



Universidade de Brasília

Instituto de Ciências Exatas  
Departamento de Ciência da Computação

## Alguma coisa de bioinformática

Gabriella de Oliveira Esteves

Monografia apresentada como requisito parcial  
para conclusão do Bacharelado em Ciência da Computação

Orientador

Prof. Dr.<sup>a</sup> Maria Emília Machado Telles Walter

Brasília  
2016

Universidade de Brasília — UnB  
Instituto de Ciências Exatas  
Departamento de Ciência da Computação  
Bacharelado em Ciência da Computação

Coordenador: Prof. Dr. Rodrigo Bonifácio de Almeida

Banca examinadora composta por:

Prof. Dr.<sup>a</sup> Maria Emília Machado Telles Walter (Orientador) — CIC/UnB  
Prof. Dr. Professor I — CIC/UnB  
Prof. Dr. Professor II — CIC/UnB

### **CIP — Catalogação Internacional na Publicação**

Esteves, Gabriella de Oliveira.

Alguma coisa de bioinformática / Gabriella de Oliveira Esteves. Brasília  
: UnB, 2016.

45 p. : il. ; 29,5 cm.

Monografia (Graduação) — Universidade de Brasília, Brasília, 2016.

1. Biologia Molecular, 2. Bioinformática

CDU 004.4

Endereço: Universidade de Brasília  
Campus Universitário Darcy Ribeiro — Asa Norte  
CEP 70910-900  
Brasília-DF — Brasil



Universidade de Brasília

Instituto de Ciências Exatas  
Departamento de Ciência da Computação

## Alguma coisa de bioinformática

Gabriella de Oliveira Esteves

Monografia apresentada como requisito parcial  
para conclusão do Bacharelado em Ciência da Computação

Prof. Dr.<sup>a</sup> Maria Emília Machado Telles Walter (Orientador)  
CIC/UnB

Prof. Dr. Professor I    Prof. Dr. Professor II  
CIC/UnB                      CIC/UnB

Prof. Dr. Rodrigo Bonifácio de Almeida  
Coordenador do Bacharelado em Ciência da Computação

Brasília, 26 de Abril de 2016

# Dedicatória

Dedicatória

# Agradecimentos

Agradecimento

# Resumo

Resumo em português

**Palavras-chave:** Biologia Molecular, Bioinformática

# Abstract

Abstract in english

**Keywords:** molecular biology, bioinformatics

# Sumário

|          |                                     |           |
|----------|-------------------------------------|-----------|
| <b>1</b> | <b>Introdução</b>                   | <b>1</b>  |
| 1.1      | História da Genética . . . . .      | 1         |
| 1.1.1    | Origens da Vida . . . . .           | 3         |
| 1.1.2    | Análise do Núcleo Celular . . . . . | 3         |
| 1.1.3    | Estudo do genoma . . . . .          | 3         |
| 1.2      | Sequenciamento genético . . . . .   | 3         |
| 1.3      | Definição do Problema . . . . .     | 3         |
| 1.4      | Justificativa . . . . .             | 4         |
| 1.5      | Objetivo . . . . .                  | 4         |
| 1.6      | Descrição dos Capítulos . . . . .   | 4         |
| <b>2</b> | <b>Biologia Molecular</b>           | <b>5</b>  |
| 2.1      | Proteína . . . . .                  | 5         |
| 2.2      | Ácidos Nucléicos . . . . .          | 6         |
| 2.2.1    | DNA . . . . .                       | 7         |
| 2.2.2    | RNA . . . . .                       | 7         |
| 2.2.3    | Transcrição e tradução . . . . .    | 8         |
| 2.3      | Genética Molecular . . . . .        | 8         |
| 2.3.1    | Cromossomos . . . . .               | 8         |
| 2.3.2    | Gene . . . . .                      | 8         |
| 2.3.3    | Genoma . . . . .                    | 8         |
| <b>3</b> | <b>Bioinformática</b>               | <b>9</b>  |
| <b>4</b> | <b>Implementação</b>                | <b>10</b> |
| <b>5</b> | <b>Resultados</b>                   | <b>11</b> |
| <b>6</b> | <b>Conclusão</b>                    | <b>12</b> |
| <b>7</b> | <b>Trabalhos Futuros</b>            | <b>13</b> |
| <b>8</b> | <b>Cronograma</b>                   | <b>14</b> |
|          | <b>Referências</b>                  | <b>15</b> |



# Lista de Figuras

|     |                             |   |
|-----|-----------------------------|---|
| 1.1 | Pai da bio . . . . .        | 3 |
| 2.1 | Adaptado de : [2] . . . . . | 6 |
| 2.2 | Adaptado de : [3] . . . . . | 8 |

# Lista de Tabelas

|     |                      |    |
|-----|----------------------|----|
| 8.1 | Cronograma . . . . . | 14 |
|-----|----------------------|----|

# Capítulo 1

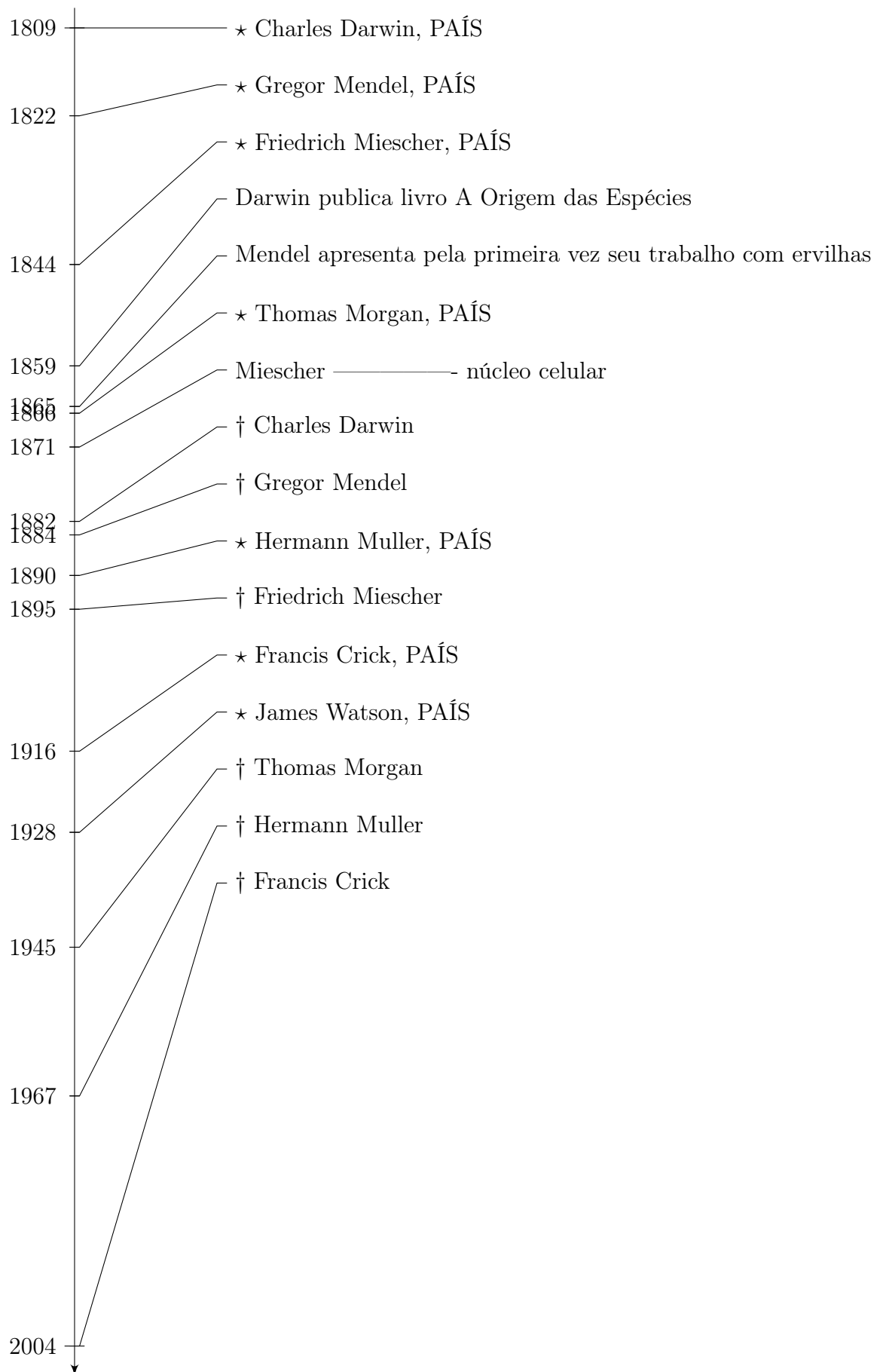
## Introdução

### 1.1 História da Genética

O estudo do núcleo celular começou no século XIX com o objetivo de ««««»». pesquisa era do ««««»»NACIONALIDADE««««»» Friedrich Miescher, ««««»». Duas teorias divergentes marcaram esta época: enquanto Charles Darwin afirmava que a evolução era um evento demorado e o ser ««««»»sobrevivente««««»» era escolhido aleatoriamente, Gregor Mendel insistia que a evolução acontecia a cada geração e, ainda, podia ser controlada. Foi apenas nas pesquisas do grupo de Thomas Morgan que essas duas teorias vieram a se ««««»»unir««««»».

A partir do século XX, o foco passou a ser a compreensão da produção de proteínas nos seres vivos e, para isso, ««««»».

A linha do tempo abaixo tem o objetivo de auxiliar na localização temporal da história da biologia molecular ao passo que apresentam as datas de nascimento e morte dos principais pesquisadores da área, cujos trabalhos serão descritos neste capítulo.



### 1.1.1 Origens da Vida

...

### 1.1.2 Análise do Núcleo Celular

...



Figura 1.1: Pai da bio

### 1.1.3 Estudo do genoma

...

## 1.2 Sequenciamento genético

...

## 1.3 Definição do Problema

Problema.

## 1.4 Justificativa

Motivação.

## 1.5 Objetivo

O objetivo. Os objetivos específicos são:

- .
- .
- .

## 1.6 Descrição dos Capítulos

No Capítulo 1 fez-se uma breve introdução aos ... No Capítulo 2 são estabelecidas as principais definições utilizadas neste trabalho mais profundamente, tais como ... Ainda, são apresentados ... Também são descritos ... O Capítulo 3 faz referência à implementação do...

# Capítulo 2

## Biologia Molecular

Neste capítulo serão descritos os conceitos básicos da biologia molecular. A seção 2.1 explica o que são, do que são constituídas e como são arranjadas as proteínas. A seção ?? ...

### 2.1 Proteína

As proteínas são macromoléculas orgânicas, ou biomoléculas, com diversas responsabilidades no corpo dos seres vivos. Se fizerem parte do grupo de proteínas fibrosas, como o colágeno, irão compor a estrutura do corpo e para isso precisam ser resistentes e insolúveis em água. Caso estejam no grupo de proteínas globulares, como a hemoglobina, realizarão processos dinâmico pelo corpo tais como transportações e cataliações [1]. Cada tarefa é realizada por um proteína com uma estrutura específica e otimizada pra tal.

As proteínas são polímeros, macromoléculas formadas por estruturas menores chamadas monômeros, que neste caso são os aminoácidos. Aminoácidos são moléculas que possuem cinco componentes: amina ( $\text{NH}_2$ ), carbono (C), hidrogênio (H), ácido carboxílico ( $\text{COOH}$ ) e uma cadeia lateral que funciona como identificador de cada um dos 20 tipos de aminoácidos presentes nos seres vivos. A maneira como eles são criados será explicada com mais detalhes na seção 2.2, pois envolve um processo complexo de síntese de proteína executado pelo ribossomo. A ligação, ou polimerização, de dois aminoácidos é feita unindo a amida de um com o ácido carboxílico do outro, liberando uma molécula de água ( $\text{H}_2\text{O}$ ) e formando uma cadeia chamada de dipeptídeo. Como houve liberação de água na ligação, o dipeptídeo não é formado por aminoácidos, mas sim resíduos dos mesmos. Nesse sentido, cadeias peptídicas de 100 à 5000 diferentes resíduos aminoácidos, ou cadeia polipeptídicas, constituem a proteína.

Existem quatro estruturas para caracterização de uma proteína [4]. A mais simples é chamada de estrutura primária e é composta por uma sequência linear de resíduos aminoácidos. A estrutura secundária é tridimensional e estabiliza-se por meio de ligações de hidrogênio na cadeia principal, chamada de *backbone*. Dependendo da disposição dos resíduos de aminoácidos, esta cadeia pode se dar forma de hélice ou em forma de folha. A estrutura terciária é dada pela união de várias estruturas secundárias e, por fim, a estrutura quaternária é composta de múltiplas estruturas terciárias [2]. A figura 2.1 ilustra os quatro tipos de proteínas descritos.

Artigo: [[[ Mesmo da subsection do topico ]]]

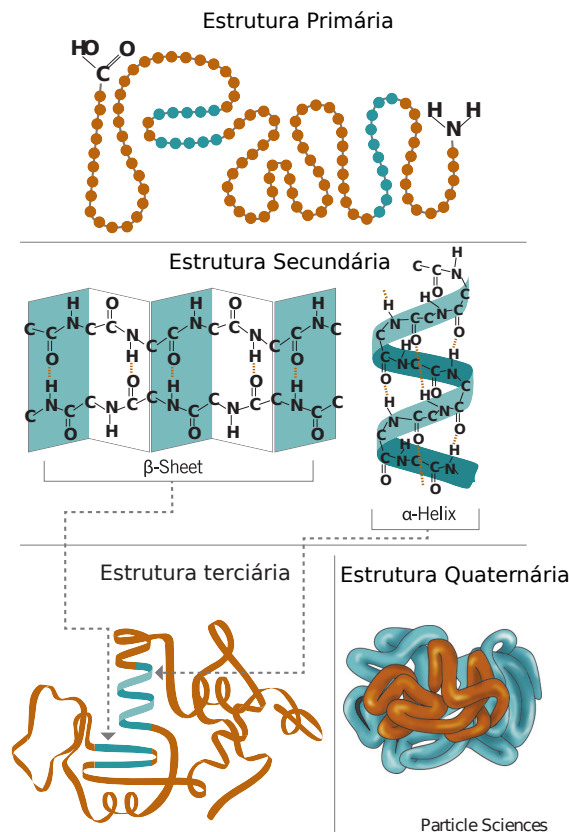


Figura 2.1: Adaptado de : [2]

Transcrição e tradução, ácidos nucleicos resumo

## 2.2 Ácidos Nucleicos

Os ácidos nucleicos são biomoléculas responsáveis pelo armazenamento, transmissão e tradução das informações genéticas dos seres vivos. Isto é possível devido ao processo de síntese de proteínas que permite, assim, a base da herança biológica. Assim como as proteínas, os ácidos nucleicos são polímeros cujos monômeros são nucleotídeos. Nucleotídeos são compostos de três elementos: um radical fosfato ( $\text{HPO}_4$ ), uma pentose, ou seja, um monossacarídeo formado por cinco átomos de carbono, e uma base nitrogenada. Existem cinco tipos de bases nitrogenadas que podem compor um nucleotídeo: Adenina(A), Timina(T), Citosina(C), Guanina(G) e Uracila(U).

**imagem de um nucleotídeo e das bases nitrogenadas. Mostrar backbone da pentose 1'...5'**

Na figura **acima !!!**, observa-se que no *backbone* do nucleotídeo existe uma numeração de 1' à 5', que representam os carbonos presentes na pentose. Para a criação de uma fita de ácido nucleico, no processo de polimerização formar-se uma ligação fosfodiéster entre o carbono da posição 5' do *backbone* de um nucleotídeo e o carbono de posição 3' do *backbone* de outro [4]. Por definição o sentido da leitura de uma fita de ácido nucleico é



$5' \rightarrow 3'$ , o que é deve ser levado em consideração ao se fazer interpretação de dados do material genético.

### **imagem da polimerização**

Dois tipos de ácidos nucleicos são encontrados nos seres vivos: ácido desoxirribonucleico (DNA) e ácido ribonucleico (RNA). Eles diferenciam-se tanto na estrutura do *backbone* e nas bases nitrogenadas, quanto em suas funções. A seguir serão apresentadas as definições de DNA e RNA.

## **2.2.1 DNA**

Os DNAs são as biomoléculas que armazenam as informações referentes ao funcionamento de todas as células dos seres vivos de maneira específica: sequências de pares de bases nitrogenadas. Nesse sentido, além de haver a ligação fosfodiéster entre os nucleotídeos, cada um também se liga a partir de suas bases nitrogenadas, formando assim um eixo helicoidal tridimensional chamada de dupla hélice [4]. Esta estrutura foi descoberta em 1953, pelo biólogo James Watson e pelo físico Francis Crick [3], porém os ácidos nucleicos já eram estudado desde 1869 na Suíça pelo químico-fisiológico Friedrich Miescher.

Em relação à estrutura dos monômeros do DNA, a pentose dos nucleotídeos é uma desoxirribose, indicada na figura 2.2. Para a formação da dupla hélice, os pares são feitos com uma base nitrogenada do grupo de purinas, composto orgânico que possui um anel duplo de carbono, e outra base do grupo de pirimidinas, composto orgânico que possui um anel simples de carbono. No caso do DNA, somente quatro das cinco bases são empregadas: as purinas Adenina(A) e Guanina(G), que se ligam com as pirimidinas Timina(T) e Citosina(C) respectivamente. Desta forma, A e T são bases complementares de G e C. Uma fita de DNA pode conter centenas de milhões de nucleotídeos.

A representação do DNA, seja nos livros ou computacionalmente, é dada por um par em paralelo de strings de letras A, T, G e C. Como explicado na seção 2.2, o sentido padrão da leitura de uma fita é de  $5' \rightarrow 3'$ , mas no caso do DNA, as hélices são dispostas de maneira antiparalela, ou seja, uma é lida de  $5' \rightarrow 3'$  e a outra, de  $3' \rightarrow 5'$ . Observa-se que a partir de uma hélice, pode-se inferir a sequência de sua hélice complementar. Seja, por exemplo, uma hélice H1 igual a AGTAAGC; então H2 em seu sentido oposto é H2' igual a TCATTGC, e no sentido regular, igual a GCTTACT. A figura 2.2 apresenta a estrutura do DNA como explicada nesta seção.

## **2.2.2 RNA**

Os RNAs são biomoléculas semelhantes ao DNA, porém contam com três diferenças básicas. A primeira é a estrutura do *backbone* dos nucleotídeos, que é composta por uma ribose ao invés de um desoxirribose. A segunda diferença é em relação às bases nitrogenadas, onde a pirimidina Uracila(U) substitui a Timina(T). Por fim, o RNA é formado por apenas uma hélice tridimensional.

Existem três tipos de RNAs presentes no citoplasma, ou seja, espaço entre a membrana plasmática e o núcleo da célula. Cada um possui funções específicas que serão detalhadas na subseção seguinte. Em suma, O RNA mensageiro (mRNA) é responsável pela transferência de informação do DNA para o RNA ribossômico (rRNA), que por sua

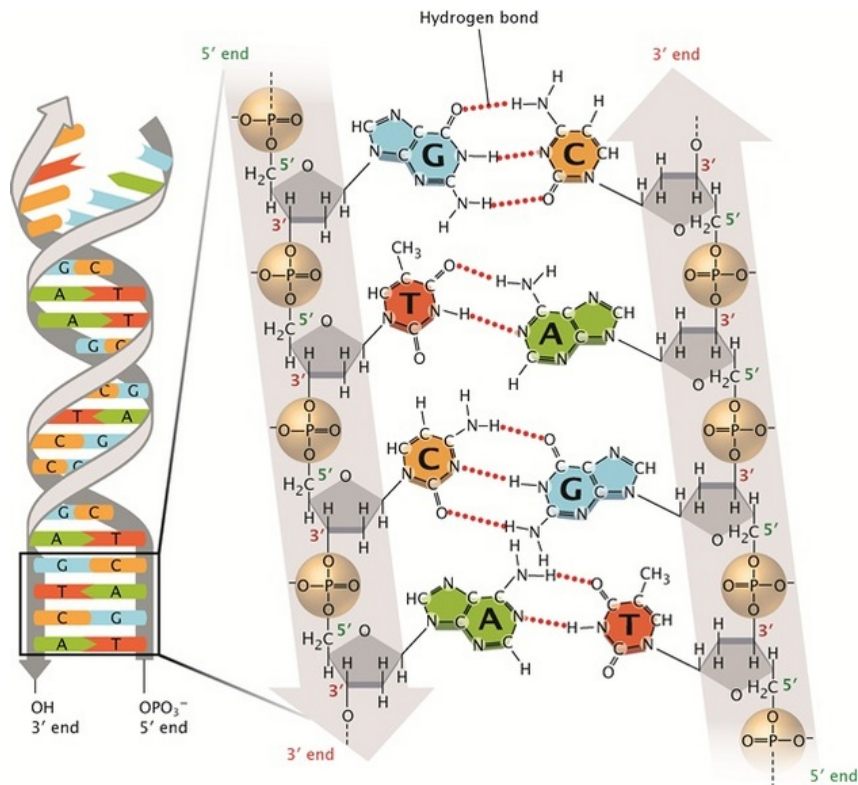


Figura 2.2: Adaptado de : [3]

vez irá desanexar a proteína do RNA transportador (tRNA) combinando-o com o rRNA, executando assim, a síntese de proteína.

### 2.2.3 Transcrição e tradução

rRNA, mRNA, tRNA

## 2.3 Genética Molecular

...

### 2.3.1 Cromossomos

...

### 2.3.2 Gene

...

### 2.3.3 Genoma

...

# Capítulo 3

## Bioinformática

Conceitualização do algoritmo.

Artigos: Introdução [introduction to bioinformatics for computer scientists]

# Capítulo 4

## Implementação

Implementação do algortimo.

# Capítulo 5

## Resultados

Neste capítulo serão apresentados os primeiros resultados experimentais obtidos.

# Capítulo 6

## Conclusão

Neste capítulo serão apresentadas as considerações finais do trabalho, assim como as limitações e dificuldades encontradas.

# Capítulo 7

## Trabalhos Futuros

A partir deste trabalho, foi possível identificar os seguintes pontos a serem melhorados:

- x

# Capítulo 8

## Cronograma

O cronograma está apresentado na Tabela a seguir, mostrando o início das atividades em Janeiro de 2016 com a revisão literária e com término previsto para Junho de 2016, juntamente com a defesa do Trabalho de Conclusão de Curso.

Tabela 8.1: Cronograma

| Atividades                  | 2016 |     |     |     |     |     |
|-----------------------------|------|-----|-----|-----|-----|-----|
|                             | Jan  | Fev | Mar | Abr | Mai | Jun |
| Revisão bibliográfica       | X    | X   |     |     |     |     |
| Familiaridade com –         |      | X   | X   |     |     |     |
| Implementação               |      |     | X   | X   | X   |     |
| Interpretação dos resultado |      |     |     | X   | X   | X   |
| Defesa                      |      |     |     |     |     | X   |



# Referências

- [1] Proteínas. <http://www.professoraangela.net/documents/proteinas.html>, visitado em 2016-01-02. 5
- [2] Protein structure, 2009. <http://www.particlesciences.com/news/technical-briefs/2009/protein-structure.html>, visitado em 2016-01-02. vi, 5, 6
- [3] Leslie A. Pray. Discovery of dna structure and function: Watson and crick, 2008. <http://www.nature.com/scitable/topicpage/discovery-of-dna-structure-and-function-watson-397>, visitado em 2016-01-15. vi, 7, 8
- [4] João Carlos Setubal and João Meidanis. *Introduction to computational molecular biology*. PWS Publishing Company, 1997. 5, 6, 7