

### 3 Fundamentação Teórica

#### 3.1. Dinâmica Molecular

Dinâmica molecular é uma técnica de simulação computacional da evolução temporal de átomos e moléculas que interagem entre si através de um potencial da posição de cada núcleo atômico ou molécula (ALDER; WAINWRIGHT, 1959). Nesta técnica, podem ser usados sistemas com todos os núcleos atômicos explícitos, partículas efetivas ou núcleos atômicos agrupados. Consequentemente, é usada de forma interdisciplinar para estudar o movimento de átomos em moléculas, líquidos, sólidos e gases.

Geralmente as forças agindo nas partículas são calculadas instantânea e discretamente durante as simulações. A energia potencial de cada partícula é descrita por potenciais de interação, também conhecidos como campos de força, calculados a partir de dados empíricos, exceto em dinâmicas moleculares *ab initio* que são realizadas a partir de primeiros princípios.

Dinâmica Molecular (DM) clássica usa as leis da mecânica newtoniana para descrever sistemas moleculares. Geralmente os modelos consideram os átomos como partículas individuais e as moléculas como átomos ligados em uma dada geometria de equilíbrio. Podem ser enumerados vários tipos de Dinâmica Molecular: Dinâmica Molecular *Ab Initio*; Dinâmica Molecular Clássica; Dinâmica Molecular Híbrida Quântico-Clássica, e Dinâmica Molecular *Coarse Grained*. Nesta dissertação, a Dinâmica Molecular Clássica é utilizada e descrita em seguida.

##### 3.1.1. Dinâmica Molecular Clássica

A dinâmica molecular clássica diminui significativamente o tempo de simulação em comparação com a dinâmica molecular *ab initio*(ALDER; WAINWRIGHT, 1959). Isto faz com que seja possível a modelagem de sistemas com tamanho e tempo maiores. Esta eficiência é atingida fundamentalmente ignorando o movimento dos elétrons e os incorporando ao núcleo, utilizando cargas parciais.

A energia do sistema é calculada unicamente com a posição dos núcleos atômicos. A posição dos núcleos depois de um tempo determinado é calculada com base na lei de movimento de Newton, resolvendo a equação:

$$\frac{d^2 X_I}{dt^2} = \frac{F_{X_I}}{M_I}$$

A equação descreve o movimento de uma partícula  $M_I$  na dimensão  $X_I$  resultado da força  $F_{X_I}$  nessa dimensão. Resolvendo esta equação para cada partícula se obtêm a velocidade e posição que resulta nas trajetórias da dinâmica molecular durante o tempo de simulação. A resolução analítica desta equação é complicada (ALDER; WAINWRIGHT, 1959). Os parâmetros para integrar esta equação por métodos matemáticos finitos devem considerar a precisão e eficiência da dinâmica molecular de acordo com os objetivos.

Geralmente as forças são atualizadas em incrementos tempo determinados (passos temporais). As forças calculadas em conjunto com a posição e velocidade das partículas em um tempo  $t$  determinam a posição e velocidade no tempo  $t + \Delta t$ , a força é atualizada e novamente a posição e a velocidade no tempo  $t + 2\Delta t$ . Distintos algoritmos podem ser usados para resolver esta equação numericamente (LESZCZYNSKI, 2012).

### 3.2.

#### Campos de Força da Dinâmica Molecular Clássica

A Mecânica Molecular descreve os sistemas moleculares desde um ponto de vista clássico (Mecânica Newtoniana). A mecânica molecular pode ser considerada como uma extensão da forma clássica de representar os sistemas químicos com ligações e átomos, e com as interações entre átomos não ligados. Na maioria dos campos de força a energia potencial das moléculas em função das coordenadas é representada pela soma de várias contribuições, normalmente:

$$\Delta E = \sum El + \sum Ea + \sum Et + \sum Enl$$

onde os termos são definidos como:

$\sum El$  Energia em função do comprimento de ligação

$\sum Ea$  Energia em função do ângulo de ligação

$\sum Et$  Energia em função da torção própria dos ângulos

$\sum Enl$  Energia em função das interações de átomos não ligados (Coulombianas e de van der Waals).

### 3.2.1. Interações entre átomos ligados

A maioria dos campos de força define as interações entre átomos ligados somente para ligações covalentes, assim as interações resultantes do compartilhamento de elétrons ou sobreposição de orbitais moleculares são modeladas com um conjunto de funções com o objetivo de manter a geometria das moléculas. Na maioria dos campos de força clássicos as ligações de hidrogênio, ligações iônicas, etc. não são consideradas como “ligações”. A energia derivada da conformação geométrica de um conjunto de partículas é determinada pelo comprimento (duas partículas), ângulo (três partículas) e torção (quatro partículas). É importante notar que normalmente os campos de força são agnósticos com respeito ao tipo de ligação, portanto não há diferença visual entre uma ligação simples, dupla ou tripla. A única diferença entre elas está no valor das constantes de interação.

#### 3.2.1.1. Energia Ligante

A energia das ligações covalentes em função das duas partículas ligadas é geralmente descrita por um termo harmônico simples. O afastamento do comprimento de referência resulta em uma energia potencial:

$$E_{(ligação)} = \sum_{ligação} k_l (l_l - l_l^{(0)})^2$$

Onde  $k_l$  é definida como a constante de ligação e determina a força da ligação,  $l_l^{(0)}$  é o comprimento de referência da ligação covalente. No entanto, deve ser lembrado que este modelo se ajusta muito bem somente na região perto do comprimento de referência. Quando a distância entre as partículas ligadas afasta-se demasiadamente, a energia perde simetria com respeito à referência. No entanto, este modelo é suficientemente exato para temperaturas e pressões ordinárias. Usadas tipicamente nos sistemas biológicos estudados por meio de dinâmica molecular clássica.

#### 3.2.1.2. Energia Angular

Normalmente, a energia que resulta do desvio de um ângulo de referência das ligações entre três partículas é descrita de forma similar à energia de ligação, usando-se o mesmo modelo:

$$E_{(angular)} = \sum_{angular} k_a (\theta_a - \theta_a^{(0)})^2$$

onde  $k_a$  é chamada constante angular e  $\theta_a^{(0)}$  é o ângulo de referência.

Energia Torcional

A energia torcional é geralmente descrita usando séries de cossenos:

$$E_{(torcional)} = \sum_{torcional} k_d [1 + \cos(n\phi - d)]$$

Onde  $k_d$  é constante para cada termo podendo ser diferente ou não para cada  $\phi$ ,  $\phi_a^{(0)}$  é o ângulo diedro de referência. A descrição desta interação é a que normalmente difere entre um e outro campo de força. Estes parâmetros influenciam grandemente a geometria das moléculas, especialmente as maiores moléculas. É importante observar que a energia derivada deste ângulo pode não ser ajustada pela função usada no campo de força. Uma forma prática de resolver este problema é usar mais de um conjunto de constantes para um mesmo quarteto de partículas até que a soma destas funções se acople satisfatoriamente ao comportamento desta interação. Para fazer mais eficientemente os cálculos da energia potencial de uma partícula com vários ângulos diedros que têm uma mesma ligação como eixo comum, é possível determinar estas interações por um mesmo quarteto de partículas em lugar de considerar cada um dos quartetos que realmente estão interagindo sobre o mesmo eixo.

### 3.2.2.

#### Interações entre átomos não ligados

Na maioria dos campos de força consideram-se duas interações não resultantes de ligações covalentes, as interações de van der Waals e as interações eletrostáticas. No entanto, a omissão ou não das interações de um par de partículas (normalmente par de partículas adjacentes) pode variar entre campos de força. Também podem ser omitidos algumas interações para ajustar o comportamento de algumas moléculas.

Assim vários campos de força (Gromos entre eles) omitem as interações de van der Waals entre átomos ligados covalentemente, porque estas interações de van der Waals estão normalmente incluídas implicitamente nos parâmetros do comprimento de ligação; nos vizinhos de segundo grau, porque normalmente estas interações estão descritas nas interações de ligação, incluindo comprimento, ângulo e outros. As interações de Coloumb entre alguns pares de átomos podem ser omitidas quando é conhecido que possuem pouca influência,

em relação às interações de ligação, e precisa-se um maior rendimento de simulação. Alguns algoritmos podem agrupar estas interações e assim diminuir substancialmente o número de cálculos numéricos feitos; por exemplo, no uso de grupos de carga para o cálculo de interações de Coulomb.

### 3.2.2.1.

#### Interações de van der Waals

Geralmente as interações de van der Waals são descritas por termos empíricos que atuam sobre os pares de partículas. Estas interações são muito utilizadas nos campos de força através do uso da função Lennard-Jones 12-6:

$$E_{vdW}(R) = \sum_{I < J} 4\epsilon_{IJ} \left[ \left( \frac{\sigma_{IJ}}{R_{IJ}} \right)^{12} - \left( \frac{\sigma_{IJ}}{R_{IJ}} \right)^6 \right]$$

Onde a constante  $\sigma_{IJ}$  é o parâmetro de colisão e a constante  $\epsilon_{IJ}$  o parâmetro da profundidade das interações. Estes parâmetros são diferentes para pares de partículas diferentes. Alternativamente pode ser usado outro modelo similar, a função de Lennard-Jones 9-6:

$$E_{vdW}(R) = \sum_{I < J} 4\epsilon_{IJ} \left[ \left( \frac{\sigma_{IJ}}{R_{IJ}} \right)^9 - \left( \frac{\sigma_{IJ}}{R_{IJ}} \right)^6 \right]$$

Uma das vantagens deste modelo é o baixo custo computacional necessário para a estimativa da energia potencial resultante destas interações devido a utilização de somente dois parâmetros.

Para diminuir o número de parâmetros definidos, ao invés de definir os parâmetros para cada par de tipos de partículas possíveis ( $(\epsilon_{AA}, \epsilon_{AB}, \epsilon_{AC}, \dots, \epsilon_{BB}, \epsilon_{BC}, \dots, \epsilon_{ZZ}, \sigma_{AA}, \sigma_{AB}, \sigma_{AC}, \dots, \sigma_{BB}, \sigma_{BC}, \dots, \sigma_{ZZ})$ ) são definidos somente os parâmetros singulares ( $(\epsilon_A, \epsilon_B, \epsilon_C, \dots, \epsilon_Z, \sigma_A, \sigma_B, \sigma_C, \dots, \sigma_Z)$ ) e os parâmetros usados são derivados por uma série de regras; por exemplo, a regra *Lorentz-Berthelot* onde os parâmetros das interações de van der Waals entre partículas de  $I$  e  $J$  são definidos por:

$$\sigma_{IJ} = \frac{(\sigma_I + \sigma_J)}{2}$$

$$\epsilon_{IJ} = \sqrt{\epsilon_I \epsilon_J}$$

onde  $\sigma_{IJ}$  é a média aritmética dos parâmetros  $\sigma_I$  e  $\sigma_J$ , e  $\epsilon_{IJ}$  é a média geométrica entre os parâmetros  $\epsilon_I$  e  $\epsilon_J$ . Uma alternativa usada comumente para determinar estes parâmetros é o uso da média geométrica em ambos termos.

Outros modelos são usados alternativamente. Alguns destes modelos são:

(potencial de *Buckingham*)

$$E_{vdw}(R) = \sum_{I < J} \left[ A_{IJ} \exp(-B_{IJ} R_{IJ}) - \left( \frac{C_{IJ}}{R_{IJ}^6} \right) \right]$$

(E o potencial de *Born–Mayer–Huggins*)

$$E_{vdw}(R) = \sum_{I < J} \left[ A_{IJ} \exp(B_{IJ}(\sigma_{IJ} - R_{IJ})) - \frac{C_{IJ}}{R_{IJ}^6} + \frac{D_{IJ}}{R_{IJ}^8} \right]$$

Nestes modelos a energia é função unicamente do comprimento entre o par de partículas.

### 3.2.2.2.

#### Interações eletrostáticas

Na maioria dos campos de força para dinâmica molecular clássica não se consideram o fluxo de elétrons nem os elétrons como partículas, são usadas cargas parciais. Normalmente as cargas parciais atribuídas aos átomos permanecem fixas durante a simulação. Geralmente estas cargas parciais são posicionadas no centro das partículas e contribuem para parte de suas propriedades. Não obstante existe a possibilidade de colocar uma carga parcial em uma partícula sem massa, unificar várias cargas parciais ou uma mistura de ambas opções.

No cálculo das interações de Coulomb geralmente são omitidas as interações entre partículas que estão separadas por uma ou duas ligações. Com esta regra a energia das interações eletrostáticas s partículas  $I$  e  $J$  é calculada com as cargas parciais utilizando a lei de Coulomb:

$$E_{coul} = \sum_I \sum_J \frac{q_I q_J}{4\pi\epsilon_0 R_{IJ}}$$

Onde  $q_I$  é carga parcial da partícula  $I$ ,  $q_J$  é a carga parcial da partícula  $J$ , e é a constante dielétrica e  $R_{IJ}$  e a distância entre a partícula  $I$  e a partícula  $J$ .

Normalmente é usada a soma de Ewald, onde as interações eletrostáticas dividem-se em dois termos:

$$\varphi(r) \stackrel{\text{def}}{=} \varphi_{sr}(r) + \varphi_{lr}(r)$$

onde  $\varphi_{sr}(r)$  são as interações de curto alcance e  $\varphi_{lr}(r)$  são as interações de longo alcance.

### 3.3.

#### Algoritmos da dinâmica molecular

Nos campos de força clássicos o Hamiltoniano  $H$  que descreve o sistema pode ser dividido em energia potencial  $V$  e energia  $K$  cinética

$$H(\mathbf{p}, \mathbf{r}; m, s) = K(\mathbf{p}; m) + V(\mathbf{r}; s)$$

Normalmente a energia cinética é independente das coordenadas e a energia potencial do momento, assim precisa-se somente das coordenadas e das funções que descrevem as interações para calcular a energia potencial, e das velocidade e das massas das partículas para calcular a energia cinética.

A configuração do sistema depois de um certo tempo decorrido determinado é calculada resolvendo a equação diferencial proveniente da segunda lei de Newton

$$\frac{d^2 X_I}{dt^2} = \frac{F_{X_I}}{M_I}$$

Que descreve o movimento da partícula  $I$  de massa  $M_I$  com a influência de uma força  $F_{X_I}$  na dimensão  $X$ . A resolução analítica deste problema em sistemas utilizando dinâmica molecular é um problema de vários corpos. O movimento das partículas está acoplado as suas coordenadas, e portanto, sempre que uma partícula se movimenta ocorre uma mudança na magnitude das interações com as outras partículas. Consequentemente mudará a força que atua sobre ela e a força que atua sobre as outras partículas. Para resolver este problema, métodos numéricos são usados.

A integração deste sistema é realizada dividindo o tempo em uma série de intervalos de tempo. Durante um certo tempo decorrido dividido em intervalos de tempos consecutivos, os valores de força e velocidade são considerados constantes. A aceleração das partículas é calculada a partir das coordenadas, velocidades e interações no tempo  $t$ . Simultaneamente, as coordenadas e velocidades nesse tempo  $t$ , são calculadas e as novas coordenadas e velocidades são calculadas tempo  $t + \Delta t$ , Subsequentemente com os novos valores de coordenada e velocidade, a aceleração é calculada no tempo  $t + \Delta t$  e as coordenadas e velocidades no tempo  $t + 2\Delta t$ .

A forma clássica e simples de integrar as coordenadas e velocidades do sistema é através do uso de séries de Taylor:

$$R(t + \Delta t) = R(t) + \Delta t V(t) + \frac{1}{2} \Delta t^2 A(t) + \frac{1}{6} \Delta t^3 B(t) + \frac{1}{24} \Delta t^4 C(t) + \dots$$

$$V(t + \Delta t) = V(t) + \Delta t A(t) + \frac{1}{2} \Delta t^2 B(t) + \frac{1}{6} \Delta t^3 C(t) + \dots$$

$$A(t + \Delta t) = A(t) + \Delta t B(t) + \frac{1}{2} \Delta t^2 C(t) + \dots$$

onde,  $R$ ,  $V$  e  $A$  são respectivamente, a coordenada, a velocidade e a aceleração no tempo  $t + \Delta t$ ,  $B$  e  $C$  são a terceira e quarta derivadas da posição com respeito ao tempo. Para omitir o uso dos termos  $B$  e  $C$  uma variedade de algoritmos são usados, cada um deles com uma série de vantagens e desvantagens

característicos de cada algoritmo. Entre os mais comuns estão o algoritmo de *Verlet*, o algoritmo *Leap-Frog* e o algoritmo de velocidade de *Verlet*.

### 3.3.1. Algoritmo Verlet

O algoritmo de Verlet (VERLET, 1967) é um dos algoritmos mais usados nas simulações que usam dinâmica molecular clássica. Este algoritmo usa a posição no tempo  $t$  e no tempo  $t - \Delta t$  além da aceleração para determinar a posição no tempo  $t + \Delta t$ , assim o cálculo é expresso da seguinte forma:

$$R(t + \Delta t) = 2R(t) - 2R(t - \Delta t) + \Delta t^2 A(t)$$

A velocidade não é calculada explicitamente. Ela é normalmente estimada com a diferença de posições entre os passos  $t - \Delta t$  e  $t + \Delta t$  dividido pelo intervalo de tempo transcorrido,  $2\Delta t$ , como pode ser visto na equação:

$$V(t) = \frac{|R(t + \Delta t) - R(t - \Delta t)|}{2\Delta t}$$

Uma alternativa é o uso de um diferencial de tempo menor e a estimação da velocidade no ponto médio do intervalo de tempo. Assim, a posição e a velocidade são determinadas nos tempos  $t$  e velocidade

$$V\left(t + \frac{1}{2}\Delta t\right) = \frac{|R(t + \Delta t) - R(t)|}{\Delta t}$$

Pela omissão dos cálculos da terceira e quarta derivadas, o algoritmo de *Verlet* é mais eficiente que o uso da série de *Taylor*, não obstante traz consigo uma série de desvantagens que devem ser consideradas. Uma desvantagem tem a ver com a expressão do Hamiltoniano do sistema; isto é, o algoritmo de *Verlet* tem a desvantagem de não calcular a velocidade de forma explícita. Como resultado a energia cinética do sistema tampouco é calculada de forma explícita. Outra desvantagem é o erro relacionado com a negligência do termo  $\Delta t^2 A(t)$  que é pequeno em relação à diferença de coordenadas entre o instante  $t - \Delta t$  e  $t + \Delta t$ , mas diminui a precisão da estimativa. Uma pequena desvantagem é a necessidade de ter a posição no tempo  $t - \Delta t$ , que no caso de  $t = 0$  não se tem, sendo necessária para a estimativa da posição virtual no tempo  $t - \Delta t$ . Com todas estas desvantagens o algoritmo de *Verlet* é uma das melhores opções quando a energia cinética não é prioritária e a prioridade é a velocidade dos cálculos.

### 3.3.2. Algoritmo *Leap-Frog*

A partir do algoritmo de *Verlet* uma série de variações foram desenvolvidas. Entre elas, um dos algoritmos mais utilizados é o algoritmo “*Leap-Frog*”(HOCKNEY, 1970) que foi desenvolvido para diminuir o erro associado à diferença de coordenadas, relativamente grande em comparação ao termo da aceleração.

No algoritmo *Leap-Frog* a velocidade é calculada no tempo  $t + \frac{1}{2}\Delta t$  com a equação:

$$V\left(t + \frac{1}{2}\Delta t\right) = V\left(t - \frac{1}{2}\Delta t\right) + \Delta t A(t)$$

Posteriormente as coordenadas no tempo  $t + \Delta t$  são estimadas usando as velocidades médias do intervalo de tempo usando a equação

$$R(t + \Delta t) = R(t) + \Delta t V\left(t + \frac{1}{2}\Delta t\right)$$

Desta forma elimina-se o uso da diferença das coordenadas que é grande em magnitude. Além disso, a velocidade é agora calculada explicitamente para o tempo  $t + \frac{1}{2}\Delta t$ , que é menos dependente das coordenadas estimadas. Embora a velocidade seja calculada explicitamente, as energias cinéticas e potenciais estimadas não estão sincronizadas. Como o Hamiltoniano não é explícito para cada instante  $t$ , é usada adicionalmente alguma outra forma de estimar. Também continua-se existindo a necessidade de calcular dados de um tempo anterior ao tempo inicial  $t_0$  e novamente é necessário o uso da série de *Taylor*. Contudo, o algoritmo *Leap-Frog* é usualmente preferido em relação ao algoritmo de *Verlet* especialmente pela obtenção explícita da velocidade e a diminuição substancial do erro.

### 3.3.3. Algoritmo de velocidade de *Verlet*

Quando a obtenção explícita da posição e velocidade no mesmo instante é prioridade, é preferível o uso de outra variação do algoritmo de *Verlet*. Com este objetivo foi proposto o algoritmo de velocidade de *Verlet*(SWOPE, 1982). Este método consta de três etapas que difere do algoritmo original de *Verlet* ou do algoritmo *Leap-Frog*. Primeiramente, a velocidade no tempo  $t + \frac{1}{2}\Delta t$  é calculada através da velocidade em  $t + \frac{1}{2}\Delta t$  por meio da aceleração no tempo  $t$  utilizando a seguinte equação:

$$V\left(t + \frac{1}{2}\Delta t\right) = V(t) + \frac{1}{2}\Delta t A(t)$$

Posteriormente, as coordenadas no tempo  $t + \Delta t$  são calculadas de forma análoga ao algoritmo *Leap-Frog*:

$$R(t + \Delta t) = R(t) + V\left(t + \frac{1}{2}\Delta t\right)$$

Com as coordenadas tempo  $t + \Delta t$  e as interações, as acelerações no tempo  $t + \Delta t$  são calculadas. Posteriormente, as velocidades no tempo  $t + \Delta t$  são calculadas através das acelerações no tempo  $t + \Delta t$  e das velocidades no tempo  $t + \frac{1}{2}\Delta t$  utilizando a seguinte equação:

$$V(t + \Delta t) = V\left(t + \frac{1}{2}\Delta t\right) + \frac{1}{2}\Delta t A(t + \Delta t)$$

Este método possui quatro etapas. Não obstante, a expressão da velocidade em  $t_0 + \frac{1}{2}\Delta t$  e da velocidade em  $t + \frac{1}{2}\Delta t$  podem ser incluídas implicitamente diminuindo-se para três etapas. Em resumo, o algoritmo de velocidade de Verlet consiste nas seguintes etapas:

$$R(t + \Delta t) = R(t) + \Delta t V(t) + \frac{1}{2}\Delta t^2 A(t)$$

$$A(t + \Delta t)$$

$$V(t + \Delta t) = V(t) + \frac{1}{2}\Delta t (A(t) + A(t + \Delta t))$$

Obtém-se assim explicitamente as energias potencial e cinética no mesmo instante  $t$ , elimina-se a necessidade de coordenadas em um tempo anterior ao inicial. Desta forma, obtém-se uma maior precisão do que o algoritmo de *Verlet* original com um custo computacional não muito maior do que no algoritmo *Leap-Frog*.

### 3.4. Algoritmos de Minimização de Energia

É importante iniciar a dinâmica molecular com um sistema com a estrutura desejada e com a energia potencial minimizada. Também, se a configuração inicial tiver uma energia alta, é provável que o sistema colapse. Com o objetivo de diminuir o tempo usado para equilibrar o sistema e evitar o colapso do mesmo, algoritmos de otimização de coordenadas são utilizados ao invés dos algoritmos de dinâmica molecular.

O algoritmo *steepest descent* é o método de gradiente mais simples para obter uma estrutura otimizada. Ele consiste em alterar as coordenadas das

partículas na direção de gradiente negativo da energia resultante das interações das partículas com as coordenadas iniciais:

$$d_n = - \frac{\partial E(c_n)}{\partial c}$$

$$c_{n+1} = c_n + \Delta_n d_n$$

Onde  $E(c)$  é a energia em função das coordenadas  $c$ ,  $c_n$  são as coordenadas no passo  $n$ ,  $\Delta_n$  é sempre positivo para a energia sempre diminuir. Este algoritmo é simples e robusto. O mínimo local é sempre garantido de ser encontrado. No entanto, o valor de  $d_n$  diminui consideravelmente próximo ao mínimo e a convergência é lenta. Existe a possibilidade de voltar as coordenadas anteriores, diminuindo a eficiência do método. Quando coordenadas iniciais inadequadas são utilizadas no algoritmo, existe a possibilidade do resultado do método oscilar entre duas coordenadas.

### 3.5. Ensembles

Um estado macroscópico é resultado da coletividade de microestados, que interagem de modo a chegarem a um equilíbrio dinâmico estatístico. A este conjunto de microestados que estatisticamente equivale a um macroestado é chamado ensemble estatístico.

Em termodinâmica estatística um ensemble é um conjunto de microestados (estado de um elemento particular, não do conjunto) que resultam em propriedades termodinâmicas como pressão, temperatura, energia interna, etc. (propriedades do conjunto, não particulares). Estes microestados podem ser definidos usando mecânica clássica ou mecânica quântica. Utilizando mecânica clássica para o estado da partícula, o Hamiltoniano de uma partícula normalmente é definido pela posição e momento em cada dimensão. Quando se analisa um ensemble estacionário, ou seja, quando suas propriedades termodinâmicas (do conjunto) permanecem constantes, existe a liberdade de analisar a evolução desses microestados (a evolução dos estados da partícula) sem modificar a média do conjunto (do ensemble). Isto é muito útil no caso de analisar sistemas em equilíbrio (termodinâmica do equilíbrio). Mas quando se observam evolução de macroestados como caso de processos irreversíveis ou desequilíbrio de temperatura sobre um material, e dizer termodinâmica de não equilíbrio, aqui o, a análise volta-se mais complexa. O uso de ensembles fica limitado, não obstante é possível usar os mesmos ensembles (termodinâmica do equilíbrio) quando se analisam pequenas transições nas propriedades

termodinâmicas (termodinâmica de quase equilíbrio, ou termodinâmica de não equilíbrio como se for termodinâmica de equilíbrio).

Com as características e limitações das simulações computacionais em dinâmica molecular os ensembles comumente usados são:

- o ensemble canônico (NVT), onde o número de partículas, o volume e a temperatura permanecem constantes.
- o ensemble isotérmico isobárico (NPT), onde o número de partículas, a pressão e a temperatura permanecem constantes.

### 3.5.1. NVT

Para analisar este ensemble por dinâmica molecular é necessário monitorar e controlar a energia cinética média para manter constante a temperatura, sendo necessário retirar ou adicionar energia cinética. Este ensemble é útil para calcular a energia livre de Helmholtz.

O único termo que tem que ser controlado neste ensemble é a temperatura para a energia cinética média  $\langle K \rangle$  ser constante. No caso da dinâmica molecular clássica, a energia cinética de uma partícula é determinada pelo momento e pode ser calculada através da velocidade por  $\frac{1}{2}mv^2$ . Assim, a energia cinética média para N partículas similares é dada por  $\langle K \rangle = \frac{\frac{1}{2}\sum_{i=1}^N m_i V_i^2}{N}$ .

Ao mesmo tempo, a energia média geralmente pode ser relacionada com a temperatura através da constante de Boltzmann e do número de graus de liberdade pela equação  $\langle K \rangle = \frac{1}{2} N_{gl} K_b T$ . Consequentemente, o ensemble é definido por  $N, V, T$  ou  $\frac{\sum_{i=1}^N m_i V_i^2}{N N_{gl} K_b}$  constante.

A temperatura na dinâmica molecular pode ser controlada de forma simples usando termostatos, não obstante nem todos os termostatos representam muito bem um ensemble.

### 3.5.2. NPT

Este tipo de ensemble é o de maior interesse em química porque a maioria dos experimentos dos sistemas analisados em química ocorrem a pressão e temperatura constantes. Este ensemble é utilizado para o cálculo da energia livre de Gibbs. No entanto, em dinâmica molecular a pressão não é uma

variável muito fácil de controlar, especialmente em sistemas de tamanho reduzido.

A pressão é calculada da diferença entre a energia cinética  $K$  e a virial  $\Xi$ , que é o somatório do produto entre a posição e as forças interatômicas em cada dimensão para cada partícula. Seu tensor está definido como  $-\frac{1}{2}\sum_{i<j}^N r_{ij}\otimes F_{ij}$ . Como resultado a pressão é definida como  $P = \frac{2}{V}(K - \Xi) = \frac{2}{V}(\frac{1}{2}\sum_{i=0}^N m_i v_i \otimes v_i + \frac{1}{2}\sum_{i<j}^N r_{ij}\otimes F_{ij})$ . Note que negligenciando as interações interatômicas resulta-se na equação do gás ideal.

Para manter este ensemble é necessário o uso de dois acoplamentos um de pressão e um de temperatura. Além disso, a pressão em termos da energia cinética usualmente está interligada ao acoplamento de temperatura, resultando em uma acumulação de erro ou incerteza nesta propriedade.

### 3.6. Termostatos

A maioria dos sistemas de interesse é analisada em uma temperatura determinada, resultado de colisões entre partículas e interações entre elas. Portanto, a energia cinética do sistema pode variar, modificando assim sua temperatura. Nestes casos, a temperatura do sistema deve ser mantida constante no decorrer da simulação. Com este objetivo uma série de algoritmos para controlar a temperatura (termostatos) foram desenvolvidos.

No início da dinâmica molecular, é preferível ter um termostato muito restrito, aplicado em cada passo da simulação para manter a temperatura o mais próximo possível do alvo. Posteriormente, para melhorar o rendimento da simulação, é preferível aplicar um termostato menos restrito depois de um número de passos determinado. Portanto, para não influenciar demasiadamente as velocidades, tolera-se uma maior diferença de temperatura.

#### 3.6.1. Termostato de redimensionamento

Um dos métodos mais simples de manter uma temperatura determinada durante o tempo de simulação é o método de reescalonamento ou redimensionamento da velocidade (BUSSI; DONADIO; PARRINELLO, 2007), onde as velocidades das partículas são redimensionadas por um fator de reescalonamento  $\lambda$ , calculado por:

$$\lambda = \sqrt{T_{alvo}/T_0}$$

onde  $T_{alvo}$  é a temperatura desejada e  $T_0$  é a temperatura inicial antes de aplicar o termostato. A velocidade é redimensionada com o fator de reescalonamento.

$$V_{final} = \lambda V_{inicial}$$

Normalmente, o termostato é aplicado depois de um certo número de passos determinados  $\tau$ , Utiliza-se uma tolerância de diferença de temperatura  $\Delta T$  e redimensiona-se só uma fração da diferença de temperatura,  $f_r$ , em cada passo do algoritmo. Uma das deficiências deste termostato é que ele redimensiona depois de um número de passo fixo, e neste período, a temperatura pode ter sido modificada significativamente.

### 3.6.2. Termostato Berendsen

Para ter um redimensionamento mais dinâmico é preferível não ter um tempo de reescalonamento fixo. No caso do termostato de Berendsen (BERENDSEN et al., 1984) o reescalonamento é feito quando a taxa de mudança de temperatura for proporcional a diferença de temperatura entre  $T_{alvo}$  e  $T(t)$  (temperatura no tempo  $t$ ), como pode ser visto na equação:

$$\frac{dT}{dt} = \frac{1}{\tau} (T_{alvo} - T(t))$$

que ao ser integrada resulta em

$$\Delta T = \frac{\Delta t}{\tau} (T_{alvo} - T(t))$$

Esta última diferença de temperatura é usada para reescalar as velocidades através de

$$\lambda^2 = 1 + \Delta T/T_0$$

A desvantagem deste termostato é que não corresponde a nenhum ensemble específico.

### 3.6.3. Termostato Nosé–Hoover

Para ter um termostato correspondente a um ensemble usa-se o termostato *Nosé-Hoover* (HOOVER, 1985; NOSÉ, 1984). Neste termostato, considerando um banho de energia proveniente de um reservatório, adiciona-se um grau de liberdade  $\lambda$ , correspondente à energia potencial do reservatório:

$$E_{pot} = (f + 1)k_B T \ln(s)$$

Onde  $f$  é número de graus de liberdade do sistema e  $s$  é o grau de liberdade adicional ao Hamiltoniano. A energia cinética do reservatório é representada por:

$$E_{cin} = \frac{Q}{2} \left( \frac{ds}{dt} \right)^2$$

onde  $Q$  é a massa fictícia das variáveis de Nosé-Hoover ou magnitude do acoplamento.

Com estas modificações no hamiltoniano, o sistema é redimensionado por:

$$\gamma = -\frac{1}{\tau_{NH}} \left( \frac{f + 1}{f} \frac{T_{alvo}}{T} - 1 \right)$$

Os dados que precisam ser introduzidos para o uso deste algoritmo são:

$T_{alvo}$ , : a temperatura alvo do sistema.

$Q$ , massa fictícia das variáveis de Nosé-Hoover.

$\tau_{NH}$ , ou constante de amortecimento.

### 3.7.

#### Barostatos

Outra variável que normalmente precisa ser controlada é a pressão. Para manter a pressão constante no decorrer da simulação, uma série de algoritmos, os barostatos, são usados. Não obstante, esta propriedade apresenta flutuações maiores do que a temperatura para o ensemble NVT ou do que na energia para o ensemble NVE. Desta forma, a medida da pressão muda rapidamente, mas a média da pressão é normalmente mantida.

Da mesma forma que no termostato, é necessário no barostato um maior controle sobre a pressão na etapa inicial da simulação, e posteriormente, pode-se diminuir a influência do barostato.

#### 3.7.1.

##### Barostato Berendsen

Seguindo a mesma ideia do termostato de Berendsen, (BERENDSEN et al., 1984) Berendsen também publicou seu barostato. A taxa de flutuação da pressão é comparada com a diferença entre a pressão no tempo  $t$  e a pressão de referência, utilizando  $\tau_p$  como constante do tempo no barostato que ao ser integrada resulta em

$$\frac{dP}{dt} = \frac{1}{\tau_p} (P_{alvo} - P(t))$$

Integrando:

$$\Delta P = \frac{\Delta t}{\tau} (P_{alvo} - P(t))$$

O volume é redimensionado com a constante  $\mu$  da seguinte forma:

$$\mu = \left[ 1 - \frac{dt}{\tau_p} (P - P_{alvo}) \right]^{\frac{1}{3}}$$

Então, as coordenadas de cada partícula são escalonadas por:

$$R_I(t + \Delta t) = \mu^{\frac{1}{3}} R_I(t)$$

Para a utilização deste termostato os seguintes valores devem ser fornecidos:

$P_{alvo}$ , a pressão alvo; e

$\tau_p$ , constante de tempo do barostato.

Também, a pressão pode ser acoplada de forma independente para cada dimensão, provendo também valores de pressão alvo e constante de tempo para cada grupo de dimensões.

### 3.7.2.

#### Barostato Nosé-Hoover

De forma similar ao termostato Nosé-Hoover (HOOVER, 1985; NOSÉ, 1984) criou-se um barostato, onde adiciona-se ao Hamiltoniano um grau de liberdade  $\eta$  correspondente a um pistão. O uso do termostato de Nosé-Hoover é necessário para a utilização deste barostato. Existem várias formas de implementar este barostato, que geralmente redimensiona o volume total do sistema.

Ele requer o conhecimento dos parâmetros:

$P_{alvo}$ , pressão alvo

$T_{alvo}$ , temperatura alvo

$\tau_p$ , constante de tempo do barostato

$\tau_T$ , constante de tempo do termostato

$Q$ , a massa fictícia das variáveis de Nosé-Hoover.

O algoritmo Parrinello- Rahman permite utilizar o barostato de Nosé-Hoover em sistemas não isotrópicos (APOL et al., 2012).

### 3.8.

#### Modelos de água

Utilizar um bom modelo de água é fundamental para obter resultados de alta qualidade. É de grande importância especialmente em bioquímica molecular, química atmosférica e ambiental e em geologia. Portanto, as moléculas de água são uma das poucas modeladas normalmente com mais pontos de interação do que átomos.

Os modelos de água baseados em cálculos *ab initio* ou empíricos sobre moléculas isoladas não apresentam bons resultados na fase condensada. Por exemplo, o momento dipolo da água na fase condensada é 25% maior do que em moléculas isoladas.

Quando não é a água o objeto de pesquisa na dinâmica molecular, os modelos de água mais usados são os modelos rígidos de 3 pontos de carga. O primeiro modelo com estas características foi o modelo BF, não obstante os resultados se afastavam significativamente dos resultados experimentais. Em consequência foi ajustado à densidade e energia experimental, resultando em um modelo de água utilizado comumente em sistemas biológicos e similares. Posteriormente, foi ajustada à difusividade experimental no modelo, tentando sempre utilizar o menor número de pontos de interação e evitando utilizar a polarizabilidade para diminuir o custo computacional associado.

O modelo de ponto de carga simples (*single point charge*, sigla SPC), publicado por Berendsen (BERENDSEN et al., 1981) é usado principalmente com seus campos de força. Já o modelo TIP3P, publicado por Jorgensen (JORGENSEN et al., 1983), é usado principalmente com os campos de força CHARMM e AMBER. No entanto, os campos de força são relativamente independentes do modelo de água, podendo-se escolher modelos diferentes em concordância com os objetivos da pesquisa. Apesar do esforço colocado no desenvolvimento dos modelos de água nenhum deles reproduz fielmente as características da água. Do um ponto de vista prático, a escolha de um modelo de água deve ser baseada tendo-se em mente o custo computacional e a fidelidade do modelo.

O modelo de água (SPC) foi publicado por Berendsen em 1981 (BERENDSEN et al., 1981) com o objetivo de ter um modelo de água com resultados satisfatórios em simulações de Monte Carlo e dinâmica molecular de proteínas solvatadas sem consumir muitos recursos computacionais.

A molécula de água neste modelo é constituída por três partículas, correspondentes a dois hidrogênios e um oxigênio, ligados de forma rígida, ou seja, não está permitido a modificação dos comprimentos e/ou ângulos das ligações. O ângulo entre as ligações é de  $109,5^\circ$  que corresponde ao ângulo ideal de uma estrutura tetraédrica e não ao ângulo medido experimentalmente,  $104,47^\circ$ .

Utiliza-se três cargas, exatamente na mesma posição da massa de cada partícula. Estas cargas são constantes e não é modelada nenhuma polarizabilidade. O fato de as cargas serem postas na posição do centro de

massa traz menor custo computacional por precisar somente de um conjunto de 3 coordenadas.

O modelo SPC utiliza somente as interações de Lennard Jones entre os oxigênios. Por praticidade, usa-se o mesmo valor experimental da expressão de atração, e portanto, somente são ajustados o parâmetro de repulsão e os pontos de carga.

O parâmetro de repulsão e as cargas foram ajustadas ao valor experimental da energia de interação,  $48,1 \text{ KJ mol}^{-1}$  usando simulações de Monte Carlo e dinâmica molecular a uma temperatura de  $300 \text{ K}$  e densidade  $1000 \frac{\text{Kg}}{\text{m}^3}$ . Também, foram obtidos gráficos de função de distribuição radial espacial com formato similar ao experimental. Destas simulações resultou-se um valor da carga do hidrogênio de  $0,41 e$  e um valor de parâmetro de repulsão de  $0,3428 \text{ mm}(\text{KJ mol}^{-1})^{-12}$ . Estes parâmetros resultaram em um modelo de água com baixo custo computacional, e na época de publicação, apresentou os resultados mais próximos aos experimentais entre modelos de água de três pontos para a temperatura e densidade de interesse biológico. No entanto, apresentou valores de constante de difusão ( $3,6 \times 10^{-9} \text{ m}^2 \text{ s}^{-1}$ ) maior do que o experimental ( $2,30 \times 10^{-9} \text{ m}^2 \text{ s}^{-1}$ ).

Outra propriedade importante que se afasta dos resultados experimentais é a energia derivada das interações intermoleculares. Estes potenciais de interação efetivos entre pares são mais importantes quando moléculas polares são modeladas. Já para moléculas pouco polares ou apolares, as interações de Lennard Jones são mais importantes e o modelo de água SPC traz a vantagem de não modificar o parâmetro da atração medido com dados experimentais.

Embora o modelo SPC apresente resultados satisfatórios na maioria dos propósitos, é possível melhorar sem muitas complicações seus resultados de densidade, função de distribuição radial e constante de difusão. Estas três propriedades apresentadas para este modelo foram melhoradas baseando-se na polarizabilidade da molécula de água e nas resultantes cargas efetivas. Isto conduziu a um novo modelo, o modelo estendido de cargas pontuais simples (SPC/E).

### **3.9. Campos de Força Gromos e Gromos53A6**

O campo de força Gromos foi desenvolvido focando-se em matéria condensada para sistemas biológicos. Após a publicação de sua primeira versão (SCOTT et al., 1999) o campo de força Gromos tem sido refinado e

reparametrizado, fato que foi impulsionado pelo grande aumento de pesquisas com este campo de força que tem se tornado em um campo de força muito popular. O grupo desenvolvedor deste campo de força tem publicado várias versões ou parametrizações importantes. Iniciando em 1980 tem-se desenvolvido um programa computacional, chamado Gromos, que possui as atualizações Gromos87, Gromos96, Gromos05 e a mais recente Gromos11. Paralelamente e de forma independente foram publicados uma série de parâmetros, entre os quais, os de maior uso são: Gromos37C4(1985), Gromos43A1(1996), Gromos43A3(2001), Gromos53A5(2004)(OOSTENBRINK et al., 2004), Gromos53A6(2004)(OOSTENBRINK et al., 2004), Gromos54A7(2011)(SCHMID et al., 2011) e Gromos54B7(2011)(SCHMID et al., 2011). Normalmente o uso de um conjunto de parâmetros ocorre alguns anos após a publicação depois de uma série de pesquisas de validação e suas respectivas publicações. Por causa o bom desenvolvimento em sistemas biológicos, eficiência, robustez e simplicidade na parametrização o campo de força Gromos53A6 foi utilizado nesta dissertação.

A filosofia com que se construiu o campo de força e que norteia seu desenvolvimento é a obtenção de um campo de força balanceado entre a obtenção de descritores precisos e operacionalmente simples. Os descritores das interações ligantes podem ser obtidos de dados cristalográficos e espectroscópicos A obtenção dos parâmetros das interações entre átomos não ligados é mais difícil do que as outras. Com relação a outros campos de força (AMBER, CHARMM, OPLS), o campo de força Gromos no conjunto de parâmetros Gromos53A5 e Gromos53A6 utiliza a energia de solvatação para ajustar os dados de grupos funcionais neutros de aminoácidos e carbonos alifáticos. O conjunto de parâmetros Gromos53A6 usa a entalpia livre de hidratação de compostos polares. Ele também dá maior atenção aos solventes mais usados comumente (água, clorofórmio, metanol, dimetilsulfóxido e tetracloreto de carbono). Este enfoque e esta metodologia resultaram em um campo de força muito confiável na pesquisa de importantes processos biomoleculares como o enovelamento de proteínas, a associação e reconhecimento de biomoléculas, a formação de micelas e membranas e o transporte através destas últimas. Assim o campo de força pode ser usado tanto em pesquisa com proteínas e/ou membranas, resultando em dados altamente confiáveis em temperatura e pressões fisiológicas.

### 3.9.1. Interações

No campo de força Gromos, o Hamiltoniano que descreve um sistema está definido pelo momento e posição em conjunto com a massa e as interações das partículas. Este Hamiltoniano pode ser dividido em duas partes independentes uma da outra, e pode ser ilustrado na seguinte equação:

$$H(\mathbf{p}, \mathbf{r}; m, s) = K(\mathbf{p}; m) + V(\mathbf{r}; s)$$

Onde o Hamiltoniano  $H$  é função do momento  $p$ , coordenada  $r$ , massa  $m$  e o conjunto de interações  $s$ , e dividido em energia cinética  $K$ , função do momento  $p$  e da massa  $m$ , independente das interações  $s$  e das coordenadas  $r$ , e na energia potencial  $V$ , função das coordenadas  $r$  e do conjunto de interações  $s$ , independente do momento  $p$  e da massa  $m$ .

A energia cinética pode ser descrita pela velocidade como mostra a seguinte equação:

$$K(\mathbf{p}; m) = \sum_{i=1}^N \frac{p_i^2}{2m_i} = \sum_{i=1}^N \frac{1}{2} m_i v_i^2$$

Utilizando uma série de massas predefinidas em concordância com o tipo de átomo, limitando assim os tipos de átomos listados nos parâmetros, as massas podem ser modificadas ou pode-se adicionar novos tipos de átomos. No entanto, devido as validações e comparações com os experimentos, estas alterações ou modificações não são práticas recomendadas.

A força é calculada através da derivada da energia potencial respeito a sua coordenada:

$$f_i = -\frac{\partial}{\partial r_i} V(r_1, r_2, \dots, r_N; s)$$

Inicialmente a energia potencial é separada em duas categorias, energia potencial física, derivada dos parâmetros do campo de força, e energia potencial especial, usada em determinadas situações com objetivos mais específicos como no uso de restrição de posição ou outros métodos geralmente usados para manter uma geometria ou subestrutura específica. Esta divisão resulta em:

$$V(\mathbf{r}; s) = V^{fis}(\mathbf{r}; s) + V^{esp}(\mathbf{r}; s)$$

Adicionalmente a energia potencial física é dividida de forma clássica, em energia potencial derivada de interações relacionadas as ligações covalentes e energia potencial derivada de interações relacionadas as interações que não incluem ligações covalentes, resultando na equação:

$$V^{fis}(\mathbf{r}; s) = V^{lig}(\mathbf{r}; s) + V^{nlig}(\mathbf{r}; s)$$

### 3.9.2. Interações entre átomos ligados

As interações entre átomos ligados são descritas pelo comprimento de ligação, ângulo de ligação, torção do ângulo diedro e do ângulo diedro impróprio. Estas interações são definidas da seguinte forma:

$$V^{lig}(\mathbf{r}; s) = V^{lig}(\mathbf{r}; K_b, b_0) = \sum_{n=1}^{N_b} \frac{1}{4} K_{b_n} [b_n^2 - b_{0_n}^2]^2$$

$$V^{ang}(\mathbf{r}; s) = V^{ang}(\mathbf{r}; K_\theta, \theta_0) = \sum_{n=1}^{N_\theta} \frac{1}{2} K_{\theta_n} [\cos \theta_n - \cos \theta_{0_n}]^2$$

$$V^{prop}(\mathbf{r}; s) = V^{prop}(\mathbf{r}; K_\varphi, \delta, m) = \sum_{n=1}^{N_\varphi} K_{\varphi_n} [1 + \cos(\delta_n) \cos(m_n \varphi_n)]$$

$$V^{improp}(\mathbf{r}; s) = V^{improp}(\mathbf{r}; K_\xi, \xi_0) = \sum_{n=1}^{N_\xi} \frac{1}{2} K_{\xi_n} [\xi_n - \xi_{0_n}]^2$$

É recomendado usar somente um conjunto de átomos para cada diedro, não usar ângulos em anéis planares.

### 3.9.3. Interações entre átomos não ligados

O campo de força Gromos separa as interações entre átomos não ligados em interações de van der Waals e interações eletrostáticas:

$$V^{nlig}(\mathbf{r}; s) = V^{LJ}(\mathbf{r}; s) + V^{Columb}(\mathbf{r}; s)$$

omitindo todas as interações entre átomos não ligados, entre pares de átomos com ligações covalentes e vizinhos de segunda ordem. Ele também omite as interações de átomos de vizinhança 1-3 e 1-4 que são partes ou estão ligadas diretamente a anéis aromáticos. Além disso, divide as interações entre átomos não ligados em três categorias: as de curto alcance a uma distancia  $R_p$ , são listadas a cada  $N_l$  passos; as interações resultantes de membros da lista que são calculadas em cada passo; e as interações de partículas com uma distância entre  $R_p$  e  $R_l$  que são calculadas com a construção da lista de curto alcance. Os resultados são mantidos constantes durante  $N_l$  passos. Os pares com comprimento maior a  $R_l$  são introduzidos como contribuições às constantes dielétricas.

### 3.9.3.1. Interações de van der Waals

As interações de van der Waals são calculadas com o potencial de Lennard-Jones 12-6:

$$V^{LJ}(\mathbf{r}; s) = V^{LJ}(\mathbf{r}; C12, C6) = \sum_{par\ i,j} \left( \frac{C12_{ij}}{r_{ij}^{12}} - \frac{C6_{ij}}{r_{ij}^6} \right)$$

Os parâmetros dos pares de tipos de átomos são determinados por médias geométricas dos parâmetros individuais:

$$C12_{ij} = \sqrt{C12_{ij} \cdot C12_{ij}}$$

$$C6_{ij} = \sqrt{C6_{ij} \cdot C6_{ij}}$$

Para os parâmetros de repulsão  $C12$  se definem três tipos,  $C12(I)$ ,  $C12(II)$  e  $C12(III)$ .  $C12(I)$  é o parâmetro padrão,  $C12(II)$  é utilizado quando existe possibilidade de ligação de hidrogênio. E  $C12(III)$  é usado com pares com pelo menos um membro de grupos totalmente carregados. Ocasionalmente, é necessário o uso de parâmetros específicos em pares com interações de ligações covalente 1-3 e 1-4.

### 3.9.3.2. Interações Eletrostáticas

As interações eletrostáticas são definidas como:

$$V^C(\mathbf{n}; s) = V^C(\mathbf{r}; q) = \sum_{par\ i,j} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1 r_{ij}}$$

Interações entre pares com distância de afastamento maior que  $R_l$  podem ser definidas com distintos modelos dependendo do nível de precisão e do custo computacional requisitado.

É importante notar que o campo de força Gromos não distingue entre tipos de ligação e se limita a obter a energia potencial a partir da geometria e às funções descritoras das interações de ligação. Assim, a única diferença entre uma ligação simples, dupla, tripla, aromática, etc. são os parâmetros que as descrevem. Também é desconhecido a existência de ligações  $\pi$ -cátion, interações  $\pi - \pi$ , etc. Para descrever as ligações de hidrogênio e ligações iônicas utiliza-se o artifício de mudar o valor da constante de repulsão das interações de van der Waals. Também não existe transferência de elétrons ou modificações da carga parcial das partículas, e não existe rompimento ou

criação de ligações covalentes nem modificações dos parâmetros de ligação durante o tempo de simulação.

### 3.10. PME

Um dos maiores inconvenientes na dinâmica molecular é o tratamento das interações de longo alcance. A solução típica de eliminar as interações até uma distância especificada pode ser suficiente no caso das interações de van der Waals. No caso das interações eletrostática terem grande influência, o resultado se afasta muito da realidade ao aplicar esta solução na dinâmica molecular, especialmente quando existe muitos momentos dipolos e se usa condições periódicas de contorno. Para as dinâmicas moleculares representarem melhor sistemas reais é normalmente usado o método de Ewald.

O método de Ewald ou somatória de Ewald foi desenvolvido por Paul Peter Ewald.(ESSMANN et al., 1995; EWALD, 1921). Ele desenvolveu um método para computar as interações de dipolo em células unitárias de cristais. A forma simples de calcular a energia resultante das imagens especulares da partícula é feita utilizando a seguinte equação:

$$E = \frac{f}{2} \sum_{n_x} \sum_{n_y} \sum_{n_z} \sum_i^N \sum_j^N \frac{q_i q_j}{r_{ij,n}}, (n_x, n_y, n_z) = n \neq 0$$

onde  $n$  é o número de imagens da célula unitária para cada coordenada. O uso desta função não é prático na dinâmica molecular porque converge muito lentamente. Além disso, seria necessário a utilização de um grande número de imagens da células unitárias. Para fazer esta função Ewald convergir mais rápido divide-se a função em três partes independentes: uma de curto alcance, onde é calculada de forma direta, outra onde se presume uma distribuição gaussiana sobre a energia, utilizando uma transformada de Fourier, e uma terceira que é adicionada para manter a neutralidade de carga da célula unitária.

$$E_{tot} = E_{curto} + E_{longe} + E_{propria}$$

$$E_{curto} = \frac{f}{2} \sum_{i,j}^N \sum_{n_x} \sum_{n_y} \sum_{n_z} \frac{q_i q_j}{r_{ij,n}} \operatorname{erfc}(\beta r_{ij,n})$$

$$E_{longe} = \frac{f}{2\pi V} \sum_{i,j}^N q_i q_j \sum_{m_x} \sum_{m_y} \sum_{m_z} \frac{\exp(-\left(\frac{\pi m}{\beta}\right)^2 + 2\pi i m \cdot (r_i - r_j))}{m^2}$$

$$E_{propria} = -\frac{f\beta}{\sqrt{\pi}} \sum_i^N q_i^2$$

Onde  $\beta$  é o peso do parâmetro direto e indireto da soma,  $V$  é o volume da célula unitária,  $m$  é o número de células unitárias, e  $q$  é a carga da partícula  $i$ .

Este somatório permite utilizar um menor número de células unitárias, no entanto não é facilmente utilizável em cálculos em paralelo. Para este fim divide-se o termo de longe alcance em uma série de setores e interpolam-se os *splines*. Assim, em lugar de somar cada termo somam-se os *splines* que podem ser computados em diferentes processadores, resultando na otimização do somatório de Edwald por paralelização sem obter muitas incertezas. Não obstante, deve-se ter o cuidado de não usar *splines* muito pequenos que resultam na omissão de interações entre partículas pouco separadas.

### 3.11. LINCS

Alguns periodos de vibração de ligações são menores do que os passos de tempo utilizados na simulação. Quando a posição atualizada das partículas afasta-se muito do comprimento ideal de ligação, ocorre geração de energia que acaba esquentando o sistema. Este fenômeno é mais comum em partículas de pouca massa pois o momento médio para o sistema resulta em um grande movimento das partículas.

A solução trivial para este problema é utilizar intervalos de tempo menores, mas isto aumenta significativamente o custo computacional. A solução é restabelecer o comprimento ideal da ligação tendo o cuidado de não modificar a direção dos novos vetores de forças nem de velocidades. Com estes objetivos, usa-se o algoritmo solucionador de restrição linear LINCS (LINEar Constraint Solver)(HESS et al., 1997). O enfoque principal no seu desenvolvimento é melhorar a paralelização quando é implementado. Além disso, pode ser usado tanto com o algoritmo Leap-Frog como com o algoritmo de Verlet.

O método LINCS restabelece o comprimento de ligação em cada novo cálculo de posição e corrige a modificação na rotação que resulta do restabelecimento do comprimento de ligação. O método introduz ao potencial  $V(r)$  para uma restrição  $k$ , um multiplicador de Langrange  $\lambda_k(t)$  com uma função de restrição  $g_k = (r_a - r_b)^2 - d_k^2$ , onde  $r_a$  e  $r_b$  são as posições das partículas participantes da restrição e  $d_k$  é a distância ideal da ligação. As forças de cada partícula resultam em

$$-m_i \frac{dr_i^2}{dt^2} = \frac{d}{dr} (V(r_i(t)) - k(t) \cdot g_k)$$

O uso de um multiplicador de Lagrange tem a vantagem de não modificar a energia total do sistema.

Em consequência da restrição, a posição não corrigida  $r_i$  no tempo  $t + \Delta t$  é usada para calcular a posição corrigida  $r_i^*(t + \Delta t)$  com a equação

$$r_i^*(t + \Delta t) = r_i(t + \Delta t) + \sum_{k=1}^n \lambda_k \frac{\partial g_k(t)}{\partial r_i} (\Delta t)^2 m_i^{-1}$$

Para resolver a nova posição de  $N$  partículas com  $n$  restrições é comum usar análise numérica. Utiliza-se neste caso o método de Newton–Raphson nas simulações, construindo matrizes independentes para cada molécula. O multiplicador de Lagrange  $\lambda_k(t)$  é resolvido com uma matriz  $J_g$  para cada molécula, assim o multiplicador de Lagrange  $\lambda_k(t)$ , para as partículas  $i$  e  $j$  com uma molécula de  $n$  restrições resulta em

$$\lambda_k(t) = -J_g^{-1} \left[ \left( r_i(t + \Delta t) - r_j(t + \Delta t) \right)^2 - d_k^2 \right]$$

$$J_g = \begin{bmatrix} \frac{dg_1}{d\lambda_1} & \dots & \frac{dg_1}{d\lambda_n} \\ \vdots & \ddots & \vdots \\ \frac{dg_n}{d\lambda_1} & \dots & \frac{dg_n}{d\lambda_n} \end{bmatrix}$$

É evidente aqui que o erro aumenta com o tamanho da molécula e o número de restrições nela. Antigamente usava-se um método iterativo para calcular as novas coordenadas, mas LINCS utiliza uma expansão em série do inverso da matriz jacobiana  $J_g$

$$(I - J_g)^{-1} = I + J_g + J_g^2 \dots$$

Assim não é necessário calcular o inverso da matriz, somente a série. Cada elemento implica em um nível de acoplamento entre as restrições sendo a série truncada até uma precisão aceitável, resultando em um método mais eficiente. Mas é implementável só em moléculas muito rígidas como modelos de água rígida e anéis aromáticos e outras moléculas de pouco tamanho.

### 3.12. Gromacs

Com base no programa de computador Gromos desenvolvido por van Gunsteren e Berendsen nos 80's (VAN GUNSTEREN; BERENDSEN, 1987), o grupo de Berendsen na Universidade de Groningen na Holanda modificou o código que deixou de ser desenvolvido em Fortran para ser desenvolvido em C++(BERENDSEN; VAN DER SPOEL; VAN DRUNEN, 1995), chamado de Gromacs. O objetivo foi distribuir um pacote de programa de computador para o

uso em dinâmica molecular, sendo versátil e rápido, especialmente planejado para sistemas biológicos em ambientes aquosos ou membranas.

Atualmente, o Gromacs é mantido por uma comunidade de pesquisadores espalhados por todo o mundo. Até a versão 4.5 era distribuído com a licença GNU-GPL e a partir da versão 4.6 passou a ser distribuído com a licença GNU-LGPL. Basicamente a diferença entre as licenças é que os usuários tem total liberdade de modificação e distribuição na versão GNU-GPL, não obstante a licença GNU-LGPL não dá a liberdade de modificar alguns pacotes, mas sim de distribuir.

No início da pesquisa a versão estável mais recente era a versão 4.5.4. Atualmente, a versão estável mais recente é a versão 4.6.5, e a versão beta aberta mais recente é a versão 5.0. Além da diferença de licenças, a diferença entre a versão 4.5.4 e a versão 4.6.5 é a compatibilidade com o uso de unidades de processamento gráfico além das unidades de processamento central, tratando de aproveitar a velocidade com que as unidades de processamento gráfico efetuam alguns métodos numéricos específicos. Mudanças menores foram feitas em alguns pacotes e adicionaram-se alguns novos aplicativos.

O pacote de programas de computador GROMACS não tem campo de força próprio, mas pelo seu objetivo de ser versátil é compatível com vários campos de força (GROMOS, OPLS, AMBER...). É atualmente o pacote de programas de computador para o uso de dinâmica molecular clássica mais rápida. Também funciona para dinâmica molecular híbrida com o auxílio algum *software* de Química Computacional que possui metodologia *ab initio*. A distribuição com licença GNU-GPL fez com que uma grande comunidade vem mantendo e modificando o pacote. Desta forma, o desenvolvimento do GROMACS tem sido rápido e robusto, e as novas implementações são adicionadas em concordância com as necessidades da comunidade científica que usam o pacote.