

COMPUTAÇÃO EM QUÍMICA TEÓRICA: INFORMAÇÕES TÉCNICAS<sup>#</sup>

Nelson Henrique Morgon\*

Departamento de Físico-Química, Instituto de Química, Universidade Estadual de Campinas, CP 6154, 13083-970 Campinas - SP

Recebido em 4/7/00; aceito em 29/11/00

COMPUTATION IN THEORETICAL CHEMISTRY: TECHNICAL INFORMATION. The purpose of this work is to demonstrate the usefulness of low cost high performance computers. It is presented technics and software packages used by computational chemists. Access to high-performance computing power remains crucial for many computational quantum chemistry. So, this work introduces the concept of PC cluster, an economical computing platform.

Keywords: computers; PC cluster; theoretical chemistry.

## INTRODUÇÃO

Considerando-se que a química por excelência é uma ciência experimental, do ponto de vista do química teórico, a frase a seguir antecipa em aproximadamente 100 anos a expressão puramente matemática ( $\hat{H}\Psi = E\Psi$ ) dada por E. Schroedinger (1926) na descrição da estrutura eletrônica de sistemas e interações moleculares.

*Quanto mais as ciências físicas progridem, mais elas tendem a entrar no domínio da Matemática, que é um tipo de centro para onde elas convergem. Nós podemos julgar o grau de perfeição que a ciência tem alcançado, pela facilidade com que ela pode ser submetida a CÁLCULO.*- A. Quetelet (1828).

No entanto, M. Planck, um dos fundadores da Mecânica Quântica que propiciou o desenvolvimento da Eq. de Schroedinger, assinala que:

*Experimentos são os únicos meios de conhecimento a nossa disposição, o resto é poesia, imaginação.*- M. Planck (1901).

Esta aparente contradição permite assinalar uma das grandes virtudes atuais da química teórica, a de atuar como ferramenta de apoio na análise e interpretação de dados experimentais, através de informações que muitas vezes não são possíveis de serem obtidas diretamente dos experimentos<sup>1-4</sup>, ou na previsão de propriedades diversas.

Na primeira parte do trabalho é apresentado um panorama sucinto da química teórica (de estrutura eletrônica e mecânica clássica), enfatizando aplicações e evolução. Posteriormente, serão abordados aspectos técnicos relacionados a computadores como instrumento de trabalho utilizados pelos químicos teóricos, destacando-se informações básicas de seus principais componentes e a descrição de alguns programas de cálculos e sistema operacional, enfatizando-se o caráter gratuito dos mesmos. Finalmente, será apresentada a idéia de computação de alto desempenho a baixo custo, trazendo informações recentes sobre construção de *cluster* e aspectos relacionados a eles.

Não se objetiva citações tão brilhantes quanto as precedentes, mas tão somente que ao final do trabalho tenha ficado algo, sob o ponto de vista do autor, como:

**O que diferencia uma nação tecnologicamente desenvolvida, de uma em desenvolvimento é a propriedade**

**do conhecimento tecnológico. E, o que pode igualar os povos destas nações é a capacidade intelectual de aprendizado de seus membros.**

Preende-se, deste modo, fornecer subsídios iniciais para uma análise de que, não existe limitação de ordem financeira, dentro de um mínimo aceitável, que impeça através de material disponível no mercado, a pesquisadores, principalmente iniciantes e de instituições de ensino e pesquisa não tão abonadas, a aquisição de equipamentos e o desenvolvimento e execução de projetos de pesquisa na área de química teórica computacional, com qualidade.

A química teórica, basicamente compreendida em quântica (métodos *ab initio*, semi-empírico, funcional de densidade, ...) e clássica (mecânica dinâmica - determinístico e monte carlo - estocástico; ...) tem-se tornado bastante popular nas últimas décadas. Só a título de informação, considerando-se os números de trabalhos teóricos na última década de dois dos principais eventos para os químicos teóricos no país, a Reunião Anual da Sociedade Brasileira de Química (RASBQ) e o Simpósio Brasileiro de Química Teórica (SBQT) observa-se pela Tabela 1 que, neste período, o número de trabalhos nestes encontros tem aumentado significativamente, proporcional é claro, ao aumento do número de participantes nestes eventos.

**Tabela 1.** Número de trabalhos em Química Teórica<sup>(a)</sup> da Reunião da SBQ e SBQT.

Ano	Nº de Trabalhos	
	RASBQ <sup>(b)</sup>	SBQT <sup>(c)</sup>
1.991 <sup>(d)</sup>	-	124
1.992 <sup>(d)</sup>	-	-
1.993	22	189
1.994	24	-
1.995	37	216
1.996	35	-
1.997	25	300
1.998	23	-
1.999	32	386
2.000	44	-

<sup>(a)</sup>Fonte: Livros de Resumo; <sup>(b)</sup>Especificamente na Seção de Química Teórica, embora existam outras seções que utilizem cálculos teóricos; <sup>(c)</sup> Encontro Bianual; <sup>(d)</sup> Nos anos anteriores a 1.993, os trabalhos de Química Teórica na RASBQ constavam na Seção de Físico-Química.

Para explicar o porquê deste aumento, alguns fatores têm contribuído, como:

<sup>#</sup> Parcialmente apresentado no curso "Química Computacional" da 23ª Reunião Anual da Sociedade Brasileira de Química, Poços de Caldas, MG, Maio/2000.

\* e-mail: morgon@iqm.unicamp.br - http://canario.iqm.unicamp.br

1. o crescimento da comunidade científica;
2. a popularização de pacotes computacionais de química teórica, como por exemplo a série de programas Gaussian<sup>5</sup>;
3. a implementação de algoritmos matemáticos e métodos teóricos mais eficientes; e
4. equipamentos (computadores e periféricos) com melhor desempenho em processamento e transmissão de dados, e a custos menores.

Estes fatores refletem principalmente na qualidade dos resultados calculados de propriedades moleculares (comparáveis à precisão química) e na possibilidade de estender os cálculos a sistemas de interesse químico (experimental), destacando-se aqueles sistemas de tamanho médio, principalmente empregando metodologias de alto nível (*ab initio* e funcional de densidade). Isto é fato atualmente, visto que muitos dos trabalhos apresentados nestes encontros são frutos de colaboração entre teóricos e experimentais.

Um resumo da evolução e desenvolvimento computacional de métodos de química quântica utilizados no estudo de propriedades de estrutura eletrônica de átomos e moléculas foi apresentado com bastante propriedade recentemente nesta revista<sup>6</sup>.

No caso da implementação de métodos, os dois exemplos a seguir são marcantes. O primeiro refere-se à Teoria de Perturbação (MPn), fundamentada nos trabalhos de Møller e Plesset<sup>7</sup> da década de 30, mas que apenas a partir da década de 70 teve cálculos de energias e gradientes implementados eficientemente nos programas computacionais. O segundo exemplo, é mais recente, trata-se da Teoria do Funcional de Densidade, formulada a partir dos trabalhos de Kohn, Hohenberg e Sham<sup>8,9</sup>, que se tornou popular a partir de meados da década de 80<sup>10</sup>, ou seja 20 anos depois. E, com relação a algoritmos mais eficientes destacam-se implementações de cálculo direto de integrais eletrônicas, métodos de otimização global, de integração numérica (DFT), cálculos analíticos de gradientes de energia, da matriz hessiana (frequência e intensidades), busca na superfície de energia potencial de estado de transição, entre outros.

No caso dos equipamentos, o entendimento fica claro considerando-se a própria evolução histórica do processo. De acordo com Clementi<sup>11</sup>, a primeira geração de computadores (1946-1955) começa com o computador ENIAC e termina com o modelo IBM-704 (a primeira máquina comercial) capaz da proeza em obter desempenho de aproximadamente 5 kFlop/s<sup>12,13</sup>. Ainda, segundo Clementi, deve-se estar agora na sétima geração (199X-2010), cuja principal característica é a de ser definida por máquinas multiprocessadas<sup>14</sup>, que alcançam desempenho de TFlop/s<sup>15,16</sup> ou seja, isto equivale a uma evolução em Flop/s da ordem de 200 milhões de vezes (num período de meio século), usando-se como parâmetro o que há de equipamento mais avançado. No entanto, mesmo computadores pessoais conseguem processamento bastante significativo, na faixa de MFlop/s<sup>17</sup>. Alguns projetos científicos estão começando a pensar em computadores PetaFlop/s, são máquinas que irão necessitar de um número expressivo de processadores trabalhando em paralelo num mesmo problema<sup>18,19</sup>.

Um outro avanço que disseminou o uso de cálculos teóricos foi o surgimento da Internet, marcante principalmente por facilitar: a) a divulgação de conhecimento muito rápido; b) consultas *on line* de bancos de dados; e c) acesso a revistas científicas.

Graças aos avanços citados acima, observa-se uma forte interação entre químicos teóricos e experimentais, e com uma frequência crescente o surgimento, em bancadas dos laboratórios, de estações de trabalho e computadores pessoais voltados exclusivamente para cálculos teóricos. Aplicações rotineiras podem ser encontradas, nas diversas áreas da Química, como:

**Físico-Química** - no cálculo de propriedades termodinâmicas de gases; na interpretação de espectro molecular; na determinação de propriedades estruturais (comprimentos e ângulos de ligação); na obtenção de diferenças de energias conformacionais e de barreiras de energias rotacionais; na caracterização de estados de transição e estimativa de constantes de velocidade;

**Química Orgânica** - no estudo de estabilidade relativa de isômeros; na caracterização de intermediários, úteis no estabelecimento e entendimento de mecanismos de reação; no estudo de aromaticidade de compostos; na obtenção e análise de espectros de RMN;

**Química Inorgânica** - no uso da teoria do campo ligante - método quântico aproximado; na utilização do estudo de íons de complexos de metais de transição; em catálises homogênea e heterogênea; em processos de adsorção;

**Química Analítica** - no uso de métodos espectroscópicos de análise (frequências e intensidades de espectros); no estudo de compostos de interesse ambiental; e em

**Bioquímica** - na análise conformacional de grandes sistemas moleculares de importância biológica (macromoléculas, proteínas, enzimas); no estudo da interação enzima-substrato; em processos sob efeito de solventes. Especificamente no caso da bioquímica, a potencialidade nesta área é muito grande, como por exemplo no planejamento racional de fármacos. Enfim, o espectro de aplicações transcende os exemplos enumerados acima.

Há ainda um novo cenário surgindo com o aparecimento de uma nova estratégia de cálculos, fruto do casamento entre métodos de Química Quântica (estrutura eletrônica) e Mecânica/Dinâmica Molecular, conhecida por: *Quantum Mechanics/Molecular Mechanics* - QM/MM, ou métodos híbridos. As vantagens de uns sanando desvantagens de outros, por exemplo, é sabido que métodos de mecânica molecular falham na descrição de propriedades onde há a necessidade explícita da participação de elétrons, como na quebra e formação de ligações químicas, mas são extremamente úteis em sistemas moleculares grandes, para os quais existam parâmetros<sup>20</sup>. Assim, uma possibilidade é descrever partes do sistema por um ou outro método, mas isto será objeto de um outro trabalho.

A discussão a seguir aborda aspectos da evolução tecnológica, que propiciou esta difusão de pesquisa na área teórica, porém tendo como princípio apresentar, do ponto de vista técnico, um estudo que possibilite a criação de um ambiente computacional. Isto será feito através de informações e ferramentas úteis para iniciantes e para quem se aventura na área de cálculos teóricos. A principal característica que este trabalho ambiciona, é a de dar informações para a construção deste ambiente, e que seja um ambiente de alto desempenho computacional a um baixo custo financeiro, assim toda a ênfase será dada, de modo a que o interessado gaste somente na aquisição de equipamentos (dentro da ótica de menor custo), e que toda a "infraestrutura" de *software* seja sem nenhum custo adicional. Deste modo, quem é o público-alvo deste trabalho? a) jovens pesquisadores; b) pesquisadores de instituições com limitados orçamentos destinados à pesquisa; e c) aqueles que optam por não ficar dependentes de pacotes comerciais.

Um primeiro passo são os computadores pessoais, conectados à rede mundial, esquematizado na Figura 1. São equipamentos de baixo custo, na faixa de R\$ 2.000,00 a R\$ 5.000,00 e que podem desempenhar uma variedade de tarefas: cálculos para uma grande diversidade de sistemas e propriedades (em grandes sistemas, métodos de Mecânica Clássica - Mecânica/Dinâmica Molecular e Monte Carlo; para sistemas intermediários e/ou baixa precisão, métodos semiempíricos e/ou Hartree-Fock com conjuntos de funções de base modestos, e para pequenos sistemas - alguns átomos, é possível cálculos mais precisos com correlação eletrônica e conjuntos de base estendidos). Podem ser utilizados também para edição e preparação de artigos, projetos, relatórios científicos, produção de material didático (textos, hipertextos, gráficos, imagens) e conectados à internet, que permite, entre tantas outras utilidades, acessar bibliotecas virtuais (*Web of Science*<sup>21</sup>, ProBE<sup>22</sup> e SciELO<sup>23</sup>), o que possibilita a constante atualização e aquisição tanto de conhecimento científico quanto de novos *softwares*.

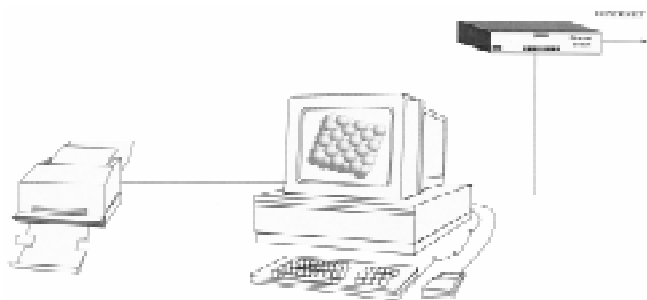


Figura 1. Equipamento de uso pessoal como infraestrutura mínima necessária para desenvolver pesquisas em química computacional.

Abaixo estão alguns aplicativos com descrições básicas e informações técnicas, bastante úteis para o químico teórico.

1. Equipamento computacional (características mínimas necessárias): **PC** com processador (a partir de 200MHz), podendo ser simples ou dual, do tipo Intel (Pentium Pro, II, III e IV), Cyrix, AMD (Duron, Athlon K7), ...; placa de rede (10Mb/s) para acesso à internet; disco (5GB) IDE ou SCSI com controladora (melhor desempenho), memória RAM (128MB) e placa de vídeo (2MB) para aplicações envolvendo visualização gráfica, monitor e impressora.
2. Sistema Operacional: **UNIX**<sup>24</sup>, trata-se de um sistema concebido a partir do projeto MULTICS (*MULTiplexed Information and Computing Service*) e desenvolvido por Massachusetts Institute of Technology (MIT), Bell Labs e General Electric (GE), na década de 60. Origina-se como um sistema operacional experimental e específico para computador GE635. Foi criado para ser flexível e interativo, posteriormente Ken Thompson e Dennis Ritchie alteraram o complexo sistema operacional inicial e criaram um sistema de arquivos simples denominado UNICS (*UNIplicated Information and Computing Service*) - UNIX. Como características gerais, destaca-se por ser multiusuário e multitarefa, possuir sistema de arquivos (e a própria concepção de arquivo), ter facilidade em ser usado em rede e possuir um conjunto de utilitários. Existe uma grande variedade de versões comerciais (AIX, Ultrix, Sun-OS, Convex, SGI, Cray, OSF) e de domínio público (**FreeBSD**<sup>25,26</sup> e **LINUX**<sup>27,28</sup>).
3. Compiladores para fortran, C, C++: GNU (**gcc**, **g77**), **egcs**, **f2c**<sup>29</sup>.
4. Editores de texto científico: **latex**<sup>30</sup>, é um conjunto de macros para TeX, um sistema de processamento de texto de alta qualidade voltado para a produção de documentos técnicos e científicos. Vale salientar que revistas científicas internacionais<sup>31,32</sup> aceitam artigos preparados com LaTeX, dispondo de estilos próprios facilmente implementados.
5. Construtores de estruturas e geração de figuras: **xfig**<sup>33</sup>
6. Visualizadores de imagens e textos: **xv**<sup>34</sup>, **gv**<sup>35</sup>, **xdvi**, ...
7. Web Pages: **latex2html**<sup>36</sup>, é conversor de alta qualidade de documentos preparados em LaTeX para HTML e de grande utilidade na preparação de material técnico, científico e educacional a ser disponibilizado em rede.
8. Conversor de formatos de arquivos: **babel**<sup>37</sup>, é um programa voltado para interconverter inúmeros formatos muito usados em modelagem molecular. Por exemplo, arquivos em coordenadas cartesianas (xyz) para matriz-Z.
9. Pacotes Gráficos:
  - \* **gnuplot**<sup>38</sup>, pacote gráfico iterativo que plota funções, conjuntos de dados em 2D e superfícies (3D);
  - \* **xmgr**<sup>39</sup>, pacote para plotar em 2D usado em sistema X Window e que utiliza interface OSF/Motif.
  - \* **glace**<sup>40</sup>, sucessor do xmgr.

10. Pacotes Gráficos para edição e visualização de estruturas e propriedades moleculares:

- \* **molden**<sup>41</sup>, programa de pré- e pós-processamento gráfico, utilizado na visualização de estrutura eletrônica e edição de estruturas moleculares. Capaz de permitir a visualização de orbitais moleculares, densidades eletrônica, potencial eletrostático, entre outras propriedades e potencialidades. Na Figura 2 tem-se uma amostra do editor de estruturas moleculares.

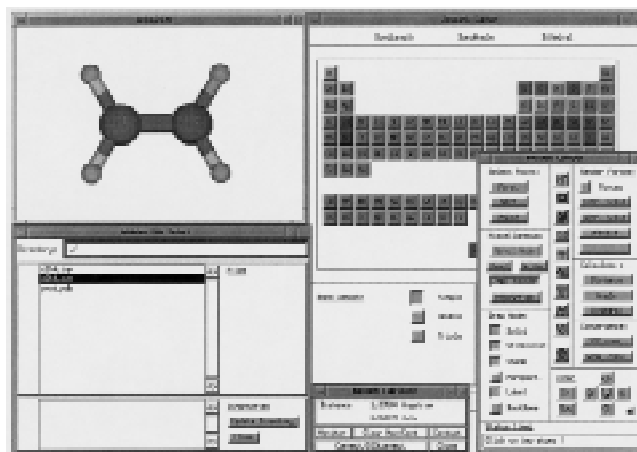


Figura 2. Editor e visualizador de estruturas moleculares do pacote Molden.

Outros pacotes gráficos de edição e visualização molecular são:

- \* **rasmol**<sup>42</sup>, manipula com bastante rapidez e qualidade de imagem gráfica, estruturas com muitos átomos, por exemplo, proteínas, como mostra a Figura 3. Lê diversos formatos de dados, principalmente pdb (*Protein Data Bank*).
- \* **garlic**<sup>43</sup>
- \* **qmol**<sup>44</sup>
- \* **xcrysdn**<sup>45</sup>, programa de visualização de estrutura e densidade molecular e cristalina

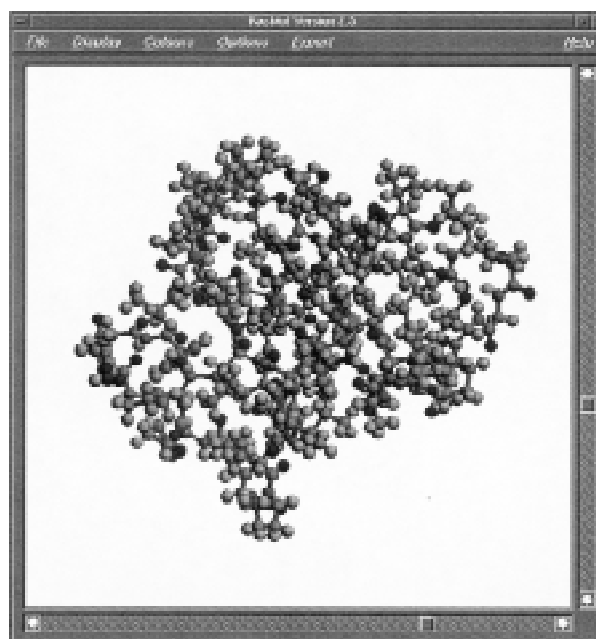


Figura 3. Estrutura da proteína crambin encontrada em sementes e composta por 45 grupos, 752 átomos e 1003 ligações, visualizada pelo pacote gráfico Rasmol.

## 11. Pacotes de programas de cálculo teórico:

- \* **Gamess**<sup>46</sup>: pacote geral de química quântica para cálculos *ab initio*, funcional de densidade e semi-empírico (MNDO, AM1 e PM3).
- \* **Defft**<sup>47</sup>: é um *software* de mecânica quântica computacional, baseado na teoria do funcional de densidade. Utiliza funcionais de densidade do tipo gaussianos.
- \* **Molfdir**<sup>48</sup>: código de química quântica que faz cálculos de sistemas moleculares multi-eletrônicos usando formalismo de Fock-Dirac e cálculos adicionais de correlação.
- \* **Moldy**<sup>49</sup>, é um programa de propósitos gerais, voltado para simulação de dinâmica molecular. É suficientemente flexível, devendo ser útil para uma grande faixa de cálculos de simulação de sistemas atômicos, iônicos e moleculares.
- \* **Dalton**<sup>50</sup>: programa de química quântica, cujo forte está nas áreas de propriedades elétricas e magnéticas e no estudo de superfícies de energia potencial para ambas as investigações estática e dinâmica.
- \* **Dirac**<sup>51</sup>: código para cálculos moleculares relativísticos baseados no hamiltoniano Dirac-Coulomb.
- \* **Tinker**<sup>52</sup> pacote de modelagem molecular. É concebido para ser um sistema de uso fácil e flexível empregado em mecânica e dinâmica molecular.

## 12. Outros aplicativos.

Existe uma grande quantidade de aplicativos. Para obter outros programas e informações adicionais consultar: <http://sal.engnux.ufsc.br/index.shtml>.

Como exemplo de aplicação geral, fez-se um cálculo de estrutura eletrônica *ab initio* para a molécula de  $C_2H_4$ , usando-se o programa Gamess. A partir da função de onda obtida a nível HF/4-31G, é possível, usando o programa Molden, visualizar algumas propriedades eletrônicas deste sistema. Nas Figuras 4, 5 e 6 estão, respectivamente, as imagens do orbital molecular correspondente à ligação dupla, do contorno da distribuição eletrônica total sobre a molécula e do potencial eletrostático (e as cargas) derivado da análise populacional de Mülliken.

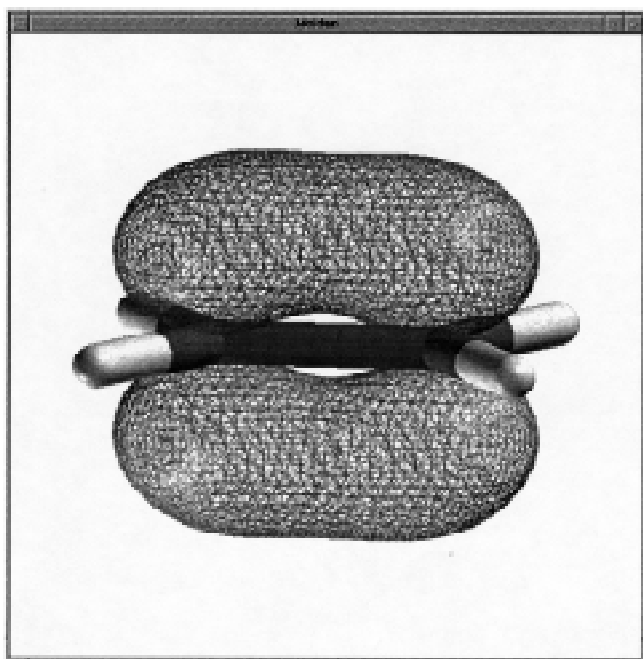


Figura 4. Orbital Molecular para  $C_2H_4$ , obtido pelo Gamess e visualizado pelo Molden.

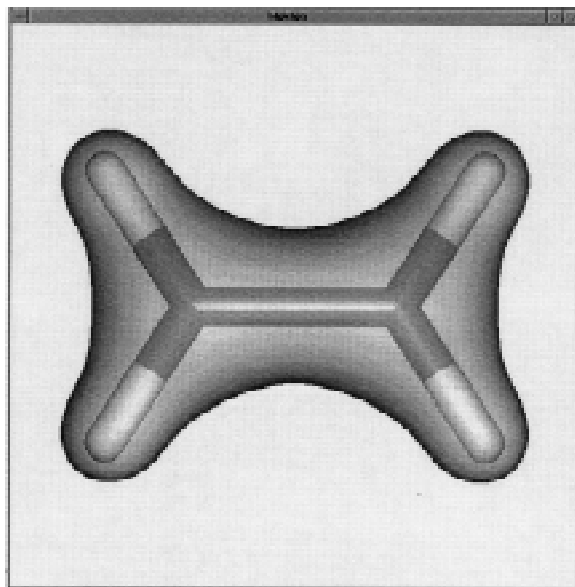


Figura 5. Densidade eletrônica para  $C_2H_4$ , calculada pelo Gamess e visualizada pelo Molden.

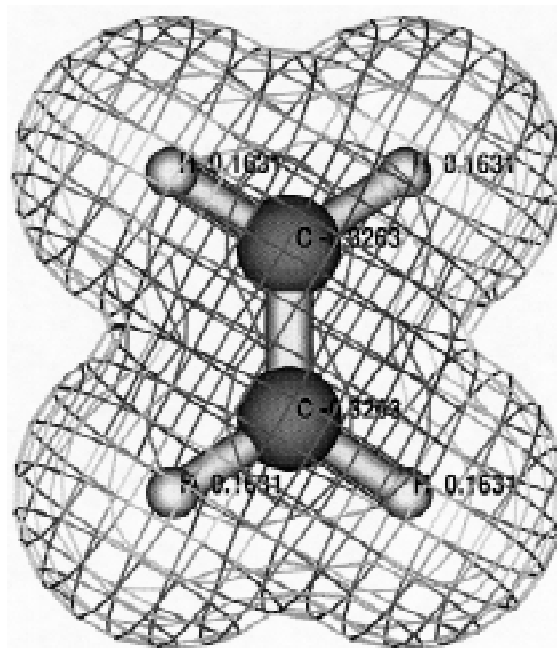


Figura 6. Potencial eletrostático para  $C_2H_4$ , calculado pelo Molden a partir da análise populacional obtida pelo Gamess.

As vantagens da internet são bem conhecidas, e algumas já foram enumeradas e comentadas anteriormente, no entanto um conceito particular da conexão em rede entre computadores e que será abordado a seguir, envolve o processo de transferência de dados entre computadores "próximos". Este procedimento permite a criação de um ambiente com mais de um processador trabalhando em paralelo numa determinada tarefa. Tais equipamentos são denominados de *clusters*. São máquinas com grande poder de processamento criando um ambiente de alto desempenho em computação paralela<sup>14</sup>.

Com o desempenho e o baixo custo atualmente dos PCs e a disponibilidade de conexão por ethernet (fastethernet: 100Mbit/s ou gigathernet: 1000Mbit/s), tornou-se possível combiná-los e construir ambientes de computação paralela de alto desempenho, da ordem de GFlop/s, a um custo reduzido,

se comparado a máquinas multiprocessadas comerciais. Com versões livres de UNIX e pacotes de *software* de domínio público, nenhum sistema de computação paralela, disponível comercialmente, pode competir com o preço de tais equipamentos. O argumento contrário a esta concepção de arquitetura, é freqüentemente apontado como a não existência de um centro de suporte quando um problema surgir, mas existe uma grande variedade de informações disponíveis em *sites* de *ftp*, *web* e *newgroups*<sup>53</sup>, que auxiliam. Além do aspecto da formação de recursos humanos, isto é, permite a educação de estudantes e a criação de uma cultura em ambiente de processamento de alto desempenho<sup>54</sup>.

Deste modo, tem se tornado muito populares os *clusters* de PCs ao redor do mundo, com dezenas e centenas de processadores e com inúmeras aplicações úteis em química computacional<sup>55</sup>. Uma versão modesta foi construída no IQ/UNICAMP em 1.997, usando 4 unidades, sendo cada uma composta por Pentium Pro 200 MHz, com 128 MB de memória RAM e 1GB de disco SCSI, placa Ethernet 100 Mbit/s e conectados via rede através de um *hub* de 8 portas. Equipamentos com no mínimo 16 processadores, são uma classe especial de *cluster* denominados *Beowulf*<sup>56</sup>. Eles têm se tornado muito populares em muitos centros de pesquisa no exterior como uma opção barata e de alta *performance*.

A conexão dos processadores pode ser feita por *hubs* ou *switches*, e placas de rede ethernet (*Gigabit Ethernet* e *FastEthernet*), que são relativamente de baixo custo, comparadas a outras tecnologias como fibra ótica (*Fiber Distributed Data Interface - FDDI*) e ATM (*Asynchronous Transfer Mode*). A principal diferença entre *hubs* e *switches* está em como as mensagens são distribuídas entre os processadores. No caso do *hub*, para uma mensagem ser enviada a partir de um dado processador para um outro processador, é feito um *broadcast* para todos os computadores da rede, ao passo que com *switch*, as mensagens são trocadas em canais exclusivos estabelecidos pelos dois processadores, em questão. Isto permite às *switches* suportarem comunicação denominada *full duplex*, ou seja, elas têm a habilidade em dobrar a velocidade de cada *link*, por exemplo de 100Mb/s a 200Mb/s. Hoje existe uma grande variedade de *switches* disponíveis, contendo de 8 a 100 portas, que podem alcançar desempenho na transferência de dados da ordem de Gigabit/s. Elas têm-se tornado tão baratas que não existem motivos para construir *clusters* interconectando-se os processadores usando *hubs* ou outra conexão.

*Clusters* de PCs em rede podem ser usados para processamento de múltiplos cálculos em seqüencial e podem também dar suporte para processamento em paralelo. Na Figura 7 tem-se um exemplo esquemático de um ambiente para execução de programas em processamento paralelo. A maioria dos programas de

química computacional tem implementação paralela disponível, principalmente para máquinas com memória distribuída (como os *clusters*). Eles trabalham tanto em ambiente seqüencial quanto em paralelo, como por exemplo, o pacote Games<sup>46</sup>. Eles utilizam-se de bibliotecas específicas envolvidas nas trocas (envio e recebimento) de mensagens entre os processadores na rede (conexão local). As mais populares são: *mpich*<sup>57</sup> e *pvm*<sup>58</sup>. São aplicativos que também permitem o desenvolvimento de programas paralelos.

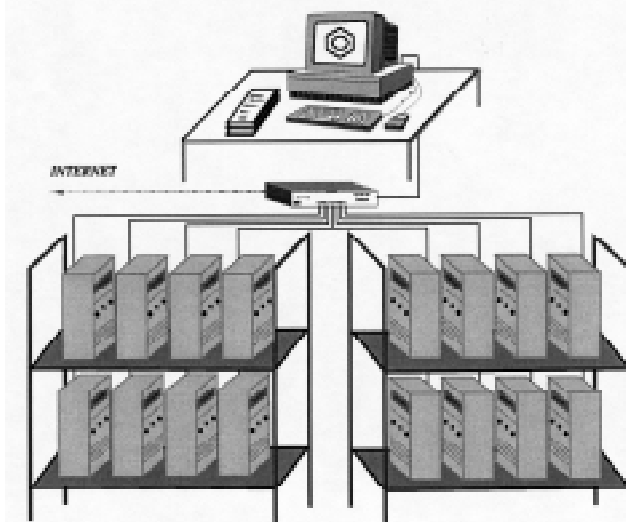


Figura 7. Cluster de PC's.

De um modo geral, na estrutura descrita pela Figura 7, todo o sistema de gerenciamento de contas de usuários e localização de programas e arquivos, é feita por um computador mais robusto (o servidor). Isto simplifica muito a manutenção e atualização dos programas e o próprio gerenciamento e controle do ambiente. Além de uma substancial economia no armazenamento de dados. As áreas comuns são propagadas do servidor via sistema de arquivos para os processadores (clientes) (*Network File System - NFS*).

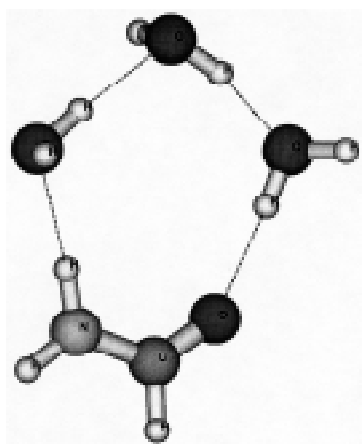
Na Tabela 2 estão colocados os tempos de processamento para os cálculos de otimização de geometria e freqüência vibracional harmônica, para o sistema formamida "solvatada" por três moléculas de água (representadas por EFP - *effective fragment potentials*), como mostra a Figura 8, no nível HF e otimização de geometria a MP2, usando-se conjuntos de funções de base de Dunning, DH(p,d). Apenas como comparação foram feitos testes usando

**Tabela 2.** Análise do tempo de processamento (em s) nos cálculos de otimização de geometria e freqüência vibracional harmônica para o sistema formamida solvatada por três moléculas de água (descritas por EFP), usando-se metodologias HF e otimização de geometria a MP2, com conjuntos de funções de base DH(p,d). Os cálculos foram feitos em ambientes de *cluster* de PCs e de estações de trabalho do CENAPAD/SP.

Método	Cálculo	Nº de PCs <sup>(a)</sup>		CENAPAD/SP <sup>(b),(c)</sup>		
		1	4	Thin66	Thin120	PWR3
HF	Otimização	5.758,3 (1.411,3)	1.547,2	3.066,5	966,4	1.065,5
	Freqüência	3.695,5 (829,4)	1.001,0	1.825,0	592,5	656,7
MP2	Otimização	8.002,0 (5.160,3)	4.215,1	8.545,3	2.553,2	3.005,3

<sup>(a)</sup>Pentium Pro 200@MHz (em parênteses, AMD Athlon(tm) 700@MHz) com 128MB de memória (RAM) e 1GB de disco local. Os processadores estão em uma rede ethernet de 100Mb/s; <sup>(b)</sup>Thin66: 4 IBM RISC/6000-370 66@MHz com 256MB RAM e 2GB de disco cada. Thin120: 4 IBM RISC/6000-Power2 Super 120@MHz com 512MB RAM e 4,5GB de disco cada. Interface SPS - 150Mb/s bi-direcional. PWR3: 1 CPU PowerPC 630 200@MHz (com 2 processadores) com 256MB de memória RAM, 4MB de cache L2 e 6GB de disco e Interface de rede Fast Ethernet (100Mb/s)<sup>59</sup>; <sup>(c)</sup>O CENAPAD/SP possui máquinas mais rápidas, como: IBM RISC/6000-Power2 Super 160@MHz com 1GB RAM e 9GB de disco conectados por interface SPS - 150Mb/s bi-direcional e SILVER - 1 CPU com 4 processadores 200@MHz com 2GMB de memória RAM, e 40GB SSA de disco não usadas no teste (ver texto).

PC e *cluster* de PCs do IQ/UNICAMP e estações de trabalho disponíveis no Centro Nacional de Processamento de Alto Desempenho em São Paulo (CENAPAD/SP)<sup>59</sup>.



**Figura 8.** Estrutura molecular para a formamida solvatada por três moléculas de água.

Analisando-se os resultados da Tabela 2 observa-se que otimização de geometria e frequência usando-se método HF tem taxa de paralelização bastante alta da ordem de 93%, enquanto otimização com MP2 esta taxa fica abaixo de 50%. A quantidade de memória influencia o desempenho de processamento, pois comparando-se o ambiente Thin120 (120@MHz e 512MB) com os 4 processadores PPro (200@MHz e 128MB) observa-se que o primeiro tem um desempenho superior. Outras análises podem ser feitas a partir da Tabela 2, como processamento do tipo SPMD (máquinas com mais de 1 CPU)<sup>14</sup>, tipo de interface de rede, memória *cache*, discos, etc. Versões mais novas de processadores, como AMD Athlon(tm) - 700@MHz, apresentam desempenho próximo ao de 4 processadores PPro 200@MHz. Embora, não se pretenda analisar exclusivamente custos de PCs x Estações de Trabalho, um dado interessante refere-se à disponibilidade das máquinas. Testes envolvendo máquinas mais rápidas do CENAPAD/SP (como as denominadas por Thin160 e SILVER, ver Tabela 2) não foram possíveis de serem feitos, pois os *Jobs* submetidos ficaram nas filas de submissão por vários dias, inviabilizando-se a análise. Isto é um dado interessante, pois com o *cluster* de PCs tem-se uma maior exclusividade.

De um modo geral, a preocupação em fornecer informações técnicas a respeito de processamento computacional para os químicos teóricos deve-se ao fato dos mesmos serem responsáveis por grande demanda de tempo de processamento. Para se ter uma idéia do que isto representa, o CENAPAD/SP tem aproximadamente 180 usuários cadastrados atualmente, sendo que os Químicos representam 40% destes usuários e no entanto, são responsáveis por 80% do tempo de processamento disponibilizado pelo Centro. Isto representa algo em torno de 10 GFlop/s.

Por que a necessidade de tanta demanda computacional ?

Basicamente pela natureza dos cálculos envolvidos. Considere-se os seguintes aspectos básicos nos exemplos ilustrativos de cálculos teóricos:

a) usando-se o método Hartree-Fock-Roothaan (método bastante utilizado na solução aproximada da Equação de Schroedinger), tem-se a necessidade de cálculo de um grande número de integrais de 1 e 2 elétrons, oriundos de um grande número de funções de base usadas na construção dos orbitais atômicos, combinados na formação dos orbitais moleculares. E, este número aumenta significativamente com o aumento do número de átomos;

b) usando-se métodos correlacionados (usados para minimizar os efeitos da aproximação do método HFR), além da etapa precedente, há a necessidade de calcular um grande número de configurações, geradas como função do número de orbitais necessários na construção da função de onda; e

c) já para os métodos de Mecânica, Dinâmica ou Simulação de Monte Carlo, o que existe é um grande número de moléculas componentes do sistema e/ou são sistemas macromoleculares;

Assim,

*“o que limita o uso de cálculos teóricos, não é o sistema estudado ou a metodologia empregada, mas o equipamento disponível.”*

## CONCLUSÃO

Com este trabalho procurou-se apresentar alternativas de baixo custo, bom desempenho de processamento e um roteiro básico inicial, para quem pretende utilizar-se de cálculos teóricos como mecanismo de apoio ao entendimento de aspectos relacionados à estrutura eletrônica e molecular de sistemas químicos.

Os equipamentos podem ser desde um PC, até um *cluster*. O custo para a construção do *cluster* de PCs, depende apenas do número de unidades (cada uma em torno de R\$ 2.000,00 e R\$ 5.000,00, variando-se o tipo e quantidade de disco, processador e memória, placas de rede e de vídeo, ...) e da conexão destes processadores que pode ser feita por *hubs* ou *switches*, que são relativamente de baixo custo (de R\$ 1.000,00 e R\$ 2.000,00 dependendo também do número de “portas”). Comparativamente aos custos de aquisição de estações de trabalho, os valores seriam dependentes de outros fatores, como marca (fabricante), modelo, suporte técnico, sistema operacional, bibliotecas matemáticas, compiladores, etc.

As grandes vantagens de equipamentos como os apresentados neste manuscrito é o de serem uma alternativa de baixo custo e bom desempenho e propiciar a formação de recursos humanos em atividades outras, que não a de meros executores de programas. E por desvantagem, se se considerar isto desvantagem, é o da manutenção, instalação de pacotes, etc.

## AGRADECIMENTOS

O autor gostaria de agradecer ao Instituto de Química da UNICAMP pelas facilidades computacionais e ao Conselho Nacional de Desenvolvimento Científico (CNPq) pelo suporte financeiro. Um agradecimento especial aos Profs. Pedro Vazquez e Rogério Custódio do IQ-UNICAMP pelas informações e discussões relacionadas a sistemas operacionais, redes e programas de cálculos teóricos.

## REFERÊNCIAS

- Morgon, N. H.; Custódio, R.; Riveros, J. M.; *Chem. Phys. Lett.* **1995**, 235, 436.
- Morgon, N. H.; Argenton, A. B.; Silva, M. L. P.; Riveros, J. M.; *J. Amer. Chem. Soc.* **1997**, 119, 1708.
- Morgon, N. H.; Riveros, J. M.; *J. Chem. Phys. A* **1999**, 102, 10399.
- Morgon, N. H.; Xavier, L. A.; Riveros, J. M.; *Int. J. Mass Spectrom.* **2000**, 196, 363.
- Frisch, M. J. et al. *“Gaussian/94 - Revision D.2”*, Gaussian Inc.—Pittsburgh PA, 1994.
- Freitas, L. C. G.; *Quim. Nova* **1999**, 22, 293.
- Møller, C.; Plesset, M. S.; *Phys. Rev.* **1934**, 46, 618.
- Hohenberg, P.; Kohn, W.; *Phys. Rev.* **1964**, 136, B864.
- Kohn, W.; Sham, L. J.; *Phys. Rev.* **1965**, 140, A1133.
- Morgon, N. H.; Custódio, R.; *Quim. Nova* **1995**, 18, 44.
- Clementi, E.; *Int. J. Quantum Chem.* **1992**, 42, 547.

12. 5000 operações de ponto flutuante por segundo, uma grandeza que “mede” a velocidade do processador em executar qualquer operação que envolva números fracionários.
13. McDaniel, G., Ed.; *The IBM Dictionary of Computing*; McGraw-Hill: , Tenth ed.; 1993.
14. Morgon, N. H.; *Quim. Nova* **1995**, *18*, 481.
15. 1 Tera =  $10^{12}$ .
16. “TOP500: Site dos 500 sistemas computacionais de maior desempenho”, <http://www.top500.org>.
17. *Benchmarks: Standard Performance Evaluation Corporation - SPEC*; centro para avaliação de desempenho computacional”, <http://www.specbench.org>.
18. *National Science Foundation*, <http://www.nsf.gov>.
19. Taubes, G.; *Discover* **1997**, *18*, 76.
20. Coelho, L. W.; Junqueira, G. M. A.; Herrera, J. O. M.; Machado, S. D.; Machado, B. D.; *Quim. Nova* **1999**, *22*, 396.
21. “A *Web of Science (WoS)* é uma base de dados produzida pelo *Institute for Scientific Information (ISI)*, com informações sobre artigos publicados”, <http://webofscience.fapesp.br>.
22. “ProBE (Programa Biblioteca Eletrônica da Fapesp) procura contribuir para o desenvolvimento da pesquisa, por meio da aquisição de revistas eletrônicas com textos completos”, <http://www.probe.br>.
23. “O *Scientific Electronic Library Online - SciELO* é uma biblioteca eletrônica virtual que cobre uma coleção selecionada de revistas científicas brasileiras”, <http://www.scielo.br>.
24. Salus, P. A.; *A quarter century of UNIX*; Addison-Wesley Publishing Company, Inc.: New York, 1994.
25. Hubbard, J. K. “FreeBSD, é um avançado sistema operacional BSD UNIX para computadores ‘PC-compatíveis’, desenvolvido e mantido por um grande número de colaboradores”, <http://www.freebsd.org>.
26. Brandi, E. Um guia “FreeBSD - Primeiros Passos” pode ser encontrado em: <http://freebsd.ag.com.br/>.
27. Torvalds, L. “Linux é um sistema operacional gratuito baseado em Unix, originalmente criado por Linus Torvalds”, <http://www.linux.org>.
28. Existem várias versões de LINUX, como: Red Hat, Debian, S.u.S.E. e Slackware, algumas sendo comerciais.
29. “GNU project. Compiladores: fortran, C, C++”, <http://gcc.gnu.org/>.
30. Lampion, L. “LaTeX: conjunto de macros para TeX (sistema de preparação de documentos de alta qualidade, frequentemente usado pela comunidade científica e técnica)”, <http://www.latex-project.org>.
31. “Editora Elsevier”, <http://www.elsevier.nl/homepage/authors>.
32. “*The Journal of Chemical Physics*, revista do Instituto Americano de Física”, <http://www.aip.org/pubservs/compuscript.html>.
33. Autor original: Supoj Sutanthavibul, “xfig: construção iterativa de figuras sob X11”, [ftp://ftp.x.org/contrib/applications/drawing\\_tools/xfig](ftp://ftp.x.org/contrib/applications/drawing_tools/xfig).
34. Bradley, J. “xv: visualizador iterativo de imagens para sistema X Window”, <ftp.cis.upenn.edu>.
35. Plass, J. “gv: pré-visualizador de PostScript e PDF”, <http://www.thep.physik.uni-mainz.de/~plass/gv>.
36. Drakos, N. “LaTeX2HTML: tradutor de LaTeX para formato HTML”, <http://saftack.fs.uni-bayreuth.de/~latex2ht>.
37. Walters, P.; Stahl, M. “Babel: conversor de formatos de arquivos”, <http://www.eyesopen.com/babel.html>.
38. Williams, T.; Kelley, C. “GnuPLOT: programa iterativo para construção de gráficos em 2D e superfícies”, <http://www.gnuplot.org>.
39. Turner, P. J. “Xmgr: ACE/gr é um pacote para plotar em 2D em sistema X Window, que utiliza interface Motif”, <http://plasma-gate.weizmann.ac.il/Xmgr>.
40. Turner, P. J.; Stambulchik, E. “Grace é um conjunto de ferramenta WYSIWYG para gráficos 2D em sistema X Window com interface OSF/Motif. É o sucessor do xmgr”, <http://plasma-gate.weizmann.ac.il/Grace>.
41. Schaftenaar, G. “Molden - Programa para pré e pós processamento de estruturas moleculares e eletrônicas”, CAOS/CAMM Center, Holanda, <http://www.caos.kun.nl/schaft/molden>.
42. Sayle, R. “Rasmol: programa para visualização gráfica molecular”, <http://www.umass.edu/microbio/rasmol>.
43. Zucic, D. “Garlic: programa de visualização molecular”, <http://pref.etfos.hr/garlic>.
44. Gans, J. “Qmol: programa para visualização de estrutura molecular e animação de trajetórias moleculares”, <http://lancelot.bio.cornell.edu/jason/qmol.html>.
45. Kokalj, A. “XCrySDen: programa de visualização de estrutura e densidade molecular e cristalina”, <http://www-k3.ijs.si/kokalj/xc/XCrySDen.html>.
46. Schmidt, M. W.; Baldrige, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. “GAMESS: programa *ab initio* de estrutura eletrônica”, 1993 <http://www.msg.ameslab.gov/GAMESS/GAMESS.html>.
47. St-Amant, A. “DeFT é um *software* de mecânica quântica computacional, baseado na teoria do funcional de densidade”, <http://www.chem.uottawa.ca/DeFT.html>.
48. Visscher, L.; Visser, O.; Aerts, P. J. C.; Merenga, H.; Nieuwpoort, W. C.; *Comput. Phys. Commun.* **1994**, *81*, <http://theochem.chem.rug.nl/~broer/Molfdiir/Molfdiir.html>.
49. Refson, K. “Moldy, é um programa de simulação de dinâmica molecular, bastante flexível e útil numa grande faixa de cálculos de simulação para sistemas atômicos, iônicos e moleculares”, <http://www.earth.ox.ac.uk/~keith/moldy.html>.
50. Helgaker, T.; *et. ali.*, “Dalton, programa de química quântica para cálculo de propriedades moleculares com funções de onda SCF, MP2 ou MCSCF”, <http://www.kjemi.uio.no/software/dalton>.
51. Saue, T.; Enevoldsen, T.; Helgaker, T.; Jensen, H. J. A.; Laerdahl, J.; Thyssen, J.; Visscher, L. “Dirac, programa de estrutura eletrônica completamente relativístico”, <http://dirac.chem.sdu.dk>.
52. Ponder, J. W. “Tinker, conjunto de *softwares* para modelagem molecular”, <http://dasher.wustl.edu/tinker>.
53. Lindheim, J. “The Beowulf Project at CACR”, <http://Vazquez, P. A. M.; Morgon, N. H.; 17ª Reunião Anual da SBQ, Caxambu, MG, 1994>.
54. Vazquez, P. A. M.; Morgon, N. H.; *17ª Reunião Anual da SBQ*, Caxambu, MG, 1994.
55. Tirado-Rives, J.; Jorgensen, W. L.; *J. Comput. Chem.* **1996**, *17*, 1385.
56. Sterling, T.; Becker, D. “The Beowulf Project”, <http://www.beowulf.org/>.
57. “MPICH implementação portátil de bibliotecas MPI - interface padrão de ‘message-passing’”, Argonne National Laboratory, <http://www-unix.mcs.anl.gov/mpi/mpich>.
58. “PVM - *Parallel Virtual Machine*, pacote de *software* para trocas de mensagens numa rede heterogênea de computadores”, Oak Ridge National Laboratory, <http://www.epm.ornl.gov/pvm/>.
59. “CENAPAD/SP - Centro Nacional de Processamento de Alto Desempenho em São Paulo”, <http://www.cenapad.unicamp.br>.