

UNIVERSIDADE FEDERAL FLUMINENSE

ISRAEL CALHEIROS GRÜNEVALD

**Um framework para a criação e gerenciamento de  
estruturas de dados para representação de proteínas**

NITERÓI

2012

UNIVERSIDADE FEDERAL FLUMINENSE

ISRAEL CALHEIROS GRÜNEVALD

# Um framework para a criação e gerenciamento de estruturas de dados para representação de proteínas

Dissertação de Mestrado submetida ao Programa de Pós-Graduação em Computação da Universidade Federal Fluminense como requisito parcial para a obtenção do título de Mestre. Área de concentração: Matemática Computacional.

Orientador:

D.Sc Mauricio Kischinhevsky

NITERÓI

2012

# Um framework para a criação e gerenciamento de estruturas de dados para representação de proteínas

Israel Calheiros Grünevald

Dissertação de Mestrado submetida ao Programa de Pós-Graduação em Computação da Universidade Federal Fluminense como requisito parcial para a obtenção do título de Mestre.

Aprovada por:

---

Prof. D.Sc. Mauricio Kischinhevsky / DCC-UFF  
(Orientador)

---

Prof. D.Sc. Carlile Campos Lavor / IMECC-Unicamp

---

Profa. D.Sc. Helena Cristina da Gama Leitão / DCC-UFF

---

Prof. D.Sc. José Henrique Carneiro de Araújo / DCC-UFF

Niterói, 18 de junho de 2012.

*Dedico esta dissertação a minha filha Letícia Calheiros Grünevald.*

# Agradecimentos

Ao professor orientador Maurício Kischinhevsky por todo apoio e conhecimento que foram essenciais para conclusão desta dissertação. Ao meu amigo, padre Luis Madero López. Aos meus pais Jared Calheiros da Silva e Lúcia Calheiros da Silva e principalmente a minha esposa Jaqueline Mayworm Grünevald.

# Resumo

O objetivo desta dissertação é a representação em uma estrutura de dados denominada grafo das proteínas do banco de dados público *protein Databank*. Também foi parametrizada a distância entre átomos-ligantes a fim de que fossem avaliados os resultados desta representação. O framework pode ser ampliado para atender outros tipos de campos de forças atualmente em uso pela modelagem mecânica molecular tais como o campo de força AMBER.

**Palavras-chave:** Biologia Molecular; Mecânica Molecular; Amber

# Abstract

This text discusses the representation of information contained in public domain's Protein DataBank (PDB). This repository stores protein data such as atoms (and/or groups of atoms) and their spatial location for a wide variety of macromolecules such as DNAs and RNAs. Moreover, during the creation of the data structure the parametrization of distances between pairs of atoms was performed. The resulting framework can be adapted to accommodate the other force fields often employed in Molecular Mechanics, e.g. in AMBER formula.

**Keywords::** Molecular Biology, Molecular Mechanics, Amber.

# Palavras-chave

1. Modelagem Mecânica Molecular
2. Enovelamento
3. Parametrização do Campo de Força
4. Amber



# Glossário

$i, j, k$	:	vetores unitários ortogonais pertencem aos eixos x, y, z.
$\phi, \theta, \psi$	:	Ângulos de Euler
$r, \theta, \phi$	:	coordenadas polares esféricas
$\lambda$	:	multiplicador de Lagrange
ALA	:	Alanina
ARG	:	Arginina
ASN	:	Asparagina
ASP	:	Asparagina
ASP	:	Ácido Aspártico
BLAST	:	Basic Local Alignment Search Tool
CYS	:	Cisteína
GLN	:	Glutamina
GLU	:	Ácido Glutâmico
HIS	:	Histidina
IC	:	Instituto de Computação;
LEU	:	Leucina
LYS	:	Lysina
MET	:	Metionina
PHE	:	Fenilalanina
PDB	:	Protein Data Bank
PRO	:	Prolina
SER	:	Serina
TRP	:	Triptofano
TYR	:	Tirosina
UFF	:	Universidade Federal Fluminense;
VAL	:	Valina

# Sumário

<b>Conceitos Básicos de Biologia Celular</b>	<b>3</b>
2.1 A lógica da vida . . . . .	4
2.2 A produção e o consumo de energia no metabolismo . . . . .	4
2.3 Composição química dos organismos vivos . . . . .	7
2.4 Grupos funcionais determinam propriedades químicas . . . . .	8
2.5 Aminoácidos . . . . .	8
2.6 Estrutura Tridimensional . . . . .	9
<b>Estrutura da proteína e enovelamento</b>	<b>12</b>
3.1 Introdução . . . . .	12
3.2 Estrutura da Proteína . . . . .	14
3.2.1 Estrutura Primária . . . . .	14
3.2.2 Estrutura Secundária . . . . .	15
3.2.2.1 $\alpha$ -hélice . . . . .	15
3.2.2.2 Folhas $\beta$ . . . . .	16
3.2.3 Estrutura Terciária e Quaternária . . . . .	16
<b>Mapeamento do Campo de Força</b>	<b>17</b>
4.1 Modelo Quântico . . . . .	18
4.2 Modelo Mecânico Molecular . . . . .	18
4.3 Ligações químicas . . . . .	20
4.4 Comprimento da ligação . . . . .	21
4.5 Ângulo de ligação . . . . .	23
4.6 Ângulo de torção . . . . .	24
4.7 Forças de Van der Waals . . . . .	24
4.8 Juntando as peças . . . . .	26
4.8.1 Hibridação . . . . .	27
4.8.1.1 Hibridação $sp^3$ . . . . .	28
4.8.1.2 Hibridação $sp^2$ . . . . .	29

---

4.8.1.3	Determinação do estado da Hibridação . . . . .	30
4.9	Tipo de Átomo . . . . .	31
4.10	Protein Databank . . . . .	31
<b>Hyle</b>		<b>34</b>
5.1	Massa de dados . . . . .	34
5.2	Importação da massa de dados . . . . .	35
5.3	Roteiro de instalação do software Hyle3D . . . . .	38
5.4	Modelo de dados do programa Hyle3D . . . . .	40
5.4.1	Modelo Entidade-Relacionamento . . . . .	40
5.4.2	Mapeamento Lógico-Relacional . . . . .	41
5.4.2.1	Tabela Molecule . . . . .	41
5.4.2.2	Tabela Atom . . . . .	41
5.4.2.3	Tabela Bond . . . . .	43
5.4.2.4	Tabela Molecule type . . . . .	43
5.4.2.5	Tabela Monomer . . . . .	43
5.4.2.6	Tabela Parameter Bond . . . . .	44
5.4.2.7	Tabela Residue . . . . .	44
<b>Conclusões e Perspectivas</b>		<b>46</b>
<b>Índice Remissivo</b>		<b>50</b>
<b>Apêndice A – Script para criar as tabelas do programa Hyle3D</b>		<b>51</b>
<b>Referências</b>		<b>56</b>

# Lista de Figuras

2.1	Bases Nitrogenadas . . . . .	6
2.2	Radical $\alpha$ do Aminoácido . . . . .	8
2.3	Convenção e nomenclatura Lisina . . . . .	9
2.5	Stereo isômeros de $\alpha$ -aminoácidos . . . . .	10
2.4	20 aminoácidos das proteínas . . . . .	11
3.1	Ângulos representados pelas letras gregas . . . . .	15
3.2	Gráfico de Ramachandran . . . . .	16
4.1	Modelo Massa-Mola . . . . .	19
4.2	Tipos de Forças . . . . .	20
4.3	Comprimento da ligação . . . . .	21
4.4	Variação da energia em razão da distância interatômica . . . . .	22
4.5	Ângulo átomos . . . . .	23
4.6	Ângulo de torção entre 4 átomos. . . . .	24
4.7	Torção átomos . . . . .	25
4.8	Representação das Forças de Van der Waals . . . . .	25
4.9	Van der Waals . . . . .	26
4.10	Molécula de Propano . . . . .	26
4.11	Distribuição dos elétrons nos orbitais do elemento químico carbono (Valence bond theory: sp	
4.12	Hibridação $sp^3$ (Valence bond theory: sp, sp <sup>2</sup> , and sp <sup>3</sup> hybrid orbitals 2012).	28
4.13	Arranjo tetraédrico da hibridação $sp^3$ possibilita a formação de quatro li- gações (Valence bond theory: sp, sp <sup>2</sup> , and sp <sup>3</sup> hybrid orbitals 2012).	28
4.14	Átomo de carbono conectado a 4 átomos de hidrogênios através da hibri- dação $sp^3$ (Valence bond theory: sp, sp <sup>2</sup> , and sp <sup>3</sup> hybrid orbitals 2012).	28
4.15	Hibridação $sp^2$ (Valence bond theory: sp, sp <sup>2</sup> , and sp <sup>3</sup> hybrid orbitals 2012).	29
4.16	Hibridação $sp^2$ (Valence bond theory: sp, sp <sup>2</sup> , and sp <sup>3</sup> hybrid orbitals 2012).	29
4.17	Hibridação $sp^2$ . . . . .	29
4.18	Exemplo de arquivo do <i>Protein Databank</i> . . . . .	33
5.1	Screenshot do programa Hyle3D . . . . .	38
5.2	Diagrama Entidade-Relacionamento Hyle3D . . . . .	40

# Lista de Tabelas

4.1	Tabela de Comprimento de Ligação e Energia . . . . .	21
4.2	Tabela de Ângulos e Energia . . . . .	23
5.1	Tabela com mapeamento lógico-relacional da entidade <i>Molecule</i> . . . . .	41
5.2	Tabela com mapeamento lógico-relacional da entidade <i>Atom</i> . . . . .	42
5.3	Tabela com mapeamento lógico-relacional da entidade <i>Bond</i> . . . . .	43
5.4	Tabela com mapeamento lógico-relacional da entidade <i>Molecule_Type</i> . . .	43
5.5	Tabela com mapeamento lógico-relacional da entidade <i>Monomer</i> . . . . .	44
5.6	Tabela com mapeamento lógico-relacional da entidade <i>Bond</i> . . . . .	44
5.7	Tabela com mapeamento lógico-relacional da entidade <i>Residue</i> . . . . .	45
6.1	Tabela de distância entre átomos-ligantes com 50 Proteínas. . . . .	48
6.2	Tabela de distância entre átomos-ligantes com 50 Proteínas . . . . .	49

# Lista de Equações

4.1 Campo de Força Amber . . . . .	19
4.2 Energia potencial Morse . . . . .	22
4.3 Lei de Hooke . . . . .	23
4.4 Energia Potencial torsional . . . . .	24

# Capítulo 1

## Introdução

As proteínas são vitais para o funcionamento de todos os organismos vivos. Entre alguns exemplos gerais que poderiam ser citados sobre o papel da proteína servem como estruturas para o suporte mecânico dos tecidos<sup>1</sup>; algumas são capazes de gerar movimentos em células e tecidos<sup>2</sup>; algumas podem armazenar pequenas moléculas ou íons, transportar sinais de uma célula à outra<sup>3</sup>, detectar sinais e transmiti-los à maquinaria de resposta celular<sup>4</sup>, regular a transcrição gênica<sup>5</sup> e, mais importante, assumir o papel de catalisadora das reações orgânicas (Alberts et al. 2006). No capítulo 2 será apresentada uma introdução à biologia celular.

A forma como uma proteína é caracterizada é determinada por sua sequência de aminoácidos. Na modelagem molecular, a determinação da estrutura tridimensional da proteína a partir da sua sequência de aminoácidos é denominada como problema do enovelamento ou dobramento<sup>6</sup>. Esse tópico será visto com mais detalhe no capítulo 3. Uma abordagem utilizada é a mecânica molecular, que considera átomos e ligações como um conjunto de pontos interconectados por molas.

Hyle3D é um framework em C++ que permite carregar arquivos de proteína do *proteinDatabank* em estruturas de dados em memória denominada grafo. Cada átomo e cada ligação de uma molécula são representados, respectivamente, como vértices e arestas de um grafo. A viabilidade desta estrutura em memória possibilita diversas operações bási-

---

<sup>1</sup>No meio extracelular: o colágeno e a elastina formam fibras em tendões e ligamentos. No meio intracelular: a tubulina forma microtúbulos longos e rígidos e a actina forma filamentos que acompanham e sustentam a membrana plasmática.

<sup>2</sup>A miosina em células dos músculos esqueléticos, provê força motriz necessária para movimentação.

<sup>3</sup>Na circulação sanguínea a albumina sérica transporta lipídeos e a hemoglobina transporta oxigênio.

<sup>4</sup>A rodopsina na retina detecta a luz.

<sup>5</sup>O repressor da lactose, em bactéria, silencia o gene para a enzima responsável pela degradação da lactose; várias outras proteínas agem como interruptores genéticos.

<sup>6</sup>Folding problem

cas da modelagem molecular tais como a parametrização dos valores de distância, ângulos e torção entre átomos ligantes, além das interações entre átomos não-ligantes como Van der Waals e Eletrostática.

Uma aplicação dos valores parametrizados é a utilização no cálculo de energia potencial para um sistema molecular; uma fórmula muito utilizada é o campo de força *AMBER*.

$$v(r^n) = \sum_{bound} \frac{k_i}{2} (l_i - l_{i,0})^2 + \sum_{angle} \frac{k_i}{2} (\Theta_i - \Theta_{i,0})^2 + \sum_{torsion} 1 + \cos(\eta\omega - \gamma) \\ + \sum_{i=1}^N \sum_{j=1}^N (4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}})$$

Onde  $v$  representa a energia potencial aplicada sobre  $r^n$  pares de átomos ligantes. A explicação detalhada desta fórmula será descrita no capítulo 4.

Também foi parametrizado o valor da distância entre átomos ligantes. A parametrização oferece uma forma de se obter o valor referencial  $l_{i,0}$  da fórmula AMBER. Compreende o valor em que a resultante das forças seja nulo. Em outras palavras: o valor que não comprime e nem estende a mola. Os resultados desta parametrização se aproximou de outros artigos publicados tal como (Allen et al. 1987).

Para um trabalho futuro sugere-se a parametrização de outros valores como ângulo de rotação e torção assim como simulações utilizando a modelagem molecular.



## Capítulo 2

### Conceitos Básicos de Biologia Celular

## 2.1 A lógica da vida

A bioquímica procura explicar a origem da vida em termos químicos. Estes princípios não são novas leis da física ou forças, mas aquelas mesmas que já são conhecidas por toda a ciência (Nelson e Cox 2010).

Todos os organismos são constituídos de componentes mais simples. A maior parte dos seres vivos é composta de cadeias de carbonos interligadas por ligações covalentes com átomos de hidrogênio, oxigênio e nitrogênio. Como veremos, a grande estabilidade das cadeias carbono-carbono propiciaram moléculas altamente estáveis, além de uma grande variedade de moléculas, algumas até compartilham propriedades em comum.

Um fato interessante é que grande parte do desenvolvimento da bioquímica se deve ao estudo da bactéria *Escherichia Coli*, que contém mais de 6000 diferentes tipos de componentes orgânicos incluindo 3000 diferentes tipos de proteínas e um número similar de ácido nucléicos; ela se tornou o laboratório perfeito para o entendimento dos aspectos biológicos envolvidos no processo de tradução do DNA (Nelson e Cox 2010, p.5).

A maior parte dos aspectos biológicos aprendidos com estudo da *E.Coli* pôde ser generalizada para os mais diversos organismos, até mesmo o homem. É notável também que o processo de tradução proteica seja exatamente igual em todos os organismos, sem haver praticamente nenhuma exceção conhecida até o momento. Esta similaridade permitiu que os aspectos bioquímicos fossem compreendidos em organismos mais simples para daí então ser testados e generalizados para organismos mais complexos.

## 2.2 A produção e o consumo de energia no metabolismo

A variação de energia é um aspecto extremamente importante para ser analisado em todas as reações químicas. Os organismos vivos tentam a todo instante lutar contra o equilíbrio energético armazenando energia em processos endotérmicos e liberando energia em processos exotérmicos. A manutenção da temperatura é uma característica importante da maior parte das células.

As reações químicas em organismos biológicos ocorrem em temperatura constante. A velocidade em que essas reações ocorrem também assume um papel essencial na manutenção do equilíbrio dinâmico dos componentes químicos da célula.

É conhecido que uma reação química pode ser acelerada de diversas formas, por

exemplo: através do aquecimento, aumento de pressão, etc.. Entretanto essas variáveis devem ser constantes nos seres orgânicos, por isso, é necessário um outro mecanismo para acelerar os processos químicos.

O principal mecanismo utilizado pelos seres biológicos são os **catalisadores**. Eles aceleram uma reação química propiciando em média um aumento de velocidade da ordem  $10^{14}$ , o que nos faz perceber a importância dos catalisadores nos processos químicos biológicos.

Nos organismos vivos são as enzimas que realizam o papel de catalisador. Além disso, o equilíbrio químico do corpo humano é dinâmico; isto quer dizer que a proporção de concentrações de reagentes e de produtos ocorre através de reações reversíveis que são equilibradas através de um sofisticado sistema de interrupção ou aceleração de reações químicas a partir das enzimas.

Por exemplo: a *treonina* é precursora da *isoleucina*; a produção dela ocorre em cinco etapas que conta com a participação de cinco enzimas. Caso a concentração de isoleucina esteja baixa, novas enzimas são introduzidas propiciando aumento da concentração, mas quando a concentração está alta ocorre interrupção da fabricação da enzima da etapa inicial ocasionando diminuição da concentração da isoleucina.

Os seres vivos possuem diversos mecanismos para manter a concentração dos componentes químicos em limites aceitáveis (Nelson e Cox 2010, p.12).

A economia deste avançado processo em todos os organismos vivos é realizada pelo DNA.

Macromoléculas são polímeros feitos de um número de blocos que oferecem uma funcionalidade específica. A diversidade de blocos habilita as biomoléculas realizarem um extenso número de funções que são vitais para o funcionamento da célula. No caso das proteínas, estas subunidades são os vinte aminoácidos.

Para que as interações ou as reações que ocorrem durante a ação dos bioligantes aconteçam, deve haver complementaridade estrutural entre a molécula do bioligante e a biomacromolécula. Esta complementaridade permite o estabelecimento de interações mais ou menos específicas, como interações dipolo-dipolo e íon-íon, ligações hidrogênio e forças de dispersão, que contribuem para a energia de interação entre o bioligante e a biomacromolécula. É justamente neste ponto que a modelagem molecular pode dar sua contribuição para o desenvolvimento de bioligantes, pois através dela pode-se ter uma descrição detalhada da estrutura, das interações intermoleculares e, se for o caso, das

reações químicas entre o bioligante e a biomacromolécula (SantAnna 2009).

A função biológica pode ser definida pela forma estrutural que esta se encontra em temperatura ambiente. Proteínas são responsáveis por diversos papéis dentro da célula. Elas podem agir como moléculas de comunicação transferindo informações e instruções de uma forma extremamente específica. Elas também podem agir como catalisadores ou como máquinas moleculares que consomem energia para produção de algum processo mecânico (Stryer 2006).

Os elementos variáveis no DNA são as quatro bases nitrogenadas - Adenina(A), Guanina(G), Citosina(C) e Timina(T), que carregam a informação genética. Elas são interligadas por uma cadeia de açúcar-fosfato para formar um canal polinucleotídeo (Stryer 2006).

Ácidos desoxiribonucleícos são compostos de apenas quatro diferentes tipos de subunidades de desoxiribonucleotídeos, enquanto que os ácidos ribonucleícos são compostos de quatro tipos diferentes de ribonucleotídeos.

Proteínas são compostas de 20 tipos diferentes de aminoácidos que são idênticos em todos os organismos vivos. Os nucleotídeos não servem apenas como subunidades dos ácidos nucleícos mas também como moléculas de transporte energético.

Os aminoácidos são subunidades das moléculas de proteínas e também precursores de hormônios, neurotransmissores, pigmentos e muitos outros tipos de moléculas biológicas.

O DNA consiste de duas cadeias dupla-hélices que são conectadas por pontes de hidrogênio emparelhando cada base nitrogenada da sequência com uma outra base nitrogenada do outro par formando a hélice.

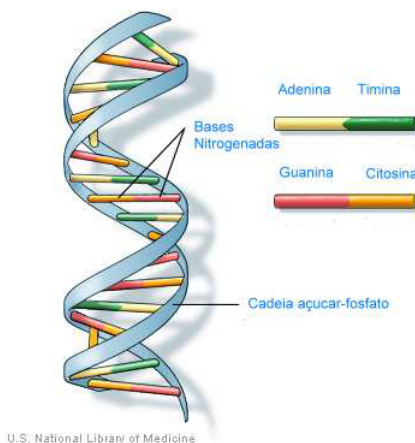


Figura 2.1: Gráfico que mostra como as bases nitrogenadas estão interligadas.

Esta sequência é complementar com adenina sempre pareada com uma timina e guanina com citosina e vice-versa (Stryer 2006).

A molécula de RNA é muito similar ao DNA em composição química, com apenas uma mudança da base nitrogenada timina para uracila. Entretanto as cadeias de RNA são muito mais instáveis que o DNA e rapidamente se degeneram (Schlick et al. 2011).

Na natureza, uma simples cadeia de RNA frequentemente se enovela para formar estruturas complexas em forma de loop. O papel mais importante do RNA na célula é agir como um mensageiro entre o DNA, interno ao núcleo da célula, e a proteína.

Alguns vírus no entanto carregam sua informação genética em RNA, ao invés de DNA. Por serem quimicamente reativos, o RNA também pode agir algumas vezes como enzima. Sabe-se que formas de vida mais simples foram originárias de moléculas de RNA anterior ao surgimento das proteínas e do próprio DNA (Stryer 2006).

Um único aminoácido é especificado pela sequência de 3 bases denominadas *codon*. Como existem 64 possibilidades de codons para 20 aminoácidos existe uma redundância considerável no processo de tradução. Nem todo o DNA no núcleo codifica proteína. De fato, a maior parte do material genético ainda não tem uma função conhecida e tem sido chamada de "DNA lixo", enquanto que outras partes agem como sítio ativos para ação das proteínas e controle da expressão gênica. O RNA mensageiro é sintetizado a partir de um modelo de DNA pela proteína RNA polimerase durante o processo de transcrição. Inicialmente, RNA polimerase se une próximo a fita dupla de DNA e se espalha sobre a hélice até detectar um sítio promotor onde a transcrição se inicia.

## 2.3 Composição química dos organismos vivos

A composição química dos organismos vivos difere notavelmente do mundo inanimado, como pode demonstrar Lavoisier (1743-1794). Os seres vivos são ricos em carbono, oxigênio, nitrogênio e fósforo. O desenvolvimento da química orgânica em grande parte se deve ao desenvolvimento da bioquímica.

Apenas 30 dos 90 compostos químicos são essenciais para os organismos vivos. A maior parte destes elementos possuem baixo peso atômico. De forma geral, os elementos mais leves formam ligações mais fortes.

A química dos organismos vivos está organizada a partir do elemento químico carbono que corresponde a mais da metade do peso de uma célula desidratada. No metano  $\text{CH}_4$  o

carbono compartilha 4 pares de elétrons com quatro átomos de hidrogênio, cada par de elétron compartilhado forma uma ligação. Os átomos de carbono podem formar ligações entre si bem estáveis. A ligação carbono-carbono pode ser dupla ou tripla. Os compostos orgânicos possuem formas e dimensões específicas. As quatro ligações covalentes simples podem ser formadas por um arranjo tetraédrico e um ângulo médio de 109,5 graus. Existe um ângulo de rotação livre para cada ligação simples carbono-carbono; algumas vezes este ângulo pode ser restrito. A ligação covalente dupla carbono-carbono é mais curta, em média 0,134 nm, e rígida e permite pouco espaço para rotações sobre um eixo. Nenhum outro elemento químico pode formar moléculas com tantos formatos diferentes e tamanhos tão distintos ou com tantos grupos funcionais como o carbono (Nelson e Cox 2010, p.58).

## 2.4 Grupos funcionais determinam propriedades químicas

A maior parte das moléculas de organismos vivos são provenientes de hidrocarbonetos compostos de uma estrutura de átomos ligantes. As ligações carbono-carbono são extremamente estáveis.

Os átomos de hidrogênio podem ser substituídos por uma infinidade de grupos funcionais. Famílias típicas de componentes orgânicos são os álcoois, que possuem um ou mais grupos hidroxilo. Aldeídos e cetonas que possuem grupos carboxílico, além de outros. Cada um desses componentes possui propriedades em comum por isso designamos famílias; além disso cada família reage de forma semelhante (Nelson e Cox 2010, p.60).

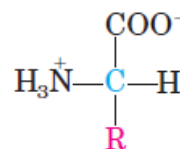


Figura 2.2: Esta estrutura é comum a todos os  $\alpha$ -aminoácidos. A exceção é a prolina que é um aminoácido cíclico. O grupo R em vermelho é substituído por um radical (Nelson e Cox 2010, p.76)

## 2.5 Aminoácidos

Proteínas são polímeros de aminoácidos e cada qual é unido ao seu vizinho por um tipo específico de ligação covalente no carbono  $\alpha$ . É uma união em que ocorre desidratação desses elementos pela perda de hidrogênio e oxigênio, que se unem na forma de molécula

de água e são liberados no meio.

Proteínas também podem ser decompostas pela hidrólise de seus componentes constituintes, os aminoácidos. Existem 20 aminoácidos que são comumente encontrados em proteínas e se diferem pela presença do radical (veja a figura 2.2.).

O primeiro a ser descoberto foi a asparagina em 1806; o último dos 20 é a treonina, que só foi identificada em 1938. Os 20 aminoácidos possuem um nome geralmente relacionado ao meio em que foram encontrados; por exemplo: asparagina foi encontrada no aspargo, glutamato no gluten e assim por diante. O carbono  $\alpha$  é denominado centro quiral.

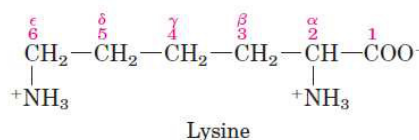


Figura 2.3: Lisina (Nelson e Cox 2010, Figura 3.2a)

Devido ao arranjo tetraédrico das ligações com o carbono  $\alpha$ , os 4 diferentes grupos podem ocupar duas distribuições espaciais; portanto os aminoácidos possuem duas possibilidades de estereoisômeros como exemplificado na figura 2.5. Se elas não são sobrepostas elas são denominadas enantiômeros e possuem a característica de ser opticamente ativas, ou seja, rotacionam num plano polarizado.

Somente aminoácidos do tipo *L* são encontrados em proteínas; a forma *D* não é muito comum e existem somente em algumas bactérias e peptídeos de alguns antibióticos. Células são capazes de sintetizar apenas os isômeros *L* de aminoácidos porque os sites ativos das enzimas são assimétricos, o que leva a concluir que eles são estereo-específicos (Nelson e Cox 2010, p.77).

## 2.6 Estrutura Tridimensional

Embora a descrição funcional seja importante, a estrutura tridimensional e a distribuição no espaço dos átomos das moléculas biológicas é crucialmente vital (Nelson e Cox 2010, p.60). Componentes derivados de carbono frequentemente existem em duas ou mais formas químicas tridimensionais, mas apenas uma é biologicamente ativa. Todos os 20 tipos de aminoácidos são  **$\alpha$ -aminoácidos**. Eles possuem um **grupo carboxila** e um grupo amino ligado ao mesmo carbono. Eles diferem um dos outros pelos radicais, como pode ser visualizado na figura 2.2.

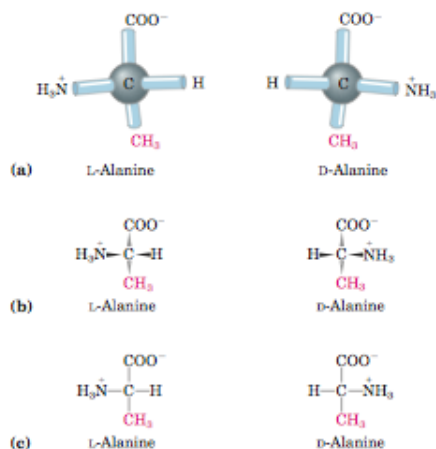


Figura 2.5: Stereo isômeros de  $\alpha$ -aminoácidos (Nelson e Cox 2010, Fig 3.3)

Outra característica é que as ligações moleculares, em função da diferença de **eletronegatividade**<sup>1</sup>, podem exercer diferentes tipos de atrações. Uma ligação entre átomos com eletronegatividade igual é denominado **apolar** e se for diferente é denominado **polar**. A polaridade de uma molécula determina se ela será hidrófila ou hidrófoba afetando a estrutura tridimensional da proteína.

Existem algumas convenções que são usadas para identificar os carbonos nos aminoácidos. Segundo a IUPAC, os carbonos são enumerados com letras gregas

$\alpha, \beta, \gamma, \delta, \epsilon$  e assim por diante. Outra forma é numerá-los com numerais iniciando no grupo carboxila com 1 e assim por diante. A figura 2.3 ilustra essas duas situações. O arquivo do *protein databank* utiliza letras gregas. A dificuldade é que as letras gregas são ambíguas quando existem ciclos. Os arquivos do *protein databank*, notadamente os registros *ATOM* e *HETEROATOM* utilizam assim mesmo letras gregas. Para resolver o problema da ambiguidade em ciclos é necessário consultar os arquivos de monômeros.

De acordo com a documentação do PDB (Protein Databank 2012) os registros heteroátomos (*HETEROATOM*) representam coordenadas químicas de substâncias não-poliméricas<sup>2</sup> e coordenadas não padronizadas, tais como água. Os heteroátomos também possuem sua localização definida neste registro.

Os registros *ATOM* apresentam as coordenadas atômicas para aminoácidos padronizados e nucleotídeos, assim como outras informações, tais como localização e temperatura.

<sup>1</sup>Eletronegatividade é o alcance da capacidade de atrair elétrons. Ela cresce em função do aumento da capacidade de atrair elétrons.

<sup>2</sup>Non-polymer



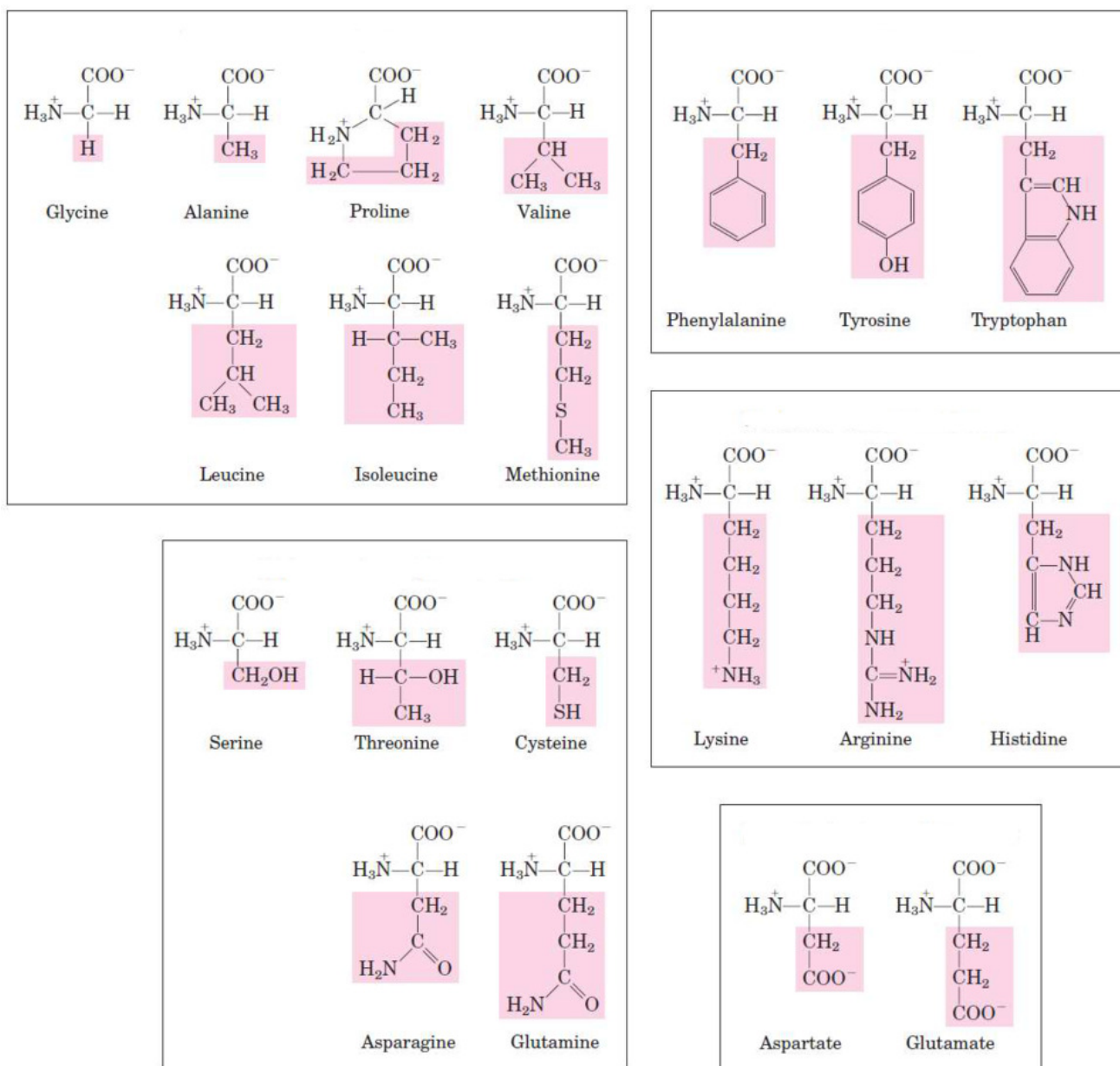


Figura 2.4: Esta estrutura é comum a todos os  $\alpha$ -aminoácidos. A exceção é a prolina que é um aminoácido cíclico. O grupo R em vermelho é substituído por um radical (Nelson e Cox 2010, Fig 3.5)

# Capítulo 3

## Estrutura da proteína e enovelamento

### 3.1 Introdução

Peptídeos e proteínas são polímeros construídos a partir de sequência de aminoácidos. Eles realizam diversas funções essenciais à vida. Existem cerca de 10 aminoácidos comuns que ocorrem naturalmente na natureza, como já foi apresentado no capítulo anterior. Os aminoácidos são unidos podendo formar longas cadeias estáveis. Todos os aminoácidos que ocorrem naturalmente na natureza possuem a mesma estereoquímica.

Uma proteína pode ser composta de vários aminoácidos que formam uma estrutura lateral que pode possuir diversos tamanhos, formas, capacidade de formar pontes de hidrogênio e distribuição de cargas, que habilita a proteína a formar um vasto universo de funções biológicas requerido pelos organismos vivos (Leach 2001, p.509).

A biosíntese da proteína é algo extremamente complexo. As sequências de aminoácidos são determinadas pela sequência de DNA do gene correspondente. Cada aminoácido é codificado por 3 bases de DNA adjacentes. Contudo, DNA não é utilizado diretamente na biosíntese da proteína. Ao invés disto, uma cópia do DNA é feita produzindo um modelo de RNA (RNA mensageiro ou mRNA) que realiza o trabalho no processo de biosíntese de um DNA na etapa denominada **transcrição**. Numa etapa subsequente ocorre o processo de **tradução** do qual o RNA mensageiro é lido por um RNA de transferência ou RNA transportador (tRNA). Este processo de duas etapas foi proposto por Francis Crick e também é conhecido como "Dogma Central"<sup>1</sup>

A função biológica da proteína ou de um peptídeo está intimamente ligada com sua

---

<sup>1</sup>Atualmente não é mais um dogma já que existem alguns vírus retrovirais que possuem a capacidade de transformar RNA em DNA dentro das células hospedeiras tais como a AIDS.

conformação estrutural. Em contraste com outros tipos de polímeros, onde moléculas individuais podem adotar diversas formas, uma proteína normalmente só existe em um único estado ativo. Este estado ativo corresponde na maior parte das vezes no estado (Ph, temperatura, etc.) em que se encontra a célula.

Proteínas podem ser desenoveladas modificando o Ph, desidratando ou usando altas temperaturas. Essa é uma reação reversível, de modo que as proteínas podem ser enoveladas de volta ao seu estado natural em laboratório.

Os métodos mais comuns para mapeamento estrutural de uma proteína são a cristalografia por raio-X<sup>2</sup>, a ressonância magnética nuclear <sup>3</sup> e a microscopia eletrônica. Infelizmente, a taxa em que novas proteínas são sequenciadas excede a taxa em que as estruturas das proteínas são determinadas experimentalmente. Além disso, nem todo o DNA codifica proteína e em muitos genes a informação está contida em unidades distintas denominadas **exons** (as regiões que não codificam são denominadas **introns**). A **função genômica** é um conceito que está relacionado com a caracterização da proteína expressa pelo genoma e a sua função biológica. É possível atribuir uma função somente através da sequência de aminoácidos. Entretanto, a íntima relação entre a forma tridimensional e a função da proteína traz o fato de que a estrutura é uma peça fundamental para o entendimento dos processos bioquímicos e não pode ser ignorado.

A dificuldade em obter a estrutura de uma proteína por técnicas experimentais faz com que haja um considerável interesse em métodos teóricos para predição como um problema a ser resolvido<sup>4</sup>.

Tal interesse também pode ser exemplificado pelo número de proteínas ligados a diversos tipos de doenças, como os males de Alzheimer, Parkinson, encefalopatia espongiforme, diabetes tipo II, fibrose cística e alguns tipos de câncer (Trevizani 2011, p.1), daltonismo, pseudohermafroditismo masculino, hiperparatireoidismo neonatal, deficiência do hormônio somatotrófico, hipergonadismo, obesidade extrema e outras (Allen 2004).

A compreensão da relação sequência-estrutura-função, como cita (Walsh 2005), produz um enorme interesse na pesquisa de engenharia de proteínas, facilitando a modificação de estruturas conhecidas e possibilitando o desenvolvimento de biofármacos com sequências de aminoácidos modificadas, tais como anticorpos, hormônios e interferons (Walsh 2005).

---

<sup>2</sup>X-Ray crystallography

<sup>3</sup>NMR

<sup>4</sup>protein folding problem

A utilização de modelos computacionais para direcionar a modificação de estruturas conhecidas foi reportado pela comunidade científica (Cheng-Yu 2009) e (Murphy 2009), ainda existe o maior desafio e objetivo da engenharia de proteínas que é o design de moléculas totalmente novas. Uma proteína pode ser considerada nova porque, de fato, não existe ou apenas não teve a sua estrutura caracterizada pela biologia estrutural (Trevizani 2011, p.3).

## 3.2 Estrutura da Proteína

As primeiras estruturas em raio-x revelaram que as proteínas não adotam uma estrutura regular ou simétrica, mas, ao contrário, eram extremamente complexas. Contudo, foi observado que certas estruturas de proteínas eram muito frequentes. Estas constituem a estrutura secundária de uma proteína; a estrutura primária é a própria sequência de aminoácido.

Se ignorar pequenas variações nos ângulos definidos por duas ligações covalentes consecutivas, então a conformação de um resíduo de um aminoácido numa proteína pode ser classificada de acordo com os ângulos de torção. Existem 3 ângulos torsionais apelidados de  $\psi$ ,  $\phi$  e  $\omega$ . A conformação das cadeias laterais é caracterizada pelo ângulo de torção  $\xi_1, \xi_2$ . Os amidos possuem uma barreira energética relativamente alta para rotação; portanto  $\omega$  raramente desvia muito de 0 ou 180 graus. Além disso, a ligação entre o carbono da carbonila e o nitrogênio não é livre para girar porque apresenta um caráter parcial de dupla ligação. Por isso, os ângulos  $\psi\phi$  são geralmente considerados suficientes para reconstrução de uma estrutura (Holmes 2004).

### 3.2.1 Estrutura Primária

Proteínas são moléculas orgânicas formadas por um grupo de 20 aminoácidos unidos de forma contínua por uma ligação denominada ligação peptídica.

A ligação peptídica é uma ligação covalente formada por uma reação química entre dois aminoácidos liberando uma molécula de água. A desidratação dos aminoácidos na formação da ligação peptídica faz com que o termo resíduo seja preferencialmente utilizado para se referir a um aminoácido em um peptídeo (Trevizani 2011).

Em uma sequência, denomina-se esqueleto peptídico a parte comum a todos os resíduos: o grupamento amina ligada a um carbono- $\alpha$  e este, por sua vez, ligado à uma

carbonila e um átomo de hidrogênio. A identidade do resíduo é definida então pela cadeia lateral ligada ao carbono- $\alpha$  que determina suas propriedades como carga, hidropatia, hidrofobicidade, polaridade, etc. (Trevizani 2011)

### 3.2.2 Estrutura Secundária

Os aminoácidos se distribuem em estruturas formadas por padrões repetidos de pontes de hidrogênio, denominadas estruturas secundárias e constituídas em sua maioria por  $\alpha$ -hélices<sup>5</sup> e folhas- $\beta$ <sup>6</sup>. Existem determinadas composições entre os valores dos ângulos diedrais da cadeia principal para que a região da sequência assumam um destes arranjos. Tais valores foram determinados por Ramachandran a partir de modelos com esferas rígidas, o que permite, por exclusão estérica, eliminar valores do par  $(\psi, \phi)$  que sejam desfavoráveis à formação de estruturas secundárias.

Linus Pauling previu que as estruturas  $\alpha$ -hélices são extremamente estáveis; sua predição foi baseada numa análise cuidadosa dos modelos mecânicos existentes após observação da geometria dos peptídeos unidos a estruturas cristalizadas de moléculas menores e pode ser considerado um exemplo clássico do poder de previsibilidade da modelagem mecânica molecular.

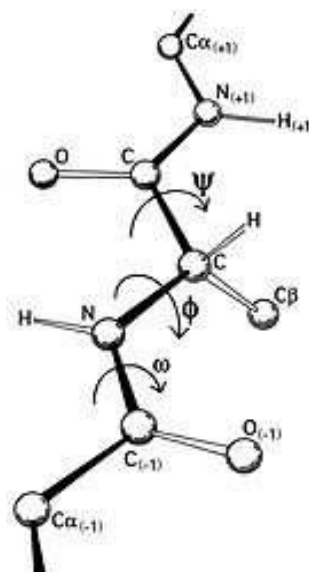


Figura 3.1: Ângulos representados pelas letras gregas. Neste caso, as letras gregas não possuem relação com a nomenclatura IUPAC dos elementos químicos de uma ligação, descrita na seção 2.6.

#### 3.2.2.1 $\alpha$ -hélice

São estruturas helicoidais, geralmente do sentido horário e estabilizadas por pontes de hidrogênio entre os grupamentos C=O e N-H de resíduos relativamente próximos. A parte interna do cilindro delimitado por  $\alpha$ -hélices é formada por pontes de hidrogênio (Creighton 1992).

<sup>5</sup>  $\alpha$ -helix

<sup>6</sup>  $\beta$ -strands

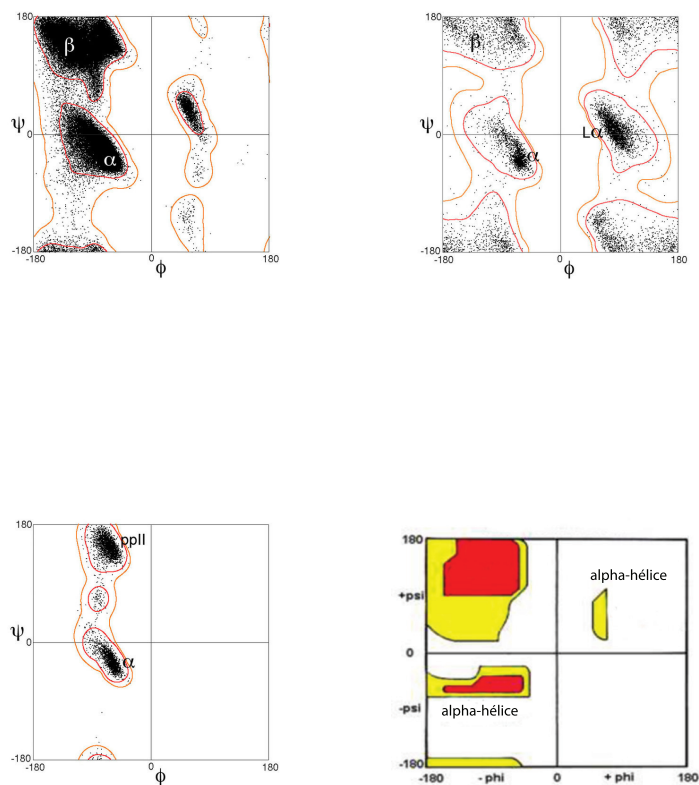


Figura 3.2: Gráfico de Ramachandran mostra as regiões favoráveis(vermelho) e permitidas em caso de relaxamento estérico(amarelo) para formação de estruturas secundárias (Trevizani 2011, p.5). No topo esquerdo é ilustrado o caso geral. Ao lado a glicina e na parte inferior esquerda temos a prolina. Repare a quantidade de folhas betas e alfa-hélices indicada em cada uma das imagens.

### 3.2.2.2 Folhas $\beta$

As folhas- $\beta$ s frequentemente formam estruturas estáveis e planas unidas uma sobre a outra através de pontes de hidrogênio entre os grupamentos amina e carbonila de resíduos distantes da sequência sobre um fluxo paralelo ou anti-paralelo. Uma fita- $\beta$  não é uma estrutura estável, sendo estabilizada somente pelas interações com outras fitas

### 3.2.3 Estrutura Terciária e Quaternária

Refere-se à condição tridimensional de uma só cadeia peptídica, onde as estruturas secundárias são compactadas e rearranjadas pelo processo de enovelamento.

# Capítulo 4

## Mapeamento do Campo de Força

"Aqui, sem dúvida, os leitores duvidarão, e virão muitas objeções a seu espírito; rogo-lhes que avancem a passos lentos comigo e que não formulem seu juízo antes que tenham lido tudo." (Baruch Spinoza)

"A force field can contain a large number of parameters, even if it intended for calculations on only a small set of molecules. Parametrisation of a force field is not a trivial task. A significant amount of effort is required to create a new force field entirely from scratch, and even the addition of few parameters to an existing force field in order to model a new class of molecules can be a time-consuming procedure." (Leach)

## 4.1 Modelo Quântico

A Mecânica Quântica usa equações de física quântica para calcular as propriedades de uma molécula a partir dos elétrons e núcleo. O movimento dos elétrons é considerado mais rápido e independente do núcleo e, uma vez que os elétrons giram em torno do núcleo, é possível descrever a energia eletrônica independente da energia nuclear. Nos cálculos de mecânica quântica também são feitas aproximações baseadas em dados empíricos e podem ser subdivididos em dois métodos: *ab initio* e *semi-empírico*.

Método *ab initio* pode ser aplicado apenas a moléculas pequenas e, apesar de mais preciso e não necessitar de dados armazenados, requer grande quantidade de memória e tempo de cálculo, até mesmo quando ignorado os elétrons, como no modelo quântico semi-empírico.

O método *semi-empírico* é menos exato, porém mais rápido e pode ser utilizado na minimização de energia e otimização de moléculas que variam de 10 a 120 átomos. A equação de Schrödinger pode fornecer a energia utilizada através dos parâmetros armazenados. Em alguns casos, porém, os campos de forças<sup>1</sup> podem fornecer respostas semelhantes em menos de uma fração de tempo que daria se utilizasse a mecânica quântica (Leach 2001, p.165) (Carvalho\*, Pupo e Bernardes 2003).

As propriedades da mecânica molecular são verdadeiras devido à validade de diversas outras propriedades. A primeira delas é a aproximação de *Born-Oppenheimer*, que seria impossível contemplar escrevendo energia como função de coordenadas nucleares (Leach 2001, p.165).

## 4.2 Modelo Mecânico Molecular

Na mecânica molecular as moléculas são descritas como um conjunto de "esferas conectadas", ao invés de núcleos e elétrons, como acontece nos modelos quânticos. O modelo clássico da mecânica molecular considera um sistema como um conjunto de esferas (átomos) unidas por molas, como exemplificado na figura 4.1.

O modelo da mecânica molecular é justificável porque os parâmetros associados a conjuntos de átomos permanecem razoavelmente constantes entre estruturas diferentes, desde que o tipo e a hibridação dos átomos envolvidos sejam os mesmos (SantAnna 2009). Na seção 4.9 será explicado melhor o conceito de hibridação.

---

<sup>1</sup>Force-Field



Contudo, a mecânica molecular não pode oferecer respostas evidentemente quando as propriedades dependem da distribuição eletrostática ao redor da molécula (Leach 2001, p.165).

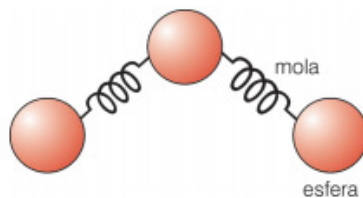


Figura 4.1: Modelo Massa-mola, onde cada esfera representa um átomo e a mola, as ligações intermoleculares.

Existe uma função que descreve como ocorrem as mudanças energéticas numa molécula; este modelo é descrito através das interações entre átomos ligantes (ligações covalentes) e não-ligantes (Van der Waals, Lei de Coulomb). Existem modelos mais sofisticados, mas eles sempre terão essas quatro componentes: ligações entre átomos, ligações entre moléculas, ângulos e torções. Considera-se aqui a fórmula a seguir (Leach 2001, p.166):

$$v(r^n) = \sum_{bonds} \frac{k_i}{2} (l_i - l_{i,0})^2 + \sum_{angle} \frac{k_i}{2} (\Theta_i - \Theta_{i,0})^2 + \sum_{torsion} 1 + \cos(\eta\omega - \gamma) \quad (4.1)$$

$$+ \sum_{i=1}^N \sum_{j=1}^N (4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}})$$

Onde  $v(r^N)$  representa a configuração  $R$  de  $N$  átomos e  $v$  representa a energia potencial do sistema.

Esta fórmula representa a conjunção de várias partes distintas de um sistema; no primeiro termo é visualizada a interação entre pares de átomos ligantes<sup>2</sup>, modelado com o potencial harmônico, que descreve o crescimento da energia em função da distância  $l_0$  em relação a um valor referencial  $l_{i,0}$ . O valor  $l_{i,0}$  não é conhecido à priori e necessita ser descoberto (parametrizado). Ele representa o valor referencial ou valor de repouso. O valor cuja resultante de forças é nulo. É possível usar valores parametrizados previamente pela comunidade científica como descrito no artigo (Allen et al. 1987). Desta forma pode ser definido, a priori, o valor referencial da distância.

---

<sup>2</sup>Bonded Atoms

A segunda parte é a soma de todas as energias potenciais referentes aos ângulos de ligação entre três átomos conectados; novamente utiliza-se o potencial harmônico. O valor  $\Theta_{i,0}$  representa o valor angular referencial e, como já descrito, descreve o valor cuja resultante de forças é nulo. Este valor também precisa ser parametrizado ou previamente definido.

O terceiro termo é a torção<sup>3</sup>, que modela a variação de energia em função da rotação dos átomos ligantes<sup>4</sup>; desta forma envolve sempre a interação entre 4 átomos. O valor  $\gamma$  representa o valor referencial e da mesma forma precisa ser parametrizado ou previamente definido.

A quarta contribuição é de átomos não-ligantes<sup>5</sup>. Ela é calculada entre pares de todos os átomos do sistema e representa, respectivamente, os termos potenciais de Lennard-Jones e lei de Coulomb para interações eletrostáticas.

O objetivo da mecânica molecular é prever a energia associada com determinada conformação de uma molécula. Porém, como afirma (Carvalho\*, Pupo e Bernardes 2003), não expressa quantidades absolutas, apenas diferenças de energia entre duas ou mais conformações (Patrick 1995).

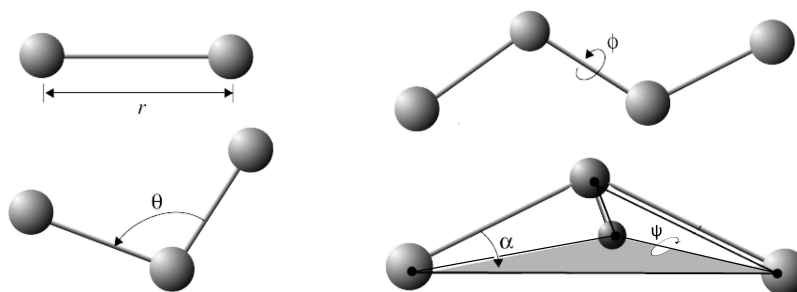


Figura 4.2: Gráfico que mostra os parâmetros avaliados pela fórmula AMBER.

## 4.3 Ligações químicas

São atrações estabelecidas entre átomos que permitem a formação de substâncias químicas que contem um ou mais átomos. A ligação é causada pela força de atração eletromagnética entre cargas diferentes. Cargas opostas se atraem via força eletromagnética, a carga

---

<sup>3</sup>Torsional potentials

<sup>4</sup>Bonded Atoms

<sup>5</sup>Non-bonded term

negativa do elétron e a carga positiva do núcleo se atraem mutuamente.

## 4.4 Comprimento da ligação

O primeiro parâmetro a ser determinado na fórmula *AMBER* tem o nome de **bond-stretching** ou simplesmente **bond** e pode ser compreendido como a energia de ligação entre pares de átomos ligantes. As ligações podem ser estendidas ou comprimidas assim como uma mola. Existe um ponto de repouso onde a resultante das forças nos dois sentidos é nula. Este valor é representado pelo gráfico 4.4 na região que a curva traz o valor de  $v(l)$  mínimo. Este valor mínimo ou **referencial** representa o **valor parametrizado de distância**.

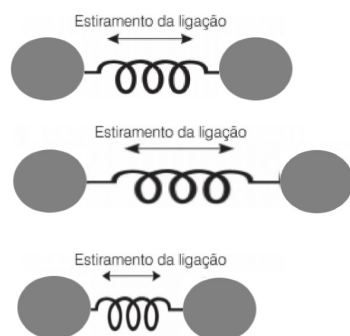


Figura 4.3: O comprimento da ligação. A compressão produz uma forte força repulsiva, enquanto que o estiramento produz uma força que vai se reduzindo em razão da distância.

Tabela 4.1: Tabela de Comprimento de Ligação e Energia

Ligação	$l_0$	Energia Potencial em K(Kcal $mol^{-1}deg^{-1}$ )
$Csp^3 - Csp^3$	1.523	317
$Csp^3 - Csp^2$	1.497	317
$Csp^3 = Csp^2$	1.337	690
$Csp^3 = O$	1.208	777
$Csp^3 = Nsp^3$	1.438	367
$Csp^3 = N(amide)$	1.345	719

A tabela 4.1 mostra que o estiramento e a compressão produzem diferentes valores de energia potencial. Isto significa: se um átomo estiver muito próximo de outro átomo haverá uma forte força de repulsão. Isso ocorre devido ao fato de existirem dois elétrons com mesmo spin em regiões de alta probabilidade de ocorrência. Da mesma forma existe uma força de atração devido à ligação molecular.

A ligação molecular é caracterizada pelo compartilhamento de pares de elétrons entre átomos ligantes. A estabilidade das forças repulsivas e atrativas entre átomos quando eles compartilham elétrons é conhecida como ligação covalente. É o principal tipo de ligação entre os compostos orgânicos. A força dessas ligações é maior que das interações intermoleculares e comparável à da ligação iônica.

A curva de energia potencial para uma ligação molecular<sup>6</sup> assume a forma descrita na figura 4.4. A função que descreve essa curva é definida pela energia potencial de Morse.

$$v(l) = D_e \{1 - \exp[-a(l - l_0)]\}^2 \quad (4.2)$$

Podemos visualizá-la através do gráfico 4.4.

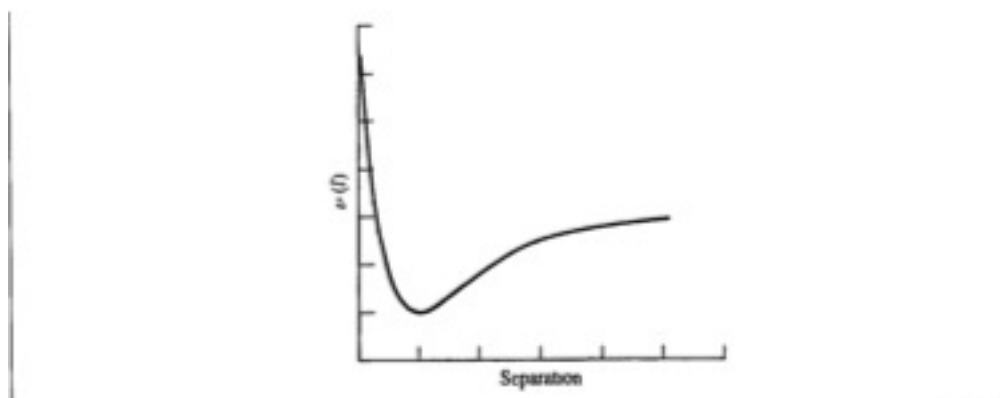


Figura 4.4: Variação da energia de ligação em razão da distância interatômica (Leach 2001, p.170)

Neste caso,  $D_e$  representa o potencial de energia mínimo e  $a = \omega \sqrt{\frac{\mu}{2D_e}}$ , onde  $\mu$  é a massa reduzida e  $\omega$  representa a frequência de vibração,  $\omega = \sqrt{\frac{k}{\mu}}$ ,  $k$  é a constante da ligação (Nelson e Cox 2010, p.170).

De forma geral, o potencial Morse é computacionalmente mais caro que o potencial harmônico (Hooke), embora seja mais preciso. O potencial Morse tem outra desvantagem: ele permite que a ligação se estenda até um valor não-realístico, visto que a força torna-se muito pequena em razão da distância, como é possível visualizar o gráfico 4.2. A desvantagem do potencial harmônico é que as forças para distâncias mais longas são extremamente altas provocando instabilidade no sistema (Nelson e Cox 2010, p.170)..

<sup>6</sup>Bond-stretching

A aproximação mais usada é a lei de Hooke, também conhecida por ser a lei que descreve as energias envolvidas em problemas massa-mola. O grande problema da utilização da lei de Morse é que ela exige um custo computacional elevado quando existe um grande número de átomos a ser computado.

## 4.5 Ângulo de ligação

Este tipo de força também pode ser simplificado com a aplicação da lei de hooke ou potencial harmônico (Leach 2001, p. 173).

$$v(\theta) = \frac{k}{2}(\theta - \theta_0)^2 \quad (4.3)$$

A contribuição de cada ângulo é caracterizada por uma força constante e um valor referencial. A variação do ângulo, em relação ao seu valor de referencial, promove valores de energia distintos.

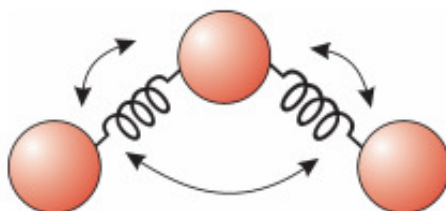


Figura 4.5: Ângulo átomos

A tabela 4.2 mostra valores de energia para um conjunto de ângulos de rotação.

Tabela 4.2: Tabela de Ângulos e Energia

Ângulo entre ligações	$\theta_0$	Energia potencial em K(Kcal $mol^{-1}deg^{-1}$ )
$Csp^3 - Csp^3 - Csp^3$	109.47	0.0099
$Csp^3 - Csp^3 - H$	109.47	0.0079
$H - Csp^3 - H$	109.47	0,0070
$Csp^3 - Csp^2 - Csp^3$	117.2	0.0099
$Csp^3 - Csp^2 = Csp^2$	121.4	0.0121
$Csp^3 - Csp^2 = O$	122.5	0.0101

## 4.6 Ângulo de torção

O valor denominado torção envolve o ângulo do plano formado por três átomos sobre um quarto átomo ligante. Na figura 4.7, a esfera A1 exerce torção sobre a esfera A4 através de A2 e A3, da mesma forma, mas em sentido contrário a esfera A4 exerce torção sobre a esfera A1 através de A3 e A2.

Possui pouco grau de liberdade e um valor de energia muito alto deve ser fornecido para provocar mudanças em relação ao seu valor referencial (Leach 2001, p. 173).

A energia potencial para torção geralmente é expressa em termos da série de *Fourier*:

$$E_w = \sum \frac{1}{2} [V_1(1 + \cos w) + V_2(1 + \cos 2w) + V_3(1 + \cos 3w) + \dots] \quad (4.4)$$

O somatório geralmente é truncado no terceiro termo. A variação do ângulo produz diferentes valores de energia como se pode constatar no gráfico 4.7.

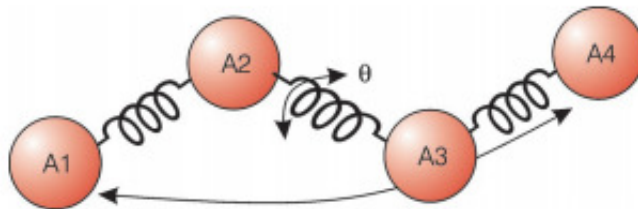


Figura 4.6: Ângulo de torção entre 4 átomos

## 4.7 Forças de Van der Waals

É um tipo de força elétrica relativamente fraca que age entre moléculas orgânicas e sólidas. Este tipo de força pode surgir entre as moléculas de alguns materiais, que embora eletricamente neutro, podem apresentar dipolos elétricos permanentes.

Devido à distribuição da carga elétrica na estrutura de algumas moléculas, um lado da molécula será um pouco positivo e o outro será um pouco negativo. A tendência de formação de dipolos permanentes resulta numa rede de forças atrativas.

A presença de moléculas que são dipolos permanentes distorce a carga eletrônica de alguma outra molécula próxima polar ou não-polar.

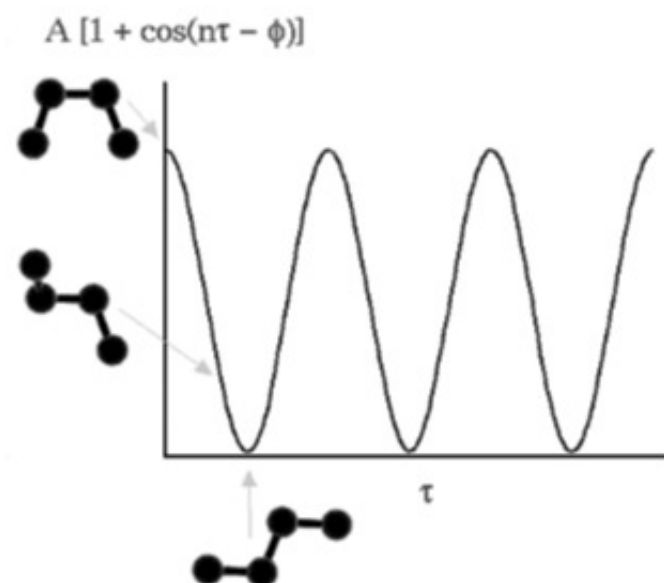


Figura 4.7: Observe através do gráfico que a variação do ângulo torsional provoca uma grande variação de energia.

Observe o deslocamento da nuvem eletrônica na figura 4.8 fazendo com que uma parte seja eletricamente positiva e a outra negativa para cada um dos átomos.

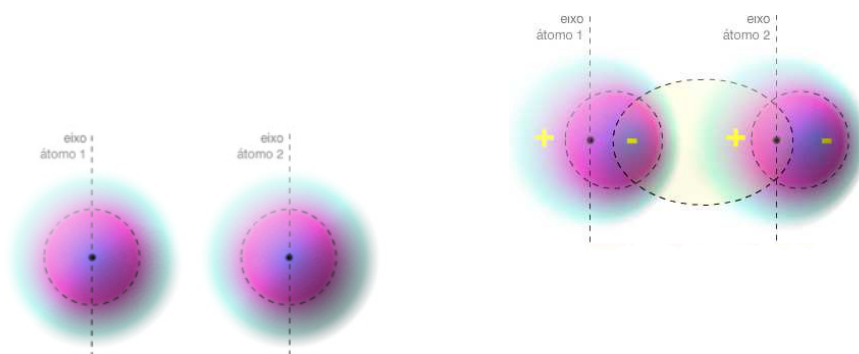


Figura 4.8: Representação das forças de Van der Waals.

As interações eletrostáticas não podem ser responsáveis por todos os tipos de interações entre átomos não-ligantes. Existe um segundo tipo de interação denominado *Forças de Van der Waals*. Ela ocorre entre átomos neutros, átomos de gases nobres (He, Ne, Ar, Kr, Xe, Rn), entre moléculas não-polares e, de forma geral, em todos os líquidos e sólidos (Leach 2001, p. 204).

A figura 4.9 ilustra como o raio de Van der Waals de um átomo representa seu tamanho real. Se dois átomos não ligados são aproximados, a atração de van der Waals entre eles aumenta. Quando a distância entre eles é igual à soma dos raios de van der

Waals a atração é máxima. Se os átomos são aproximados ainda mais, ocorre uma forte repulsão (RODRIGUES 2001).

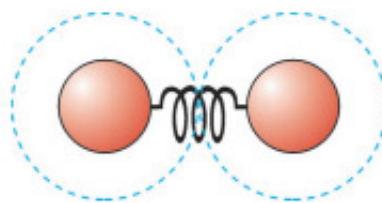


Figura 4.9: Van der Waals

## 4.8 Juntando as peças

A mecânica molecular considera um modelo onde existem diversas contribuições energéticas individuais que, em conjunto, produzem um valor energético resultante. Veja um exemplo de um esquema onde se ilustra essas diversas contribuições. Considere a molécula de propano:

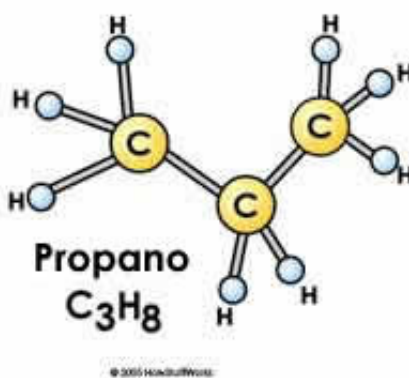


Figura 4.10: Molécula de Propano

- 10 ligações entre átomos ligantes<sup>7</sup>: duas C-C e 8 ligações C-H
- 18 valores de rotação<sup>8</sup> considerando 3 pares de átomos: 1 ângulo C-C-C, 10 ângulos C-C-H e 7 ângulos H-C-H
- 18 valores torsionais<sup>9</sup>: 12 torsões H-C-C-H e 6 torsões H-C-C-C

<sup>7</sup>Bond-Stretching ou Ligações covalentes

<sup>8</sup>Angle bending

<sup>9</sup>Bend-rotation ou torção



- 27 interações entre átomos não-ligantes<sup>10</sup>: 21 interações H-H e 6 interações C-H
- As contribuições eletrostáticas são contabilizadas pela Lei de Coulomb
- Van der Waals são contabilizadas utilizando o potencial Lennard-Jones de cada átomo com  $\epsilon_{ij}$  e  $\sigma_{ij}$  apropriado

Um número muito grande de variáveis é incluído até mesmo num simples cálculo de uma molécula simples de propano. Não existem ligações iônicas na molécula de propano. Todas as ligações nessa molécula são moleculares.

As ligações moleculares, por serem mais fortes, afetam abruptamente o resultado do cálculo da energia potencial, enquanto que as ligações de van der waals não exercem grande influência.

De qualquer forma é evidente que este problema envolve um número muito menor de calculo do que pelo método quântico (Leach 2001, p.167).

### 4.8.1 Hibridação

O modelo mecânico não pode ignorar completamente os elétrons (Leach 2001, p.169). Embora a carga do spin e a multiplicidade não sejam realmente requeridos, a hibridação provoca consequências na geometria das moléculas. Desta forma, a hibridação altera os valores médios de ângulos das ligações moleculares (valor referencial). A hibridação ocasiona uma deformação geométrica entre ângulos de átomos ligantes.

Uma segunda consequência da hibridação é que ela torna possível a formação de múltiplas ligações. Também torna possível a existência de ligações duplas e triplas. No caso de proteínas puras, não existe um caso de hibridação com ligações triplas, nem duas ligações duplas para o mesmo átomo porque não existem aminoácidos com essas características. As proteínas puras são aquelas formadas exclusivamente pelos 20 aminoácidos descritos na figura 2.4.

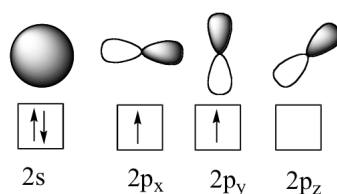


Figura 4.11: Distribuição dos elétrons nos orbitais do elemento químico carbono (Valence bond theory: sp, sp<sup>2</sup>, and sp<sup>3</sup> hybrid orbitals 2012).

<sup>10</sup>non-bonded term

### 4.8.1.1 Hibridação $sp^3$

A figura 4.11 demonstra que o carbono com distribuição eletrônica  $1s^2 2s^2 2p_x^1 2p_y^1$  é capaz de formar duas ligações com outros átomos devido à presença de dois orbitais vazios  $2p_z^1 2p_y^1$ . Entretanto observa-se, experimentalmente, que o carbono é capaz de realizar quatro ligações com outros átomos. Também se verifica um arranjo tetraédrico e que as ligações possuem a mesma força e comprimento.

O conceito denominado hibridação define como um átomo de carbono consegue realizar mais de duas ligações. Na hibridação, o orbital  $s$  cede um dos seus elétrons para o orbital  $2p_z^1$  formando o orbital híbrido  $1s^2 2s^2 2sp^3$

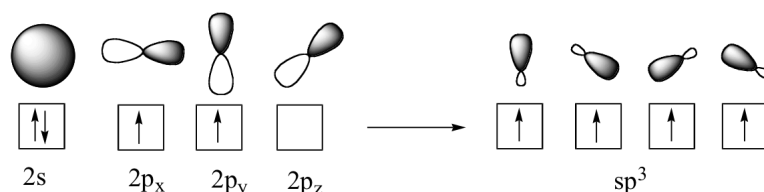


Figura 4.12: Hibridação  $sp^3$  (Valence bond theory: sp, sp<sup>2</sup>, and sp<sup>3</sup> hybrid orbitals 2012).

Esse tipo de hibridação possibilita que o átomo de carbono se relacione com mais quatro outros átomos através de ligações simples.

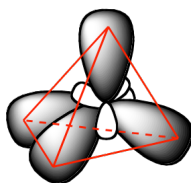


Figura 4.13: Arranjo tetraédrico da hibridação  $sp^3$  possibilita a formação de quatro ligações (Valence bond theory: sp, sp<sup>2</sup>, and sp<sup>3</sup> hybrid orbitals 2012).

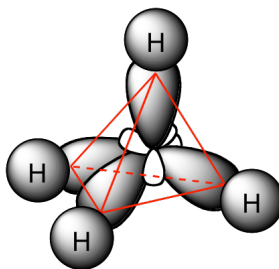


Figura 4.14: Átomo de carbono conectado a 4 átomos de hidrogênios através da hibridação  $sp^3$  (Valence bond theory: sp, sp<sup>2</sup>, and sp<sup>3</sup> hybrid orbitals 2012).

### 4.8.1.2 Hibridação $sp^2$

Este tipo de hibridação ocorre quando um orbital  $s$  combina com dois orbitais  $p$  conforme a figura 4.16, 1 elétron é cedido para o último orbital. Diferente da hibridação  $sp^3$ , a hibridação  $sp^2$  é planar.

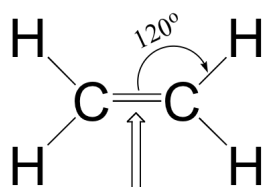


Figura 4.15: Hibridação  $sp^2$  (Valence bond theory: sp, sp2, and sp3 hybrid orbitals 2012).

A figura 4.15 mostra um exemplo de hibridação do tipo  $sp^2$  na molécula de eteno. A hibridação do tipo  $sp^2$  também permite a formação de ligações duplas.

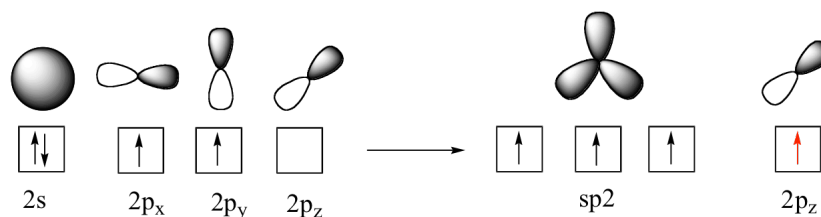


Figura 4.16: Hibridação  $sp^2$  (Valence bond theory: sp, sp2, and sp3 hybrid orbitals 2012).

Esse tipo de ligação é mais forte e altera o comprimento da ligação entre átomos conectados por ligações covalentes.

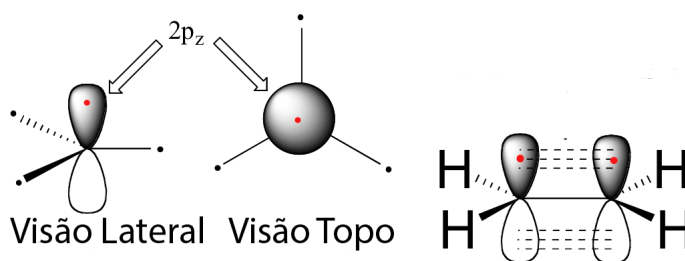


Figura 4.17: Duas imagens que representam a hibridação  $sp^2$ . Ao lado esquerdo a visão lateral e superior do átomo no estado hibridizado  $sp^2$ . À direita, a molécula de eteno com uma ligação dupla entre os dois átomos de carbono (Valence bond theory: sp, sp2, and sp3 hybrid orbitals 2012).

### 4.8.1.3 Determinação do estado da Hibridação

Para determinar o estado de hibridação de um átomo é necessário observar os ângulos de conexão com outros átomos. Entretanto, como proteínas puras só possuem os 20 aminoácidos conhecidos é possível pressupor o estado de hibridação de um átomo somente pelo número de ligações que eles formam.

Essa otimização é possível porque não existem nas proteínas puras aminoácidos com ligações triplas ou duas ligações duplas consecutivas, nem hibridação  $sp^1$ .

Examinar o número de conexões de 1 átomo é muito menos custo computacionalmente que avaliar o ângulo entre 3 átomos. O número de conexões de cada átomo já é uma informação existente na estrutura do Grafo e não precisa ser calculada. Desta forma, apenas analisando o número de ligações de um determinado átomo é possível afirmar qual o estado de hibridação deste elemento.

A distinção do estado de hibridação é necessária porque altera os ângulos e o comprimento das ligações, alterando os valores referenciais da parametrização.

---

**Algorithm 1** Algoritmo para determinação do estado de hibridação de um átomo

---

```

numeroAtómicoCarbono  $\leftarrow$  6
hibridacaoSP3  $\leftarrow$  4
hibridacaoSP2  $\leftarrow$  3
hibridacaodesconhecido  $\leftarrow$  0
for proteína = 0 to totalProteínas do
  for átomo = 0 to totalÁtomos(proteína) do
    if ( numeroAtómico(getNomeÁtomo(proteína,átomo)) = numeroAtómicoCarbono ) then
      numeroConexões  $\leftarrow$  getNumConexões(proteína,átomo)
      if ( numeroConexões = 3 ) then
        SetHibridacao(proteína,átomo,hibridacaoSP2)
      else if ( numeroConexões = 4 ) then
        SetHibridacao(proteína,átomo,hibridacaoSP2)
      else
        SetHibridacao(proteína,átomo,hibridacaodesconhecido)
      end if
    end if
  end for
end for

```

---

O algoritmo 1 deve ser aplicado somente com proteínas puras para dedução do estado de hibridação.

## 4.9 Tipo de Átomo

Um conceito comum na maioria dos campos de forças é o conceito do **tipo de átomo** (Leach 2001, p.169). Quando se preparam os valores para o cálculo através da mecânica quântica é necessário especificar não só o número atômico, como também a geometria, carga e multiplicidade do spin. Para um campo de força entretanto a carga e a multiplicidade do spin não é explicitamente exigida, mas frequentemente será necessário assinalar o tipo de cada átomo no sistema.

O tipo de átomo é mais do que o número atômico, ele contém informações como hibridação<sup>11</sup> e algumas vezes informações locais. Por exemplo, na maioria dos campos de forças é necessário distinguir entre o carbono hibridizado  $sp^3$  que denota a geometria tetraédrica,  $sp^2$  que conduz a geometria tridiagonal, a hibridação  $sp^1$  não ocorre em proteínas puras.

Cada parâmetro de um campo de força é expresso em termos destes tipos de átomos, então o ângulo de referência  $\Theta_0$  para o carbono tetraédrico será 109,5 graus enquanto que para o tridiagonal será 120 graus.

Os tipos de átomos refletem os campos de força do ambiente vizinho. Por exemplo, *MM2*, *MM3* e *MM4* e *Allinger* que são amplamente utilizados para o cálculo de moléculas pequenas. Estes campos distinguem os tipos de átomos:  $sp^3$ ,  $sp^2$ ,  $sp$ , carbonila, ciclopropano e íon carbônico.

## 4.10 Protein Databank

A determinação da geometria da molécula é de vital importância para o entendimento das estruturas químicas das moléculas. A maior parte dos dados experimentais são obtidos a partir da *cristalografia por raio-X*<sup>12</sup> e *difração de neutrons*<sup>13</sup>, *espectroscopia por microondas*<sup>14</sup> e *difração de elétrons*<sup>15</sup>.

Inicialmente essas publicações eram esporádicas mas nos últimos anos essas publicações se tornaram mais densas e frequentes. O resultado é que existem hoje grandes bancos de dados sequenciais com informação de mais de 50000 compostos orgânicos (CSD Cam-

---

<sup>11</sup>hybridisation

<sup>12</sup>X-Ray crystallography

<sup>13</sup>Neutron diffraction

<sup>14</sup>Microwave spectroscopy

<sup>15</sup>Electron diffraction

bridge) (Berman et al. 1999, p.235-p.242). Em 2002, três bancos de dados sequenciais de proteínas - o *Protein Information Resource* no *National Biomedical Research Foundation* de Georgetown, *SWISS-PROT* e o *TrEMBL* do *Swiss Institute of Bioinformatics* de Genebra coordenaram seus esforços para formar o consórcio *UniProt*. Os parceiros desta iniciativa compartilham o banco de dados, mas continuam a oferecer ferramentas separadas de acesso à informação (Lesk 2008, p.144).

Já os bancos de dados estruturais anotam e distribuem conjuntos de coordenadas atômicas. O principal banco de dados de estrutura de macromoléculas biológicas é o *protein Databank* (PDB). Ele contém estruturas de proteínas e ácidos nucleicos.<sup>16</sup>

A figura 4.18 mostra um exemplo de arquivo PDB. Um arquivo PDB pode conter:

- Informações sobre a proteína tais como grupos funcionais, fórmula
- Quem determinou a estrutura da proteína
- Detalhes experimentais sobre a determinação estrutural da proteína
- Sequência de aminoácidos
- Co-fatores, inibidores e solvente, como moléculas de água
- Atribuições como estrutura secundária: hélice e folha
- Pontes dissulfeto
- Coordenadas atômicas

---

<sup>16</sup><http://www.pdb.org>

```

HEADER      TRANSFERASE                                18-DEC-08    2ZWV
TITLE       CRYSTAL STRUCTURE OF THERMUS THERMOPHILUS 16S RRNA
TITLE       2 METHYLTRANSFERASE RSMC (TTHA0533)
(...)
COMPND      2 MOLECULE: PROBABLE RIBOSOMAL RNA SMALL SUBUNIT
COMPND      3 METHYLTRANSFERASE;
(...)
SOURCE      2 ORGANISM_SCIENTIFIC: THERMUS THERMOPHILUS;
SOURCE      3 ORGANISM_TAXID: 300852;
(...)
AUTHOR      H.WANG,M.KAWAZOE,T.KAMINISHI,A.TATSUGUCHI,C.NAOE,T.TERADA,
AUTHOR      2 M. SHIROUZU,C.TAKEMOTO,S.YOKOYAMA,RIKEN STRUCTURAL
(...)
REMARK 200  EXPERIMENTAL DETAILS
REMARK 200  EXPERIMENT TYPE                      : X-RAY DIFFRACTION
REMARK 200  DATE OF DATA COLLECTION              : 03-MAR-07
REMARK 200  TEMPERATURE                          (KELVIN) : 100
REMARK 200  PH                                    : 8.9
REMARK 200  NUMBER OF CRYSTALS USED               : 1
(...)
ATOM        1  N   SER A   2           45.623  19.967  23.987  1.00  52.48
ATOM        2  CA  SER A   2           46.685  19.984  22.942  1.00  50.33
ATOM        3  C   SER A   2           46.300  20.908  21.788  1.00  51.12
ATOM        4  O   SER A   2           46.924  21.952  21.577  1.00  52.96
ATOM        5  CB  SER A   2           46.920  18.565  22.417  1.00  47.98
ATOM        6  OG  SER A   2           45.715  17.994  21.938  1.00  49.64
ATOM        7  N   LEU A   3           45.267  20.521  21.045  1.00  44.14
ATOM        8  CA  LEU A   3           44.806  21.318  19.917  1.00  39.44
ATOM        9  C   LEU A   3           44.067  22.570  20.356  1.00  41.34
ATOM       10  O   LEU A   3           43.262  22.543  21.286  1.00  45.49
(...)
ATOM      2855  C   GLY A  374          36.607  38.031  32.464  1.00  35.87
ATOM      2856  O   GLY A  374          36.908  38.301  33.645  1.00  38.56
TER        2857          GLY A  374
HETATM 2858  N   SAH A  376          24.272  14.684  22.476  1.00  25.40
HETATM 2859  CA  SAH A  376          23.776  13.382  22.005  1.00  24.17
HETATM 2860  CB  SAH A  376          24.908  12.612  21.323  1.00  27.40
(...)

```

Figura 4.18: Exemplo de arquivo do *Protein Databank*

# Capítulo 5

## Hyle3D

Framework desenvolvido em Visual C++ com rotinas de importação, exportação e parametrização da distância entre átomos ligantes de arquivos de proteínas puras da base de dados pública, denominada protein Databank.

Cada arquivo do protein Databank representa uma molécula no programa, como ilustra a figura (??). Cada molécula é representada pelo framework como uma estrutura de dados denominada grafo. Os átomos e conexões representam, respectivamente, vértices e arestas deste grafo. O programa não acessa diretamente os arquivos; existe uma etapa inicial de importação para o banco de dados. Todos os dados ficam armazenados em tabelas do banco de dados conforme ilustrado no diagrama entidade-relacionamento 5.2.

### 5.1 Massa de dados

A massa de dados utilizada nesta dissertação são arquivos compactados no formato zip, obtidos a partir do protein databank. Esses arquivos são atualizados periodicamente e novos arquivos são adicionados.

Inicialmente foi realizada a extração de todas as proteínas no conjunto de arquivos PDB para remoção de moléculas de DNA e RNA. Na segunda etapa selecionou-se proteína que fosse composta exclusivamente dos 20 aminoácidos (proteínas puras). Trabalhar com proteínas puras torna a massa de dados mais homogênea. O resultado da parametrização da distância entre átomos ligantes demonstrou que não houve muita diferença de valor na amostra selecionada. A tabela 6.2 apresenta os resultados da parametrização.



## 5.2 Importação da massa de dados

A informação da conexão é adquirida em parte pelo arquivo PDB que registra através de um campo denominado CONECT. Entretanto o arquivo PDB só registra a conexão de heteroátomos. Para estabelecer conexões entre átomos foi utilizada uma segunda fonte de dados, que são os arquivos de monômeros.

Os arquivos de monômeros foram obtidos no mesmo repositório dos arquivos do protein Databank. Ele consta de diversos arquivos nomeados com 3 letras que representam o nome de algum monômero. Esta sigla está representada no campo NAME do arquivo PDB. Esta correlação foi feita dinamicamente a fim de evitar múltiplos acessos ao banco. Se cada molécula tivesse que representar todos os elementos químicos tornaria o arquivo PDB muito grande desnecessariamente.

A metodologia usada para filtrar estes resultados foi a avaliação do campo ATOM do arquivo PDB<sup>1</sup>. Proteínas são compostas de aminoácidos e a presença de qualquer um dos 20 aminoácidos no registro ATOM indica que a molécula descrita no arquivo em questão trata-se de proteína. Veja este trecho obtido a partir do arquivo de uma molécula de proteína *pdb1mzg.ent* :

```
ATOM 1 N LEU A 3 40.996 27.179 86.089 1.00 38.26 N
ATOM 2 CA LEU A 3 40.801 28.633 85.766 1.00 36.90 C
ATOM 3 C LEU A 3 39.742 29.231 86.682 1.00 35.10 C
ATOM 4 O LEU A 3 39.349 28.626 87.692 1.00 34.56 O
ATOM 5 CB LEU A 3 40.349 28.794 84.309 1.00 38.33 C
ATOM 6 CG LEU A 3 41.073 29.802 83.410 1.00 38.73 C
ATOM 7 CD1 LEU A 3 40.293 29.915 82.110 1.00 38.80 C
ATOM 8 CD2 LEU A 3 41.187 31.162 84.085 1.00 39.32 C
ATOM 9 N LEU A 4 39.265 30.420 86.323 1.00 31.97 N
ATOM 10 CA LEU A 4 38.234 31.064 87.128 1.00 27.89 C
ATOM 11 C LEU A 4 36.864 30.875 86.483 1.00 26.81 C
ATOM 12 O LEU A 4 36.722 30.882 85.251 1.00 24.38 O
ATOM 13 CB LEU A 4 38.534 32.562 87.298 1.00 28.13 C
ATOM 14 CG LEU A 4 39.839 32.923 88.040 1.00 29.66 C
ATOM 15 CD1 LEU A 4 40.013 34.431 88.043 1.00 30.77 C
ATOM 16 CD2 LEU A 4 39.796 32.381 89.467 1.00 25.51 C
```

<sup>1</sup>Para mais informações sobre o formato do arquivo PDB consulte <http://www.rcsb.org/pdb/home/home.do>

```
ATOM 17 N PRO A 5 35.840 30.661 87.309 1.00 23.48 N
ATOM 18 CA PRO A 5 34.501 30.481 86.740 1.00 22.89 C
ATOM 19 C PRO A 5 34.002 31.792 86.142 1.00 20.42 C
ATOM 20 O PRO A 5 34.521 32.849 86.476 1.00 19.78 O
ATOM 21 CB PRO A 5 33.675 30.044 87.949 1.00 22.73 C
ATOM 22 CG PRO A 5 34.388 30.738 89.104 1.00 21.96 C
ATOM 23 CD PRO A 5 35.829 30.477 88.771 1.00 23.07 C
```

Observe que no texto, a coluna indicada em negrito, corresponde ao atributo resíduo que contém os aminoácidos Leucina (LEU) e Prolina (PRO). Isso indica que o arquivo em questão é de uma proteína. Ao contrário da molécula *pdb1amd.ent* que possui os campos DT que indica uma timina de um ácido desoxiribonucléico (DNA):

```
ATOM 1 O5' DT A 1 30.742 21.350 17.515 1.00 23.97 O
ATOM 2 C5' DT A 1 31.166 20.901 16.230 1.00 23.64 C
ATOM 3 C4' DT A 1 31.575 19.410 16.330 1.00 23.10 C
ATOM 4 O4' DT A 1 30.576 18.613 16.981 1.00 22.32 O
ATOM 5 C3' DT A 1 31.989 18.679 15.051 1.00 23.29 C
ATOM 6 O3' DT A 1 33.407 18.521 15.211 1.00 24.92 O
ATOM 7 C2' DT A 1 31.073 17.415 15.038 1.00 21.79 C
ATOM 8 C1' DT A 1 30.553 17.265 16.466 1.00 20.16 C
ATOM 9 N1 DT A 1 29.157 16.799 16.638 1.00 0.00 N
ATOM 10 C2 DT A 1 28.889 15.524 17.123 1.00 0.00 C
ATOM 11 O2 DT A 1 29.765 14.674 17.269 1.00 0.00 O
ATOM 12 N3 DT A 1 27.558 15.268 17.426 1.00 0.00 N
ATOM 13 C4 DT A 1 26.509 16.158 17.276 1.00 0.00 C
```

O algoritmo 2 descreve os passos para seleção de proteínas puras a partir de arquivos PDB.

---

**Algorithm 2** Algoritmo para seleção de proteínas puras
 

---

```

while Existir arquivo do
  verificaProteina  $\leftarrow$  1
  for linha = 0 to totalLinhas(arquivo) do
    if tipoRegistro(arquivo[linha], 'ATOM') then
      if ( tipoResiduo(arquivo[linha], "Adenina", "Guanina", "Timina", "Cito-
        sina","Uracila" ) ) then
        verificaProteina  $\leftarrow$  0
      end if
      residuo = ['Asp', 'Glu', 'Ala', 'Arg', 'Asn', 'Cys', 'Phe', 'Gly', 'Gln', 'His', 'Ile',
        'Leu', 'Lys', 'Met', 'Pyl', 'Pro', 'Ser', 'Sec', 'Tyr', 'Thr', 'Trp', 'Val']
      if ( not tipoResiduo(arquivo[linha], aminoacidos ) ) then
        verificaProteina  $\leftarrow$  0
      end if
    else if tipoRegistro(arquivo[linha], 'HETATM') then
      if ( not tipoResiduo(arquivo[linha], 'HOH') ) then
        verificaProteina  $\leftarrow$  0
      end if
    end if
    if not verificaProteina then
      break
    end if
  end for
  if verificaProteina then
    carregar proteína
  end if
end while

```

---

## 5.3 Roteiro de instalação do software Hyle3D

Execute o script de criação de tabelas, veja o apêndice A<sup>2</sup>. O framework possui um sistema de conversão da massa de dados do protein Databank para arquivos CSV. Execute o comando LOAD DATA IN FILE do mySql selecionando o arquivo csv convertido correspondente à tabela<sup>3</sup>.

Importe os arquivos de monômeros a partir do site do Protein Databank. Estes arquivos possuem ligações internas de todos os monômeros que são utilizados pelo arquivo PDB. Os arquivos de monômeros são necessários porque o PDB não traz informação das conexões internas; apenas a informação dos heteroátomos é informada no PDB.

Após a importação a última etapa é a parametrização da distância entre átomos ligantes que é feita a partir da média de todos os valores de distância entre átomos ligantes levando em conta o número de ligações. O resultado desta query se encontra disponível na tabela 6.2.

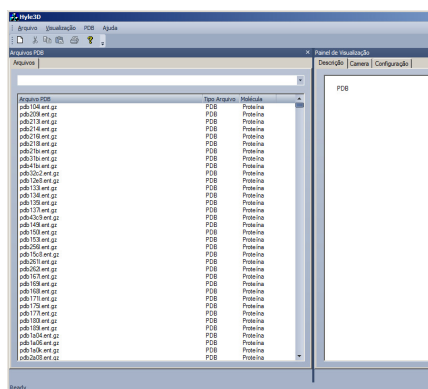


Figura 5.1: Esta figura mostra a tela principal do programa Hyle3D.

- **Painel de arquivos PDB** Painel onde são listados os arquivos do programa que serão usados na parametrização
- **Painel de visualização** Painel que disponibiliza informações estatísticas sobre cada molécula, uma aba para visualização da câmera para visualização em ambiente

<sup>2</sup>É necessário possuir os requisitos mínimos de memória e processador para execução do programa. Sistema operacional Windows 7 64 bits com quantidade de memória igual ou superior a 10 gigas para que o mySQL tenha um desempenho satisfatório.

<sup>3</sup>O comando LOAD DATA IN FILE possui desempenho 7 vezes superior ao INSERT.

tridimensional e uma aba de configuração.

No menu PDB, temos:

- **Importar PDB** Realiza a importação dos arquivos PDB para um arquivo CSV. Todos os arquivos PDB importados são condensados para um único arquivos CSV.
- **Importar Monômeros** Importa os arquivos de monômeros para o banco de dados. Esta operação só precisa ser realizada uma única vez.
- **Parametrizar** Realiza a parametrização dos valores de comprimento da ligação.

## 5.4 Modelo de dados do programa Hyle3D

### 5.4.1 Modelo Entidade-Relacionamento

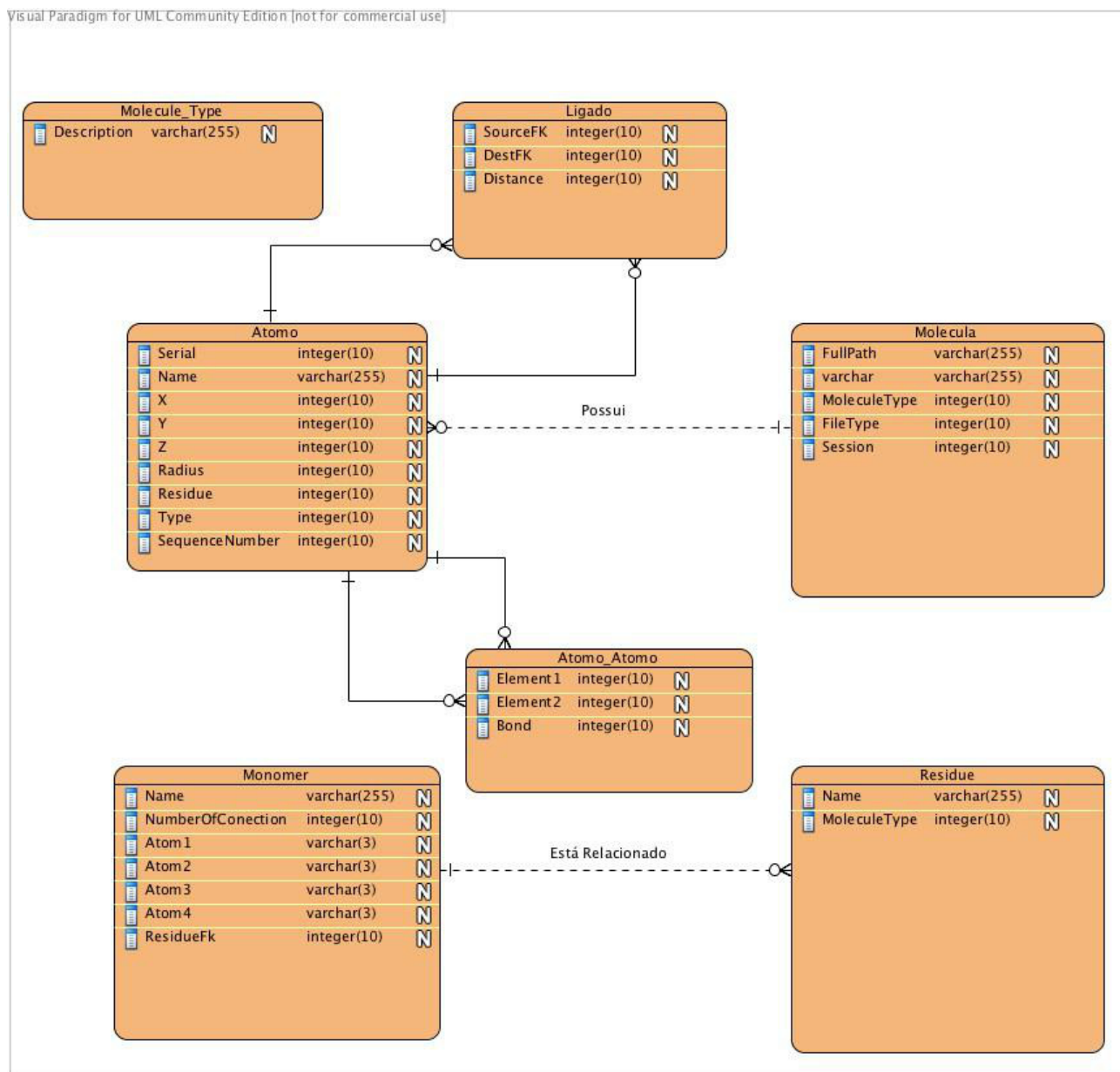


Figura 5.2: Diagrama Entidade-Relacionamento Hyle3D

## 5.4.2 Mapeamento Lógico-Relacional

### 5.4.2.1 Tabela Molecule

Tabela com informações sobre o arquivo PDB. Possui informações sobre localização do arquivo importado, o tipo de molécula e se a importação foi realizada com sucesso.

Tabela 5.1: Tabela com mapeamento lógico-relacional da entidade *Molecule*

	Atributo	Tipo	Descrição
1	ID	Inteiro	Chave-primária
2	Fullpath	Texto	Caminho com a localização dos arquivos PDBs.
3	Filename	Texto	Nome do arquivo.
4	MoleculeType	Inteiro	Chave-estrangeira para MoleculeType. 0=UNKNOWN, 1=RNA, 2=DNA, 3=PROTEIN
5	FileType	Inteiro	Tipo de arquivo. Compactado ou explodido.
6	Session	Inteiro	Garante que dados processados não sejam reprocessados em caso de falha.

### 5.4.2.2 Tabela Atom

Tabela que informa a posição cartesiana dos átomos. A unidade de medida é angstroms.

Tabela 5.2: Tabela com mapeamento lógico-relacional da entidade *Atom*

	Atributo	Tipo	Descrição
1	ID	Inteiro	Chave-primária
2	Serial	Inteiro	Representa um identificador do átomo no arquivo PDB, é um número sequencial para cada arquivo PDB e as conexões entre os átomos são feitas a partir deste identificador.
3	Name	Texto	Possui o nome do elemento químico no arquivo PDB mais a posição na cadeia. Utiliza a nomenclatura da <i>IUPAC</i> com algumas modificações.
4	Element	Inteiro	Número atômico do elemento químico.
5	X	Ponto-flutuante	Posição cartesiana do átomo no eixo X.
6	Y	Ponto-flutuante	Posição cartesiana do átomo no eixo Y.
7	Z	Ponto-flutuante	Posição cartesiana do átomo no eixo Z.
8	Radius	Ponto-flutuante	Distância em relação ao centro. Obsoleto <sup>4</sup> .
9	Molecule_FK	Inteiro	Chave-estrangeira para tabela Molecule. Relaciona um átomo a sua molécula correspondente.
10	Residue	Texto	Possui um texto informando qual o resíduo correspondente. No caso de DNAs e RNAs corresponde as siglas das bases purinas e pirimidinas. No caso de proteínas pode ser quaisquer um dos aminoácidos, além da água.
11	Type	Inteiro	Informa se o tipo de átomo foi adicionado dinamicamente. O arquivo PDB representa a molécula no estado não-protonado por isso alguns átomos não são representados. este campo informa se algum átomo faz parte da descrição do PDB.
12	SequenceNumber	Inteiro	Número sequencial do resíduo.



### 5.4.2.3 Tabela Bond

Esta tabela representa as ligações entre heteroátomos. O arquivo PDB não registra as ligações entre átomos de um mesmo resíduo. Para saber de que forma eles se ligam basta consultar a tabela de monômeros.

Tabela 5.3: Tabela com mapeamento lógico-relacional da entidade *Bond*

	Atributo	Tipo	Descrição
1	ID	Inteiro	Chave-primária
2	Molecule_FK	Inteiro	Chave-estrangeira para a tabela Molécula.
3	Source_FK	Inteiro	Chave-estrangeira para a tabela Atom. Representa um átomo participante de uma ligação molecular.
4	Dest_FK	Inteiro	Chave-estrangeira para a tabela Atom. Representa um átomo participante de uma ligação molecular.
5	Distance	Ponto-flutuante	Representa a distância molecular entre dois átomos ligantes.
6	Stretching	Ponto-flutuante	Informações de hibridação. Obsoleto <sup>5</sup> .

### 5.4.2.4 Tabela Molecule type

Tabela 5.4: Tabela com mapeamento lógico-relacional da entidade *Molecule\_Type*

	Atributo	Tipo	Descrição
1	ID	Inteiro	Chave-primária
2	Descrição	Texto	Informa o tipo de molécula. 0=UNKNOWN, 1=RNA, 2=DNA, 3=PROTEIN.

### 5.4.2.5 Tabela Monomer

Contém informações sobre todos os átomos dos resíduos presentes nas proteínas. O arquivo PDB só representa as ligações dos heteroátomos; entretanto é necessário fazer as conexões entre átomos ligantes.

Esta tabela informa todos os pares de ligações moleculares.

Tabela 5.5: Tabela com mapeamento lógico-relacional da entidade *Monomer*

	Atributo	Tipo	Descrição
1	ID	Inteiro	Chave-primária
2	Name	Texto	Nome do monômero.
3	NumberOfConection	inteiro	Número de conexões.
4	Atom1	Texto	Nome do átomo ligante no resíduo.
5	Atom2	Texto	Nome do átomo ligante no resíduo.
6	Atom3	Texto	Nome do átomo ligante no resíduo.
7	Atom4	Texto	Nome do átomo ligante no resíduo.
8	Residue_FK	Inteiro	Chave-estrangeira para o resíduo.

#### 5.4.2.6 Tabela Parameter Bond

Para obter o resultado da parametrização da distância entre dois átomos ligantes basta obter o valor médio das distâncias representadas entre dois átomos. A coluna Element1 e Element2 também informa o estado de hibridação para sejam distinguidos elementos químicos comuns, mas com estados de hibridação diferentes.

Tabela 5.6: Tabela com mapeamento lógico-relacional da entidade *Bond*

	Atributo	Tipo	Descrição
1	ID	Inteiro	Chave-primária
2	Element1	Texto	Número do primeiro Elemento Químico.
3	Element2	Texto	Número do segundo Elemento Químico.
4	Bond	Ponto-Flutuante	Distância entre dois elementos químicos.

#### 5.4.2.7 Tabela Residue

Esta tabela contém todos os resíduos utilizados pelas proteínas. É necessário consultar esta tabela porque o arquivo PDB representa a proteína no estado não-protonado. Isto significa que algumas vezes o elemento químico hidrogênio não é representado na molécula. O hidrogênio é um átomo extremamente pequeno e na maior parte das vezes a cristalografia por raio-x é incapaz de reconhecê-lo.

Tabela 5.7: Tabela com mapeamento lógico-relacional da entidade *Residue*

	<b>Atributo</b>	<b>Tipo</b>	<b>Descrição</b>
1	ID	Inteiro	Chave-primária
2	Name	Texto	Nome do resíduo
3	Molecule_Type	Inteiro	Tipo de molécula

## Capítulo 6

### Conclusões e Perspectivas

Nesta dissertação foi implementado um framework em C++ denominado *Hyle3D* para representação de proteínas puras em estrutura de dados denominada Grafo, a partir dos arquivos do banco de dados público do *proteinDatabank*. Foi realizado um teste com a parametrização do valor do comprimento da ligação entre átomos ligantes sobre um conjunto de 50 proteínas.

A classe *Molecule* fornece uma abstração de um arquivo PDB e contém rotinas para adição, remoção, pesquisa e navegação nos átomos da molécula. Todos os átomos são representados como vértices e as ligações como arestas de um Grafo. O método *ForceFieldParameterization* da classe *MoleculeManager* realiza a parametrização da distância entre átomos ligantes.

Ao comparar os resultados com os valores publicado em artigo (Allen et al. 1987) verifica-se uma boa aproximação. Existe possibilidade de estender a parametrização para outros valores como torção, rotação. Os resultados desta parametrização podem ser utilizados como valores referenciais da fórmula AMBER, que calcula a energia potencial do sistema.

Uma aplicação deste framework é a construção de uma ferramenta de visualização de moléculas de proteínas em ambiente gráfico. Já existe a incorporação de visualização 3D, embora não exista recursos de navegação sobre essa estrutura.

Desta forma existe possibilidade de estender o framework para trabalhar com simulações envolvendo a modelagem mecânica molecular. Outro desafio seria a programação distribuída já que a maioria dos cálculos de modelagem molecular podem ser paralelizados.

Tabela 6.1: Tabela de distância entre átomos-ligantes com 50 Proteínas.

Conexões	Elemento1	Elemento2	Conexões	Hyle3D
1	H	C	2	0,947302128
1	H	C	3	0,950651713
1	H	C	4	0,935460983
1	H	N	2	0,948977323
1	H	N	3	0,86824243
1	H	N	4	0,873206772
1	H	O	2	0,924117015
1	H	O	3	1,114125184
1	H	S	2	1,039E+16
1	H	S	4	1,039E+16
1	O	C	1	1,131004543
1	O	C	3	1,131023187
2	O	H	1	0,989536681
2	O	C	1	1,135085257
2	O	C	3	1,113287469
2	O	C	4	1,341870129
2	S	H	1	1,039E+16
2	S	C	1	1,576198376
2	S	C	4	1,61822808
3	C	H	1	0,944871289
3	C	C	1	1,357975134
3	C	C	2	1,35800289
3	C	C	3	1,346336016
3	C	C	4	1,359851188
3	C	N	1	1,234300222
3	C	N	2	1,258525626
3	C	N	3	1,255410896
3	C	N	4	1,239236978
3	C	O	1	1,131285365
3	C	O	2	1,130974323
3	C	O	3	1,166799787
3	N	H	1	0,868285558
3	N	C	1	1,32995403

Tabela 6.2: Tabela de distância entre átomos-ligantes com 50 Proteínas

Conexões	Elemento1	Elemento2	Conexões	Hyle3D
3	N	C	3	1,235463023
3	N	C	4	1,330007505
4	C	H	1	0,934949394
4	C	C	1	1,375720516
4	C	C	2	1,380657363
4	C	C	3	1,370535382
4	C	C	4	1,375692078
4	C	N	2	2,78338E+16
4	C	N	3	1,305029424
4	C	N	4	1,311574736
4	C	O	1	1,341870129
4	C	O	2	1,341870129
4	C	S	1	1,610601841
4	C	S	2	1,615914703
4	C	S	3	1,0007E+15
4	C	S	4	1,619878975
4	N	H	1	0,886506567
4	N	C	1	1,326713942
4	N	C	4	1,326713942

# Índice Remissivo

- apolar, 10
- bond-stretching, 23
- Born-Oppenheimer, 19
- Campo de força
  - Allinger, 32
  - MM2, 32
  - MM3, 32
  - MM4, 32
- catalisador, 5
- centro quiral, 9
- cristalografia, 13
- difração
  - de elétrons, 32
  - de neutrons, 32
- eletronegatividade, 10
- enântiometro, 9
- espectroscopia por microondas, 32
- grupo carboxila, 10
- Hibridação, 32
- hidrófila, 10
- hidrófoba, 10
- isomeria, 9
- lei de hook, 23
- ligações moleculares, 23
- NMR, 13
- polar, 10
- ressonância magnética nuclear, 13
- site ativo, 9
- Tipo de átomo, 32
- tradução, 12
- transcrição, 12



## APÊNDICE A - Script para criar as tabelas do programa Hyle3D

```
# SQL Manager 2007 for MySQL 4.3.4.1
# -----
# Host      : localhost
# Port      : 3306
# Database  : hyle

/*!40101 SET @OLD_CHARACTER_SET_CLIENT=@@CHARACTER_SET_CLIENT */;
/*!40101 SET @OLD_CHARACTER_SET_RESULTS=@@CHARACTER_SET_RESULTS */;
/*!40101 SET @OLD_COLLATION_CONNECTION=@@COLLATION_CONNECTION */;
/*!40101 SET NAMES utf8 */;

SET FOREIGN_KEY_CHECKS=0;

DROP DATABASE IF EXISTS `hyle`;

CREATE DATABASE `hyle`
  CHARACTER SET 'latin1'
  COLLATE 'latin1_swedish_ci';

USE `hyle`;

DROP TABLE IF EXISTS `molecule`;

CREATE TABLE `molecule` (
  `ID` INTEGER(10) UNSIGNED NOT NULL AUTO_INCREMENT,
  `Fullpath` VARCHAR(256) COLLATE latin1_swedish_ci DEFAULT NULL,
  `FileName` VARCHAR(120) COLLATE latin1_swedish_ci DEFAULT NULL,
  `MoleculeType` INTEGER(10) UNSIGNED DEFAULT NULL,
  `FileType` INTEGER(10) UNSIGNED DEFAULT NULL,
  `Session` INTEGER(11) DEFAULT NULL,
  PRIMARY KEY (`ID`),
  KEY `MOLECULE_MOLECULETYPE_FK` (`MoleculeType`),
  KEY `MOLECULE_FILETYPE_FK` (`FileType`)
)ENGINE=InnoDB
AUTO_INCREMENT=48877 CHARACTER SET 'latin1' COLLATE 'latin1_swedish_ci';

DROP TABLE IF EXISTS `atom`;

CREATE TABLE `atom` (
  `ID` INTEGER(10) UNSIGNED NOT NULL AUTO_INCREMENT,
  `serial` INTEGER(10) UNSIGNED DEFAULT NULL,
  `name` CHAR(4) COLLATE latin1_swedish_ci DEFAULT NULL,
  `Element` INTEGER(4) NOT NULL,
  `X` FLOAT NOT NULL,
  `Y` FLOAT NOT NULL,
  `Z` FLOAT NOT NULL,
  `RADIUS` FLOAT DEFAULT NULL,
  `MOLECULE_FK` INTEGER(10) UNSIGNED NOT NULL,
  `RESIDUE` CHAR(4) COLLATE latin1_swedish_ci DEFAULT NULL,
  `type` INTEGER(3) DEFAULT NULL COMMENT 'If the atom is heteroatom or atom 1=atom
2=heteroatom',
  `sequenceNumber` INTEGER(10) UNSIGNED DEFAULT NULL,
  PRIMARY KEY (`ID`),
  KEY `atom_ibfk_1` (`RESIDUE`),
  KEY `ATOM_MOLECULE_FK` (`MOLECULE_FK`),
  CONSTRAINT `ATOM_MOLECULE_FK` FOREIGN KEY (`MOLECULE_FK`)
    REFERENCES `molecule` (`ID`) ON DELETE CASCADE ON UPDATE CASCADE
)ENGINE=InnoDB
AUTO_INCREMENT=251106774 CHARACTER SET 'latin1' COLLATE 'latin1_swedish_ci';
```

```
DROP TABLE IF EXISTS `bond`;

CREATE TABLE `bond` (
  `ID` INTEGER(10) UNSIGNED NOT NULL,
  `Molecule_FK` INTEGER(10) UNSIGNED DEFAULT NULL,
  `Source_FK` INTEGER(10) UNSIGNED NOT NULL,
  `Dest_FK` INTEGER(10) UNSIGNED NOT NULL,
  `Distance` FLOAT DEFAULT NULL,
  `Stretching` FLOAT DEFAULT NULL,
  PRIMARY KEY (`ID`),
  KEY `BOND_ATOM_DEST_FK` (`Dest_FK`),
  KEY `BOND_ATOM_SOURCE_FK` (`Source_FK`),
  KEY `BOND_MOLECULE_FK` (`Molecule_FK`),
  CONSTRAINT `BOND_ATOM_DEST_FK` FOREIGN KEY (`Dest_FK`) REFERENCES `atom` (`ID`)
    ON DELETE CASCADE ON UPDATE CASCADE,
  CONSTRAINT `BOND_ATOM_SOURCE_FK` FOREIGN KEY (`Source_FK`) REFERENCES `atom` (`ID`)
    ON DELETE CASCADE ON UPDATE CASCADE,
  CONSTRAINT `BOND_MOLECULE_FK` FOREIGN KEY (`Molecule_FK`) REFERENCES `molecule` (`ID`)
    ON DELETE CASCADE ON UPDATE CASCADE
)ENGINE=InnoDB
CHARACTER SET 'latin1' COLLATE 'latin1_swedish_ci';

DROP TABLE IF EXISTS `hybridization`;

CREATE TABLE `hybridization` (
  `ID` INTEGER(11) UNSIGNED NOT NULL AUTO_INCREMENT,
  `RESIDUE` CHAR(4) COLLATE latin1_swedish_ci DEFAULT NULL,
  `NAME` CHAR(4) COLLATE latin1_swedish_ci DEFAULT NULL,
  `TYPE` INTEGER(11) DEFAULT NULL COMMENT 'Type of hybridization. 0 = UNDEFINIED 1=sp1 2=
sp2 3=sp3 and 4 = sp3d 5=sp3d2',
  PRIMARY KEY (`ID`)
)ENGINE=InnoDB
AUTO_INCREMENT=42 CHARACTER SET 'latin1' COLLATE 'latin1_swedish_ci';

DROP TABLE IF EXISTS `molecule_type`;

CREATE TABLE `molecule_type` (
  `ID` INTEGER(11) UNSIGNED NOT NULL,
  `DESCRIPTION` VARCHAR(20) COLLATE latin1_swedish_ci DEFAULT NULL,
  PRIMARY KEY (`ID`)
)ENGINE=InnoDB
CHARACTER SET 'latin1' COLLATE 'latin1_swedish_ci';

DROP TABLE IF EXISTS `residue`;

CREATE TABLE `residue` (
  `ID` INTEGER(10) UNSIGNED NOT NULL,
  `NAME` CHAR(3) COLLATE latin1_swedish_ci NOT NULL DEFAULT '',
  `MOLECULE_TYPE` INTEGER(11) DEFAULT NULL,
  PRIMARY KEY (`ID`),
  UNIQUE KEY `NAME` (`NAME`)
)ENGINE=InnoDB
CHARACTER SET 'latin1' COLLATE 'latin1_swedish_ci';
```

```

DROP TABLE IF EXISTS `monomer`;

CREATE TABLE `monomer` (
  `ID` INTEGER(10) UNSIGNED NOT NULL,
  `NAME` CHAR(4) COLLATE latin1_swedish_ci DEFAULT NULL,
  `NUMBEROFCONNECTION` TINYINT(4) DEFAULT NULL,
  `ATOM1` CHAR(4) COLLATE latin1_swedish_ci DEFAULT NULL,
  `ATOM2` CHAR(4) COLLATE latin1_swedish_ci DEFAULT NULL,
  `ATOM3` CHAR(4) COLLATE latin1_swedish_ci DEFAULT NULL,
  `ATOM4` CHAR(4) COLLATE latin1_swedish_ci DEFAULT NULL,
  `RESIDUE_FK` INTEGER(10) UNSIGNED DEFAULT NULL,
  PRIMARY KEY (`ID`),
  KEY `RESIDUE_FK` (`RESIDUE_FK`),
  CONSTRAINT `monomer_fk` FOREIGN KEY (`RESIDUE_FK`) REFERENCES `residue` (`ID`) ON DELETE
  CASCADE ON UPDATE CASCADE
)ENGINE=InnoDB
CHARACTER SET 'latin1' COLLATE 'latin1_swedish_ci';

DROP TABLE IF EXISTS `parameter_angle`;

CREATE TABLE `parameter_angle` (
  `ID` INTEGER(11) UNSIGNED NOT NULL,
  `Element1` TINYINT(4) UNSIGNED DEFAULT NULL,
  `Element2` TINYINT(4) UNSIGNED DEFAULT NULL,
  `Element3` TINYINT(4) UNSIGNED DEFAULT NULL,
  `Hybridization12` TINYINT(4) DEFAULT NULL,
  `Hybridization23` TINYINT(4) UNSIGNED DEFAULT NULL,
  `Angle` FLOAT(9,3) DEFAULT NULL,
  PRIMARY KEY (`ID`)
)ENGINE=InnoDB
CHARACTER SET 'latin1' COLLATE 'latin1_swedish_ci';

DROP TABLE IF EXISTS `parameter_bond`;

CREATE TABLE `parameter_bond` (
  `ID` INTEGER(11) UNSIGNED NOT NULL,
  `Element1` CHAR(3) COLLATE latin1_swedish_ci NOT NULL DEFAULT '',
  `Element2` CHAR(3) COLLATE latin1_swedish_ci NOT NULL DEFAULT '',
  `bond` FLOAT(9,3) NOT NULL,
  PRIMARY KEY (`ID`),
  UNIQUE KEY `atom1atom2saturation` (`Element1`, `Element2`)
)ENGINE=InnoDB
CHARACTER SET 'latin1' COLLATE 'latin1_swedish_ci';

DROP TABLE IF EXISTS `parameter_torsion`;

CREATE TABLE `parameter_torsion` (
  `ID` INTEGER(11) UNSIGNED NOT NULL,
  `Element1` TINYINT(4) UNSIGNED DEFAULT NULL,
  `Element2` TINYINT(4) UNSIGNED DEFAULT NULL,
  `Element3` TINYINT(4) UNSIGNED DEFAULT NULL,
  `Element4` TINYINT(4) UNSIGNED DEFAULT NULL,
  `Hybridization12` TINYINT(4) DEFAULT NULL,
  `Hybridization23` TINYINT(4) UNSIGNED DEFAULT NULL,
  `Hybridization34` TINYINT(4) DEFAULT NULL,
  `Torsion` FLOAT(9,3) DEFAULT NULL,
  PRIMARY KEY (`ID`)
)ENGINE=InnoDB
CHARACTER SET 'latin1' COLLATE 'latin1_swedish_ci';

DROP TABLE IF EXISTS `session`;

```

```
CREATE TABLE `session` (  
  `ID` INTEGER(11) UNSIGNED NOT NULL,  
  `SESSION_STATUS` INTEGER(11) DEFAULT NULL,  
  `DESCRIPTION` VARCHAR(200) COLLATE latin1_swedish_ci DEFAULT NULL,  
  PRIMARY KEY (`ID`),  
  UNIQUE KEY `SESSION_STATUS` (`SESSION_STATUS`)  
)ENGINE=InnoDB  
CHARACTER SET 'latin1' COLLATE 'latin1_swedish_ci';
```

```
/*!40101 SET CHARACTER_SET_CLIENT=@OLD_CHARACTER_SET_CLIENT */;  
/*!40101 SET CHARACTER_SET_RESULTS=@OLD_CHARACTER_SET_RESULTS */;  
/*!40101 SET COLLATION_CONNECTION=@OLD_COLLATION_CONNECTION */;
```

# Referências

Alberts et al. 2006 ALBERTS et al. *Fundamentos da Biologia Celular*. 2. ed. [S.l.]: artmed, 2006.

Allen et al. 1987 ALLEN, F. H. et al. Tables of bond lengths determined by x-ray and neutron diffraction. part 1. bond lengths in organic compounds. v. 1, p. 19, 1987.

Allen 2004 ALLEN, S. M. Inherited diseases involving g proteins and g protein-coupled receptors. *Annu Rev Med*, 55:27-39, 2004. Disponível em: <<http://dx.doi.org/10.1146/annurev.med.55.091902.103843>>.

Berman et al. 1999 BERMAN, H. M. et al. The protein databank. *Nucleic Acid Research*, v. 28, p. 235–242, 1999.

Carvalho\*, Pupo e Bernardes 2003 CARVALHO\*, I.; PUPO, M. T.; BERNARDES Áurea D. L. Borges e L. S. C. Introdução a modelagem molecular de fármacos no curso experimental de química farmacêutica. *Química Nova*, v. 26, n. 3, p. 428–438, 2003.

Cheng-Yu 2009 CHENG-YU. Computational structure-based redesign of enzyme activity. p. 3764–3769, Mar 2009. Disponível em: <<http://dx.doi.org/10.1073/pnas.0900266106>>.

Creighton 1992 CREIGHTON, T. E. *Proteins: Structures and Molecular Properties*. Second. [S.l.]: W. H. Freeman, 1992.

Holmes 2004 HOLMES, J. B. Some fundamental aspects of building protein structures from fragment libraries. Jun 2004. Disponível em: <<http://dx.doi.org/10.1110/ps.03494504>>.

Leach 2001 LEACH, A. R. *Molecular Modeling Principles and Applications*. Second. [S.l.]: Pearson Education Limited, 2001. 165-246 p.

Lesk 2008 LESK, A. M. *Introdução à bioinformática*. 2. ed. [S.l.]: artmed, 2008. 384 p.

Murphy 2009 MURPHY, P. M. Alteration of enzyme especificity by computational loop remodeling and design. Jun 2009. Disponível em: <<http://dx.doi.org/10.1073/pnas.0811070106>>.

Nelson e Cox 2010 NELSON, D.; COX, M. *Lehninger Principles of Biochemistry*. Fifth. [S.l.]: Worth Publishers, 2010.

Patrick 1995 PATRICK, G. *An introduction to medicinal chemistry*. Oxford University Press, 1995. ISBN 9780198558729. Disponível em: <<http://books.google.com.br/books?id=FSa5QgAACAAJ>>.

Protein Databank 2012 PROTEIN DATABANK. *PDB Format*. June 2012. Disponível em: <<http://www.wwpdb.org/index.html>>.

RODRIGUES 2001 RODRIGUES, C. R. Processos modernos no desenvolvimento de fármacos. *química nova na escola*, são paulo, n. 3, p. 43-49. 2001.

SantAnna 2009 SANTANNA, C. M. R. *Métodos de modelagem molecular para estudo e planejamento de compostos bioativos: Uma introdução*. UFF, 2009. Disponível em: <<http://www.uff.br/RVQ/index.php/rvq/article/view/13/32>>.

Schlick et al. 2011 SCHLICK, T. et al. RAG: An update to the RNA-as-graphs resource. *BMC Bioinformatics*, BioMed Central, v. 12, p. 219, May 2011. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3123240>>.

Stryer 2006 STRYER, L. *Biochemistry*. [S.l.]: W.H. Freeman and Company, 2006.

Trevizani 2011 TREVIZANI, R. *Biblioteca de fragmentos para predição de estruturas de proteínas*. Tese (Doutorado) — Laboratório Nacional de Computação Científica, 2011.

Valence bond theory: sp, sp<sup>2</sup>, and sp<sup>3</sup> hybrid orbitals 2012 VALENCE bond theory: sp, sp<sup>2</sup>, and sp<sup>3</sup> hybrid orbitals. June 2012. Disponível em: <<http://chemwiki.ucdavis.edu/>>.

Walsh 2005 WALSH, G. Biopharmaceuticals: recent approvals and likely directions. 2005. Disponível em: <<http://dx.doi.org/10.1016/j.tibtech.2005.07.005>>.