W. *Physical Chemistry*. quinta. Oxford: Oxford University Press, 1994.
- [30] BARDEN, C. J.; III, H. F. S. Quantum chemistry in the 21 century. *Pure and Applied Chemistry*, v. 72, p. 1405–1423, 2000.
- [31] FREITAS, L. C. G. Prêmio nobel de química em 1998: Walter kohn e john pople. *Química Nova*, v. 22, p. 293–298, 1999.
- [32] BOHR, F.; HENON, E. Comparison of various quantum chemistry methods for the computation of equilibrium constants. *Journal of Physical Chemistry A*, v. 102, p. 4857–4862, 1998.
- [33] BOTTONI, A. et al. New computational and experimental evidence for the mechanism of the sakurai reaction. *Journal of American Chemical Society*, v. 119, p. 12131–12135, 1997.

- [34] BRAGA, J. P. O colapso da equipartição de energia. *Química Nova*, v. 24, p. 693–699, 2001.
- [35] DEDE, D. A.; III, W. A. G. First principles of protein folding rates. *Journal of Molecular Biology*, v. 294, p. 619–625, 1999.
- [36] FUKUDA, S.; AKIYOSHI, Y.; HORI, K. Computer-aided reaction design. development of a new facile procedure to synthesize 2-mercapto-3-alkoxycarboxylate on the basis of a *ab initio* molecular orbital calculations. *Journal of Organic Chemistry*, v. 64, p. 4768–4774, 1999.
- [37] ROSSI, I.; VENTURINI, A.; ZEDDA, A. Modeling hindered-amine light stabilizer-promoted polymer stabilization: Computational insight into the mechanism for nitroxyl radical regeneration from aminoethers. *Journal of the American Chemical Society*, v. 121, p. 7914–7917, 1999.
- [38] STERN, H. A. et al. Fluctuating charge, polarizable dipole, and combined models: Parametrization from *ab initio* quantum chemistry. *Journal of Physical Chemistry B*, v. 103, p. 4730–4737, 1999.
- [39] SIMPLIFIED Introduction to Ab-initio Basis Set: J. k. labanowski. Disponível em: <<http://www.ccl.net/cca/documents/basis-sets/basis.html>>. Acesso em: 24 de junho de 2005.
- [40] PERÄKYLÄ, M.; KOLLMAN, P. A. A simulation of the catalytic mechanism of aspartylglucosaminidase using *ab-initio* quantum mechanics and molecular dynamics. *Journal of American Chemical Society*, v. 119, p. 1189–1196, 1997.
- [41] ZAVODINSKY, V. G.; KUYANOV, I. A.; ZAVODINSKAYA, O. M. Atomic and electronic structures of nanometer sized silica particles. *Journal of Non-Crystalline Solids*, v. 243, p. 123–136, 1999.
- [42] DUFRESNE, G. et al. Thermochromic and solvatochromic conjugated polymers by design. *Macromolecules*, v. 33, p. 8252–8257, 2000.
- [43] ZHANG, L.; QIN, Q. Computational studies on the reaction pathways of *cf3br* with *o*(¹*d*,³*p*) atoms. *Journal of Physical Chemistry A*, v. 105, p. 215–218, 2001.
- [44] YASUDA, M.; CHIBA, K.; BABA, A. An *ab-initio* study on the reaction of organotin enolates: Comparison of highly coordinated tin reagent with noncoordinated reagent. *Journal of American Chemical Society*, v. 122, p. 7549–7555, 2000.

- [45] BARTLETT, K. L.; GOLDBERG, K. I.; BORDEN, W. T. Computational study of reductive elimination reactions to form $c-h$ bonds from platinum(ii) and platinum(iv) centers with strongly coordinating trimethylphosphine ligands. *Organometallics*, v. 20, p. 2669–2678, 2001.
- [46] PONDER, J. W.; RICHARDS, F. M. An efficient newton-like method for molecular mechanics energy minimization of large molecules. *Journal of Computational Chemistry*, v. 8, p. 1016–1024, 1987.
- [47] DUAN, Z. H.; KRASNY, R. An adaptive treecode for computing nonbonded potential energy in classical molecular systems. *Journal of Computational Chemistry*, v. 22, p. 184–195, 2001.
- [48] ELCOCK, A. H. et al. Computer simulation of protein-protein association kinetics: Acetylcholinesterase-fasciculin. *Journal of Molecular Biology*, v. 291, p. 149–162, 1999.
- [49] SPINNATO, P. F.; ALBADA, G. D. van; SLOOT, P. M. A. Performance analysis of parallel N-body codes. *Proceedings of the sixth ASCI*, p. 213–220, 2000.
- [50] MCLEOD, A. S. Monte carlo simulations of heterogeneous catalytic reactions on highly dispersed supported metal catalysts. *Catalysis Today*, v. 53, p. 289–302, 1999.
- [51] FLORIAN, J.; WARSHEL, A. Langevin dipoles model for *ab initio* calculations of chemical processes in solution: Parametrization and application to hydration free energies of neutral and ionic solutes and conformational analysis in aqueous solution. *Journal of Physical Chemistry B*, v. 101, p. 5583–5595, 1997.
- [52] LYATSKAYA, Y. et al. Designing compatibilizers to reduce interfacial tension in polymer blenders. *Journal of Physical Chemistry*, v. 100, p. 1449–1458, 1996.
- [53] SELVES, J. L.; ABRAHAM, M. H.; BURG, P. A new method for the explanation of liquids properties in terms of molecular interactions. *Fluid Phase Equilibria*, v. 148, p. 69–82, 1998.
- [54] QSAR and drug design: D. r. bevan. Disponível em: <<http://www.netsci.org/science/compchem/feature12.html>>. Acesso em: 4 de fevereiro de 2004.
- [55] AN introduction to QSAR methodology: A. b. richon and s. s. young. Disponível em: <<http://www.netsci.org/science/compchem/feature19.html>>. Acesso em: 4 de fevereiro de 2004.

- [56] KATRITZKY, A. R.; KARELSON, M.; LOBANOV, V. S. Qspr as a means of predicting and understanding chemical and physical propoerties in terms os structure. *Pure and Applied Chemistry*, v. 69, p. 245–248, 1997.
- [57] FAMINI, G. R.; WILSON, L. Y. *Using theoretical descriptors in quantitative structure activity relationships and linear free energy relationships*. [S.l.], 2004.
- [58] WALTERS, R. N. *Molar group contributions to the heat of combustion*. [S.l.], 10 de novembro de 2003.
- [59] MARANAS, C. D. Optimal computer-aided molecules design: A polymer design case study. *Industrial & Engineering Chemistry Research*, v. 35, p. 3403–3414, 1996.
- [60] CRAY Inc.: the world's most powerful supercomputers. Disponível em: <<http://www.cray.com/>>. Acesso em: 15 de maio de 2003.
- [61] FUJITSU do Brasil. Disponível em: <<http://www.fujitsu.com.br>>. Acesso em: 15 de maio de 2003.
- [62] IBM Brasil. Disponível em: <<http://www.ibm.com/br/>>. Acesso em: 15 de maio de 2003.
- [63] INTEL: Hyper-threading technology. Disponível em: <<http://www.intel.com/technology/hyperthread>>. Acesso em: 6 de dezembro de 2005.
- [64] SUN Microsystems Inc. Disponível em: <<http://www.sun.com.br>>. Acesso em: 15 de maio de 2003.
- [65] HP Brasil. Disponível em: <<http://www.hp.com.br>>. Acesso em: 15 de maio de 2003.
- [66] INTEL: Intel technology.
- [67] AMD: Advanced micro devices. Disponível em: <<http://www.amd.com.br>>. Acesso em: 18 de novembro de 2003.
- [68] VIA Technologies, Inc. Disponível em: <<http://www.viatech.com>>. Acesso em: 15 de maio de 2003.
- [69] MICROSOFT Brasil. Disponível em: <www.microsoft.com.br>. Acesso em: 21 de maio de 2003.

- [70] APPLE: Apple computer, inc. Disponível em: <www.apple.com.br>. Acesso em: 21 de maio de 2003.
- [71] CONECTIVA Linux. Disponível em: <<http://www.conectiva.com.br>>. Acesso em: 15 de maio de 2003.
- [72] BEBIAN: Portal debian no brasil. Disponível em: <<http://www.debian.com.br/>>. Acesso em: 15 de maio de 2003.
- [73] RED Hat: Linux, embedded linux and open source solutions. Disponível em: <www.redhat.com.br>. Acesso em: 21 de maio de 2003.
- [74] THE FreeBSD Project. Disponível em: <www.freebsd.org>. Acesso em: 21 de maio de 2003.
- [75] THE NetBSD Project. Disponível em: <www.netbsd.org>. Acesso em: 21 de maio de 2003.
- [76] APPLE: MacOS x - darwin. Disponível em: <developer.apple.com/darwin/>. Acesso em: 21 de maio de 2003.
- [77] PLAN 9 from Bell Labs: Fourth edition. Disponível em: <<http://cm.bell-labs.com/plan9dist>>. Acesso em: 21 de maio de 2005.
- [78] MPI: Message passing interface forum. Disponível em: <<http://www.cray.com/products/software/>>. Acesso em: 21 de maio de 2003.
- [79] MPICH: A portable implementation of mpi. Disponível em: <<http://www-unix.mcs.anl.gov/mpi/mpich/>>. Acesso em: 2 de dezembro de 2005.
- [80] LAM/MPI Parallel Computing. Disponível em: <<http://www.lam-mpi.org/>>. Acesso em: 2 de dezembro de 2005.
- [81] PVM: Parallel virtual machine. Disponível em: <http://www.csm.ornl.gov/pvm/pvm_home.html>. Acesso em: 21 de maio de 2003.
- [82] THE World Wide Web Virtual Library: Concurrent systems. Disponível em: <<http://www.afm.sbu.ac.uk/concurrent/>>. Acesso em: 21 de maio de 2003.
- [83] HPF: The high performance fortran. Disponível em: <www.crpc.rice.edu/HPFF/>. Acesso em: 21 de maio de 2003.

- [84] MAISIE Programming Language. Disponível em:
<<http://may.cs.ucla.edu/projects/maisie/>>. Acesso em: 21 de maio de 2003.
- [85] OPENMOSIX: An open source linux cluster project. Disponível em:
<openmosix.sourceforge.net/>. Acesso em: 21 de maio de 2003.
- [86] ADAPTOR: Gmd's high performance fortran compilation. Disponível em:
<http://www.scai.fraunhofer.de/EP-CACHE/adaptor/www/adaptor_home.html>. Acesso em: 21 de maio de 2003.
- [87] SHPF: A hpf compilation system. Disponível em:
<www.vcpc.univie.ac.at/information/software/shpf>. Acesso em: 22 de maio de 2003.
- [88] OSCAR: Open source cluster application resources. Disponível em:
<<http://oscar.openclustergroup.org/>>.
- [89] THE MathWorks: Matlab and simulink for technical computing. Disponível em:
<<http://www.mathworks.com/>>. Acesso em: 2 de dezembro de 2005.
- [90] SCILAB: A free scientific software package. Disponível em:
<<http://www.scilab.org/>>. Acesso em: 2 de dezembro de 2005.
- [91] DALTON: Quantum chemistry program. Disponível em:
<<http://www.kjemi.uio.no/software/dalton/dalton.html>>. Acesso em: 20 de junho de 2005.
- [92] GAMESS: General atomic and molecular electronic structure system. Disponível em: <<http://www.msg.ameslab.gov/GAMESS/GAMESS.html>>. Acesso em: 20 de junho de 2005.
- [93] NWCHEM. Disponível em: <<http://www.emsl.pnl.gov/docs/nwchem/>>. Acesso em: 22 de novembro de 2005.
- [94] COMPUTATION, Computers, Information and Mathematics Center. Disponível em: <<http://www.cs.sandia.gov/>>. Acesso em: 28 de maio de 2003.
- [95] THE Linpack Benchmark. Disponível em: <www.netlib.org/linpack>. Acesso em: 2 de dezembro de 2005.
- [96] SPEC: Standard performance evaluation corporation. Disponível em:
<www.spec.org>. Acesso em: 6 de julho de 2005.

- [97] FLOPS: Al aburto. Disponível em: <ftp://ftp.sunet.se/pub/benchmark/aburto>. Acesso em: 21 de maio de 2005.
- [98] MCCALPIN, J. D. *STREAM: Sustainable Memory Bandwidth in High Performance Computer*. [S.l.], 20 de maio de 2005.
- [99] BONNIE: A comprehensive collection of disk i/o benchmarks. Disponível em: <http://www.acnc.com/benchmarks.html>. Acesso em: 20 de maio de 2002.
- [100] YOUNG, D. C. *Computational Chemistry*. New York: John Wiley & Sons, Inc., 2001.
- [101] MOPAC: A semiempirical quantum mechanics program. Disponível em: <www.ccl.net/cca/software/SOURCES/FORTRAN/mopac7/>. Acesso em: 22 de novembro de 2005.
- [102] GAUSSIAN: The official gaussian website. Disponível em: <http://www.gaussian.com/>. Acesso em: 06 de julho de 2005.
- [103] MORGON, N. H.; CUSTODIO, R. Teoria do funcional de densidade. *Química Nova*, v. 18, p. 44–55, 1995.
- [104] CIRINO, J. J. V.; BERTRAN, C. A. Estudo da geometria da uréia por métodos *ab initio* e simulação computacional de líquidos. *Química Nova*, v. 25, p. 358–363, 2002.
- [105] SADUS, R. J. *Molecular Simulation of Fluids*. [S.l.]: Elsevier, 1999.
- [106] MOLECULAR Dynamics Simulation: Elementary Methods. [S.l.]: John Wiley and Sons, 1997.
- [107] MAZUR, A. K. Common molecular dynamics algorithms revisited: Accuracy and optimal time steps of störmer-leapfrog integrators. *arXiv:physics/9707008 v1*, p. 1–15, 1997.
- [108] MUKHERJEE, A.; BAGCHI, B. Nonideality in binary mixtures: Correlations between excess volume, excess viscosity and diffusion coefficients. *Journal of Physical Chemistry*, v. 105, p. 9581–9585, 2001.
- [109] VIJAYENDRAN, R. A.; LIGLER, F. S.; LECKBAND, D. E. A computational reaction-diffusion model for the analysis of transport limited kinetics. *Analytical Chemistry*, v. 71, p. 5405–5412, 1999.

- [110] AIMPAC: Atoms in molecules analysis program. Disponível em: <<http://www.chemistry.mcmaster.ca/aimpac/>>. Acesso em: 20 de junho de 2005.
- [111] AMPAC: Semiempirical quantum mechanics. Disponível em: <<http://www.semichem.com/ampac/default.php>>. Acesso em: 20 de junho de 2005.
- [112] AMSOL: Free energies of solvation of molecules and ions in solution. Disponível em: <<http://comp.chem.umn.edu/amsol/>>. Acesso em: 20 de junho de 2005.
- [113] DEFT. Disponível em: <<http://server.ccl.net/cca/software>>. Acesso em: 20 de junho de 2005.
- [114] ECEPPAK: The ecepp package. Disponível em: <<http://cbsu.tc.cornell.edu/software/eceppak/>>. Acesso em: 22 de novembro de 2005.
- [115] MOBOSOL. Disponível em: <<http://ftp.ccl.net/cca/software>>. Acesso em: 22 de novembro de 2005.
- [116] POLYRATE: Computer program for the calculation of chemical reaction rates for polyatomics. Disponível em: <<http://comp.chem.umn.edu/polyrate/>>. Acesso em: 25 de novembro de 2005.
- [117] ALCMD: Ames lab classical molecular dynamics for massively parallel computers. Disponível em: <http://www.cmp.ameslab.gov/cmp/CMP_Theory/cmd/alcmd>. Acesso em: 20 de junho de 2005.
- [118] AMBER: Assisted model building with energy refinement. Disponível em: <<http://amber.scripps.edu/>>. Acesso em: 20 de junho de 2005.
- [119] B: A java-based, on-line biomolecular modeling package. Disponível em: <<http://www.scripps.edu/mb/case/Biomer/>>. Acesso em: 20 de junho de 2005.
- [120] CCP5: Collaborative computational projects 5. Disponível em: <<http://www.ccp5.ac.uk/>>. Acesso em: 20 de novembro de 2005.
- [121] BROOKS, B. R. et al. Charmm: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, v. 4, p. 187–217, 1983.

- [122] THE Encyclopedia of Computational Chemistry. In: . [S.l.]: John Wiley & Sons, 1998. cap. CHARMM: The Energy Function and Its Parameterization with an Overview of the Program, p. 271–277.
- [123] FRENKEL, D.; SMIT, B. *Understanding Molecular Simulation: From Algorithms to Applications*. New York: Academic Press, 2002.
- [124] GROMACS: The world's fastest molecular dynamics. Disponível em: <<http://www.gromacs.org/>>. Acesso em: 20 de junho de 2005.
- [125] MINDY: A 'minimal' molecular dynamics program. Disponível em: <<http://www.ks.uiuc.edu/Development/MDTools/mindy/>>. Acesso em: 22 de novembro de 2005.
- [126] MOIL: A public domain molecular modeling software. Disponível em: <<http://cbsu.tc.cornell.edu/software/moil/>>.
- [127] MOLDY: A general-purpose molecular dynamics code. Disponível em: <<http://www.earth.ox.ac.uk/~keithr/moldy.html>>. Acesso em: 22 de novembro de 2005.
- [128] NAMD: Scalable molecular dynamics. Disponível em: <<http://www.ks.uiuc.edu/Research/namd/>>. Acesso em: 22 de novembro de 2005.
- [129] RICHARD Sadus: Home page. Disponível em: <<http://www.it.swin.edu.au/personal/rsadus/>>. Acesso em: 25 de novembro de 2005.
- [130] SOLVATE. Disponível em: <<http://www.ccl.net/cca/software>>. Acesso em: 10 de novembro de 2005.
- [131] TINKER: Software tools for molecular design. Disponível em: <<http://dasher.wustl.edu/tinker/>>. Acesso em: 25 de novembro de 2005.
- [132] TOWHEE: Monte carlo molecular simulation code. Disponível em: <<http://towhee.sourceforge.net/>>. Acesso em: 25 de novembro de 2005.
- [133] ABSOLV: Pharma algorithms. Disponível em: <<http://www.ap-algorithms.com/absolv.htm>>. Acesso em: 20 de junho de 2005.

- [134] ASAD: Chemical integration software. Disponível em: <<http://www.atm.ch.cam.ac.uk/acmsu/asad/>>. Acesso em: 20 de junho de 2005.
- [135] GEPOL93. Disponível em: <<http://www.ccl.net/cca/software>>. Acesso em: 22 de novembro de 2005.
- [136] KINTECUS: Welcome to kintecus.org. Disponível em: <www.kintecus.org>. Acesso em: 22 de novembro de 2005.
- [137] MELLER, J.; ELBER, R. Linear optimization and a double statistical filter for protein threading protocols. *Proteins, Structure, Function and Genetics*, v. 45, p. 241–261, 2001.
- [138] TEODORESCU, O. et al. Enriching the sequence substitution matrix by structural information. *Proteins, Structure, Function and Genetics*, v. 54, p. 41–48, 2004.
- [139] PROTARCH: Hierarchical prediction of protein structure. Disponível em: <<http://cbsu.tc.cornell.edu/software/protarch/>>.
- [140] ACDLABS: Advanced chemistry development. Disponível em: <<http://www.acdlabs.com/>>. Acesso em: 20 de junho de 2005.
- [141] OPEN Babel: File translation for computational chemistry and nanoscience. Disponível em: <<http://openbabel.sourceforge.net/>>. Acesso em: 20 de junho de 2005.
- [142] BIOCORE: A biological collaborative research environment. Disponível em: <<http://www.ks.uiuc.edu/Research/biocore/>>. Acesso em: 20 de junho de 2005.
- [143] CAMBRIDGESOFT: Chemical drawing and chemical databases. Disponível em: <<http://www.cambridgesoft.com/>>. Acesso em: 22 de novembro de 2005.
- [144] HALGREN, T. A. Merck molecular force field. i. basis, form, scope, parameterization and performance of mmff94. *Journal of Computational Chemistry*, v. 17, p. 490–519, 1996.
- [145] MOLDA: The world of cybermolecules - molecular modeling and computational chemistry. Disponível em: <<http://www.molda.org/>>. Acesso em: 22 de novembro de 2005.

- [146] NOORDIK, G. S. e J. Molden: A pre and post processing program for molecular and electronic structures:. *Journal of Computer Aided Molecular Design*, v. 14, p. 123–134, 2000.
- [147] MOLMOL: Molecule analysis and molecule display. Disponível em: <<http://hugin.ethz.ch/wuthrich/software/molmol/>>. Acesso em: 22 de novembro de 2005.
- [148] RASMOL: Molecular graphics visualisation tool. Disponível em: <<http://www.openrasmol.org/>>. Acesso em: 25 de novembro de 2005.
- [149] CRDB: Emsl computational results database. Disponível em: <<http://www.emsl.pnl.gov/proj/crdb/>>. Acesso em: 6 de julho de 2005.
- [150] CCCBDB: Computational chemistry comparison and benchmark database. Disponível em: <<http://srdata.nist.gov/cccbdb/>>. Acesso em: 6 de julho de 2005.
- [151] QCLDB: Quantum chemistry literature database. Disponível em: <<http://qcldb2.ims.ac.jp/>>.
- [152] MCDONALD, C. M.; FLOUDAS, C. A. Glopeq: A new computational tool for the phase and chemical equilibrium problem. *Computers Chemical Engineering*, v. 21, p. 1–23, 1997.
- [153] BIOCHEMISTRY. terceira. New York:: Freeman, 1988.
- [154] CHARMM: Chemistry at harvard molecular mechanics. Disponível em: <<http://www.charmm.org/index.shtml>>.
- [155] LOOP: Learning, observing and outputting protein patterns. Disponível em: <<http://cbsu.tc.cornell.edu/software/loopp/>>. Acesso em: 22 de novembro de 2005.

Índice Remissivo

- A, 18
- C_v , 14
- χ_α , 79
- ψ , 15
- Ensemble*, 95
- ab-initio, 7
- Adaptor, 39
- Algoritmo, 8
- AM1, 83
- AMBER, 105
- ASOG, 30
- ATM, 36
- B3LYP, 22, 80
- BLYP, 80
- Broadcast, 36
- CC, 68
- CCS, 69
- CCSD, 69
- CCSDT, 69
- CFF, 106
- CHARMM, 105
- CHEAT, 106
- Chipset, 50
- CI, 21, 67
- CIS, 69
- CISC, 35
- CISD, 69
- Cluster, 33
- CNDO, 81
- ComFa, 108
- CPU, 34
- Código-fonte, 39
- Dalton, 40
- Descritores Moleculares, 7
- DFT, 7
- Dinâmica Molecular, 7
- E, 13
- ECEPP, 106
- ECP, 65
- EFF, 27
- EP, 65
- FDDI, 36
- FLOPS, 45
- Fortran, 39
- Full-duplex, 36
- Funcional, 75
- GAMESS, 40
- Green Chemistry, 11
- GROMOS, 106
- GTO, 23, 61
- GVB, 67
- h, 14
- Hartrees, 66
- HF, 18
- HFS, 80
- HPF, 39
- Hückel, 81

- Inria, 40
- IP, 41
- IRC, 11, 23
- IUPAC, 7
- Kernel, 37
- LCAO, 12
- LD, 27
- LDA, 80
- Linpack, 46, 47
- M-TFD, 79
- Matriz Z, 56
- MBPT, 21
- MC, 25, 88
- MCSCF, 21, 67
- MD, 25, 88
- MIMD, 34
- MINDO, 82
- MM, 25
- MMFF, 107
- MNDO, 82
- MPI, 39
- MPLOPS, 48
- MP_n, 67
- MS- χ_{α} , 79
- Métodos Clássicos, 7
- Métodos Correlacionais, 7
- Métodos Híbridos, 8
- Métodos Quânticos, 7
- NMR, 10, 63
- NO, 71
- NWCHEM, 40
- Openmosix, 39
- OPLS, 107
- PES, 11, 23
- PFortran, 39
- PM3, 83
- PM3/TM, 83
- PRDDO, 82
- PVM, 39
- QC, 5
- QCLDB, 127
- QM/MM, 8
- QRRK, 23
- QSAR, 29
- QSAR 3D, 111
- QSAR-3D, 29
- QSPR, 29, 111
- R, 14
- RAM, 46
- RISC, 35
- SAM1, 83
- SCF, 17, 28
- Scilab, 40
- semi-empíricos, 7
- SHPF, 39
- SIMD, 34
- Simulação de Monte-Carlo, 7
- SPEC, 46
- STO, 23, 61
- STREAM, 50
- Subrotina, 39
- TF, 78
- TLSER, 31
- TPB, 24
- UFF, 108
- UNIFAC, 30

x86, 35

ZINDO, 82

Apêndice - Programas analisados

Os programas estão citados em ordem alfabética, com a descrição das principais funcionalidades, as versões analisadas e algumas observações sobre a implementação e uso dos mesmos.

1. Absolv

Descrição: Predição e cálculo de fenômenos de solvatação, segundo o método dos parâmetros de solvatação de Abraham. Os parâmetros estimados pelo programa são: caráter ácido da ligação de hidrogênio, caráter básico da ligação de hidrogênio, coeficiente de partição água-hexadecano, polaridade, polarizabilidade, excesso de refração molar e volume molar segundo McGowan [133].

Versão: 14.0, disponível apenas para avaliação do programa.

Funcionalidade: Existe um banco de dados destas propriedades, que também podem ser estimadas por mais de 100 modelos, além de permitir a inclusão de modelos desenvolvidos pelo usuário, desde que sejam matematicamente similares ao modelo de Abraham.

Licença: Comercial.

2. Acdlabs

Descrição: Pacote para desenho e manipulação de estruturas químicas em duas ou três dimensões, podendo ser utilizado de forma independente ou integrado à outras aplicações [140].

Versão: 5.0 e 8.0, disponível para uso não comercial.

Funcionalidade: Através do desenho da molécula e de um procedimento simples de otimização geométrica é possível estimar vários descritores moleculares, como peso molecular, refatividade molar, volume molar, tensão superficial, densidade, constante dielétrica e polarizabilidade.

Licença: Comercial.

Observação: Foram testadas as versões livres do programa, que não possuem diversos recursos existentes nas versões comerciais.

3. Aimpac

Descrição: Pacote que utiliza o método de Hartree-Fock para cálculo da função de onda [110].

Versão: Aimpac 95.

Funcionalidade: É possível utilizar teorias de correlação, similares ao método *MP2*, através de integração do *software* com outros pacotes computacionais, como Gaussian ou GAMESS.

Licença: Acadêmica.

Observação: A versão atual é comercial e recebe o nome de AIM2000.

4. Alcmd

Descrição: Programa que permite simulação de dinâmica molecular em grandes grupos atômicos por um longo tempo simulado, como simulação de anelamento, permitindo ainda que o resultado obtido seja utilizado em métodos *ab-initio* [117].

Versão: 3.0 e 6.4, com código em Fortran 77.

Funcionalidade: Possui diversos métodos de cálculo e um processo de otimização utilizando algoritmos genéticos. Está apto para execução em ambiente seqüencial ou paralelo.

Licença: Domínio público.

5. Allen e Tildesley

Descrição: Programas de simulação molecular extraídos do livro “Computer Simulation of Liquids”, de M. P. Allen and D. J. Tildesley, podendo ser facilmente adaptados para novas situações [6].

Versão: Versão indefinida, com código em Fortran 77.

Funcionalidade: Permite simulações de dinâmica molecular e simulações de Monte Carlo.

Licença: Domínio público.

Observação: Possui implementado apenas o método de Lennard-Jones.

6. Amber

Descrição: O Amber é um pacote de programas de simulação molecular e também um campo de força empírico, desenvolvidos para proteínas e ácidos nucleicos [118].

Versão: 8, apenas do campo de força.

Funcionalidade: O campo de força empírico está disponível em diversos programas de simulação molecular.

Licença: Versão comercial para o programa e domínio público para o campo de força empírico.

7. Ampac

Descrição: Programa de química quântica pelo método semi-empírico, desenvolvido com o objetivo de fornecer resultados exatos com facilidade de uso, sofrendo diversas melhorias ao longo dos últimos 25 anos. Foi utilizado como base para o desenvolvimento do programa Mopac [111].

Versão: 2.1, com código em Fortran 77.

Funcionalidade: Possui implementado os métodos semi-empíricos *MINDO/3*, *MNDO* e *AM1*.

Licença: Domínio público.

Observação: O programa era aberto até a versão testada. A versão atual (8.0) é comercial e possui custo de 950 e 12000 dólares para as versões acadêmica e comercial, respectivamente.

8. Amsol

Descrição: Programa quântico semi-empírico específico para predição de energia livre de solvatação em água, alcanos e diversos solventes orgânicos, tanto em fase gasosa quanto em fase condensada [112].

Versão: 6.6, com código em Fortran 77.

Funcionalidade: Pode ser utilizado também para cálculos de otimização geométrica. Os métodos de cálculo da energia livre de solvatação também estão disponíveis em diversos outros programas de química computacional.

Licença: Acadêmica.

Observação: O programa é uma expansão nas funcionalidades do pacote Ampac.

9. Asad

Descrição: O ASAD¹ é um programa para modelagem de um grande número de compostos e reações atmosféricas, onde as espécies podem ser tratadas individualmente ou em grupos. Inclui ainda tratamento para reações biomoleculares, termomoleculares, de fotólise e heterogêneas [134].

Versão: 2.0, com código em Fortran 77.

Funcionalidade: Após compilação do código fonte, o programa pode ser adaptado para diversas condições de uso sem necessitar de nova compilação do código fonte.

Licença: Acadêmica.

10. Babel ou OpenBabel

Descrição: Programa multiplataforma para conversão de diversos modelos de arquivo de entrada de programas de química computacional [141]

Versão: 1.6 e 1.100.2, com código em C.

Funcionalidade: O programa converte os dados referentes a representação molecular, mas ignora os dados referentes ao controle de simulação, quando existirem.

Licença: Domínio público.

11. Biocore

Descrição: Ambiente de trabalho que oferece uma interface para acesso a diversas tecnologias, locais ou remotas, utilizando protocolos de comunicação entre aplicativos e banco de dados, permitindo trabalhos de colaboração entre diversos centros de pesquisa biomédicas [142].

Versão: 12 de outubro de 2005.

Funcionalidade: O programa permite a execução e a visualização remota do sistema estudado.

Licença: Acadêmica.

12. Biomer

¹Do inglês *A Self-contained Atmospheric chemistry coDe*

Descrição: Pacote de modelagem biomolecular por dinâmica molecular. Por motivos de licença passou a se chamar simplesmente B [119].

Versão: 1.0, com código em Java.

Funcionalidade:s: Geração de estruturas de biomoléculas; leitura e escrita de arquivos no formato *PDB*, exportação de imagens nos formatos *jpeg*, *gif* e *ppm*; otimização geométrica; e simulações de dinâmica molecular com campo de força AMBER [118].

Licença: Domínio público.

13. CCP5

Descrição: Biblioteca gratuita de programas e documentos relacionados a química computacional, estando disponível pela internet ou através de solicitação por correio eletrônico [120].

Versão: Depende do programa utilizado.

Funcionalidade: Dinâmica molecular, simulação de Monte Carlo e outras técnicas clássicas.

Licença: Acadêmica.

14. Charmm

Descrição: Programa de simulação molecular de uso geral, baseado no campo de força empírico de mesmo nome. O campo de força está disponível em diversos outros programas de simulação molecular [121, 122, 154].

Versão: Não disponível.

Funcionalidade: Minimização energética, simulação de dinâmica molecular, análise vibracional e estimativa de propriedades termodinâmicas.

Licença: Existem versões acadêmicas e comerciais para o programa, e domínio público para o campo de força empírico.

Observação: A versão acadêmica está disponível por um custo de 600 dólares.

15. Chemoffice

Descrição: Pacote computacional com interface gráfica, desenvolvida pela Cambridge Software, para as plataformas Windows e Macintosh. Existem diversas versões, que variam em funcionalidade e preço [143].

Versão: 4.0 e 6.0.

Funcionalidade: Visualização tridimensional, medida de comprimento de ligação, ângulos e ângulos diedros, otimização geométrica através de métodos *MM2* e Huckel Extendido, interface com os programas Mopac e Gaussian, obtenção de propriedades moleculares.

Licença: Comercial.

Observação: Foram testadas as versões para avaliação do programa, que possuem funcionalidades bastante reduzidas quando comparadas as versões comerciais.

16. Dalton

Descrição:: Programa de cálculo de estrutura eletrônica molecular através de métodos *ab-initio* e do funcional de densidade, possuindo diversos recursos voltados para cálculos de propriedades de ressonância magnética [91].

Versão: 1.2.1, com código em Fortran 77.

Funcionalidade:: Obtenção de propriedades moleculares (métodos *HF*, *DFT*, *MCSCF*, e *CC*), cálculos energéticos em sistemas multireferenciais.

Licença: Acadêmica.

Observações: Existe uma lista de discussão onde é possível submeter dúvidas e trocar informações com a comunidade de usuários e desenvolvedores do programa.

17. Deft

Descrição: Programa de cálculo de funcional de densidade através da combinação linear de orbitais do tipo gaussiano, permitindo otimização geométrica, análises vibracionais e correlação de gradientes de spin locais [113].

Versão: Indefinida, com código fonte escrito em Fortran 77.

Funcionalidade: Cálculos do funcional de densidade eletrônica.

Licença: Domínio público.

Observações: Código facilmente portátil para diferentes plataformas.

18. Eceppak

Descrição: Programa para análise conformacional de peptídios, baseado em campo de força empírico de mesmo nome [114].

Versão: 3.0, com código em Fortran77.

Funcionalidade: Minimização energética, distribuição eletrostática, mapeamento energético de pares de ângulos diedros, classificação de conformação geométrica.

Licença: Acadêmica.

Observação: Possui recursos para execução em ambiente computacional paralelo.

19. Frenkel e Smit

Descrição: Programas que acompanham o livro “Understanding Molecular Simulations: From Algorithms to Applications” de Daan Frenkel e Berend Smit [123].

Versão: Indefinida, com código em Fortran 77.

Funcionalidade: Simulação molecular.

Licença: Domínio Público.

20. Gamess

Descrição: O Gamess² é um pacote de cálculos de estrutura eletrônica de uso geral [92].

Versão: Praticamente todas as versões desde 2002, com código em Fortran 77.

Funcionalidade: Cálculos pelos métodos *RHF*, *UHF*, *ROHF*, *GVB* e *MCSCF*, cálculos de correlação energética pelos métodos *DFT*, *CI*, *MP2* e *CC*, cálculos semi-empíricos pelos métodos *MNDO*, *AM1*, ou *PM3*, determinação de estados de transição e coordenadas de reação, cálculos de frequências vibracionais, cálculos de propriedade moleculares e efeitos de solvente.

Licença: Acadêmica.

Observação: O pacote inclui softwares para criação de arquivos de entrada e visualização dos arquivos de saída.

21. Gepol

²General Atomic and Molecular Electronic Structure System

Descrição: Programa para cálculo de superfície e volume molecular em função dos pontos centrais atômicos [135].

Versão: Indefinida, com código em Fortran 77.

Funcionalidade: Cálculo da superfície molecular de Van der Waals, cálculo da superfície molecular acessível e cálculo da superfície molecular não disponível ao solvente.

Licença: Domínio público.

22. Gromacs

Descrição: Programa para simulação de dinâmica molecular e minimização energética, criado inicialmente para estudos de moléculas bioquímicas, como proteínas e lipídios, mas atualmente está sendo empregado para estudos de polímeros [124].

Versão: 3.0.5, 3.1.4, 3.2.1 e 3.3, com código em C.

Funcionalidade: Cálculo de propriedades no equilíbrio (constante de ligação inibidor-enzima, potencial energético do sistema, função de distribuição radial no líquido), cálculo de propriedades dinâmicas (viscosidade do líquido, processo de difusão em membranas, dinâmica da mudança de fase).

Licença: Domínio público.

Observação: Possui recursos para execução em ambiente computacional paralelo.

23. Kintecus

Descrição: Programa para modelagem de reações em sistemas químicos, biológicos e atmosféricos [136].

Versão: kintw95, 2.71 e 3.82.

Funcionalidade: Modelagem de reações.

Licença: Comercial.

Observação: O programa foi escrito em Basic e utiliza como interface gráfica o programa Microsoft Excel.

24. Loop

Descrição: Programa para gerar coordenadas atômicas baseadas no alinhamento com padrões estruturais homólogos, utilizando para isto reconhecimento de sinais, como alinhamento seqüencial, perfil seqüencial, estrutura secundária e superfície molecular exposta [137, 138, 155].

Versão: loop2000 e suas atualizações, com código em C.

Funcionalidade: Os sinais são combinados localmente para criar modelos mistos e globalmente para fornecer pontuações gerais. As coordenadas atômicas são geradas utilizando critérios de alinhamento obtidos pela análise da pontuação geral.

Licença: Acadêmica.

25. Mindy

Descrição: Programa de simulação molecular extremamente simplificado, derivado do programa Namd, que utiliza parâmetros e estruturas compatíveis com o Xplor [125].

Versão: 1.0, com código em C++.

Funcionalidade: Simulações de dinâmica molecular.

Licença: Acadêmica.

26. MMFF

Descrição: Campo de força empírico de uso geral para simulações moleculares, desenvolvido pela Merck [144].

Versão: MMFF94.

Funcionalidade: Possui parametrização para quase todos os elementos da tabela periódica.

Licença: Domínio público.

Observação: O campo de força pode ser incorporado em diversos programas de simulação molecular.

27. Mobosol

Descrição: Versão alterada do Mopac para cálculos de energia livre e de orbitais [115].

Versão: Indefinida.

Funcionalidade: Cálculos de energia de solvatação e análise *NBO*³.

Licença: Domínio público.

28. Moil

Descrição: Programa de simulação molecular com interface gráfica que permite desenhar e imprimir estruturas moleculares individuais, conjunto de estruturas e acompanhamento da simulação de dinâmica molecular [126].

Versão: 2k e 9.0.1, com código em Fortran 77 e TCL/TK.

Funcionalidade: Cálculos de energia, trajetórias, amostragem local aperfeiçoada, modelo Landau-Zerner para dissociação e recombinação de ligantes, possuindo um algoritmo que permite a simulação de um grande intervalo de tempo.

Licença: Domínio público.

29. Molda

Descrição: Programa gráfico para criação e visualização de moléculas, em duas ou três dimensões, com suporte a diversos formatos de arquivo [145].

Versão: 1.0.

Funcionalidade: Programa escrito em Java, podendo ser utilizado nas plataformas suportadas por esta linguagem.

Licença: Acadêmica.

30. Molden

Descrição: Programa gráfico para visualização de resultados obtidos por diversos pacotes de química computacional, como Gaussian, Gamess, Mopac/Ampac, Tinker, entre outros. É possível visualizar os orbitais moleculares e a densidade eletrônica, medir distâncias interatômicas e ângulos de ligação, criar gráficos de contorno e de malha 3-D [146].

Versão: A partir da versão 3.3.

Funcionalidade: Visualização de moléculas, análise de resultados de diversos programas, geração de arquivos gráficos ou animações.

Licença: Acadêmica.

³Do inglês *Natural Bound Orbitals*

Observação: O programa também realiza cálculos de distribuição eletrônica.

31. Moldy

Descrição: Programa para simulação molecular em matéria condensada, podendo manipular muitos agrupamentos de átomos, moléculas ou íons [127].

Versão: 2.16 e 1.16E, com código em C.

Funcionalidade: Permite simulações nos agrupamentos NVE , NVT , $N\sigma H$ ou $N\sigma T$. O agrupamento não necessita ser cúbico, permitindo a simulação de sólidos e líquidos.

Licença: Domínio público.

Observação: O programa permite ainda acompanhar a consistência dos parâmetros, reiniciar simulações e traçar trajetórias.

32. Molmol

Descrição: Programa gráfico para visualização, análise e manipulação tridimensional de macromoléculas biológicas, com ênfase em proteínas e estruturas de DNA [147].

Versão: 2k.1.0 e 2k.2.0, com código fonte disponível escrito na linguagem C.

Funcionalidade: As estruturas podem ser obtidas a partir de espectros de Ressonância Magnética Nuclear.

Licença: Domínio Público.

33. Mopac

Descrição: Programa para cálculos de estrutura eletrônica através de métodos semi-empíricos para o estudo de moléculas e reações [101].

Versão: 6.0, 7.0, 7.04 e atualizações, com código fonte disponível escrito na linguagem Fortran 77.

Funcionalidade: Métodos MNDO, MINDO/3, AM1, e PM3 para obtenção de orbitais moleculares, calor de formação e geometria molecular, estimativa do espectro vibracional, parâmetros termodinâmicos e constantes de força e estado de transição.

Licença: Domínio público até a versão 7 e depois comercial.

Observação: A versão 7 possui diversos erros conhecidos.

34. Namd

Descrição: Programas para simulação molecular de alto desempenho em ambiente computacional paralelo. O paralelismo é baseado na biblioteca Charm++, podendo rodar simultaneamente em centenas de processadores [128].

versão 2.2, 2.4, 2.5 e 2.6, com código fonte disponível escrito em C.

Funcionalidade: Possui compatibilidade com os programas Amber, Charmm e Xplor.

Licença: Domínio público.

35. NWChem

Descrição: Programa de química computacional desenvolvido para alto desempenho em processamento paralelo, permitindo cálculos de estrutura eletrônica e simulação molecular clássica [93].

Versão: 4.1 e 4.7, com código fonte disponível escrito na linguagem C.

Funcionalidade: Cálculos quânticos, cálculos clássicos e abordagem híbrida.

Licença: Acadêmica.

Observação: Excelente desempenho em ambiente paralelo. Devido a grande possibilidade de cálculos, a preparação dos parâmetros de simulação são complexos.

36. Polyrate

Descrição: Programa para análise de reagentes, produtos e estados de transição de reações químicas e cálculo de constantes de reação através da teoria do estado de transição variacional (*VTST*)⁴, podendo ser aplicada em fase gasosa ou sólida e na interface gás-sólido [116].

Versão: 8.6 e 9.3.1, com código fonte disponível escrito em Fortran 77.

Funcionalidade: Análise de reações envolvendo até dois reagentes e dois produtos.

Licença: Acadêmica.

Observação: A versão 8.6 apresentou diversos erros de compilação, devido inconsistência na chamada de sub-rotinas.

⁴Do inglês *Variational Transition State Theory*

37. Protarch

Descrição: Programa para predição de estruturas de proteínas baseado em hipóteses termodinâmicas através de um método de otimização da energia da cadeia polipeptídica [139].

Versão: 1.0, com código fonte disponível escrito em Fortran 77.

Funcionalidade: Otimização geométrica; minimização energética.

Licença: Acadêmica.

38. Rasmol

Descrição: Programas para visualização gráfica de moléculas de forma rápida e intuitiva, com diversas opções de representação molecular [148].

Versão: 2.7.2.1.1, com código fonte disponível escrito em C.

Funcionalidade: Visualização de moléculas.

Licença: Domínio público.

Observação: Desenvolvido pelos laboratório farmacêutico Sandoz.

39. Sadus

Descrição: Programas do livro “Molecular Simulation of Fluids: Theory, Algorithms and Object-Orientation”, de R. J. Sadus [105, 129]

Versão: Indisponível, com código fonte escrito na linguagem C++

Funcionalidade: Simulação de Monte Carlo, simulação de dinâmica molecular, simulação mista

Licença: Domínio Público

40. Solvate

Descrição: Programas para construção de ambiente de simulação de uma macromolécula em solvente para simulação de dinâmica molecular [130].

Versão: 1.0, com código fonte disponível escrito em C.

Funcionalidade: Geração de volume de variadas formas geométricas, analisa a distância mínima ente soluto e limite da caixa de simulação, inclui íons obedecendo a distribuição de Debye-Hückel.

Licença: Domínio público.

Observação: O site do programa não está mais disponível.

41. Tinker

Descrição: Pacote de programas de uso geral para simulação molecular, com recursos adicionais para simulação de biopolímeros [131].

Versão: 3.8, 3.9, 4.1 e 4.2, com código fonte disponível escrito em Fortran 77.

Funcionalidade: Permite utilizar a maioria dos campos de força, como Amber (ff94, ff96, ff98 and ff99), CHARMM (19 e 27), Allinger MM (MM2-1991 and MM3-2000), OPLS (OPLS-UA, OPLS-AA and OPLS-AA/L), potenciais polarizáveis de Liam Dang e o seu próprio campo de força, o AMOEBA.

Licença: Acadêmica.

Observação: O usuário define quais os recursos computacionais necessários na compilação do programa.

42. Towhee

Descrição: Programa de simulação molecular pelo método de Monte Carlo para predição de equilíbrio de fases entre fluidos [132].

Versão: 4.12.6 e 4.12.7, com código fonte disponível escrito em Fortran 77.