

## 2. Técnicas Computacionais – Aproximação Clássica e Aproximação Quântica

O desenvolvimento deste capítulo nos itens 2.1-2.4 e 2.6-2.9 está baseado no manual do GROMACS (van der Spoel et al., 2005) e no trabalho de Baaden M. (Baanden M., 2003). Quando as informações apresentadas sejam tomadas de outras fontes, se fará a respectiva referência.

O comportamento da matéria em nível de partículas constituintes dos átomos e moléculas é descrito pela Mecânica Quântica, enquanto que em níveis mais complexos a matéria composta de átomos e moléculas pode ser descrita de forma aproximada pela Mecânica Clássica não relativística.

Uma aproximação clássica perde detalhes do sistema molecular em estudo, mas permite acessar propriedades macromoleculares do sistema, assim como economizar tempo de cálculo.

Os pacotes de Dinâmica Molecular baseados numa aproximação clássica estão fundamentados em duas aproximações:

1. As leis de Newton são aplicáveis.
2. O movimento dos elétrons é muito mais rápido que o movimento dos núcleos (Aproximação de Born-Oppenheimer).

Nossos modelos utilizados estão representados por sistemas massa-mola onde cada átomo é representado por sua massa, um raio e uma carga atômica e as ligações entre os átomos são representadas por molas (van der Spoel et al., 2005). As características do sistema massa-mola estão contidas no campo de força. As interações físicas entre os átomos são levadas em conta e introduzidas no termo de energia potencial. A partir do modelo físico definido pelos parâmetros do campo de forças e a representação da energia potencial, nós vamos utilizar as leis de Newton para estudar a evolução das moléculas definida pelos vetores de posição e pelos vetores de quantidade de movimento. Desta forma geraremos ensambles

estatísticos de configurações de átomos e moléculas em diversas fases (Berendsen et al., 1996).

Para se ter uma média temporal estatisticamente significativa de um sistema qualquer com a DM, seriam necessárias acessar todas as possíveis configurações do sistema e estabelecer a probabilidade de cada estado conformacional. Há que se ter em mente, porém, que neste trabalho estamos aplicando a DM sobre um sistema envolvendo uma única macromolécula e por um intervalo de tempo de poucas dezenas de nanosegundos. Embora não se possa falar neste caso de ensemble estatístico, a simulação da dinâmica de proteínas na conformação biologicamente ativa, nesta escala de tempo, é útil para obter informações sobre a estabilidade da estrutura, transições conformacionais locais e várias propriedades de interações com ligantes. Considerando que proteínas expressam sua atividade biológica em estados conformacionais próximos ao mínimo global em um funil de energia livre (Onuchic et al., 2004), espera-se que a probabilidade desses estados bio-ativos sejam bastante alta em relação ao número astronômico de possíveis conformações desta macromolécula. Mas devemos mencionar que variações na estrutura também estão relacionadas com a atividade (enzimas).

Em nossos cálculos vamos utilizar o pacote GROMACS (van der Spoel et al., 2005). Os detalhes referentes a representação de energia e parâmetros da simulação são próprios deste campo de forças.

Neste capítulo daremos a descrição clássica e explicaremos o hamiltoniano do sistema, a energia cinética, a energia potencial e a representação das forças, assim como a aplicação de restrições de distâncias, tratamentos de contorno, etc. Também veremos os métodos para a Minimização de Energia (ME), Dinâmica Molecular, tratamento eletrostático, superfícies de interação molecular, docking molecular e a geração de matrizes de correlação.

No final do capítulo veremos a aproximação quântica no cálculo de cargas através do método DFT e o formalismo RESP.

## 2.1

### **Aproximação Clássica - Função de interação e equações de movimento**

Referimos-nos a cálculos clássicos como aproximações onde são válidas as leis de Newton. Para descrever corretamente os sistemas devemos estabelecer os

graus de liberdade que serão explicitamente tratados e aqueles que serão implicitamente tratados e incorporados à função de interação.

Dentro de nosso esquema de cálculo estamos interessados nos graus de liberdade de posição das coordenadas dos átomos.

Num sistema clássico o hamiltoniano de um sistema molecular é descrito da seguinte forma:

$$H(p, r, m, g) = V(r, g) + K(p, m) \quad (1)$$

onde  $V$  é a energia potencial,  $K$  a energia cinética,  $p$  é o momento das partículas,  $r$  a posição,  $m$  a massa, e  $g$  são parâmetros presentes no campo de forças. A energia cinética está representada da seguinte forma:

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

e depende das massas e velocidades, e é independente das coordenadas  $r$  da partícula.  $V(r;g)$  corresponde a energia potencial que descreve a interação em função das coordenadas das partículas.

$$V(r; x) = V(r_1, r_2, \dots, r_N; g_1, g_2, \dots, g_N) \quad (3)$$

Em geral  $V(r;g)$  depende dos parâmetros do campo de forças indicados por  $g = (g_1, g_2, \dots, g_N)$ , e é a soma de termos que representam as distintas interações no sistema.

Uma partícula nesse sistema está submetida a uma força  $f_i$  que tem a seguinte representação:

$$f_i = -\frac{\partial}{\partial r_i} V(r_1, r_2, \dots, r_N) \quad (4)$$

Devemos lembrar que as trajetórias dependem unicamente das forças que agem sobre os átomos, e não explicitamente das energias.

Usamos as equações de Newton para descrever a evolução da posição na seguinte forma:

$$\frac{dr_i(t)}{dt} = v_i(t) \quad (5)$$

$$\frac{dv_i(t)}{dt} = \frac{f_i}{m_i} \quad (6)$$

Na simulação estas equações são integradas numericamente no tempo.

Inclui-se um termo suplementar  $V_{\text{restr}}(r)$ , que é uma função de penalidade (van der Spoel et al., 2003), dentro da função de interação do sistema que limita o movimento das partículas de tal forma que o valor simulado da partícula se aproxima ao valor desejado.

## 2.2

### O Hamiltoniano e a Função de Interação: Campo de Forças

De uma forma geral a energia potencial no campo de forças pode ser escrito da seguinte forma (van der Spoel et al., 2005: item 4.1, 4.2 e 4.3):

$$V(r; g) = V^{\text{bonded}}(r; g) + V^{\text{no bonded}}(r; g) + V^{\text{especiais}}(r; g) \quad (7)$$

Onde  $V^{\text{bonded}}$  corresponde aos termos chamados covalentes,  $V^{\text{no bonded}}$  aos termos chamados não ligados (forças eletrostáticas e de van der Waals) e  $V^{\text{especiais}}$  aos termos especiais como, por exemplo, as restrições de posição e vínculos em comprimentos e ângulos de ligação. Assim o Hamiltoniano da equação 1 queda descrito por:

$$H(p, r, m, g) = V^{\text{bonded}}(r; g) + V^{\text{no bonded}}(r; g) + V^{\text{especiais}}(r; g) + K(p, m)$$

#### 2.2.1

##### Interações entre átomos (ou grupos) ligados

Na expressão para  $V^{\text{bonded}}(r; g)$

$$V^{\text{bonded}} = V^{\text{bond}}(r; g) + V^{\text{ângulo}}(r; g) + V^{\text{diedro}}(r; g) + V^{\text{improp}}(r; g) \quad (8)$$

Esta expressão descreve a variação do comprimento das ligações, a deformação dos ângulos e as torções devido aos ângulos diedros próprios e impróprios.

##### Varição do comprimento das ligações covalentes entre átomos

A interação entre termos ligados é expressa por um potencial de tipo harmônico, com uma soma feita sobre todas as ligações covalentes  $n=1, \dots, N_b$ .

$$V_{\text{bond}} = \sum_{n=1}^{N_b} K_{b_n}^{\text{harmônico}} (b_n - b_{o_n})^2$$

No GROMACS é utilizada por razões de eficiência computacional a expressão para  $V^{\text{bond}}$

$$V_{bond} = \sum_{n=1}^{N_b} \frac{K_{b_n} (b_n^2 - b_{o_n}^2)^2}{4} \quad (9)$$

As unidades para  $b_n - b_{o_n}$  e para  $K_{b_n}$  são nm e  $\text{kJmol}^{-1}\text{nm}^{-4}$ , respectivamente.

Onde:

$$2K_{b_n} = \frac{K_{b_n}^{harmonico} b_{o_n}^{-2}}{2}$$

### Deformação dos ângulos nos átomos ligados

A seguinte relação é utilizada no GROMACS para representar o termo de deformação entre átomos ligados por uma ligação covalente (van der Spoel et al., 2005)

$$V_{angulo} = \sum_{n=1}^{N_\theta} \frac{K_{\theta_n} (\cos \theta_n - \cos \theta_{o_n})^2}{2} \quad (10)$$

Os ângulos são dados em graus e  $K_{\theta_n}$  em  $\text{kJmol}^{-1} \cdot E_{\text{KBT}}$ , a soma envolve os ângulos desde  $n=1, \dots, N_\theta$  do sistema molecular.

$E_{\text{KBT}} = 0.5961575 \text{Kcal mol}^{-1}$  é a constante de Boltzmann a 300 K.

### Termos de torção, ângulos diedros e impróprios

No campo de força GROMACS a função de interação que representa os termos harmônicos impróprios da deformação dos ângulos diedros (fora do plano, fora da configuração tetraédrica) é descrita por:

$$V^{improp} (r; g) = \sum_{n=1}^{N_\xi} K_{\xi_o} \frac{(\xi_n - \xi_{o_n})^2}{2} \quad (11)$$

A soma compreende  $n=1, \dots, N_\xi$  definidos para manter os grupos de átomos vizinhos dentro de uma configuração espacial determinada, por exemplo para manter os grupos de átomos em um plano como no caso de um anel aromático. Em suma, este potencial simula a hibridação de orbitais  $sp^2$  ou  $sp^3$ .

Os termos de torção de ângulos diedros têm a forma (Lindahl et al. 2003):

$$V^{diedro} (r; g) = \sum_{n=1}^{N_\varphi} K_{\varphi_n} [1 + \cos(\delta_n) \cos(m_n \varphi_n)] \quad (12)$$

onde a constante de força  $K_{\varphi_n}$ , o ângulo diedro  $\varphi_n$  e os valores de  $\delta_n$  e  $m_n$  estão limitados aos valores  $\delta_n=0,\pi$  e  $m_n$  é um inteiro positivo diferente de zero, por conseguinte  $\cos(\delta_n)=\pm 1$ . A soma é feita sobre todos os ângulos diedros  $n = 1, \dots, N_\varphi$ , selecionados de acordo a topologia da molécula.

## 2.2.2

### Interações não ligadas

As interações entre átomos não ligados são descritos por dois termos,  $V^{LJ}$  para as interações de Lenard-Jones e  $V^{Coul}$  para as interações eletrostáticas.

$$V^{no\ bond}(r; g) = \sum_{\text{pares não ligados}} \{V^{LJ}[r_{ij}, C_{12}(i, j), C_6(i, j)] + V^{Coul}[r_{ij}, q_i, q_j, \epsilon_o, \epsilon_1, C_{rf}, R_{rf}]\} \quad (13)$$

Esta soma é estendida a todos os pares de átomos (i,j) salvo as exceções citadas adiante (2.3.1).

### Interações de Lenard-Jones

Estas interações são chamadas também de van der Waals. Dentro do campo de forças GROMOS96 se utilizam os parâmetros  $C_6$  em  $\text{kJmol}^{-1}\text{nm}^6$  e  $C_{12}$  em  $\text{kJmol}^{-1}\text{nm}^{12}$  para a interação entre os átomos i e j a uma distância  $r_{ij}(\text{nm})$ .

$$V^{LJ} = \left[ \frac{C_{12}(i, j)}{(r_{ij})^{12}} - \frac{C_6(i, j)}{(r_{ij})^6} \right] \quad (14)$$

O primeiro termo em  $(r_{ij})^{12}$  corresponde à repulsão entre dois átomos, de tal forma que haja certo grau de superposição de suas nuvens eletrônicas, e cuja origem esta na repulsão entre os núcleos atômicos e no principio de exclusão de Pauli. O segundo termo  $(r_{ij})^6$  corresponde à dispersão atrativa de London, a combinação destes dois fatores dá o perfil de energia de interação mostrada na Fig. 2.1.

Estas equações levam em consideração as regras para determinar os parâmetros mistos entre dois átomos  $i$  e  $j$  de tipos diferentes.

$$C_{12}(i, j) = \sqrt{C_{12}(i, i) \cdot C_{12}(j, j)} \quad (15)$$

$$C_6(i, j) = \sqrt{C_6(i, i) \cdot C_6(j, j)}$$

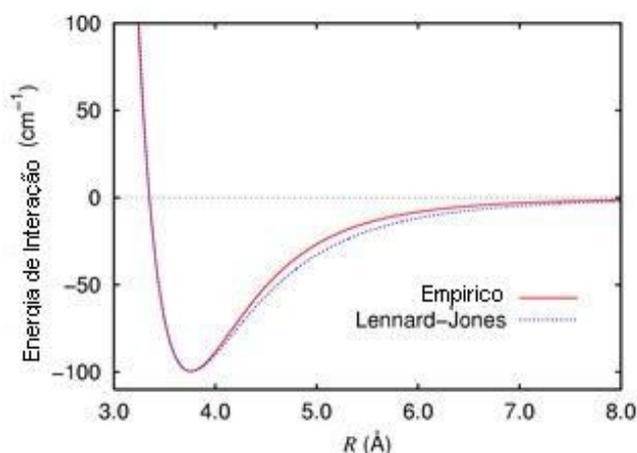


Figura 2.1 Potencial de Lennard-Jones

### Interações eletrostáticas

O primeiro termo corresponde às interações eletrostáticas coulombianas e os outros dois termos correspondem ao campo de reação, o qual simula o efeito de blindagem causada pelo solvente além do raio de corte das interações coulombianas.

$$V^{Coul\_RF} = \frac{q_i q_j}{4\pi\epsilon_o\epsilon_1} \left[ \frac{1}{r_{ij}} - \frac{C_{rf}(r_{ij})^2}{2R_{rf}^3} - \frac{1 - C_{rf}/2}{R_{rf}} \right] \quad (16)$$

O campo de reação no GROMACS é de tipo Poisson-Boltzmann que leva em conta os efeitos de força iônica. Neste modelo o exterior de uma esfera de raio  $R_{rf}$  ao redor de um soluto é modelado por um meio homogêneo polarizável.

### 2.2.3 Potenciais especiais devido a restrições

Existem potenciais “especiais” devido a restrições que são impostas sobre o sistema para se ter resultados próximos à realidade, estes podem ser usados em casos particulares. Por exemplo, restrições sobre a distância de ligação e ângulos, ou restrições sobre a posição.

Vemos o caso de restrição de posição.

$$V^{pr}(r; \mathbf{g}) = \sum_{n=1}^{N_{pr}} \frac{1}{2} K_n^{pr} (r_n - r_n^0)^2 \quad (18)$$

Onde  $pr$  é referido a uma restrição de posição (position restraints), a soma é estendida a todos os  $N_{pr}$  átomos selecionados,  $K_n^{pr}$  é a constante de força do oscilador harmônico e  $r_n^0$  é a posição de referência. As outras expressões dos termos para outros casos de forças especiais podem ser vistas na referência van Gunsteren et al., 1996.

## 2.3

### Tratamento das interações entre átomos não ligados, aproximações.

#### 2.3.1

##### Exclusão ou modificação de interações

Chamaremos de primeiros vizinhos (posições 1-2, Fig. 2.3) aos átomos  $i$  e  $j$  ligados por uma ligação covalente e de segundos vizinhos aos átomos  $i$  e  $k$ , se  $k$  está ligado a  $j$  (posições 1-3, Fig. 2.3).

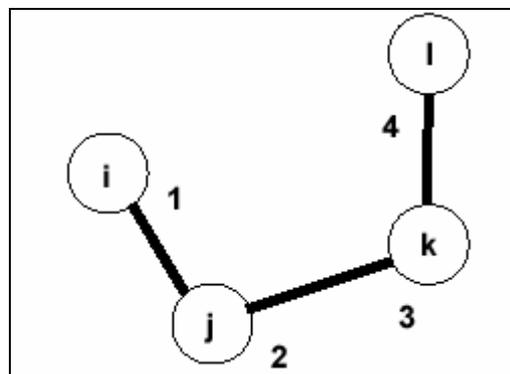


Figura 2.3. Interações entre primeiros vizinhos (1-2), segundos vizinhos (1-3) e terceiros vizinhos (1-4)

As interações entre primeiros e segundos vizinhos não serão tomadas em conta dentro da soma da equação 13, porque estas interações são do tipo harmônico e não são representados por um potencial de tipo Lenard-Jones. Essas interações serão levadas em conta na representação dos átomos ligados (2.2.1). As interações 1-4 entre terceiros vizinhos podem ser excluídas dentro de certos casos, por exemplo, átomos dentro de um anel aromático. Mas, com maior frequência,

esta interação é tratada reduzindo o diâmetro das esferas de van der Waals dos átomos envolvidos, para evitar colisões estéricas.

### 2.3.2 Distância de corte de interações não ligadas

A soma da equação 13 se estende a todos os pares de átomos exceto aqueles excluídos pelas considerações do parágrafo precedente. O número de interações aumenta com o tamanho do sistema estudado.

Grande parte do tempo de cálculo se consome em procurar os pares de átomos e avaliar suas interações. Com o objetivo de limitar o tempo computacional nós utilizamos uma dupla distância de corte (twin cutoff) além da qual as interações “não ligadas” (eletrostáticas e de van der Waals) são substituídas pelo campo de reação. No método ilustrado na Fig. 2.4, a interação dos átomos  $j$  com o átomo  $i$  no interior da zona I, limitada pelo raio  $R_c^1$ , é calculada a partir das posições instantâneas de  $j$  para cada passo da dinâmica. Se  $j$  se encontra entre  $R_c^1 < r_{ij} < R_c^2$ , suas interações com o átomo  $i$  são calculadas a partir de uma posição fixa promedio durante um número de passos  $N_c$ .

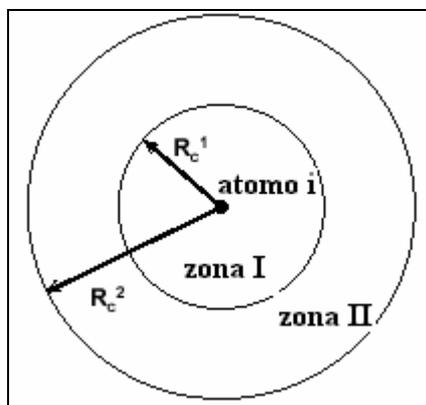


Figura 2.4 Método da dupla distância de corte.

Em geral, o método é aplicado a grupos de átomos e não a átomos isolados. Neste caso o espaço limitado pelo cut-off é igual à união das esferas ao redor de cada átomo do grupo. No GROMACS estes são chamados “grupos de cargas”, se tratando de átomos onde a soma das cargas é neutra (ou inteira, por causa dos íons).

### 2.3.3

#### Interações eletrostáticas - limites sobre as interações eletrostáticas

As interações coulombianas são inversamente proporcionais à distância  $r_{ij}$  entre os átomos  $i$  e  $j$ . Em consequência, estas interações são de longo alcance e não podem ser omitidas para as cargas atômicas parciais da ordem de 0.1 a 0.4 e até a distâncias de corte típicas de 8 a 10 Å. Dois métodos permitem reduzir os artefatos devidos ao truncamento de interações não ligadas. Primeiramente, não se consideram as componentes de alta frequência da interação coulombiana na região  $R_c^1 < r_{ij} < R_c^2$ , como descrito anteriormente. Em segundo lugar, a definição de “grupos de cargas” (átomos de H podem se unir a átomos de C e formar um “átomo” CH com uma carga que seria a soma das cargas componentes) dentro do GROMACS nos permite reduzir certas interações a tipo  $1/r^3$ .

## 2.4

### Tratamento das fronteiras

Em toda simulação deve-se levar em conta o tratamento das fronteiras. No caso de um líquido ou uma solução, os efeitos de fronteira devem ser reduzidos com a aplicação de condições periódicas de contorno.

#### 2.4.1

##### Condições periódicas

Para reduzir os efeitos de fronteira de um sistema finito, nós utilizamos as condições periódicas de contorno. Os átomos de um sistema são introduzidos dentro de uma caixa cúbica rodeada de imagens idênticas equidistantes  $R_{caixa}$  como se mostra na Fig. 2.5.

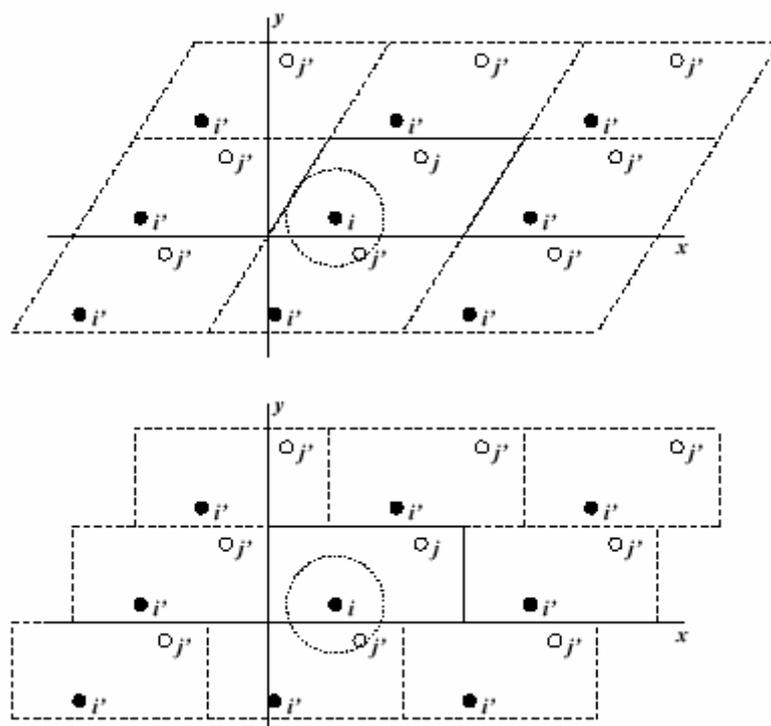


Figura 2.5 Condições periódicas em duas dimensões, mostrando a vista superior e frontal (acima e embaixo respectivamente)

O cálculo das forças sobre o átomo representado pela esfera cheia dentro da caixa central se efetua a partir das contribuições de outros átomos desta caixa e das caixas vizinhas, imagens da caixa central, que se encontram no interior da distância de corte  $R_c$ . Quando o átomo abandona a caixa por uma face, ele entra novamente pela face oposta, transladando a distância  $R_{caixa}$ , com a mesma velocidade. A aplicação das condições periódicas de fronteira lembra assim a ordem dentro de um cristal. Desta forma não há fronteiras no sistema. O artefato causado por condições indesejadas de um sistema isolado e o efeito de superfície são substituídos por condições periódicas.

Para evitar que uma molécula possa interagir com uma ou mais de suas imagens, a menor dimensão da caixa deve exceder em duas vezes a distância de corte:

$$R_{caixa} > 2R_c \quad (19)$$

Quando uma macromolécula é estudada em solução esta restrição não é suficiente. Em princípio, uma única molécula de solvente não deve interagir com ambos os lados da macromolécula. Isto significa que o comprimento de um determinado lado da caixa deve exceder o comprimento da macromolécula na

mesma direção mais duas vezes o raio de corte  $R_c$ . Uma estratégia comum é fazer a camada de solvatação o quanto menor para reduzir o custo computacional, respeitando-se o estabelecido acima.

## **2.5 Geração da topologia, coordenadas atômicas e adição do solvente**

### **2.5.1 Topologia do Sistema**

Para que os programas de simulação reconheçam as moléculas, nós devemos fornecer informações sobre as ligações entre átomos, constituição de cada molécula, massas e cargas a utilizar. Este conjunto de dados é conhecido como a topologia do sistema. No GROMACS o módulo **pdb2gmx** permite gerar topologias a partir de uma biblioteca já existente. As topologias assim geradas reproduzem valores consistentes com os experimentos quando se trata de proteínas. Este módulo não é aplicável quando o ligante não se encontra na biblioteca do GROMACS ou de complexos organometálicos (exceto o heme).

O GROMACS possui a topologia da água pré-estabelecida. Como as moléculas de água são todas idênticas, só é necessária a topologia de uma só molécula.

### **2.5.2 Colocando o soluto numa caixa de água**

Para o cálculo em solução o soluto (porfirina, HSA e HSA-porfirinas) deve ser inserido numa caixa de água e íons  $\text{Na}^+$  e  $\text{Cl}^-$  para tentar simular o meio fisiológico. Este serão reproduzido periodicamente dentro das três dimensões, tal como é mostrado na figura abaixo.

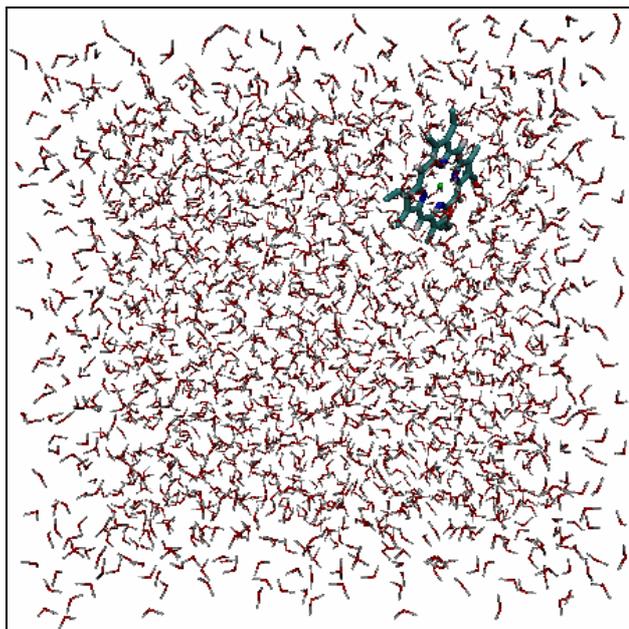


Figura 2.7 A molécula Heme numa caixa de água (Gráfico realizado com o programa VMD)

As dimensões da caixa  $R_{\text{caixa}}$  devem ser escolhidas em função das dimensões dos solutos para evitar os artefatos no cálculo. Para construir esta caixa de simulação o soluto deve ser localizado no centro de um cubo constituído de uma concatenação de  $n^3$  moléculas de solvente.

## 2.6 Mecânica e Dinâmica Molecular

### 2.6.1 Minimização de energia

A mecânica molecular nos permite minimizar a energia calculada a partir da equação (7) com o objetivo de obter as configurações do sistema no estado fundamental e de reduzir as forças iniciais muito grandes que poderiam produzir trajetórias sem sentido físico.

Mas atingir o mínimo absoluto é impraticável pelas numerosas conformações que uma proteína pode apresentar. O método *steepest-descent* permite encontrar o mínimo local mais próximo da superfície de energia potencial (Fig. 2.8), já o método dos Gradientes Conjugados encontra o mínimo de energia

de maneira mais rápida, aproximando-se mais ao mínimo absoluto. A seguir mostraremos os dois métodos:

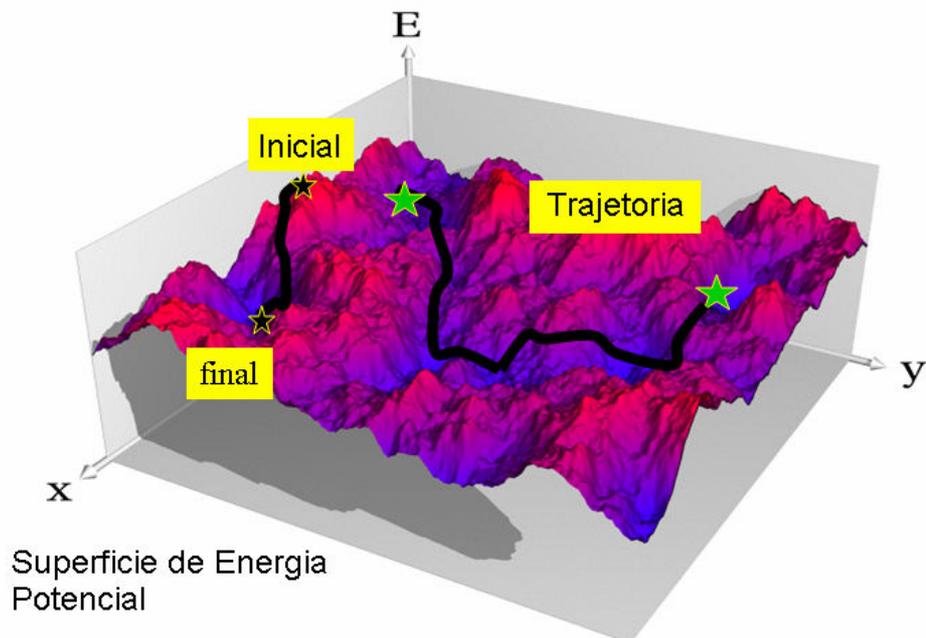


Figura 2.8 Superfície de energia potencial: Dependendo da configuração inicial da proteína ela percorre diferentes caminhos até o mínimo de energia mais próximo ao absoluto.

O primeiro método utilizado nos nossos cálculos é o método “*steepest-descent*” (Courant, 1943). Os  $n+1$  passos de minimização são feitos com base no cálculo das forças  $f(t_n)$  a partir da derivada da equação (7) para a configuração  $r(t_n)$  dada, e do cálculo da configuração seguinte  $r(t_{n+1})$  com um passo  $\Delta x$ .

$$r(t_{n+1}) = r(t_o) + N_n \Delta x f(t_n) \quad (19)$$

$N_n$  é um fator de normalização. Ao início do cálculo um valor inicial de  $\Delta x$  é dado. Nos passos da simulação, se  $V(r)$  decresce então  $\Delta x$  deve aumentar, se  $V(r)$  aumenta,  $\Delta x$  diminui. Assim, este método permite descer rapidamente para perto de um mínimo local, mas isso não assegura a convergência.

No caso do método dos *Gradientes Conjugados* (Fletcher et al., 1964), a força (o gradiente)  $f(t_{n+1})$  calculada em cada iteração é a conjugada da precedente para determinar o vetor unitário  $p(t_{n+1})$  da direção descendente sobre a hipersuperfície energética.

$$p(t_{n+1}) = f(t_{n+1}) + \beta_n p(t_n) \quad (20)$$

$$\beta_n = \frac{\langle f(t_{n+1}) | f(t_{n+1}) \rangle}{\langle f(t_n) | f(t_n) \rangle} \quad (21)$$

O passo de minimização  $s_{min}$  é escolhido levando em conta o gradiente da iteração precedente. Isto é o que permite a otimização do método com relação ao método *steepest-descent*. A nova configuração é obtida por:

$$r(t_{n+1}) = r(t_n) + s_{min}P(t_n) \quad (22)$$

O critério de convergência é feito sobre um número, a força máxima que atua por átomo. No nosso caso, esse critério ( $emtol = 100 \text{ kcal.mol}^{-1}.\text{Å}^{-1}$ , Força máxima por unidade de mol aplicada a um átomo ou grupo no Gromacs) é geralmente obtido aproximadamente em 5000 iterações, segundo a complexidade do sistema. Geralmente uma minimização de energia é feita primeiro pelo método *steepest-descent* e depois pelo método dos Gradientes Conjugados.

## 2.6.2 Dinâmica Molecular

### Princípios

As simulações de dinâmica molecular a partir de uma função de energia empírica, como dada na equação (7), são amplamente utilizadas para estudar a estrutura, a dinâmica e aspectos termodinâmicos de sistemas moleculares no equilíbrio e fora dele. As equações de movimento dos átomos são integradas no tempo. Para realizar estas integrações o GROMACS utiliza o algoritmo de *leap-frog*. Numa simulação de DM a energia total  $E$  é sempre constante. Além disso, o número de átomos  $N$  e o volume  $V$  da caixa são fixos. Assim, estas simulações se desenvolvem no formalismo de um *ensemble microcanônico* (N,V,E). Às vezes é preferível manter a temperatura constante, mais que a energia, sendo estas simulações de tipo (N,V,T). Em outros casos é vantajoso simular o sistema à pressão constante em vez de volume constante, sendo estas simulações do tipo (N,P,T). Para as simulações (N,V,T) ou (N,P,T) é necessário acoplar o sistema molecular a um banho térmico ou um banho de pressão respectivamente.

### Algoritmos de Verlet e Leap-frog

Os dois algoritmos são equivalentes, em seguida descreveremos o algoritmo de Verlet (Verlet, 1967).

A partir da diferença das séries de Taylor das velocidades  $v_i(t_n - \Delta t/2)$  a  $t=t_n$  e  $v_i(t_n + \Delta t/2)$  a  $t=t_n$  nós obtemos:

$$v_i(t_n + \Delta t/2) = v_i(t_n - \Delta t/2) + m_i^{-1} f_i(t_n) \Delta t \quad (23)$$

na qual desprezamos os termos de 3ª ordem e ordens superiores. Da mesma forma, fazendo as diferenças entre as séries de Taylor das posições em  $r_i(t_n)$  a  $t=t_n + \Delta t/2$  e  $r_i(t_n + \Delta t)$  a  $t=t_n + \Delta t/2$  dando o seguinte:

$$r_i(t_n + \Delta t) = r_i(t_n) + v_i(t_n + \Delta t/2) \Delta t \quad (24)$$

Estas duas equações permitem a integração no tempo conforme mostrado no esquema da figura 2.9.

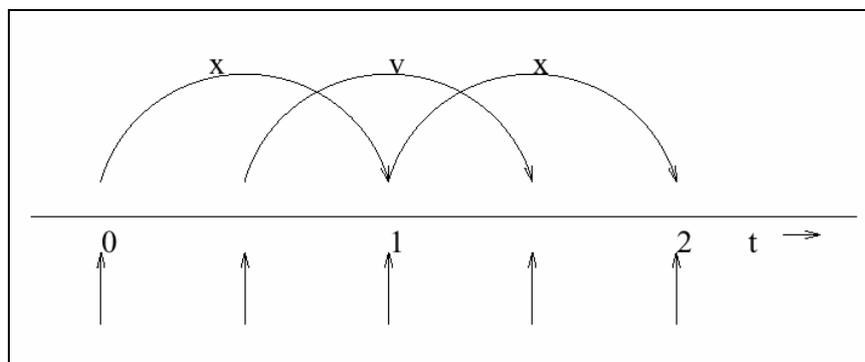


Figura 2.9 Integração das equações de movimento no algoritmo de leap-frog

Para deduzir o algoritmo de Verlet, devem-se eliminar as velocidades nas equações 23 e 24 e substituir  $t_n$  por  $t_n - \Delta t$  na equação 24, obtendo-se

$$r_i(t_n + \Delta t) = 2r_i(t_n) - r_i(t_n - \Delta t) + m_i^{-1} f_i(t_n) (\Delta t)^2 \quad (25)$$

A particularidade do algoritmo é que a velocidade dos átomos não aparece explicitamente na equação, e assim um banho térmico por acoplamento de velocidades não é possível.

O tempo de cálculo necessário é proporcional ao passo de integração  $\Delta t$ . O passo de tempo deve ser menor que o menor período de vibração do sistema para poder desta forma representar adequadamente o movimento do sistema. Em geral são as vibrações X-H que limitam  $\Delta t$  a  $10^{-15} \text{s} = 1 \text{fs}$  (as vibrações nos átomos de hidrogênio se dá em intervalos de femtosegundo).

## 2.7 Controle da temperatura

Há muitos métodos para o controle da temperatura nas simulações de DM. Um método simples que nós utilizamos no nosso sistema é o Método de Acoplamento Fraco (“weak coupling method”) (Berendsen, 1991) onde a equação de movimento dos átomos é modificada a fim de obter uma relação de primeira ordem da temperatura  $T$  perto da temperatura de referência  $T_o$ .

$$\frac{dT(t)}{dt} = \tau_T^{-1} [T_o - T(t)] \quad (26)$$

A temperatura de um ensemble de  $N_{df}$  graus de liberdade é definida em termos da energia cinética desses graus de liberdade. O controle da temperatura pode se realizar por uma modificação das velocidades dos átomos por um fator de correção  $\lambda(t)$

$$\lambda(t) = [1 + \frac{\Delta t}{\tau_T} [\frac{T_o}{T(t)} - 1]]^{1/2} \quad (27)$$

A velocidade de relaxação da temperatura é controlada pelos tempos *de relaxação da temperatura*  $\tau_T$ . Este parâmetro é ajustado em função do sistema,  $\tau_T$  deve ser suficientemente pequeno (acoplamento forte) para manter a temperatura média em  $T_o$ , mas suficientemente grande (acoplamento fraco) para evitar perturbar o sistema.

Diferentes partes de um sistema molecular podem apresentar diferentes tempos de aquecimento. Por conseguinte, é preferível acoplar as diferentes partes de um sistema separadamente aos banhos térmicos. Acoplando cada um dos três subsistemas proteína, ligante e solvente a três banhos térmicos com a mesma temperatura de referência  $T_o$  (310 K em nosso caso), garante-se uma distribuição de velocidades tal que nenhum dos subsistemas estará mais quente ou mais frio em relação à temperatura média global.

## 2.8 Controle da Pressão

O controle da pressão é realizado por um método de acoplamento fraco (Berendsen, 1984) como o controle da temperatura, mas neste caso a pressão faz o papel da temperatura e as posições atômicas das velocidades. As equações de

movimento são modificadas para obter a pressão de relaxação  $P$ , numa aproximação de primeira ordem perto do valor de referência  $P_o$ .

$$\frac{dP(T)}{dt} = \tau_P^{-1}[P_o - P(t)] \quad (28)$$

A pressão pode ser definida usando o teorema do virial,

$$P(t) = \frac{2}{3} \frac{E_{kin}(t) - W(t)}{V(t)} \quad (29)$$

Onde  $W(t)$  é o virial e  $V(t)$  o volume da caixa de simulação e  $E_{kin}(t)$  a energia cinética do sistema. O virial molecular é definido por:

$$W(t) = -\frac{1}{2} \sum_{\alpha < \beta}^{N_N} R_{\alpha\beta}(t) F_{\alpha\beta}(t) \quad (30)$$

Com  $R_{\alpha\beta}$  sendo a distância entre os centros de massa das moléculas  $\alpha$  e  $\beta$  no instante  $t$  e  $F_{\alpha\beta}$  a força sobre o centro de massa de  $\alpha$  sobre  $\beta$ . Como a pressão à temperatura constante está relacionada ao volume pela compressibilidade térmica  $\kappa_T$ , o acoplamento é efetuado por um ajuste das coordenadas atômicas e o tamanho da caixa de simulação por um fator de correção  $\mu$ . Neste caso o ajuste isotrópico é obtido por

$$\mu(t) = [1 - \kappa_T \frac{\Delta t}{\tau_P} [P_o - P(t)]]^{1/3} \quad (31)$$

O valor de  $\kappa_T$  para a água, utilizado é  $44,6 \times 10^{-6} \text{ bar}^{-1}$  [92]. Como o tempo de relaxação da pressão  $\tau_P$  é um parâmetro adaptável, não é necessário conhecer o valor de  $\kappa_T$  do sistema com exatidão.  $\tau_P$  foi escolhido considerando o valor para o água (que é o sistema com maior mobilidade).

Como a definição da pressão depende da energia cinética, o acoplamento da pressão não deverá ser mais forte que o da temperatura.

$$\tau_P \geq \tau_T \quad (58)$$

Para nosso cálculo, nós utilizamos um valor de 1.0 ps e uma pressão de referência de 1 atm.

## 2.9 Função de distribuição radial

Para a correta descrição da posição média das partículas em solução, usamos a função de distribuição radial  $g(r)$  (Function radial distribution -FDR). Calculamos unicamente a função normalizada por pares cortada pela função  $g(r)$ , a qual nos dá uma medida de como a matéria está se distribuindo ao redor do átomo ou grupos de átomos de interesse.

$$g_{ij}(r) = \frac{N(r+dr) - N(r)}{4/3 \pi \rho_j ((r+dr)^3 - r^3)} \quad (32)$$

O cálculo de  $g_{ij}(r)$  envolve a densidade dos átomos  $\rho_j$  e o número médio  $N(r)$  de átomos  $j$  dentro de uma esfera de raio  $r$  ao redor do átomo  $i$  (Fig. 2.10) e leva em conta as correlações na distribuição das moléculas provenientes das forças que elas exercem umas sobre as outras.

Quando  $g(r) = 1$  não existe correlação entre as moléculas e não se exerce influência uma sobre outra. A Energia Livre de Helmholtz pode ser escrita na forma de um potencial de Campo Médio (PCM), definido como  $W(r) = -kT \ln g(r)$  e é usado para determinar a hidropaticidade. Se  $W(r) < 0$  ( $g(r) > 1$ ) o potencial é atrativo (comportamento hidrofílico), e é repulsivo no caso contrário ( $g(r) < 1$ , comportamento hidrofóbico) (Manfred, 1996)

O número de átomos contido numa camada é obtido por integração da função  $4\pi r^2 g(r)$  na largura da camada, adicionalmente a amplitude  $g(r)$  à distância  $r$ , é necessário analisar os tipos de interação entre átomos e moléculas. Em grupos polares, por exemplo, uma ligação de hidrogênio acontecendo no limite de  $r = 0.35$  nm conectando doador e aceitador com um ângulo de tolerância hidrogênio-doador-aceitador de  $30^\circ$ .

A avaliação da FDR e os enlaces de hidrogênio mostram a possível localização das camadas de água ao redor de cada átomo presente na molécula, permitindo-nos analisar o padrão hidrofóbico em cada forma pontual e total.

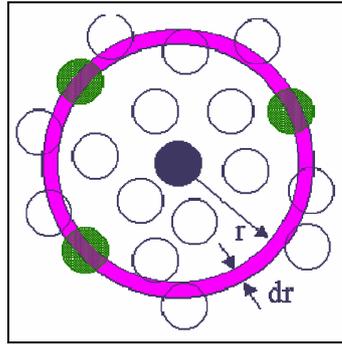


Figura 2.10 Função  $g(r)$  mostrando a distribuição da matéria em torno do átomo central (azul).

## 2.10 Modelo para a água

A água é importante em todos os processos da vida conhecida na Terra. Nós somos 80% água e, por conseguinte muito do êxito que se possa conseguir numa simulação de DM vai depender do modelo de água utilizado. Além disso, a água tem propriedades específicas que não são facilmente reproduzidas na simulação (efeito túnel nos átomos de hidrogênio a T ambiente), e dependendo dos fenômenos a estudar, eles poderiam ser desprezados. Uma dessas propriedades específicas é a dependência da densidade com a temperatura, a qual alcança um máximo perto de  $\sim 4^\circ\text{C}$ .

Os níveis de detalhe no modelo da água passam desde os modelos de solvatação contínua até os modelos de solvente explícito (Fig. 2.11).

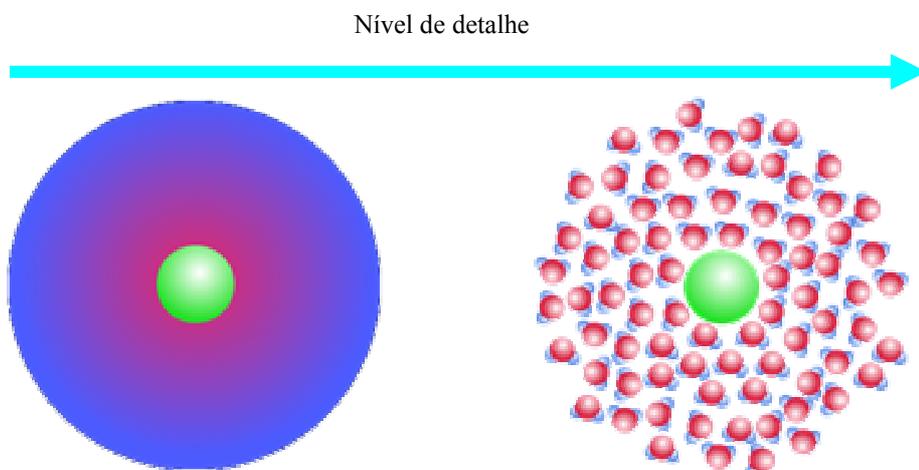


Figura 2.11 Modelos de solvente implícito (esquerda), modelos de solvente explícito (direita) (Gráfico obtido e modificado de <http://agave.wustl.edu/>)

### 2.10.1 O modelo SPC (Single Point Charge)

Para determinar as interações entre as moléculas de água existem vários modelos, uns dos mais sucedidos por sua simplicidade e eficiência computacional, usado em nosso caso, é o modelo SPC (Berendsen et al., 1981) o qual contém o efeito das cargas. No modelo SPC a molécula de água tem três centros de carga: as cargas positivas dos átomos de hidrogênio e a carga negativa do oxigênio, sendo um modelo de 3 sítios.

O fato de assumir cargas pontuais é uma aproximação que não reproduz de maneira correta o momento dipolar da molécula. Para tentar reproduzir o resultado experimental, este modelo considera o ângulo H—O—H  $109.47^\circ$ , como se mostra na Fig. 2.12.

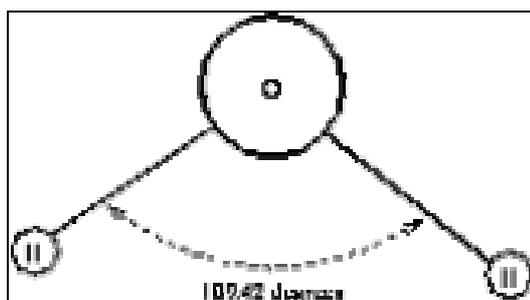


Figura 2.12 Molécula de água no modelo SPC. A carga parcial negativa do átomo de oxigênio tem duas vezes a carga parcial positiva dos átomos de hidrogênio.

Resultado da concentração da carga e do redimensionamento do ângulo H—O—H, o momento dipolar no modelo tem um valor próximo ao valor experimental. A mobilidade da água no modelo é maior que a mobilidade da água na natureza, pois não contém o par de elétrons isolados (elétrons do oxigênio). No entanto, este efeito decresce rapidamente com a temperatura.

Além do modelo SPC, existem muitos que consideram efeitos de polarização, mediante átomos “fantasmas”, para tentar reproduzir os efeitos dos orbitais eletrônicos do oxigênio, etc.

O momento dipolar efetivo de uma molécula de água em fase líquida é aproximadamente 2,6 D. Os modelos reproduzem os seguintes valores: SPC, 2.27 D e TIP4P, 2.18 D (modelo de quarto sítio, localiza uma carga negativa sobre um

átomo fantasma perto do oxigênio ao longo do bissetor do ângulo HOH) ([http://en.wikipedia.org/wiki/Water\\_model](http://en.wikipedia.org/wiki/Water_model)).

Propriedades estruturais e termodinâmicas tais como a densidade, função de distribuição radial, entalpia de vaporização, capacidade calorífica, coeficiente de difusão e constante dielétrica são usadas para ajustar os parâmetros destes modelos.

O modelo SPC é um modelo rígido, por conseguinte algumas propriedades da água não podem ser calculadas. Por exemplo, os aspectos vibracionais só poderão ser calculados se forem inseridos termos de flexibilidade na função de energia potencial.

Finalmente um fator importante no momento de escolher o modelo é o custo computacional. Deve-se tomar em conta que no processo da solvatação do sistema, uma quantidade apreciável de água é inserida no mesmo. Um modelo extremamente fino poderia tornar muito custosa a simulação.

## **2.11 Tratamento eletrostático da proteína**

O tratamento eletrostático nos dá informação a respeito dos aspectos funcionais das diferentes regiões da proteína, sobre a posição dos sítios de interesses, sobre o componente eletrostático na energia de ligação do ligante, etc.

A grande maioria dos aminoácidos que constituem uma proteína interage através de forças de longo alcance, essencialmente de natureza eletrostática. Os campos elétricos gerados por proteínas estendem-se significativamente a uma distância de 10-15 Å, dependendo de condições de temperatura, solvente e carga elétrica da proteína (em maiores distâncias, as flutuações térmicas aleatórias do solvente predominam sobre grande parte do efeito que o campo elétrico da proteína pode exercer).

Desta forma, para entender as forças de atração-repulsão entre biomoléculas é necessário compreender as leis da eletrostática que regem as interações intermoleculares. Os fundamentos da física eletrostática são bem estabelecidos e podem ser expressos concisamente por equações relativamente simples. Contudo, a aparente simplicidade destas equações esconde algumas dificuldades conceituais e numéricas da aplicação das mesmas ao estudo de sistemas complexos. Este é um

problema relevante devido à vasta quantidade de estruturas tridimensionais de biomoléculas disponíveis recentemente.

### 2.11.1

#### Aplicação da Equação de Poisson-Boltzmann para o cálculo de energias eletrostáticas em proteínas

Um sistema biomolecular sempre está rodeado de um meio o qual pode estar composto de íons, estes íons interagem eletrostaticamente com a biomolécula a qual poderia também ter carga elétrica.

Uma das aproximações que tem descrito com sucesso as interações eletrostáticas em sistemas biomoleculares é a equação de Poisson-Boltzmann (<http://bessie.che.uc.edu/tlb/rctb6/rctb6.html>).

$$\bar{\nabla} \cdot (\epsilon(\vec{r}) \bar{\nabla} \phi(\vec{r})) = -4\pi[\rho(\vec{r}) + qn_+ \exp\left(-\frac{q\phi(\vec{r})}{k_B T}\right) - v(\vec{r})] - qn_- \exp\left[\left(+\frac{q\phi(\vec{r})}{k_B T}\right) - v(\vec{r})\right] \quad (33)$$

Onde o primeiro termo a direita da igualdade representa a contribuição pelas cargas fixas na biomolécula e os outros termos representam a contribuição dos íons móveis positivos e negativos respectivamente.

$\epsilon(\mathbf{r})$  : A constante dielétrica

$\Phi(\mathbf{r})$  : Potencial eletrostático

$\rho(\mathbf{r})$  : Densidade de cargas na biomolécula

$q$ : Carga de cada íon positivo ou negativo

$n_{\pm}$  : Densidade iônica

$k_B$  : Constante de Boltzmann

$v(r)$ : é 0 ou  $\infty$  em regiões acessíveis ou inacessíveis para o íon.

Na aproximação de campo fraco, considerando-se um conjunto de cargas em um meio de constante dielétrica uniforme e de baixa força iônica, a equação 33, pode ser escrita na forma a seguir:

$$\nabla \cdot [\varepsilon(\vec{r}) \nabla \phi(\vec{r})] = \varepsilon k^2 \phi(\vec{r}) - 4\pi\rho(\vec{r}) \quad (34)$$

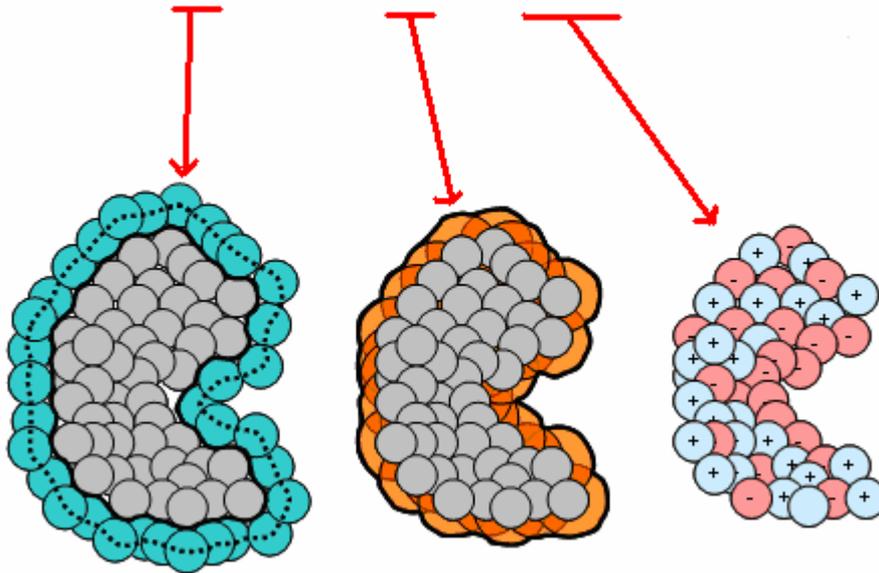


Figura 2.13 Modelo para cálculo da superfície eletrostática na equação de Poisson-Boltzmann (Gráfico obtido de <http://agave.wustl.edu/>)

Onde  $k$  é a constante de Debye-Hückel ou blindagem iônica do meio. A forma da interface dielétrica e a exclusão de íons contidos no interior da proteína são introduzidos no cálculo através da dependência das constantes  $\varepsilon$  e  $k$  do vetor  $r$ . Em condições onde  $k = 0$ , íons estão ausentes e a Equação (34) pode ser reduzida à equação de Poisson. Devido à diferença de polarizabilidade da água e do interior da proteína, duas constantes dielétricas  $\varepsilon$  e  $\varepsilon_p$  são necessárias para representar realisticamente o sistema (ver Fig. 2.14).

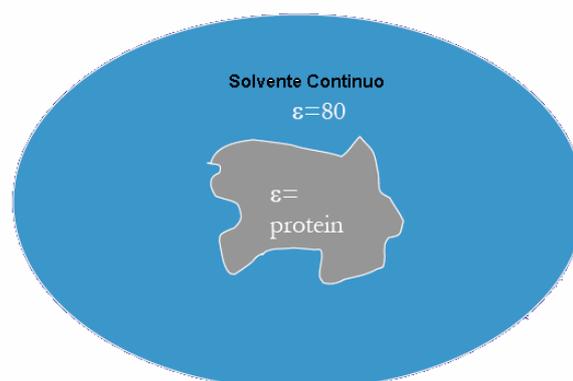


Figura 2.14 Constantes dielétricas que caracterizam os diferentes meios nos quais se aplica a equação de Poisson-Boltzmann (Gráfico obtido e modificado de <http://agave.wustl.edu/>).

## 2.12 Superfícies de interação molecular

### Superfície acessível ao solvente

A superfície acessível ao solvente (SAS) de uma biomolécula é uma forma de quantificar o efeito hidrofóbico, este conceito tem sua origem no trabalho de Lee e Richards (Lee et al., 1971) no ano 1971. A área acessível ao solvente descreve a área na qual o contato entre a proteína e o solvente ocorre.

A SAS é definida como o lugar geométrico dos centros de uma esfera de prova (representando a molécula do solvente) como se esta rodasse sobre uma superfície de van der Waals da proteína (Fig. 2.15).

### Superfície molecular (ou superfície excluída do solvente)

A superfície molecular (SM) é traçada pela frente interna da esfera de prova (Fig. 2.15). Está composta de:

- Superfície de contato: a parte da superfície de van der Waals que pode ser tocada pela esfera de prova.
- Superfície reentrante: formada pela frente interna da esfera de prova quando esta está em contato com mais de uma molécula.

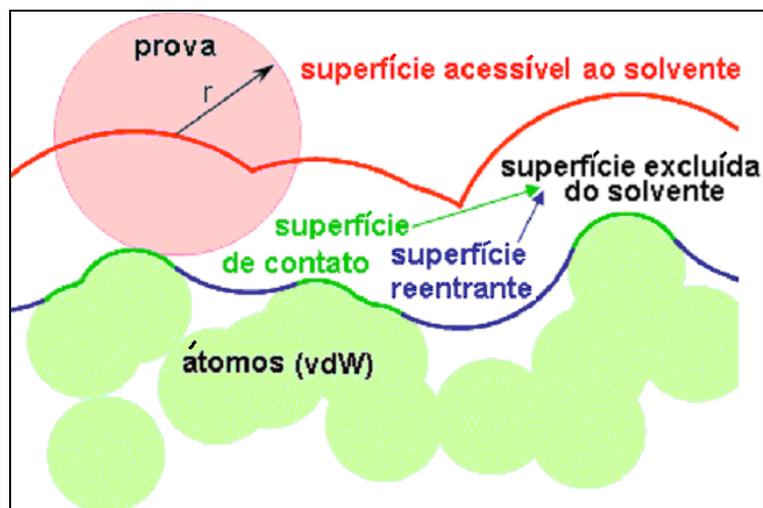


Figura 2.15 Esquema gráfico que mostra as definições de SAS e SM.

## 2.13

### **"Docking" Molecular - Algoritmo Genético**

Nos estudos de “*docking*” fizemos uso do programa AUTODOCK (Morris et al., 1998), que utiliza o Algoritmo Genético (do inglês, Genetic Algorithm, GA), para otimizar a interação e descobrir os possíveis sítios de ligação entre as porfirinas e o HSA. No programa AUTODOCK também foi implementado um algoritmo genético modificado que introduz passos de busca conformacional local entre cada passo, possibilitando que as mudanças fenotípicas (adaptações conformacionais da molécula ao ambiente) possam ser adquiridas pelas próximas gerações. Este híbrido do Algoritmo Genético é chamado de Algoritmo Genético Lamarchiano (do inglês, Lamarchian Genetic Algorithm, LGA). O método de busca local utilizado é baseado no algoritmo de otimização de Solis e Wets (SW) (Solis e Wets, 1981).

Na terminologia desses algoritmos, uma solução candidata é chamada de indivíduo e ao conjunto de indivíduos simultaneamente avaliados é dado o nome de população. Inicia-se com uma população de conformações, orientações e translações randômicas do ligante. Então o programa executa o número de avaliações energéticas e gerações (iterações) requisitadas pelo usuário e seleciona os indivíduos sobreviventes (os melhores adaptados). Uma taxa de mutação e cruzamento é especificada e introduzida no problema a fim de possibilitar uma melhor adaptação das gerações futuras (Sousa, 2005). Mais detalhes sobre o Algoritmo Genético implementado podem ser encontrados no próprio manual do AUTODOCK.

## 2.14

### **Correlação em dinâmica de biomoléculas**

Movimentos correlacionados em sistemas moleculares são essenciais para a função biomolecular, exemplos são os efeitos alostéricos e efeitos cooperativos. Muitas vezes a função energética da proteína é dominada por contribuições entrópicas que estão diretamente relacionadas a movimentos atômicos correlacionados (Lange et al., 2006). A correta compreensão destes movimentos

correlacionados é importante para o entendimento da função da proteína e poderia melhorar a interpretação dos experimentos de NMR e espalhamento de raios X.

Correlação é uma forma de medir quanto associadas ou relacionadas estão duas variáveis. O propósito de fazer correlações é fazer previsão acerca de uma variável baseada no conhecimento de outra variável

### Coeficientes de Correlação Generalizada

Os coeficientes de correlação generalizada estão baseados na formulação da Mutua Informação de Kraskov, e residem na definição fundamental de independência de variáveis aleatórias. A descrição que daremos é baseada no trabalho desenvolvido por ele no ano 2004 (Kraskov et al., 2004). Segundo este trabalho duas variáveis são independentes se e somente se a distribuição do conjunto é um produto de suas distribuições marginais,

$$P(X, Y) = P(X)P(Y) \quad (35)$$

a idéia é expressar a correlação como a diferença entre  $P(X, Y)$  e  $P(X)P(Y)$  (ver Fig 2.16).

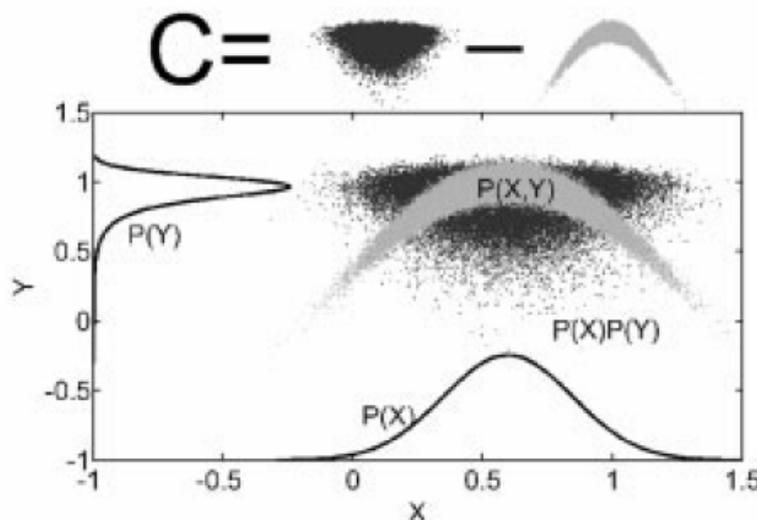


Figura 2.16 Correlação de variáveis aleatórias são definidas como a desviação de sua distribuição de probabilidade (cinza) da distribuição de probabilidades das variáveis aleatórias independentes (preto). (Gráfico obtido de Lange et al., 2006)

Isto equivale a definir a correlação  $C$  como a entropia de Shannon (correlação total) (Lange et al., 2006).

$$C[X, Y] = H[X] + H[Y] - H[X, Y] \quad (36)$$

onde  $H$  denota a entropia das variáveis aleatórias. Esta relação é equivalente a:

$$H[X, Y] \leq H[X] + H[Y] \quad (37)$$

a qual é uma igualdade se e somente se as variáveis são independentes. Pode-se observar que quando não existe correlação entre as variáveis  $X$  e  $Y$ ,  $C[X, Y]=0$ , caso contrário tem um valor positivo.

A matrix de correlação esta formado pelos coeficientes de correlação e nos forneceria informação sobre como estan relacionadas as diferentes regiones da proteina.

## 2.15

### Aproximação Quântica - Teoria do Funcional de Densidade ( DFT)

#### Porque usamos o método DFT (do inglês Density Funtional Theory) em nossos cálculos com as Porfirinas?

A aproximação de Hartree - Fock (HF) descreve a função de onda de  $N$  elétrons por um determinante simples construído de orbitais ocupados, os quais são variacionalmente determinados minimizando-se a energia. Este cálculo não incorpora a correlação eletrônica e não é adequado para descrever compostos com metais de transição (Margulis et al., 2002).

A precisão do método DFT comparado com HF é muito melhor (Ghosh et al., 2003; Himo et al., 2003, Siegbahn et al., 2000). Neste tipo de cálculo, a densidade eletrônica tem um papel central porque a energia e as propriedades moleculares são derivadas da densidade. A expressão para a energia incorpora a correlação eletrônica através do termo de correlação de intercâmbio. A aplicação do DFT precisa de um funcional de intercâmbio adequado e de uma base adequada. Na literatura há muitos funcionais desde a Aproximação de Densidade Local (local density approximation, LDA), passando por funcionais de gradiente corrigido (generalized gradient approximation, GGA) e funcionais híbridos (com mistura de HF e termos de intercâmbio) para formas mais avançadas.

Alguns destes funcionais são derivados de primeiros princípios, outros envolvem uma calibração semi-empírica ajustada a dados teóricos ou experimentais. Na prática, funcionais de gradiente corrigidos tais como BP86, BLYP e BPW91 são aplicados adequadamente a espécies enzimáticas. Sem dúvida, os funcionais híbridos como B3LYP ou B3PW91 (Ghosh et al., 2003;

Himo et al., 2003) estão entre os mais usados. Em particular, B3LYP tem tido êxito ao reproduzir valores experimentais de entalpias de formação com desvio médio absoluto de  $3 \text{ kcal.mol}^{-1}$  (Curtiss et al., 2000). B3LYP calcula muito bem um número de outras propriedades (Siegbahn et al., 2000) ainda que não reproduza bem as barreiras de reação. B3LYP é geralmente o método preferido para um cálculo DFT em sistemas biológicos e na investigação de precursores e derivados do heme (Ghosh et al., 2003; Himo et al., 2003).

### 2.15.1 Introdução à DFT

*A teoria do Funcional de Densidade está baseada na noção de que a energia total,  $E$ , de um sistema eletrônico é determinado pela distribuição da densidade eletrônica.*

Os tempos de computação nos diferentes métodos baseados em HF escalam como  $n^4$  a  $n^6$  onde  $n$  é o número de elétrons no sistema estudado. Neste aspecto, o método DFT é mais bem sucedido porque tem a vantagem de o tempo de computação escalar com  $\sim n^3$ .

Os métodos *ab initio* baseados no formalismo HF dão precisões de 0,2 Å para as distâncias de ligação, mas estes métodos têm sido menos bem sucedidos nos cálculos que envolvem metais de transição (Lüthi et al., 1982). Os parâmetros geométricos calculados por métodos DFT têm mostrado, há mais de duas décadas, uma razoável aproximação com os experimentos através das distâncias de ligação (Dunlap et al., 1986).

A Teoria do Funcional de Densidade é uma aproximação de muito êxito para a descrição de estados fundamentais de metais, semicondutores, moléculas, etc.

A idéia principal é descrever um sistema de férmions interagentes via densidade e não via função de onda de muitos corpos. Para  $N$  elétrons em um sólido, que obedecem ao princípio de Pauli, isto significa que a variável básica dos sistemas depende das coordenadas espaciais  $x$ ,  $y$ , e  $z$  em vez de  $3N$  graus de liberdade (aproximação de Thomas-Fermi). Desta forma, a energia total passa a ser escrita como um funcional  $E[\rho(\mathbf{r})]$ .

A DFT considera o termo de correlação eletrônica no potencial. Os elétrons se movem tentando evitar a interação eletrônica e assim existem zonas dentro das

quais os elétrons não podem penetrar. Isto é chamado o buraco da correlação de intercâmbio e vai dar o termo de intercâmbio (Energia de Intercâmbio) (Ziegler et al., 1991; Gritsenko et al., 1997).

### 2.15.2 Descrição da teoria

Os métodos tradicionais estão baseados nas funções de onda de muitos elétrons, uma das vantagens do método DFT é substituir a função de onda eletrônica de muitos corpos pela densidade eletrônica. A implementação mais comum da teoria é através do método de Kohn-Sham (Kohn et al., 1965), que reduz o sistema a um conjunto de elétrons interagentes num potencial efetivo. O potencial efetivo inclui o potencial externo e a interação de Coulomb. Modelar as últimas duas interações é o problema, sendo a aproximação mais simples a aproximação de densidade local (Local density approximation, LDA), a qual está baseada na energia de intercâmbio de um gás de elétrons.

No entanto, esta aproximação não é precisa o suficiente para os cálculos necessários em química quântica e é necessário adicionar outros termos ao potencial efetivo. Isto é especialmente certo quando os sistemas são dominados pela forças de van der Waals ou quando a dispersão compete significativamente com outros efeitos (em biomoléculas).

### 2.15.3 Descrição matemática

O presente item está baseado na informação disponibilizada em [http://en.wikipedia.org/wiki/Density\\_functional\\_theory](http://en.wikipedia.org/wiki/Density_functional_theory).

No formalismo DFT, a variável chave é a densidade eletrônica  $n(\vec{r})$ , que é dada por:

$$n(\vec{r}) = N \int d^3r_2 \int d^3r_3 \dots \int d^3r_N \psi^*(\vec{r}, \vec{r}_2, \dots, \vec{r}_N) \psi(\vec{r}, \vec{r}_2, \dots, \vec{r}_N) \quad (38)$$

Hohenberg e Kohn provaram em 1964 (Hohenberg et al., 1964) que a função de onda no estado fundamental  $\Psi_0(\vec{r}_1, \dots, \vec{r}_N)$  podia ser escrita como um funcional da densidade eletrônica  $n_0(\vec{r})$ . Em outras palavras que  $\Psi_0$  é um único funcional de  $n_0$ ,

$$\psi_o = \psi_o[n_o] \quad (39)$$

Disto se deriva que a energia no estado fundamental, abaixo, é um funcional de  $n_0$ , onde  $T$  e  $U$  são a energia cinética e potencial do sistema e  $V$  é o potencial externo:

$$E_o = E[n_o] = \langle \psi_o[n_o] | T + V + U | \psi_o[n_o] \rangle$$

Onde a contribuição do potencial externo é dada por  $\langle \Psi_0[n_0] | V | \Psi_0[n_0] \rangle$

O Potencial é redefinido por um potencial não interagente, onde  $s$  denota o sistema *não interagente* e  $V$  é o potencial efetivo no qual as partículas se estão movendo.  $V$  é escolhido tal que:

$$V_s = V + U + (T - T_s) \quad (40)$$

De tal forma que o funcional na equação pode ser escrito em termos de um funcional fictício de um sistema não interagente

$$E_s[n] = \langle \psi_s[n] | T_s + V_s | \psi_s[n] \rangle \quad (41)$$

Desta forma as equações de Kohn-Sham são resolvidas um sistema sem interação.

O potencial efetivo  $V_s$  pode ser escrito em mais detalhe:

$$V_s = V + \int \frac{e^2 n_s(\vec{r}')}{|\vec{r} - \vec{r}'|} d^3 r' + V_{XC}[n_s(\vec{r})] \quad (42)$$

Onde o segundo termo à direita denota o termo de Hartree descrevendo a repulsão de Coulomb elétron-elétron, e o último termo denota o potencial de correlação de intercâmbio  $V_{XC}$ .

Aqui  $V_{XC}$  inclui todas as interações de muitas partículas. O cálculo é autoconsistente pelo fato de que o termo de Hartree e  $V_{XC}$  depende de  $n(\mathbf{r})$ , o qual depende de  $\phi_i$ , que por sua vez depende de  $V_s$ .

#### 2.15.4

#### O Método RESP (Restrained Electrostatic Potential)

Como resultado de um cálculo quântico, nós obtemos o potencial eletrostático (ElectroStatic Potential - ESP). É a partir do ESP que se gera a distribuição das cargas nas moléculas.

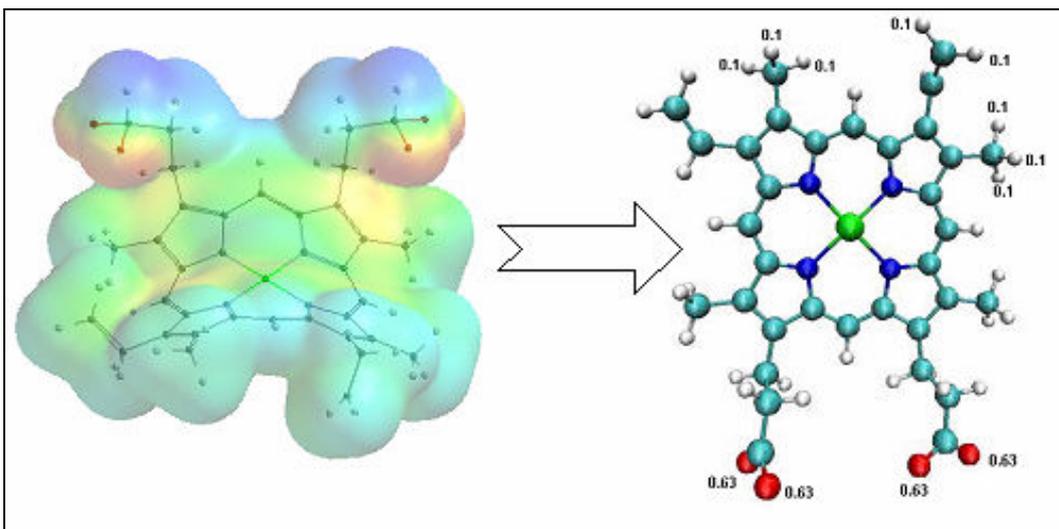


Fig 2.17 Esquema representando a obtenção das cargas a partir do ESP para o heme (as cargas mostradas são só ilustrativas, o gráfico foi gerado com o VMD).

Existem diferentes metodologias para gerar cargas através do ESP, sendo que entre as mais bem sucedidas estão:

- O método CHELPG (CHarges from ELectrostatic Potential using a Grid based method, (Breneman et al., 1990))
- O método RESP (Bayly et al., 1993)

No caso de sistemas nos quais existem átomos fortemente blindados (chamados de átomos ocultos) como o ferro no centro do anel porfirínico, o método CHELPG falha. Entre as falhas do método CHELPG estão a superestimação ou subestimação das cargas de átomos ocultos e a variação da distribuição das cargas com a conformação da molécula (Bayly et al., 1993).

Já o método RESP diminui a variabilidade da distribuição das cargas com a conformação e estima melhor as cargas dos átomos ocultos. O método RESP envolve uma aproximação em dois níveis de cálculo, onde as cargas dos átomos como os hidrogênios metil são forçadas a ser equivalentes até o segundo nível do cálculo. Neste ponto estas cargas são reajustadas, enquanto que as cargas dos outros átomos são mantidas nos valores do primeiro nível (Bayly et al., 1993).

Os resultados baseados no método RESP dão valores concordantes com a energia conformacional para pequenas moléculas e também reproduzem corretamente as energias livres de solvatação (Wang et al., 2000).