

IV Escola de Inverno do CBPF

15 a 16 de Julho de 2002

Modelagem Molecular Aplicada a Sólidos e Biomoléculas

Kleber C. Mundim

Instituto de Química

Universidade de Brasília (kcmundim@unb.br)

1. Introdução

A simulação computacional usando métodos semi-clássicos, nestes últimos anos, tem tido um papel de suma importância no desenvolvimento das mais diversas áreas de conhecimento. Podemos observar, na literatura especializada que a criação e a proposição de novos materiais tem crescido vertiginosamente após o aparecimento das técnicas computacionais em modelagem. Em particular o uso destas técnicas têm promovido o aparecimento de novas ligas, novos fármacos assim como novas estruturas em sistemas biomoleculares.

Um motivo importante para este desenvolvimento é a limitação computacional, nos tempos atuais, em resolver problemas envolvendo grande número de átomos e moléculas, a partir das equações básicas da mecânica quântica. Os tempos computacionais envolvidos em uma abordagem puramente quântica destes problemas são absolutamente proibitivos.

A alta velocidade de crescimento desta área de pesquisa tem sido possível devido ao fato de que as simulações computacionais permitem diminuir a interface entre a teoria e o experimento. A simulação computacional de experimentos tem sido usada também pelo o experimentalista como uma orientadora no sentido de prever e/ou queimar etapas desnecessárias durante a realização do experimento.

O estudo e caracterização de materiais complexos e problemas em biofísica deparam com um número imenso de graus de liberdade, nas coordenadas eletrônicas e nucleares, que inviabiliza o efetivo tratamento destes sistemas usando puramente cálculos computacionais de primeiros princípios. Técnicas alternativas envolvendo métodos híbridos clássico-quântico têm sido usadas como uma poderosa ferramenta na análise de estruturas moleculares, ligações químicas e no estudo de propriedades mecânicas, termodinâmicas, elétricas e espectroscópicas.

Neste texto discutiremos sobre algumas metodologias utilizadas na modelagem molecular, assim como a sua importância, onde e como devem ser utilizadas. No ensejo, abordaremos também algumas dificuldades inerentes às implementações computacionais.

Faremos um breve relato de como conjugar teorias, técnicas e ferramentas computacionais assim como a sua aplicação a uma diversidade de áreas do conhecimento com o intuito de aprimorar e solidificar o que denominamos um *Laboratório Virtual*.

Entendemos por Laboratório Virtual um conjunto de códigos computacionais aplicados a problemas de simulação em geral. Em particular, um Laboratório Virtual propõe-se a tratar

problemas em sistemas moleculares de interesse em Física, Química, Biologia, Ciência dos Materiais e Biotecnologia.

As metodologias para o procedimento de simulação são oriundas dos princípios gerais das dinâmicas clássica e quântica e dos princípios de extremos envolvendo grandezas físicas. Isto dá origem aos métodos Gradientes, Dinâmica Molecular (DM), os métodos híbridos clássico-quânticos e aos métodos Estocásticos (ME).

As técnicas envolvidas nestas metodologias são as seguintes: dinâmica clássica, usando campos de forças devidamente parametrizados, a partir de informações experimentais diversas acerca do sistema; cálculos quânticos usando *métodos ab-initio* e semi-empíricos para a determinação dos campos de forças com seus parâmetros; cálculos *ab-initio* e semi-empíricos para a determinação das funções potenciais a serem utilizadas nos esquemas estocásticos.

É importante ressaltar que apesar das dificuldades na implementação e utilização dos métodos quânticos em cálculos de propriedades moleculares, eles são fundamentais na parametrização dos campos de forças clássicos. Esta parametrização é feita aplicando os métodos quânticos em pequenos sistemas moleculares no sentido de caracterizar propriedades moleculares, as quais serão estendidas aos macros sistemas.

Algumas técnicas quânticas mais usadas são: a teoria do funcional da densidade (DFT), método Hartree-Fock; pseudo potencial e diversas aproximações semi-empíricas destas técnicas, tais como as aproximações AM1 e PM3.

Como já comentado anteriormente, procedimentos computacionais para cálculo de propriedades moleculares, com base em métodos puramente quânticos, tais como SCF-Hartree Fock ou SCF-Funcional da densidade, dentre outros, têm certos impedimentos computacionais. Nestes métodos, o tempo computacional cresce rapidamente com número de átomos, elétrons e funções das bases atômicas usadas na representação do sistema atômico e molecular. Isto impossibilita, na maioria dos casos, o estudo de propriedades físico-químicas e termodinâmicas de sistemas moleculares complexos como por exemplo;

- difusão de impurezas em sólidos cristalinos e ligas;
- deformação, rupturas e relaxação locais produzidas por impurezas ou forças externas;
- difusão de adsorbatos em peneiras moleculares (Zeólitas) e suas propriedades;
- estudo do processo de *docking* de fármacos em proteínas;
- estudo do enovelamento de proteínas (*protein folding*)
- Desenvolvimento e proposição de novos fármacos e estruturas de biomoléculas.

Nossa experiência no desenvolvimento e utilização de ferramentas computacionais tem mostrado que a combinação dos métodos baseados na mecânica quântica e na mecânica clássica constitui uma técnica mais satisfatória no tratamento dos problemas acima citados. Esta combinação é comumente denominada de métodos híbridos clássico-quântico. Assim, neste novo cenário vem surgindo 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. 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. Assim, uma possibilidade é descrever partes do sistema por um ou outro método.

Os métodos híbridos vêm sendo utilizado;

- 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 A estimativa de constantes de velocidade;
- 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;
- 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;
- 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.

Nestes últimos anos temos particularmente trabalhado no sentido de conjugar diferentes metodologias, as quais são reunidas em um conjunto de códigos computacionais acoplados para permitir a utilização de qualquer destas na abordagem do sistema em estudo, de acordo com as conveniências específicas do problema.

O fluxograma abaixo mostra um dos códigos, criados em nosso em nosso grupo com esse intuito.

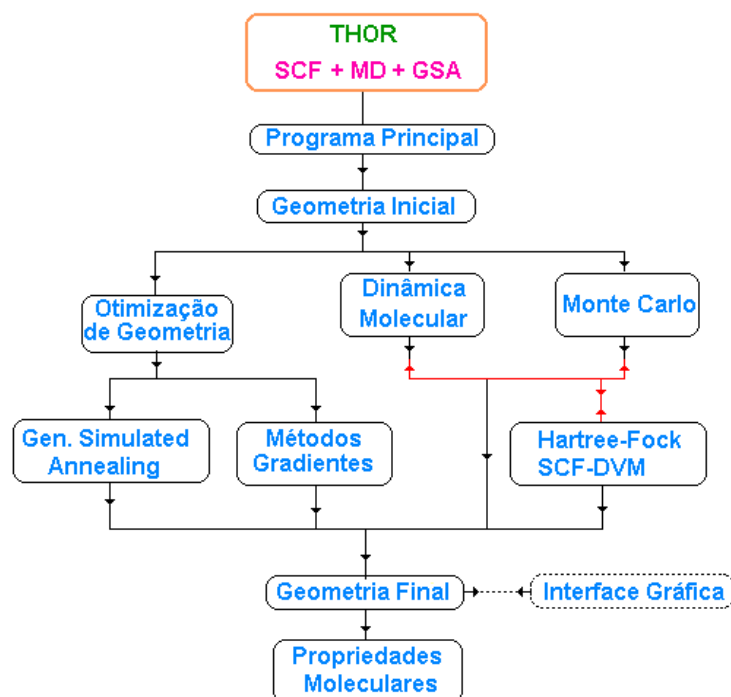


Fig.1- Fluxograma código do computacional THOR

Gostaríamos de ressaltar que mesmo tendo dificuldades de implementação computacional e utilização dos métodos quânticos, eles são de extrema importância na parametrização dos campos de forças clássicos. O desenvolvimento de campos de forças clássicos é feito inicialmente, calculando propriedades atômicas e moleculares de pequenos aglomerados, usando um método quântico apropriado.

Com o objetivo de discutir sobre as técnicas mais importantes na modelagem molecular, vamos começar a discussão com um breve relato sobre os métodos quânticos mais utilizados no estudo de propriedades moleculares. São eles o método SCF Hartree-Fock e método SCF-DFT.

2. Método Hartree-Fock - SCF

O método de Hartree-Fock tornou-se extremamente popular, entre outras coisas, pela qualidade dos resultados produzidos ao ser aplicado em cálculos de propriedades atômicas e moleculares. Entretanto, havia um outro problema extremamente importante a ser solucionado: qual deveria ser a forma matemática das funções orbitais? Enquanto que para cálculos atômicos as equações de Hartree-Fock podiam ser resolvidas numericamente, para moléculas, este mesmo procedimento demonstrava ser computacionalmente inadequado. Uma solução que tornou-se amplamente difundida e aplicada para cálculos de propriedades eletrônicas de qualquer sistema imaginável, foi o método proposto por Roothaan.

Roothaan sugeriu que funções que fossem utilizadas para representar orbitais moleculares poderiam ser obtidas em termos de funções que representassem orbitais atômicos. Se considerarmos que orbitais atômicos de sistemas multieletrônicos são funções aproximadas, a mesma idéia poderia ser utilizada para construí-los através de funções matemáticas que permitissem computacionalmente cálculos precisos de propriedades atômicas e moleculares. Este método ficou conhecido como o método de combinação linear de orbitais atômicos (linear combination of atomic orbitals - LCAO). A sugestão de Roothaan não foi a criação das combinações lineares dos orbitais atômicos, mas a sua utilização através das equações de Hartree-Fock. Genericamente, pode-se dizer que orbitais atômicos ou moleculares podem ser obtidos de forma auto-consistentes como combinações lineares de determinadas funções matemáticas ou funções de base. Assim, na aproximação LCAO (Linear Combination of Atomic Orbitals), os orbitais moleculares são construídos como uma combinação linear dos orbitais atômicos, isto é;

$$\phi_i = \sum_k C_{ik} \chi_k \quad (2.1)$$

sendo χ_k o k-ésimo orbital atômico. A equação para a energia molecular total tem forma;

$$E = \sum_i \epsilon_i^N + \sum_{i < j} \int \phi_i^2(1) \frac{1}{r_{12}} \phi_j^2(2) dv_1 dv_2 - \sum_{i < j} \int \phi_i(1) \phi_j(1) \frac{1}{r_{12}} \phi_i(2) \phi_j(2) dv_1 dv_2 \quad (2.2)$$

Condição de ortogonalidade

$$\int \phi_i \phi_j dv_1 dv_2 = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{se } i \neq j \end{cases} \quad (2.3)$$

As equações de Hartree, para os melhores orbitais, podem ser determinadas por aplicação da condição de extremos na energia com a condição auxiliar de ortogonalidade.

$$\left\{ H^N + \sum_{j=1} \left(\int \phi_j^2(1) \frac{1}{r_{12}} dv_2 \right) \right\} \phi_i(1) - \sum_{j=1} \left(\int \phi_j(2) \phi_i(2) \frac{1}{r_{12}} dv_2 \right) \phi_i(1) = \epsilon_i^{\text{SCF}} \phi_i(1) \quad (2.4)$$

ou em uma forma simplificada

$$\left(H^N + \sum_{j=1} J_j - \sum_{j=1} K_j \right) \phi_i(1) = \epsilon_i^{\text{SCF}} \phi_i(1) \quad \text{ou} \quad H^{\text{SCF}} \phi_i(1) = \epsilon_i^{\text{SCF}} \phi_i(1) \quad (2.5)$$

Esta equação pode ser reescrita em função dos orbitais atômicos por

$$\sum_n C_{in} \int \chi_m H^{\text{SCF}} \chi_n dv = \epsilon_i^{\text{SCF}} \sum_n C_{in} \int \chi_m \chi_n dv \quad (2.6)$$

ou

$$\sum_n C_{in} (H_{nn}^{\text{SCF}} - \epsilon_i^{\text{SCF}} S_{nn}) = 0 \quad (2.7)$$

Esta equação (forma matricial) apresenta características que permitem a aplicação de técnicas numéricas eficientes para determinar os coeficientes da combinação linear e as energias dos orbitais moleculares. Popularmente esta representação matricial é também denominada de equação secular.

$$\begin{pmatrix} H_{11} - \epsilon_k S_{11} & H_{12} - \epsilon_k S_{12} & \cdots & H_{1m} - \epsilon_k S_{1m} \\ H_{21} - \epsilon_k S_{21} & H_{22} - \epsilon_k S_{22} & \cdots & H_{2m} - \epsilon_k S_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ H_{m1} - \epsilon_k S_{m1} & H_{m2} - \epsilon_k S_{m2} & \cdots & H_{mm} - \epsilon_k S_{mm} \end{pmatrix} \begin{pmatrix} c_{1k} \\ c_{2k} \\ \vdots \\ c_{mk} \end{pmatrix} = 0 \quad (2.8)$$

O sistema de equação acima terá solução quando o determinante secular, abaixo, for igual a zero,

$$\det |H_{nn}^{\text{SCF}} - \epsilon_i^{\text{SCF}} S_{nn}| = 0 \quad (2.9)$$

Um aspecto extremamente importante diz respeito as possíveis soluções obtidas através da equação secular. A escolha das soluções desejadas é feita através dos orbitais que apresentam menor energia. Em outras palavras, através da solução da secular, obtém-se m diferentes orbitais moleculares. Organiza-se estes orbitais em ordem crescente de energia e escolhe-se os n orbitais de menor energia. Através da especificação dos orbitais ocupados pode-se determinar a energia eletrônica total do sistema e determinar o valor de qualquer outra propriedade de interesse.

Os orbitais desocupados ou virtuais e as respectivas energias obtidas através da solução da secular não são caracterizados corretamente. A definição do operador de Fock leva em consideração a distribuição eletrônica e conseqüentemente todos os orbitais ocupados, mais precisamente, a interação de cada um dos elétrons em um determinado orbital em relação ao campo médio de todos os outros orbitais ocupados. Desta forma, os orbitais virtuais são obtidos experimentando a interação de um elétron nesse orbital com todos os orbitais ocupados. Conseqüentemente, os orbitais virtuais apresentam uma característica mais próxima dos orbitais do íon negativo em um estado excitado, do que do sistema neutro.

Outros aspectos importantes também devem ser discutidos com relação matricial secular. A simplificação utilizada acima para apresentar as equações de Hartree-Fock-Roothaan esconde uma consideração importante que deve ser salientada. Em primeiro lugar, admitiu-se que a Eq.(2.5) pudesse ser utilizada como ponto de partida e em função dela

efetuou-se a substituição da expansão dos orbitais atômicos e/ou moleculares em termos de funções de base, chegando-se finalmente na representação da equação secular. Entretanto, uma demonstração rigorosa utiliza uma ferramenta extremamente importante denominada de ajuste variacional. Através de demonstrações rigorosas percebe-se que os coeficientes de combinação linear são obtidos em função da minimização da energia eletrônica total do sistema. Em outras palavras, ao resolver-se a Eq.(2.7) estão sendo escolhidos os coeficientes de combinação linear das funções de base que minimizam a energia eletrônica total do sistema. Estes não são os únicos parâmetros que podem ser ajustados para que a energia eletrônica seja mínima. Existem outros parâmetros que estão incorporados no tipo de função de base que está sendo utilizado. Qualquer destes ajustes que levam a energia eletrônica a um mínimo, correspondem a aplicação do *método variacional*.

Esta última consideração sugere pelo menos mais dois pontos a serem esclarecidos: a) qual deve ser o tipo de função de base que pode ser utilizado para representar apropriadamente os orbitais atômicos ou moleculares? e b) qual o número de funções de base deve ser utilizado? Estas duas perguntas correspondem ao cerne de um dos maiores problemas a ser solucionado ou pelo menos considerado quando se procura utilizar o método de Hartree-Fock-Roothaan.

Citamos abaixo algumas características positivas e negativas dos métodos quânticos:

- | | | |
|------------------------|---|---|
| Dificuldades do método | } | <ul style="list-style-type: none"> - Cálculo numérico intensivo, - Aproximação Born-Oppenheimer (núcleos fixos), - Tempo computacional cresce com a quarta potência o número de funções de base, - Limitações no tamanho do sistema molecular (da ordem de 100 átomos pesados), - Introdução do tempo na equação de Schrödinger para fazer dinâmica molecular, - Introdução da temperatura, - Utilização de muito espaço em disco rígido para armazenar as integrais de dois centros, - A implementação computacional exige algoritmos matemáticos sofisticados |
| Quando usá-lo? | } | <ul style="list-style-type: none"> - Quando as interações eletrônicas de longo e curto alcance forem importantes, - Na descrição precisa da energia, momento de dipolo, transições eletrônicas,... - Nas descrições do processo de <i>docking</i>, nas espectroscopias, nas pontes de hidrogênio, na transferência de cargas. |

Vantagens em relação aos métodos clássicos.

- Resultados teóricos compatíveis aos experimentais
- Precisão no cálculo das propriedades eletrônicas: energia, momento de dipolo, transferência de cargas,...

3. Teoria do Funcional da Densidade (DFT)

Teoria do Funcional da Densidade (DF) é uma aproximação auto-consistente, de primeiros princípios, para estrutura eletrônica a qual tem uma aplicação ampla em sólidos e moléculas. Nossos desenvolvimentos, feitos no esquema de “cluster” embebidos, permitem tratar uma expansão localizada da função de onda eletrônica, densidades e propriedades derivadas a partir do tratamento do sistema estendido em termos de fragmentos.

Esta aproximação, fornece-nos uma metodologia plausível para um tratamento auto-consistente, em sistemas moleculares grandes e com baixa simetria. Em particular, é uma técnica adequada para tratar problemas de impurezas, superfícies e interfaces.

Numa forma simplificada, podemos dizer que a carga do sistema eletrônico e a densidade de spin são decompostas em uma soma de *overlapping cluster*:

$$\rho(\mathbf{r}) = \sum_i w_i(\mathbf{r}) \rho_i(\mathbf{r} - \mathbf{R}_i) \quad (3.1)$$

As densidades dos clusters individuais são expressas em termos da função de onda DF, a uma partícula e seus números de ocupação de Fermi-Dirac:

$$\rho_i = \sum_j n_{i,j} |\psi_{i,j}(\mathbf{r})|^2 \quad (3.2)$$

A extração do volume interior de cada cluster e o alinhamento da densidade de estado equivalente, completa o procedimento autoconsistente em termos de fragmentos. Uma das formas mais eficientes de se calcular a função de onda, para clusters embebidos, é por meio do método variacional discreto denominado DVM (Discrete Variational Method), o qual é baseado na teoria da densidade local para definir os termos de troca e correlação eletrônica.

No DVM as funções de ondas mono eletrônicas para o aglomerado é representada por uma combinação linear de orbitais atômicos (LCAO). No caso de sistemas cristais e sólidos em geral é necessário definir uma região denominada *embedded* para representar o cristal como sendo uma estrutura infinita.

A seguir damos os principais passos no esquema de campo auto consistente para o caso do DVM.

a)- A função densidade eletrônica $\rho(\vec{r})$ é definida pela Eq.(3.3), a qual é calculada levando em conta todos os átomos do sistema.

$$\rho(\vec{r}) = \sum_{ij} C_{ij} R_{ij}^2 \quad (3.3)$$

onde R_{ij}^2 são as autofunções radiais usadas na construção da base LCAO e os coeficientes C_{ij} representam as amplitudes de interação entre os elementos i e j .

b)- Os termos eletrostático V_C , de troca V_{ex} e de auto correlação V_{sc} eletrônica são calculados pelas expressões;

Self-consistent DV method can be summarized in following scheme;

$$V_C = \int \frac{\rho(\vec{r})}{|\vec{r} - \vec{s}|} d\vec{s} \quad (3.4)$$

$$V_{ex} = -2 \left(\frac{\rho(\vec{r})}{|\vec{r} - \vec{s}|} \right)^{\frac{1}{3}} \quad (3.5)$$

c)- Em seguida calcula-se a matriz de troca e correlação $S_{\mu\nu} = \langle \chi_\mu | \chi_\nu \rangle$ e a energia cinética para em seguida determinar o operador hamiltoniano $H_{\mu\nu}$.

d)- Resolve-se a equação secular (Eq.3.6) para se obter informações sobre o espectro mono eletrônico e em seguida criar os novos coeficientes C_{ij} .

$$HC = EC \quad (3.6)$$

e)- Calcula-se, em seguida, o elemento de matriz $T_{\mu\nu}^\alpha = \sum_i n_i^\alpha C_{\mu i}^* C_{i\nu}$ para definir a carga pela relação $Q_{\mu\nu}^\alpha = T_{\mu\nu}^\alpha S_{\mu\nu}^\alpha$.

f)- Volta-se ao item (a) até que o critério de convergência seja atingido com a precisão desejada.

A implementação computacional da teoria do funcional da densidade (DFT), assim como a sua utilização em cálculos moleculares seguem as mesmas dificuldades apresentadas anteriormente para o método Hartree-Fock.

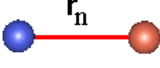
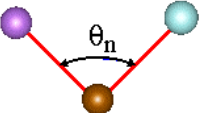
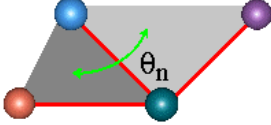
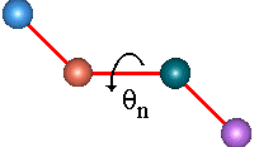
4. Modelagem Molecular Clássica

O objeto principal da modelagem molecular é buscar uma metodologia para descrever propriedades moleculares quânticas em termos de um campo de força clássico nas equações de Newton. No modelo denominado Mecânica Molecular (MM) as ligações químicas são representadas por potenciais harmônicos. Pictóricamente pode se dizer que as moléculas são consideradas como uma coleção de massas ligadas por molas ou potenciais harmônicos. Neste caso, é necessário encontrar um conjunto refinado de parâmetros, denominado campo de força, que descreva as interações quânticas com suficiente precisão. Idealmente espera-se que este refinamento do campo de força venha eventualmente ser possível unificar modelos computacionais que possam sucessivamente fazer uma mímica das propriedades moleculares observadas.

O primeiro aplicação usando dinâmica molecular for feita por Alder e Wainwright [1] os quais usaram um modelo de esferas rígidas com choques perfeitamente elásticos para representar as interações atômicas. Um dos campos de forças mais usado é o MM1 proposto por Allinger [2]. Atualmente existem incontáveis códigos computacionais usando campo de força clássico. Cada um deles usa um campo de força particular apropriado para descrever diferentes propriedades moleculares, ajustadas a algum resultado experimental.

Em mecânica molecular o átomo é representado por um corpo esférico com uma massa particular que, em geral, é igual a sua respectiva massa atômica. A energia molecular relacionada com o campo de força clássico pode ser expressa pela seguinte equação [3, 4]:

$$V = V_H + V_b + V_{\text{tor-i}} + V_{\text{tor-p}} + V_C + V_{\text{vdW}} \quad (4.1)$$

$V_H = \frac{1}{2} \sum_n^{N_b} K_{H_n} (r_n - r_o)^2$	
$V_b = \frac{1}{2} \sum_n^{N_\theta} K_{\theta_n} (\theta_n - \theta_o)^2$	
$V_{\text{tor-i}} = \frac{1}{2} \sum_n^{N_\xi} K_{\xi_n} (\xi_n - \xi_o)^2$	
$V_{\text{tor-p}} = \frac{1}{2} \sum_n^{N_\phi} K_{\phi_n} (1 + \cos(m\phi_n - \delta_n))^2$	

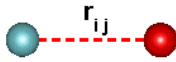
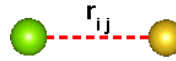
$V_C = \frac{1}{4\pi\epsilon} \sum_{i<j}^{N_{at}} \frac{q_i q_j}{r_{ij}}$	
$V_{vdw} = \sum_{i<j}^{N_{at}} \left[\frac{C_{12}(ij)}{r_{ij}^{12}} - \frac{C_6(ij)}{r_{ij}^6} \right]$	

Fig.2 – Campo de Força Clássico

Cada termo da equação (15) correlaciona um parâmetro geométrico, tal como uma ligação química, com a energia potencial. Como isto, pode-se avaliar a energia potencial associada com uma dada conformação molecular de equilíbrio. Pode-se dizer um bom campo de força possa descrever os estados conformacionais de energia mínima, local ou global, mesmo estando o sistema molecular sobre a ação de forças externas ou excitações térmicas.

Em resumo, o primeiro termo da equação (15) V_H representa a energia necessária comprimir ou alongar as ligações químicas do tipo covalentes; E_b é a contribuição na energia potencial que representa as ligações angulares; E_{tor-p} e E_{tor-i} são as contribuições torsionais que representam os movimentos harmônicos nas ligações angulares diedrais próprias e impróprias respectivamente. Os dois últimos termos E_{vdw} e E_C representam as contribuições das interações envolvendo átomos não ligados, tais como as pontes de hidrogênio, forças de van der Waals e contribuições eletrostáticas coulombianas, respectivamente. Todos estes termos são funções explícitas das posições atômicas ou conformações moleculares.

É bom ressaltar que, em geral, o valor da energia E não tem significado físico, mas diferença na entre duas conformações moleculares deve ser equivalente a energia experimental observada, isto é ela representa a barreira de energia ou a energia necessária para mudar de uma conformação molecular a outra.

Existe, atualmente, um conjunto de razões as quais garantem o sucesso dos métodos de dinâmica molecular clássica. A principal razão é tempo computacional para se calcular propriedades moleculares em sistemas macromoleculares. As pesquisas acadêmicas e os interesses na indústria de fármacos para desenvolver nos compostos ou medicamentos têm proporcionado o desenvolvimento de incontáveis códigos computacionais baseados em campo de força clássico. O tempo computacional comum em cálculos envolvendo campo de força clássico aumenta com m^2 onde m é o número de átomos na molécula. Por outro lado, o uso de métodos quânticos *ab-initio* neste tipo de sistemas moleculares é computacionalmente impraticável já que o tempo computacional para se calcular as integrais de repulsão eletrônica aumenta com n^4 , onde n é o número de funções de base usadas na descrição dos orbitais atômicos e moleculares 1s, 2s, 2p,.... Normalmente, usa-se nestes casos, várias funções por elétron, como por exemplo combinações de várias funções gaussianas. Outras razões que justificam o uso da mecânica molecular (MM) são listadas a seguir;

- A mecânica molecular clássica (equações de Newton) é computacionalmente mais simples de ser entendida do que os métodos de mecânica quântica tal como Hartree-Fock.
- Em mecânica molecular clássica é muito simples introduzir a evolução temporal
- Em mecânica molecular clássica é possível introduzir a temperatura como uma perturbação externa.

Existem duas questões que são desfavoráveis aos métodos clássicos (MM); a primeira é que não existem regras bem definidas para se determinar as constantes de força e a segunda esta relacionada com a escolha do melhor campo de força que depende, em geral, de um bom conhecimento, a priori, do sistema molecular em estudo.

Recentemente desenvolvemos um código computacional, denominado THOR [3, 4], cujo campo de força é equivalente ao apresentado na Eq. (4.1). Este campo de força tem sido usado para estudo de enovelamento de proteínas, assim como na determinação de novas estruturas em proteínas. Este código computacional contempla também potenciais apropriados para estudo de propriedades moleculares em sólidos em geral.

A modelagem molecular consiste basicamente de três metodologias, que são elas;

- Otimização de geometria ou determinação de estruturas moleculares de equilíbrio usando métodos gradientes;
- Determinação de estrutura de equilíbrio e evolução temporal do sistema molecular usando dinâmica molecular clássica, ou resolvendo as equações de Newton, e,
- Determinação de estrutura de equilíbrio usando métodos estocásticos.

A seguir trataremos cada uma destas metodologias de forma mais detalhada.

4.1 Otimização de Geometria via Métodos Gradientes

Pode-se minimizar a energia ou otimizar a geometria ajustando as coordenadas atômicas no sentido reduzir o valor da energia potencial descrita na Eq. (4.1).

Computacionalmente existem basicamente dois tipos de procedimentos para minimizar a energia de um sistema molecular. Os métodos teóricos mais comuns envolvem cálculos de derivadas da função energia potencial (Eq.4.1). Uma outra possibilidade, introduzida mais recentemente, é mapeamento da hiper-superfície de energia usando métodos estocásticos, a qual será discutida nas próximas seções. Neste último caso as geometrias são geradas de forma aleatória.

Os métodos gradientes mais importantes são o *steepest descents*, o conjugado gradiente, o de Newton-Raphson e o método de Newton truncado. Faremos um breve relato sobre os dois métodos mais importantes.

4.1.1 Método Steepest Descent

Também conhecido como Método de Cauchy ou Máximo Declive (*Steepest Descent*), foi desenvolvido por Cauchy em 1847. É bastante simples em termos computacionais, mas tem convergência lenta chegando muitas vezes a não convergir em um tempo razoável. Utiliza poucas informações, exigindo apenas as derivadas de primeira ordem para o cálculo do gradiente. Como o gradiente aponta na direção de maior crescimento da função no ponto, o método procura em cada ponto caminhar na direção oposta ao gradiente. Portanto, a direção de busca é a direção oposta ao gradiente.

Devido a sua simplicidade é bastante aplicável, porém como veremos, pode convergir muito lentamente. Onde a convergência do método Gradiente é linear.

O método gradiente é definido pelo algoritmo iterativo

$$\mathbf{r}_{i,n+1} = \mathbf{r}_{i,n} + \Delta \mathbf{r}_{i,n} \quad (4.1.1)$$

onde \mathbf{r}_i é posição do i -ésimo átomo. No método *steepest-descent*, como em outros métodos, o passo ou incremento nas coordenadas $\Delta \mathbf{r}_{i,n}$ de um átomo i , é dado na direção e sentido da força resultante sobre este átomo:

$$\Delta \mathbf{r}_{i,n}^{\rho} = k_n \frac{\mathbf{F}_{i,n}^{\rho}}{|\mathbf{F}_{i,n}^{\rho}|} \quad (4.1.2)$$

onde $\mathbf{F}_{i,n}$ é a força resultante no átomo i , no n -ésimo passo, a qual pode ser calculada pelo gradiente do potencial, isto é;

$$\mathbf{F}_{i,n}^{\rho} = -\nabla^{\rho} V(\mathbf{r}_{i,n}) \quad (4.1.3)$$

k_n é um parâmetro de ajuste do tamanho do passo e $\mathbf{F}_{i,n}^{\rho}/|\mathbf{F}_{i,n}^{\rho}|$ é o vetor unitário na direção e sentido da força resultante sobre i no passo n . O valor de que k_n é ajustado para acelerar a minimização. Partindo de um valor típico $k_0 = 0.1 \text{ \AA}$, o valor de cada passo em cada direção x , y ou z será definido pela respectiva componente do vetor unitário na Eq. (4.1.2). Se o potencial total $V_n(\{\mathbf{r}_i\})$ no passo n diminui com relação a $V_{n-1}(\{\mathbf{r}_i\})$, no passo seguinte k_{n+1} é aumentado fazendo-se por exemplo:

$$k_{n+1} = 1,2 * k_n \quad (4.1.4)$$

Se $V_n(\{\mathbf{r}_i\})$ aumenta, o sistema está se afastando do mínimo, k_n pode estar tão grande que o sistema pode ter passado pelo mínimo e estar escalando o outro lado do poço de potencial. No passo seguinte k_{n+1} é diminuído fazendo-se, por exemplo:

$$k_{n+1} = \frac{k_n}{2} \quad (4.1.5)$$

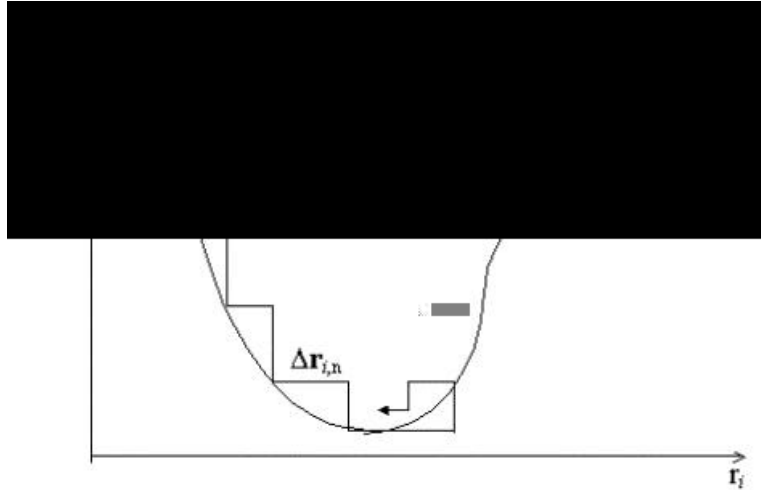


Fig.3 – Convergência do método gradiente

Em cada passo verifica-se a diferença entre os valores de $V_{n-1}(\mathbf{r}_i)$ e $V_n(\mathbf{r}_i)$. Se a mesma for menor que um fator de convergência ΔV estipulado, o processo é interrompido. Com $\Delta V = 10^{-5}$ Kcal/mol chega-se satisfatoriamente a um mínimo para muitos sistemas, embora valores menores para ΔV possam ainda diminuir a energia em alguns Kcal/mol. Com $\Delta V = 10^{-2}$ Kcal/mol as deformações em ligações, ângulos e em esferas de van der Waals são eliminadas na maioria dos sistemas moleculares, muitas vezes baixando a energia potencial para valores acessíveis a temperaturas ordinárias.

4.1.2 Método de Newton-Raphson

O Método de Newton-Raphson ou método das secantes é um caso particular do Método das Aproximações Sucessivas. Entretanto, ao invés de calcular pontos fixos, calculamos agora raízes de uma dada função, portanto este método requer tanto a avaliação da função $V(r)$ quanto a sua derivada $V'(r)$, em pontos r arbitrários. Se a derivada da função não se anular nas proximidades da raiz desejada, então o Método pode ser aplicado. Algebricamente, o método deriva da expansão de $V(r)$ em série de Taylor na vizinhança de um ponto, isto é

$$V(r + \delta) \approx V(r) + V'(r)\delta + \frac{V''(r)}{2}\delta^2 \quad (4.1.6)$$

Para pequenos valores de δ e em uma função bem comportada, tem-se que $V(r + \delta) = 0$, o que implica em;

$$\delta = -\frac{V(r)}{V'(r)}. \quad (4.1.7)$$

Devemos lembrar que o método de Newton-Raphson não está restrito a uma dimensão, isto é, ele pode ser estendido para várias dimensões. Para pontos muito afastados da raiz da função os termos de ordem superior na série são de suma importância, o que faz com que o método seja de pouca acurácia e, portanto necessita de certas correções.

Porque então o método de Newton-Raphson é tão poderoso? A resposta está em sua velocidade de convergência: Para qualquer distância pequena ϵ entorno de r a função $V(r)$ e sua derivada são aproximadamente:

$$\begin{aligned} V(r+\epsilon) &= V(r) + \epsilon V'(r) + \frac{V''(r)}{2} \epsilon^2 + \Lambda, \\ V'(r+\epsilon) &= V'(r) + \epsilon V''(r) + \Lambda \end{aligned} \quad (4.1.8)$$

De acordo com a fórmula de Newton-Raphson,

$$r_{i,n+1} = r_{i,n} - \frac{V(r_{i,n})}{V'(r_{i,n})} \quad (4.1.9)$$

ou ainda que

$$\epsilon_{i,n+1} = \epsilon_{i,n} - \frac{V(r_{i,n})}{V'(r_{i,n})} \quad (4.1.10)$$

Quando uma solução tentativa $r_{i,n}$ difere da raiz verdadeira por um fator igual a $\epsilon_{i,n}$, podemos usar as Eq. (4.1.9) em eq. (4.1.10) para calcular o novo ponto. O resultado é uma relação de recorrência para o afastamento relativo a solução tentativa,

$$\epsilon_{i,n+1} = -\epsilon_{i,n}^2 \frac{V''(r_{i,n})}{2V'(r_{i,n})} \quad (4.1.11)$$

Esta equação nos diz que o método de Newton-Raphson converge em uma forma quadrática. Isto é, para pontos próximos da raiz o número de dígitos significante dobra a cada passo. Esta forte propriedade de convergência faz do Newton-Raphson um dos melhores métodos para se determinar a raiz de qualquer função derivável e cuja derivada é contínua e diferente de zero em torno da raiz.

O método de Newton-Raphson estende a idéia do método gradiente aproveitando aproximações quadrática a $V(r)$. Aproximações quadráticas não são somente melhores que

aproximações lineares, mas ganham importância medida que se aproximam do ponto ótimo do problema.

4.2 Dinâmica Molecular Clássica (MD)

A dinâmica molecular (MD) consiste em estudar a evolução temporal de um dado sistema molecular. Neste caso os átomos estão em um movimento contínuo, as ligações químicas estão vibrando, os ângulos de ligação variando e a molécula rodando. Em MD, configurações moleculares sucessivas são geradas por integração das equações de movimento de Newton.

O resultado da integração das equações de Newton fornece uma trajetória que especifica como as posições e velocidades atômicas variam com o tempo. Esta trajetória pode ser obtida tanto por integração das equações de movimento de Newton ou resolvendo as equações de movimento de Hamilton. Com isto queremos dizer que o estado microscópico de um sistema pode ser especificado em termos das posições e momentos das partículas que o constituem. Dessa forma, a Hamiltoniana H de um sistema molecular clássico pode ser escrita como a soma das energias cinética T e potencial V , como função das séries de coordenadas generalizadas \mathbf{q}_i e de momentos generalizados \mathbf{p}_i de todos os N_{at} átomos do sistema:

$$H(\mathbf{q}_i, \mathbf{p}_i) = T(\mathbf{p}_i) + V(\mathbf{q}_i) \quad \text{onde } i = 1, 2, \dots, N \quad (4.2.1)$$

onde N é o número de átomos no sistema molecular.

A energia potencial $V(\mathbf{q}_i)$ contém os termos de interação inter e intramoleculares, de curto e longo alcance, e pode ser substituída pela função potencial $V(\mathbf{r}_i)$ da Eq.(4.2.1), tal que as coordenadas \mathbf{q}_i sejam as coordenadas cartesianas \mathbf{r}_i e \mathbf{p}_i seus momentos conjugados. A energia cinética assume a forma:

$$T(\mathbf{p}_i) = \sum_{i=1}^N \frac{p_i^2}{2m_i} \quad (4.2.2)$$

em que m_i é a massa do átomo i .

A partir de H é possível construir as equações de movimento que governam a evolução temporal do sistema e suas propriedades dinâmicas. Como a energia potencial dada pela Eq.(4.1) independe das velocidades e do tempo, H é igual a energia total do sistema e as equações de movimento de Hamilton são dadas por:

$$\frac{\partial \mathbf{q}_i}{\partial t} = \frac{\partial H}{\partial \mathbf{p}_i} \quad (4.2.3)$$

$$\frac{\partial \mathbf{p}_i}{\partial t} = -\frac{\partial H}{\partial \mathbf{q}_i} \quad (4.2.4)$$

as quais conduzem às equações do movimento de Newton:

$$\dot{\mathbf{r}}_i = \frac{\mathbf{p}_i}{m_i} = \mathbf{v}_i \quad (4.2.5)$$

$$\dot{\mathbf{p}}_i = m_i \mathbf{a}_i = -\nabla V(\mathbf{r}_i) = \mathbf{F}_i \quad (4.2.6)$$

respectivamente, em que $\dot{\mathbf{r}}_i = \mathbf{v}_i$ (ou \mathbf{v}_i) e \mathbf{a}_i são a velocidade e a aceleração do átomo i , enquanto \mathbf{F}_i é a força sobre o átomo i .

A Dinâmica Molecular consiste portanto na resolução numérica das Equações 4.2.5 e 4.2.6 e na integração das mesmas passo-a-passo no tempo, de maneira eficiente e acurada. Como resultado obtém-se energias e trajetórias para todas as partículas (ou átomos) e para o sistema como um todo, a partir das quais várias propriedades podem ser calculadas. O tempo deixa de ser contínuo, sendo discretizado nos sistemas moleculares em passos menores (geralmente 20 vezes menores) que o período das vibrações dos átomos de hidrogênio, o movimento molecular mais rápido. Em sistemas com hidrogênio usualmente aplica-se um passo de tempo de $5,0 \times 10^{-16}$ segundos. Neste procedimento é essencial que a energia potencial seja uma função contínua das posições das partículas e que as posições variem suavemente com o tempo. As forças \mathbf{F}_i sobre cada átomo, que são obtidas a partir da derivada espacial da função energia potencial como é mostrado na Eq.(4.2.6), podem desta maneira serem consideradas constantes no intervalo entre dois passos. A estabilidade da dinâmica é assim favorecida, as partículas seguem suas trajetórias clássicas mais acuradamente e a energia total do sistema tende a conservar-se.

Uma limitação para a simulação da dinâmica molecular reside portanto no fato de que para cada nano segundo de simulação são necessários dois milhões de passos com o passo de tempo acima. A simulação de um nano segundo na dinâmica de uma macromolécula com 200 átomos pode levar horas de tempo de CPU em grandes computadores, utilizando-se um algoritmo eficiente. Uma descrição e análise da eficiência de algoritmos para simulação de dinâmica molecular, pode ser encontrada na tese de doutorado de Pedro G. Pascutti [27].

4.2.1 Implementação Computacional da MD: O Algoritmo de Verlet

Os termos que compõem a energia potencial total na Eq.(4.1) caracterizam-na como uma função contínua das posições atômicas. Com o tratamento da descontinuidade dielétrica entre dois meios solventes pelo “método das imagens eletrostáticas” [27], o potencial-degrau na superfície de separação entre os dois meios é suavizado, de forma que é mantida a continuidade da energia potencial total como função das posições. Quando a energia potencial é uma função contínua das posições e o passo de tempo é pequeno o suficiente para se considerar que as posições variem suavemente com o tempo, dado um conjunto de posições atômicas num instante t , as posições no passo seguinte podem ser obtidas por uma expansão de Taylor de $\mathbf{r}_i(t)$:

$$\mathbf{r}_i^p(t + \delta t) = \mathbf{r}_i^p(t) + \mathbf{v}_i^p(t)\delta t + \frac{\mathbf{a}_i^p(t)\delta t^2}{2!} + \Delta \quad (4.2.7)$$

Do mesmo modo obtém-se as posições no passo anterior a $\mathbf{r}_i(t)$:

$$\mathbf{r}_i^p(t - \delta t) = \mathbf{r}_i^p(t) - \mathbf{v}_i^p(t)\delta t + \frac{\mathbf{a}_i^p(t)\delta t^2}{2!} - \Delta \quad (4.2.7a)$$

em que δt representa o passo de tempo. Somando-se as Equações (4.2.7 e (4.2.7a) obtém-se o algoritmo de Verlet (1967) para a propagação das posições:

$$\mathbf{r}_i^p(t + \delta t) = 2\mathbf{r}_i^p(t) - \mathbf{r}_i^p(t - \delta t) + \mathbf{a}_i^p(t)\delta t^2 \quad (4.2.8)$$

onde temos desprezado os termos de ordem superiores. A aceleração $\mathbf{a}_i^p(t)$ é obtida a partir da Eq. (4.2.6), tal que:

$$\mathbf{a}_i^p(t) = \frac{\mathbf{F}_i^p(t)}{m_i} \quad (4.2.9)$$

e

$$\mathbf{F}_i^p = -\nabla V(\mathbf{r}_i^p(t)) \quad (4.2.10)$$

No algoritmo de Verlet, como em outros métodos numéricos da Dinâmica Molecular, resolve-se as equações de movimento de Newton para cada átomo e em cada incremento no tempo. A avaliação das forças (Eq.-4.2.10) para obtenção das acelerações (Eq.4.2.9), é o processo que mais consome tempo computacional na Dinâmica Molecular. O tempo gasto no cálculo das forças depende da complexidade da função energia potencial. São tomadas as derivadas espaciais de cada termo potencial e somadas para se obter a força resultante e a aceleração, sobre cada átomo e a cada novo conjunto de coordenadas. As acelerações são inseridas num algoritmo, como o da Eq.(4.2.7), para a previsão das novas posições e em seguida o processo se repete.

Como as velocidades não aparecem na Eq.(4.2.7), elas não desempenham papel algum na determinação das trajetórias neste algoritmo. A previsão das novas posições no instante $t + \delta t$ são computadas somente a partir das posições nos instantes t e $t - \delta t$ e das forças $\mathbf{F}_i(t)$ sobre cada partícula no instante t . As velocidades, porém, são necessárias para o cálculo da energia cinética, que somada à energia potencial dá a energia total do sistema. Subtraindo-se as Equações (4.2.7) e (4.2.7a) obtém-se o algoritmo de Verlet para a propagação das velocidades:

$$\mathbf{v}_i^p(t) = \frac{\mathbf{r}_i^p(t + \delta t) - \mathbf{r}_i^p(t - \delta t)}{2\delta t} \quad (4.2.11)$$

Quando há truncamento nas interações de longo alcance, o potencial sofre variação brusca no instante em que a distância entre um par de átomos em interação ultrapassa o raio de

corde. Situações como a do potencial-degrau, devido à descontinuidade dielétrica entre solventes tratada no presente trabalho, como a do modelo de esferas rígidas ou de poços de potenciais retangulares, são outros exemplos onde o potencial varia bruscamente. Nesses casos, com a expansão de Taylor não se pode prever as posições nas vizinhanças onde o potencial é descontínuo, as partículas sofrem colisões impulsivas e as velocidades variam de maneira descontínua. Quando não é o caso de corrigir a descontinuidade no potencial, o tratamento passo-a-passo acima deve ser substituído por (ou acoplado a) uma análise colisional da dinâmica do sistema (Allen & Tildesley, 1987).

4.2.2 Implementação Computacional: Algoritmo *Summed Verlet* ou *Leapfrog*

As trajetórias obtidas a partir da Eq.(4.2.7) estão sujeitas a imprecisões numéricas pelo fato de, para cada átomo, tomar-se a diferença entre duas posições, que são quantidades grandes, com o objetivo de se obter o incremento que leva à posição seguinte, que é uma quantidade pequena. Uma forma de evitar este problema surgiu com o algoritmo de meio-passo *leapfrog*, que é uma modificação do algoritmo básico de Verlet (Berendsen et al., 1984; Allen & Tildesley, 1987; van Gunsteren & Berendsen, 1990). As velocidades nos meio-passos $t + \mathbf{dt}/2$ e $t - \mathbf{dt}/2$ podem ser definidas desmembrando-se a Eq.(4.2.12), de forma que:

$$\mathbf{v}_i(t + \delta t / 2) = \frac{\mathbf{r}_i(t + \delta t) - \mathbf{r}_i(t)}{\delta t} \quad (4.2.12)$$

e

$$\mathbf{v}_i(t - \delta t / 2) = \frac{\mathbf{r}_i(t) - \mathbf{r}_i(t - \delta t)}{\delta t} \quad (4.2.13)$$

A propagação das posições é obtida diretamente da Eq.(4.2.13), na forma:

$$\mathbf{r}_i(t + \delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t + \delta t / 2)\delta t \quad (4.2.14)$$

a qual apresenta resultados mais acurados por não lidar com diferenças entre valores grandes, comparados às pequenas variações nas posições atômicas entre t e $t + \mathbf{dt}$. Para atualizar as posições com a Eq.(4.2.15), no entanto, é necessário o conhecimento das velocidades $\mathbf{v}_i(t + \delta t / 2)$. Para obtê-las, substitui-se a Eq.(4.2.7) na Eq.(4.2.13) e em seguida a Eq.(4.2.14), conduzindo a:

$$\mathbf{v}_i(t + \delta t / 2) = \mathbf{v}_i(t - \delta t / 2) + \mathbf{a}_i(t)\delta t \quad (4.2.15)$$

em que a aceleração, no último termo, é obtida como indicado nas Eqs.(4.2.10) e (4.2.11). Para se obter a energia cinética do sistema num instante t , determina-se as velocidades em t pela média:

$$\mathbf{v}_i(t) = \frac{\mathbf{v}_i(t + \delta t/2) + \mathbf{v}_i(t - \delta t/2)}{2} \quad (4.2.16)$$

Conhecendo-se a energia potencial no instante t , é possível determinar a energia total do sistema ($H = T + V$) no instante t .

As Equações (4.2.15) e (4.2.16) compõem o algoritmo *Summed Verlet* ou *leapfrog*. Este algoritmo equivale algebricamente ao algoritmo de Verlet básico, exposto na seção anterior. Porém, além de sua menor susceptibilidade a erros numéricos o *leapfrog* ocupa menos espaço de armazenamento na memória, sendo um dos algoritmos mais estáveis, simples e eficientes, podendo ser aplicado em sistemas constituídos de fluidos simples a biopolímeros. Talvez sua maior vantagem seja o fato de as velocidades aparecerem no cálculo das novas posições, o que torna o sistema acoplável a um banho térmico por correções nas velocidades. Como a energia potencial é uma função das posições e a variação destas depende das velocidades, Eq.(4.2.15), controlar as velocidades significa controlar diretamente, além da energia cinética, também a energia potencial e conseqüentemente a energia total do sistema.

A temperatura T de um sistema pode ser aproximada em termos da energia cinética média sobre um número passos (N_{passos}):

$$\frac{3}{2} N_{at} K_B T = \left\langle \sum_{i=1}^{N_{at}} \frac{1}{2} m_i \mathbf{v}_i^2 \right\rangle_{N_{passos}} \quad (4.2.17)$$

N_{passos} varia geralmente entre 20 a 50 passos, dependendo do número N_{at} de átomos do sistema. O acoplamento a um banho térmico se dá a partir da comparação da temperatura T a cada ciclo de N_{passos} , com uma temperatura de referência T_0 . Se a diferença entre T e T_0 estiver fora de um desvio ΔT estipulado, geralmente 10% de T , então as novas velocidades são corrigidas por um fator:

$$\mathbf{v}_i(t + \delta t/2) = \mathbf{v}_i(t - \delta t/2) \left(\frac{T_0}{T} \right)^{1/2} \quad (4.2.18)$$

Quando não há forças externas atuando sobre um sistema sua energia total se conserva, podendo o mesmo ser mantido isolado, dispensando as correções nas velocidades. Porém, como no início da dinâmica há uma tendência geralmente na direção de configurações de mais baixa energia potencial a energia cinética e, portanto a temperatura tende a aumentar. Mesmo nas vizinhanças do mínimo global, ou onde a hiper-superfície de energia é rugosa o suficiente, poderá haver flutuações entre configurações levando a desvios na temperatura. Erros numéricos levam também à não conservação da energia total do sistema. Esses fatores motivam o estudo da dinâmica de macromoléculas utilizando um ensemble canônico, acoplando o sistema a um banho térmico representado pelas correções nas velocidades.

O controle das velocidades é também útil no início da dinâmica. Dado um conjunto de coordenadas iniciais otimizadas, são introduzidas velocidades iniciais por um sorteio dentre uma

distribuição de Maxwell-Boltzmann associada a um valor de temperatura igual ou inferior a 10 K. Uma dinâmica preliminar de poucos picos segundos pode ser efetuada para melhorar a distribuição e equilíbrio das velocidades. O sistema é então aquecido, ajustando-se as velocidades, a taxas em torno de 20 K por pico segundo até se alcançar a temperatura desejada. Em seguida a dinâmica desenvolve-se à temperatura constante por algumas dezenas de pico segundos para equilibrar o sistema no espaço de fase, passando à coleta de dados. Um protocolo semelhante pode ser encontrado em Brooks (1983).

Em resumo podemos dizer que esta metodologia permite que o sistema molecular ultrapasse as barreiras de potencial visitando diversas regiões de mínimos locais, mas não existe nenhuma garantia que estas geometrias de equilíbrio correspondam ao mínimo global. Desta forma, o procedimento usual para MD pode não ser a melhor forma de mapear os pontos de mínimos locais e globais na hiper-superfície de energia. Neste contexto, se a hiper-superfície de energia é não-convexa, veremos a seguir, que os métodos estocásticos oferecem um modo mais eficiente de vasculhar as regiões de mínimos locais e globais.

4.3 Otimização Estocástica – *Generalized Simulated Annealing*

Os sistemas moleculares compostos por diversas moléculas, podem existir em diferentes geometrias conformacionais. Entende-se por conformações geométricas as diversas estruturas e arranjos espaciais dos átomos na molécula. Obviamente, o conjunto de possíveis conformações de equilíbrio cresce a medida que o número de átomos contidos no sistema aumenta. Em particular, aglomerados moleculares envolvendo interação com um meio solvente, sólidos e sistemas biológicos, podem apresentar milhares de possíveis estados de equilíbrio. Isto constitui um dos grandes obstáculos no mapeamento da hiper superfície de energia, isto é determinação dos mínimos locais e globais.

As técnicas usuais em química quântica e mecânica molecular clássica, para determinar os estados de equilíbrio, geralmente são baseadas em métodos gradientes. Como estas técnicas constituem métodos de busca de mínimos locais, elas não permitem determinar ou distinguir os diferentes mínimos existentes. Neste caso, é necessário procurar outras técnicas que possam dar uma idéia mais ampla da localização dos possíveis mínimos. Atualmente, a técnica mais sofisticada para atacar tal problema é a de busca aleatória baseada no método Monte Carlo com temperatura dependente do tempo. Em nosso caso, fazemos uso do procedimento denominado, em inglês, por *Generalized Simulated Annealing* (GSA). O algoritmo usado para construir a estratégia GSA é baseado na termo-estatística, generalizada por Tsallis (referência em anexo).

Neste procedimento, os possíveis mínimos são obtidos através da comparação das energias moleculares de duas geometrias moleculares consecutivas geradas estocásticamente. Isto é, a rotina GSA gera randomicamente as novas coordenadas atômicas e o programa THOR usa-as para calcular a energia molecular correspondente. Uma representação diagramática do algoritmo computacional pode ser visto na figura abaixo.

Uma outra vantagem na utilização da dinâmica estocástica, sobre a dinâmica clássica convencional, está no fato de não ser necessário calcular a força que atua em cada átomo individualmente, o que reduz o tempo computacional drasticamente. No caso da dinâmica estocástica é necessário conhecer apenas o potencial molecular total, o qual, será minimizado durante o processo evolutivo.

Este processo termina quando a convergência é atingida para um dado critério escolhido a priori.

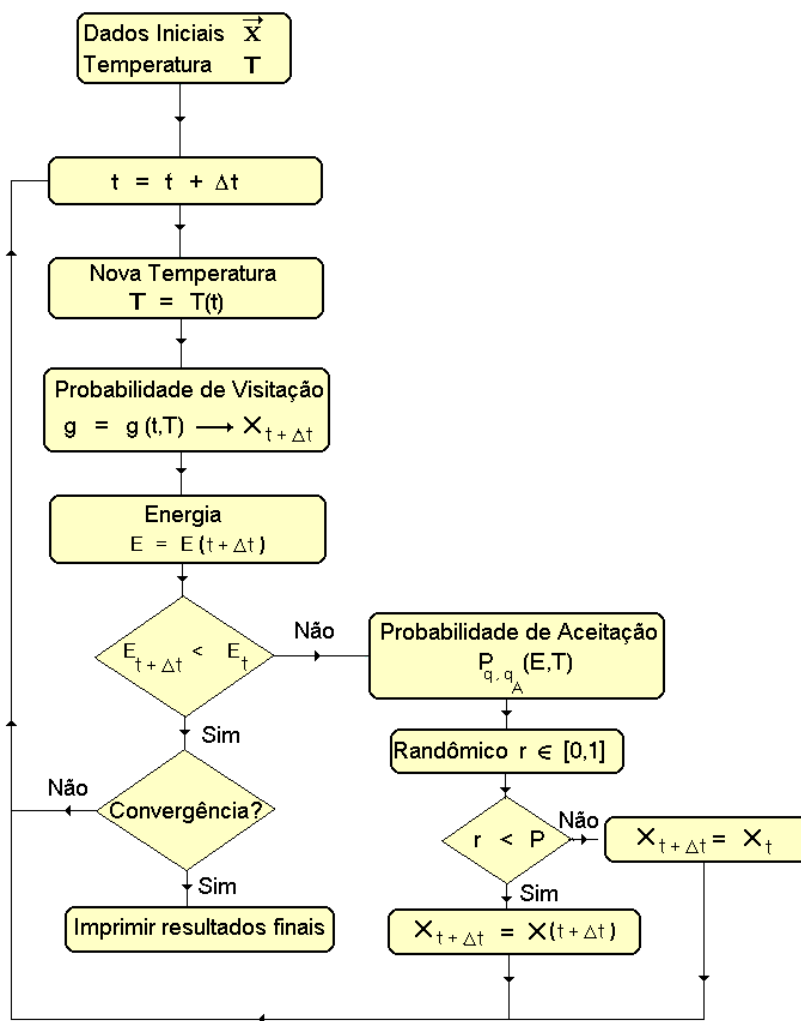


Figura -4

Uma representação esquemática do algoritmo computacional GSA

No programa THOR foi implementada uma versão adaptada do algoritmo GSA que basicamente consiste nos seguintes passos:

i) São geradas conformações aleatórias do ligante dentro da rede e é introduzida uma temperatura inicial artificial T_{q_T} a qual é gradualmente diminuída de acordo com a expressão abaixo:

$$T_{q_T}(t) = T_{q_T}(1) \frac{2^{q_T-1} - 1}{(1+t)^{(q_T-1)} - 1} \quad (4.3.1)$$

onde t é o número de ciclos; q_T é um parâmetro pré-estabelecido e $T_{q_T}(1)$ é a temperatura inicial.

ii) O procedimento de busca da conformação de mínimo global consiste em se comparar a energia de interação ligante/receptor (utilizando o procedimento de interpolação dos valores da rede) de duas conformações aleatórias consecutivas $X(t)$ e $X(t+1)$. Onde $X(t)$ é um vetor contendo os valores relativos aos graus de liberdade que descrevem a conformação do ligante no ciclo t . As geometrias do ligante em dois ciclos consecutivos são relacionadas por:

$$x(t+1) = x(t) + \Delta x \quad (4.3.2)$$

onde Δx corresponde a um vetor contendo as perturbações nos parâmetros associados aos graus de liberdade do ligante. Para cada grau de liberdade do ligante temos associado um parâmetro x e uma perturbação Δx . O vetor Δx é gerado utilizando a seguinte expressão:

$$\Delta x = g^{-1}(\omega) = \text{inversa} \left(\int_{-\infty}^{x'} g_{q_v}(x') dx \right) \quad (4.3.3)$$

x' é um número randômico entre $[0,1]$ e g_{q_v} é uma função de distribuição de probabilidades dada por:

$$g_{q_v}(x') = \left(\frac{q_v - 1}{\pi} \right)^{\frac{1}{2}} \frac{\Gamma\left(\frac{1 - 0.5(q_v - 1)}{q_v - 1}\right) [T_{q_v}(t)]^{-\left(\frac{1}{3 - q_v}\right)}}{\Gamma\left(\frac{1}{q_v - 1} - \frac{1}{2}\right) \left\{ 1 + (q_v - 1) \frac{(x')^2}{[T_{q_v}(t)]^{\frac{2}{3 - q_v}}} \right\}^{\left(\frac{1}{q_v - 1} - \frac{1}{2}\right)}} \quad (4.3.4)$$

onde q_v é um parâmetro pré-estabelecido (parâmetro de visitação).

iii) Uma vez obtidas as conformações do ligante $x(t)$ e $x(t+1)$ são calculadas as respectivas energias de interação com o receptor $E[x(t)]$ e $E[x(t+1)]$. A nova conformação $x(t+1)$ será aceita seguindo os seguinte critérios:

- se $E[x(t+1)] \leq E[x(t)]$ então a conformação é aceita: $x(t+1) \rightarrow x(t)$.
- se $E[x(t+1)] > E[x(t)]$ então gera-se um número randômico r entre 0 e 1;
- se $r < P_{q_a}$ então a conformação é aceita: $x(t+1) \rightarrow x(t)$.

P_{q_A} é função de aceitação que depende de T_{q_T} e do parâmetro pré-estabelecido q_A (parâmetro de aceitação).

$$P_{q_A}(x_t \rightarrow x_{t+1}) = \begin{cases} 1 & \text{se } E(x_{t+1}) \leq E(x_t) \\ \frac{1}{1 + \left[1 + \frac{(q_A - 1)[E(x_t) - E(x_{t+1})]}{T_{q_A}(t)} \right] \left(\frac{1}{q_A - 1} \right)} & \text{se } E(x_{t+1}) > E(x_t) \end{cases} \quad (4.3.5)$$

iv) Volta ao passo (i). O procedimento é repetido até que se atinja o número total de ciclos estipulado inicialmente.

Embora o procedimento acima tenha uma validade geral, a eficiência na sua utilização depende da escolha dos parâmetros envolvidos e na estratégia geral de sua aplicação. No estudo de conformações de pequenas proteínas e peptídeos uma estratégia que simplifica e torna factível o cálculo é considerar constantes os tamanhos das ligações covalentes e os ângulos entre duas ligações consecutivas, mantendo-se seus valores de equilíbrio. As variáveis serão então somente os ângulos de torção.

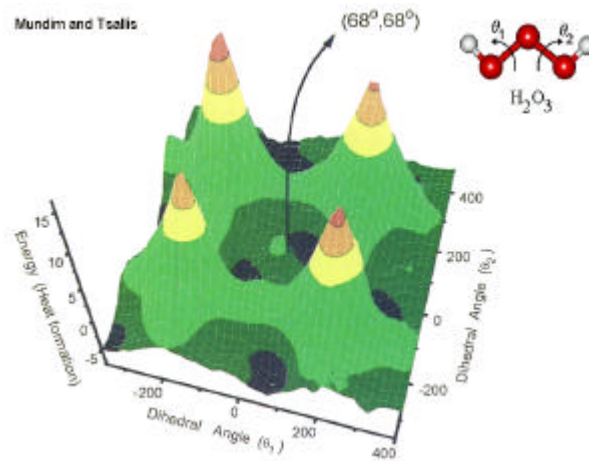


Fig.5a – Mapeamento da hiper-superfície de energia

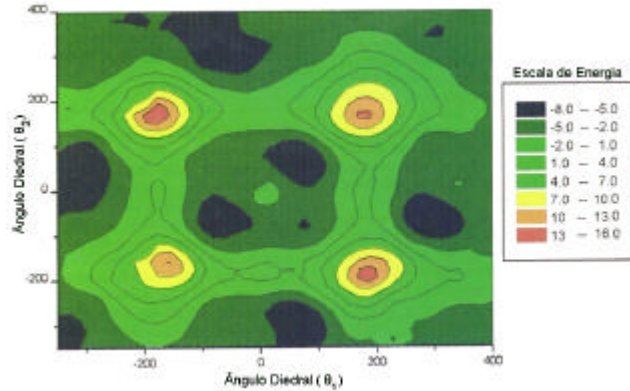


Fig.5b – Hiper-superfície de energia em curvas de níveis

5. Procedimento para Extrair Potenciais entre Pares Atômicos usando Cálculos Ab-initios

A seguir daremos os principais passos necessários na construção de potenciais entre pares atômicos a partir de cálculos *ab-initios*.

Primeiramente, calcula-se a energia coesiva da solução (A-B) em estudo pela relação;

$$E(c) = \frac{1}{2}V(0)c^2 + \frac{1}{2}[1 - 2c]V_{AA}(0) + cV_{AB}(0) \quad (5.1)$$

onde c é a fração atômica da fração de componentes do tipo B e ainda;

$$V(0) = \sum_{\mathbf{r}, \mathbf{r}'} V_{AA}(\mathbf{r}, \mathbf{r}') + V_{BB}(\mathbf{r}, \mathbf{r}') - 2V_{AB}(\mathbf{r}, \mathbf{r}') \quad (5.2)$$

$$V_{AA}(0) = \sum_{\mathbf{r}, \mathbf{r}'} V_{AA}(\mathbf{r}, \mathbf{r}') \quad (5.3)$$

$$V_{AB}(0) = \sum_{\mathbf{r}, \mathbf{r}'} V_{AB}(\mathbf{r}, \mathbf{r}') \quad (5.4)$$

O primeiro termo na Eq.(5.1) pode também ser calculado usando, em caso de ligas (*solid-solution*) com pequenas concentrações de componentes B, a seguinte relação;

$$E(c) = cE_B + (1 - c)E_A + Uc(1 - c) \quad (5.5)$$

Neste modelo E_A e E_B são as energias coesivas dos componentes A e B, respectivamente e o último termo representa a contribuição da mistura ou liga. Comparando os termos c^2 nas equações (5.2.1) e (5.2.2) obtém-se imediatamente que $V(0)/2 = -U$. Levando em conta que as contribuições energéticas de cada componente são dadas por;

$$E_A = \frac{1}{2} \sum_{\substack{i,j \\ i \neq j}} V_{AA} \left(\left| \rho_i^p - \rho_j^p \right| \right), \quad E_B = \frac{1}{2} \sum_{\substack{i,j \\ i \neq j}} V_{BB} \left(\left| \rho_i^p - \rho_j^p \right| \right) \quad (5.6)$$

pode-se obter facilmente a energia da mistura a qual está diretamente ligada com o potencial de interação entre A e B, isto é;

$$\delta E = \sum_{\substack{i,j \\ i \neq j}} V_{AB} \left(\left| \rho_i^p - \rho_j^p \right| \right) = U + E_A + E_B \quad (5.7)$$

Este resultado é muito importante pois ele permite extrair o potencial A-B usando cálculos *ab-initio*.

5.1 Ajuste do Potencial entre Pares Atômicos

■ Usando o método LMTO CPA (*Linear Muffin-Tin Orbitals Coherent Potential Approximation*) calcula-se a energia para diferentes parâmetros de rede de uma célula bcc de tungstênio com boro intersticial ou pura. Em seguida faz-se um ajuste da energia por função potencial. Em geral, as funções de ajustes são do tipo Morse, isto é;

$$E_{\text{tot}} = A + Be^{-2\lambda(a-a_0)} - 2Be^{-\lambda(a-a_0)} \quad (5.1.1)$$

Onde os dois últimos termos representam a energia coesiva dos componentes puros. Assume-se que a energia coesiva em função da distância pode ser expressa pela equação;

$$E(r) = \sum_{p=1}^{\infty} n_p V(s_p r) \quad (5.1.2)$$

Assume-se também que as separações atômicas estão agrupadas em camadas de coordenação (p) com raios $s_p r$, contendo cada uma delas n_p átomos. Em seguida é feita uma dilatação uniforme da rede variando o parâmetro r, supondo as constantes de estrutura s_p e n_p fixas. As camadas são indexadas por $s_1 < s_2 < s_3 < \dots$. Com base na Eq.(5.1.2) pode-se calcular o potencial de interação para a primeira camada usando a relação;

$$V(r) = \frac{1}{n_1} \left(E(r) - \sum_{p=2}^{\infty} n_p V(s_p r) \right) \quad (5.1.3)$$

E por recursivas substituições obtém-se a expressão geral para o potencial V(r);

$$V(r) = \frac{1}{n_1} E(r) - \sum_{p=2}^{\infty} \frac{n_p}{n_1^2} E(s_p r) + \sum_{p,q} \frac{n_p n_q}{n_1^3} E(s_q s_p r) - \Lambda \quad (5.1.4)$$

Este potencial é finalmente ajustado por uma outra função do tipo Morse que em seguida é usado para determinar as estruturas de equilíbrios. A estruturas de equilíbrios são obtidas por relação atômica de forma randômica usando o método GSA.

Este procedimento permite, por exemplo, estudar *grain-boundary* ($\Sigma_3 \langle 111 \rangle$ GB) metais. Em particular tem-se construído um excelente potencial para ligas de tungstênio e boro.

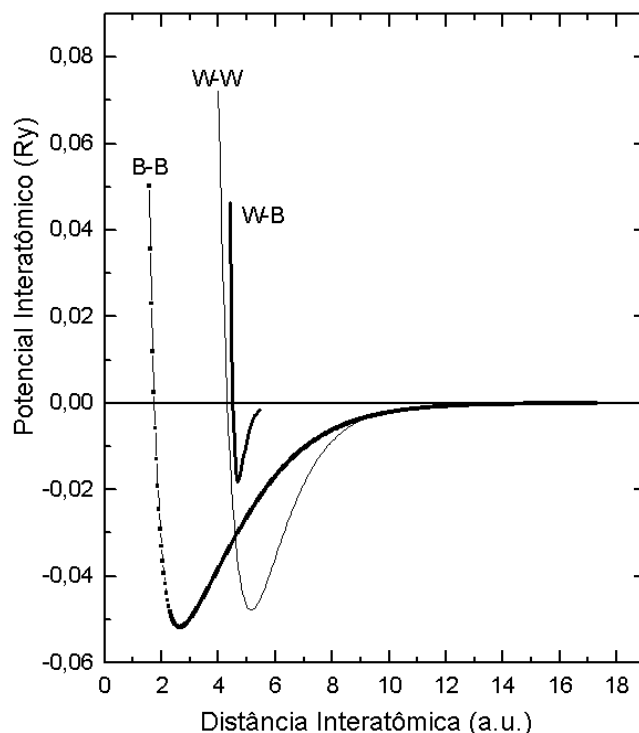


Fig. 6 – Potencial entre pares atômicos para o sistema W-B

6. Códigos Computacionais Disponíveis

Pacotes de programas de cálculo teórico:

- **Thor**: pacote de modelagem molecular usando métodos híbridos QM-MM.
- **Tinker** pacote de modelagem molecular. É concebido para ser um sistema de uso fácil e flexível empregado em mecânica e dinâmica molecular.
- **Gamess**: pacote geral de química quântica para cálculos *ab initio*, funcional de densidade e semi-empírico (MNDO, AM1 e PM3).
- **Defl**: é um *software* de mecânica quântica computacional, baseado na teoria do funcional de densidade. Utiliza funcional de densidade do tipo gaussiano.
- **Moldir**: código de química quântica que faz cálculos de sistemas moleculares multi-eletrônicos usando formalismo de Fock-Dirac e cálculos adicionais de correlação.
- **Moldy**, 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** 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.

Pacotes de visualização gráfica:

- **Rasmol**: Programa de visualização molecular em 3D. Tem código para Linux e Windows.
- **WebLabview**: Programa de visualização molecular em 3D para o Windows
- **ChemWin**: Plug-ins para browser (Netscape ou Explorer) permite a visualização e animação em 3D.

Bibliography

1. Alder, B.J and and Wainwright, T.E., J. Chem. Phys., **27**, 1208-1209 (1957); Phys. Rev. , **1**, 18-21 (1970)
2. Allinger, N.L. and Burkert,U., “Molecular Mechanics”, ACS-Monograph, **177** (1982)
3. Areas, E. P. G., Pascutti, P. G., Schreier, S., Mundim, K. C., Bisch, P. M., J. Phys. Chem., **99**, 14882-14892 (1995).
4. Areas, E. P. G., Pascutti, P. G., Schreier, S., Mundim, K. C., Bisch, P. M., Brazilian J. Mol. Biol. Res., **27**, 527-533 (1994)
5. Kirkpatrick, S., J. Stat. Phys., **34** , 975 (1984)
6. Kirkpatrick, S., Gelat, C. D., and Vecchi, M. P., Science, **220**, 671 (1983).
7. Ceperly, D., Alder, C., Science, **231**, 555 (1986).
8. Szu, H., and Hartley, R., Phys. Lett. A, **122**, 157 (1987).
9. Tsallis, C., J. Stat. Phys., **52**, 479 (1988).
10. Curado, E. M. F., and Tsallis, C., J. Phys A, **24**, L69 (1991); Corrigenda: J.Phys. A, **24**, 3187 (1991).
11. Tsallis, C., and Stariolo, D. A., Phys. A , **233** , 395 (1996).
12. Mundim, K.C., Tsallis, Int. J. Quantum Chem. , **58** , 373-381 (1996).
13. Moret, M.G., Pascutti, P.G , Bish, P.M. and Mundim, K.C., J. Comp. Chem., **19**, 647-657(1998).

14. Mundim, K. C., Lemaire, T. J. and Amin Bassrei, A., *Physica A*, **252**, 405-416 (1998).
15. Ellis, D.E. and Guo, J., in "Electronic Density Functional Theory of Molecules, Clusters, and Solids", ed. D.E. Ellis, (Kluwer, Dordrecht, 1995) p263.
16. Guo, L., Ellis, D.E, Mundim, K.C. and Hoffman, B. M. , *Journal of Porphyrins and Phthalocyanines*, **2**, 1, (1998)
17. S.Dorfman and D.Fuks, *Composite Interfaces* **3**, 431 (1996) and references therein.
18. F.Dellanay, L.Froyen, and A.Deruyttere, *Journ. of Mater. Sci.* **22**, 1 (1987).
19. U.Gangopadhyay and P.Wynblatt, *Journ. of Mater. Sci.*, **30**, 94 (1995); S.Hara, K.Nogi, and K.Ogino, in *Proceedings of Int. Conf. "High-temperature capillarity"*,
20. S.Dorfman and D. Fuks, *Composites* **27A**, 697 (1996).
21. Wyckoff, H.W., *Acta Cryst.*, **11**, 199 (1958)
22. Ellis, D.E, Mundim, K.C., Fuks, D. , Dorfman, S. and A. Berner , *Mat. Res. Soc. Symp., Proc. vol. 527* (1998) 69-74.
23. C-X Guo, D.E. Ellis, O. Warschkow, unpublished.
24. D.E. Ellis and K.C. Mundim, in "Proc. of 9th CIMTEC-World Ceramics Congress & Forum on New Materials" P. Vincenzine, Ed., Techna, Florence, 1998, to be published
25. Pascutti, P. G., Mundim, K. C., Ito, A. S. and Bisch, P.M, *J. Comp. Chem.* (to be published 1999).
26. Moret, M.A., Master-These (IF-UFBa 04/1996)

27. Pascutti, P. G., Tese de doutorado, USP (1994)