



**SAMUEL MESSIAS DE OLIVEIRA**

**INTRODUÇÃO À BIOINFORMÁTICA ESTRUTURAL: A  
CONSTRUÇÃO DE UMA EMENTA**

**INCONFIDENTES – MG  
2016**

**SAMUEL MESSIAS DE OLIVEIRA**

**INTRODUÇÃO À BIOINFORMÁTICA ESTRUTURAL: A  
CONSTRUÇÃO DE UMA EMENTA**

Trabalho de Conclusão do Curso apresentado como pré-requisito para a conclusão do curso de Graduação em Licenciatura em Ciências Biológicas do Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas Gerais – Campus Inconfidentes, para obtenção do título de Licenciado em Ciências Biológicas.

Orientador: Prof. Dr. Marcelo Augusto dos Reis

**INCONFIDENTES - MG  
2016**

**SAMUEL MESSIAS DE OLIVEIRA**

**INTRODUÇÃO À BIOINFORMÁTICA ESTRUTURAL: A  
CONSTRUÇÃO DE UMA EMENTA**

**Data de aprovação: 05/10/2016**

---

**Orientador: Prof. Dr. Marcelo Augusto dos Reis  
IFSULDEMINAS - Campus Inconfidentes**

---

**Prof. Ms. Nilton Luiz Solto  
IFSULDEMINAS - Campus Inconfidentes**

---

**Prof. Dr. Jorge Alexandre Nogueira dos Santos  
IFSULDEMINAS - Campus Inconfidentes**

## DEDICATÓRIA

*À minha família, que sempre me apoiou durante toda a minha formação.*

*À sabedoria não igualei nenhuma pedra preciosa, pois, em comparação com ela, todo o ouro é um punhado de areia e, diante dela, a prata será avaliada como lodo. (Sabedoria 7, 9)*

## **AGRADECIMENTOS**

- A Deus.
- A meu orientador, Dr. Marcelo Augusto dos Reis, por ter me mostrado um lado surpreendente e fascinante da ciência com a qual não havia tido contato anteriormente. Aos seus conselhos e auxílios durante todo o percurso até o final deste Trabalho de Conclusão de Curso.
- A todos os professores que contribuíram com minha formação.
- A meus pais, que sempre me inspiraram a estudar. Que ensinaram meus valores e por me corrigirem sempre que se faz preciso.
- A meus irmãos, que sempre estiveram ao meu lado durante toda a vida.
- Aos amigos de faculdade, com os quais compartilhei os melhores momentos que a graduação pôde nos proporcionar.
- Um agradecimento especial ao colega Rafael Sousa, pela fraternidade e amizade durante todo o tempo em que vivemos sobre o mesmo teto em terras portuguesas.
- À minha namorada e companheira de ensino médio e faculdade, Franciele de Cássia Guimarães, por tudo o que construímos em conjunto durante o período em que estamos unidos. Pelo companheirismo e conforto, pelo apreço e cuidado. Pela união que temos. Pelo auxílio ao meu crescimento pessoal.
- Ao IFSULDEMINAS, por proporcionar minha formação acadêmica.
- A todos aqueles que de alguma forma contribuíram com a minha formação.

## RESUMO

A bioinformática tem seu início em meados dos anos 1960. Nessa época tornou-se necessário a utilização de bancos de dados computacionais para armazenar a grande quantidade de informação extraída de moléculas biológicas. Nos últimos anos, com o maior acesso às tecnologias e como decorrência do Projeto Genoma, a bioinformática tem se mostrado uma ciência em desenvolvimento. No Brasil, os esforços para trabalhar com a bioinformática têm início no período entre o final do século passado e início deste. Contudo, mesmo com o desenvolver da bioinformática, o país possui obstáculos para o seu amadurecimento. A falta de profissionais capacitados para atuar na área e a falta de cursos para a formação de bioinformatas dificultam sua aprendizagem e difusão. Com isto, o intuito deste trabalho foi levar a uma parcela dos alunos licenciandos em Ciências Biológicas pertencentes ao IFSULDEMINAS – Campus Inconfidentes a participação em um curso introdutório à bioinformática. O curso possuiu duração de três meses, com encontros semanais com duração de uma hora e meia. Para a análise do curso, foi utilizada neste trabalho a metodologia de pesquisa-ação e como ferramentas principais foram utilizados o diário de campo e aulas práticas. Os dados mostraram que a maior parte dos estudantes desconhecia a bioinformática. Além disso, os alunos apresentaram dificuldades em manusear ferramentas computacionais e mostraram-se renitentes à adoção de novas tecnologias. Mesmo assim, os dados também mostram que o curso obteve sucesso em apresentar a uma parcela dos estudantes licenciandos em ciências biológicas algumas ferramentas utilizadas em bioinformática e os conteúdos relacionados à área.

**Palavras-chave:** Bancos de dados; Moléculas biológicas; Pesquisa ação; Diário de campo; Novas tecnologias.

## ABSTRACT

Bioinformatics has its beginning in the middle of the 1960 years. At this time it became necessary to use computational databases to store the big quantity of biological molecules' information created. In the last years, the more technological access and as consequence of Genome Project, bioinformatics has shown itself as a developing science. In Brazil, efforts to work with bioinformatics have begun between the end of the last century and the beginning of this century. Nevertheless, even with the bioinformatics development, the country has some obstacles against its growth. The lack of enabled professionals to work with and the lack of courses to form bioinformaticians make it hard to learn and spread the field of bioinformatics. Thus, the aim of this work was to perform a case study with a group of Biology students from IFSULDEMINAS – campus Inconfidentes in order to guide them to the bioinformatics field through an introductory course developed from scratch. The course lasted three months. It had weekly meetings of one hour and a half. To analyze the course, action research was used and the main tools used were field diary and practical classes. The data showed that most part of students didn't know what bioinformatics really is. Besides, the students presented difficulties in handling computational tools and appeared obstinate to adopt new technologies. Even so, the data also showed that the course obtained success to present a part of Biology's students some commonly used tools in bioinformatics and contents used in the field.

**Key-words:** Databases; Biological molecules; Action research; Field diary; New technologies.



## LISTA DE FIGURAS

Figura 1: Osciloscópio utilizado por Levinthal em seus experimentos.....	16
Figura 2: Representação das principais áreas da Bioinformática.....	18
Figura3: Representação do Dogma Central da Biologia .....	20
Figura 4: Esquema representativo da duplicação do DNA. ....	21
Figura 5: Representação dos passos para a transcrição de um DNA em RNA. ....	22
Figura 6: Representação do esquema de tradução do RNA em DNA.....	23
Figura 7: Níveis de estruturas nas proteínas.....	24
Figura 8: Distinção entre alinhamento local e alinhamento global .....	25
Figura 9: Matrizes de alinhamento .....	26
Figura 10: <i>Root mean square deviation</i> .....	27
Figura 11: Alinhamento estrutural de proteínas similares .....	28
Figura 12: Alinhamento estrutural de proteínas não similares: .....	28
Figura 13 : Cristalografia de Raio-X .....	29
Figura 14: Representação resumida do processo de cristalografia.....	30
Figura 15: Caminho completo utilizado pelo servidor I-TASSER para modelagem proteica .	31
Figura 16: Exemplos de aminoácidos visualizados na primeira prática.....	37
Figura 17: Estruturas secundárias .....	37
Figura 18: Alinhamento estrutural das proteínas hemoglobina (em azul) e mioglobina(em rosa) .....	38
Figura 19: Alinhamentos sequenciais realizados para a proteína Manganês Peroxidase de <i>Trametes versicolor</i> .....	39
Figura 20: Modelagem por homologia.....	40

## SUMÁRIO

<b>1. INTRODUÇÃO .....</b>	<b>1</b>
1.1 O PROBLEMA E OS OBJETIVOS DO TRABALHO .....	3
1.2 O QUE É UMA EMENTA? .....	3
1.3 O QUE É BIOINFORMÁTICA? .....	4
1.4 A BIOINFORMÁTICA NO BRASIL.....	5
1.5 BIOINFORMÁTICA: UM INÍCIO PARA O IFSULDEMINAS – Campus INCONFIDENTES .....	7
<b>2. METODOLOGIA.....</b>	<b>10</b>
2.1. PRÉ-CURSO .....	10
2.2 CURSO .....	10
2.3 CONSTRUÇÃO DAS AULAS TEÓRICAS E AULAS PRÁTICAS. ....	13
2.3.1 VMD.....	13
2.3.2. PyMOL.....	14
2.3.3 Swiss Model .....	14
2.4 COLETA DE DADOS.....	15
2.5 CONTEÚDOS .....	16
2.5.1 A história da Bioinformática. Do seu início até o primeiro banco de dados.....	16
2.5.2 Banco de dados.....	18
2.5.3 O dogma central da Biologia. Aspectos de duplicação, transcrição e tradução.....	19
2.5.4 Proteínas: Níveis de informação .....	24
2.5.5 Alinhamento .....	25
2.5.5.2 BLAST .....	27
2.5.5.3 Alinhamento estrutural .....	27
2.5.6 Cristalografia de Raio-X .....	29
2.5.7 Modelagem de proteínas por homologia.....	30
<b>3. RESULTADOS E DISCUSSÕES .....</b>	<b>32</b>
3.1 PRÁTICAS .....	34
3.2 AVALIAÇÕES .....	40
3.3 EMENTA.....	45
<b>4. CONCLUSÃO.....</b>	<b>49</b>

<b>5. REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>50</b>
<b>ANEXO I.....</b>	<b>56</b>
<b>ANEXO II.....</b>	<b>57</b>
<b>ANEXO III.....</b>	<b>58</b>
<b>ANEXO IV .....</b>	<b>59</b>
<b>ANEXO V.....</b>	<b>62</b>
<b>ANEXO VI.....</b>	<b>64</b>
<b>ANEXO VII.....</b>	<b>66</b>
<b>ANEXO VIII.....</b>	<b>68</b>

## 1. INTRODUÇÃO

Estamos rodeados de pessoas e organizações que necessitam diretamente do acesso à internet e suas tecnologias. Enviamos e-mails, acessamos notícias de última hora e redes sociais. Há hoje um conjunto de necessidades que nos fazem estar conectados a todo instante ao mundo virtual. O Acesso de milhões de pessoas à internet e às ferramentas que dela podemos utilizar cresce a cada dia. Ricoy e Couto (2014), comentam que mais do que as bibliotecas, a internet no mundo de hoje transformou-se no local preferencial para a procura de informações, sobretudo para estudantes universitários.

Segundo Neves, Santos e Gomes (2012), a tecnologia tem papel estruturante na sociedade. Com as tecnologias criadas através dos tempos os homens modificaram comportamentos e ações sociais, transformando a maneira com que o homem domina a natureza, tornando-se cada vez mais ativa a atuação do homem sobre aquilo que o cerca.

Hoje, uma das tecnologias que a maior parte das pessoas no mundo globalizado e em desenvolvimento têm acesso é a internet. Em um período curto de tempo, a internet consolidou-se como uma plataforma muito potente que modificou a maneira com a qual fazemos negócios e o modo como nos comunicamos. A Internet, como nenhum outro meio de comunicação houvera feito, deu um grande salto à globalização do mundo, tornando-se a fonte de informação universal de milhões de pessoas, seja em casa, na escola ou no trabalho.

Por sua vez, os computadores, que são mais antigos que a internet (LEINER *et al.*, 2012), popularizaram-se a partir do ano de 1984 com a introdução do computador pessoal Macintosh, da empresa *Apple* (POLSSON, 2009). Antes, os computadores ocupavam grandes espaços, como, por exemplo, o ENIAC (*Electronic Numerical Integrator and Computer*), desenvolvido durante a Segunda Guerra Mundial. Suas dimensões eram de oitenta pés de comprimento por oito pés e meio de largura. Suas dimensões eram tamanhas que o computador necessitava ser disposto em forma de U no local onde era instalado (FILHO, 2007).

Atualmente, os computadores, seus custos operacionais e suas dimensões estão cada vez mais populares. Há tempos não seria possível comunicar-se com um cidadão que estivesse em algum outro país num prazo menor que semanas ou meses. Hoje essa comunicação, graças ao advento tecnológico vivenciado ultimamente, tem ocorrido com espaços de tempo na ordem dos segundos. Destarte possuímos a necessidade de estarmos cada vez mais vinculados ao virtual, seja para uso pessoal ou profissional.

Assim também caminha a ciência, criando *networkings* entre várias áreas que há algum tempo pouco possuíam em comum, contudo, graças às novas tecnologias, cada vez mais têm estreitado seus laços de união, de forma a produzir no final um conhecimento mais alinhado a uma visão integralista de ciência.

No auge do Projeto Genoma, quando dados genômicos estavam sendo produzidos a níveis fordistas de produção e com o advento de tecnologias como a Reação em Cadeia da Polimerase (PCR) e o método didesoxi (Método de Sanger), a informação sobre o DNA e seus subprodutos cresceu em progressão geométrica. Em sequência, a informação, para não ser perdida, e por ser muito extensa, necessitou ser armazenada em locais onde nada fosse descartado e onde sempre novos produtos pudessem ser depositados.

Neste paradigma um crescente grupo de pesquisadores, alguns com viés computacional, e outros com viés biológico, uniram-se de modo a contribuir mutuamente para o progresso do conhecimento científico e progresso da maneira com que tratamos a ciência.

Desse modo, construiu-se uma rede de profissionais interessados em compreender os aspectos e funções genômicas a partir de um método e de ferramentas de utilização mais rápidas, e que abrigasse um maior número de informações sobre o genoma, transformando-os em conhecimento através de experimentos *in silico* (experimentos de simulação computacional), e não somente a partir de experimentos *in vivo* (experimentos feitos em organismos vivos) e experimentos *in vitro* (experimentos realizados em laboratórios).

## 1.1 O PROBLEMA E OS OBJETIVOS DO TRABALHO

O trabalho aqui desenvolvido focalizou em responder ao seguinte questionamento: Qual a melhor forma de introduzir alunos licenciandos em Ciências Biológicas no ramo da Bioinformática?

Com o trabalho espera-se responder a essa pergunta auxiliando de maneira positiva na aprendizagem de um novo conteúdo por parte dos licenciandos participantes. Para além, o trabalho busca auxiliar na difusão do conhecimento de novas áreas correlatas à Biologia e Biotecnologia, com especial atenção às capacidades didáticas e lúdicas que ferramentas como o PyMOL (The PyMOL Molecular Graphics System, Version 1.7.4.5 Schrödinger, LLC.) e o Swiss-Model (BIASINI *et al.*, 2014) podem oferecer ao ensino de Biologia e Bioinformática.

Por fim, na conclusão desse trabalho, pretende-se trazer à luz uma ementa para a disciplina de Introdução à Bioinformática Estrutural. Com isso, objetiva-se deixar à disposição do curso de Licenciatura em Ciências Biológicas, e dos cursos a que aprover, um documento já pronto para uma possível nova disciplina.

## 1.2 O QUE É UMA EMENTA?

A ementa refere-se a um documento composto da relação resumida de diferentes conteúdos trabalhados dentro de uma disciplina. Os conteúdos descritos no seu corpo devem ser executados durante o período em que a disciplina estiver em aplicação. Em suma, são os tópicos de conteúdos a serem executados dentro de uma dada disciplina.

Segundo Spudeit (2014) o planejamento primário coloca-se como necessário para construir a ementa, uma vez que este deve unir a reflexão pedagógica do professor e os conteúdos a serem trabalhados. Mais ainda, é importante destacar que a ementa constitui-se de uma previsão mais global acerca das atividades de uma disciplina, sendo que fatores internos ou externos podem diminuir ou acelerar o andamento da disciplina, causando modificações na forma de aplicação da ementa.

A ementa é formada pela descrição em um parágrafo que declare quais conteúdos serão trabalhados. A descrição deve ser básica e direta e a ementa deve estar de acordo com as normas do projeto político pedagógico do curso, não podendo ser modificada pelo professor sem ser previamente aprovada pelas autoridades competentes de cada universidade.

### 1.3 O QUE É BIOINFORMÁTICA?

Computadores cada vez mais rápidos e mais baratos nos permitem abordar problemas, literalmente, inimagináveis há poucos anos.

(VERLI, 2014)

Para além das tarefas comuns do dia-a-dia que possamos ter com o computador, biólogos e outros cientistas também utilizam as ferramentas computacionais para resolver problemas que são específicos da biologia, como a modelagem de proteínas por homologia, e elucidação de genomas. Esses problemas constroem aquilo que se entende por bioinformática (CLAVERIE; NOTREDAME. 2007).

Com o advento nos últimos anos das tecnologias de informação e comunicação, da maior democratização do acesso à internet e a microcomputadores, as ciências têm cada vez mais se aproximado uma das outras, com áreas distintas encontrando meios para se integrarem a outras áreas de conhecimento que a princípio não são correlatas. Este é o caso da Bioinformática, uma ciência que tem sua origem no encontro entre a computação e a Biologia em meados dos anos de 1960 (VERLI, 2014).

As áreas biológicas ao longo do tempo passaram por grandes mudanças, como a descoberta de microorganismos por Leeuwenhoek, a Seleção Natural de Darwin, a hereditariedade de Mendel e a descoberta da estrutura do DNA por Watson e Crick. Todos esses acontecimentos modificaram a maneira como o homem observava a vida e como se comportava frente à natureza. Nos últimos anos, uma nova mudança ocorreu na área de estudo da vida, a genômica. Favorecido enormemente pelo Projeto Genoma, grandes laboratórios se buscaram decompor todos os organismos na sua estrutura principal, o DNA (ácido desoxirribonucleico), buscando desvendar todo o código genético de todos os organismos, das bactérias ao homem (GERAQUE, 2003).

Com o projeto genoma e o surgimento de sequenciadores automáticos de DNA, houve uma expansão na quantidade de informação gerada nos laboratórios. Por sua vez, esta informação necessitava ser armazenada e, dessa forma, surgiu a necessidade de recursos computacionais cada vez mais avançados para esta função. Além da função de acondicionar informação, com o passar do tempo os pesquisadores passaram a utilizar os computadores para analisar a informação preservada em seus bancos de dados, tornando necessário o uso de plataformas computacionais eficientes para interpretação dos resultados obtidos. Assim, é possível dizer que a bioinformática surgiu da necessidade em analisar a grande gama de informações biológicas depositadas em bancos de dados (PROSDOCIMI, 2002).

Destarte, segundo Verli (2014, p 2-3) “... a bioinformática se refere ao emprego de ferramentas computacionais no estudo de problemas e questões biológicas...”. Tisdall (2001) também define a bioinformática como a aplicação de ferramentas e técnicas computacionais para o gerenciamento e análise da informação biológica. Prosdocimi e Santos (2004) são mais específicos para designar a bioinformática, dizendo que esta consiste principalmente na análise *in silico* de sequências de DNA, RNA e proteínas. Além, conceituam que a bioinformática é todo estudo computacional que se pode realizar com biomoléculas para organizar e obter informações biológicas. Ao contrário de Prosdocimi e Santos (2004), Setúbal (2003, *apud* ARBEX; COSTA; SILVA, 2006) vai mais além e coloca a bioinformática como uma nova ciência que busca compreender todo o mundo de seres vivos a partir de conceitos da informática. De todo modo, sempre que houver referência à palavra bioinformática neste trabalho, entenda-se: Aplicação de ferramentas computacionais para a análise de informações biológicas.

Outros nomes também são aplicados à área da bioinformática como Biologia Molecular *in silico* e Biologia Computacional. Lengauer (2001) diz que o objetivo desse campo de pesquisa é construir métodos computacionais para analisar o grande volume de dados ainda sem a devida análise, gerados em diversos projetos de sequenciação de genomas ou outras tecnologias experimentais da biologia molecular. Para finalizar, o autor diz que o ramo da biologia computacional apresenta um dos grandes desafios do nosso tempo, uma vez que, ao mesmo tempo, a bioinformática é uma forte ferramenta para a obtenção informação imediata sobre o genoma, é também é uma ciência para o futuro. Assim, a bioinformática mais do que uma ciência para uso imediato, possui um vasto campo de pesquisa para as próximas gerações.

#### 1.4 A BIOINFORMÁTICA NO BRASIL.

Ao final dos anos de 1980, o Brasil deixou de ser um simples produtor de café, principalmente pela explosão da agricultura, oferecendo ao mundo outras mercadorias como a soja, carne, aves, frutas e farinha. O crescimento da agricultura veio juntamente com um plano socioeconômico bem estruturado conduzido pelo governo brasileiro, que visava dar apoio à ciência e tecnologia. Toda esta atividade econômica crescente permitiu ao Brasil investir parte da sua economia em pesquisa científica. (NESHICH, 2007).



Com tais acontecimentos, no ano de 1997 a FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) investiu em pesquisa na área de sequenciamento genômico. Os esforços obtiveram resultados quando no ano 2000 um grupo de pesquisadores brasileiros publica na revista *Nature* o trabalho *The genome sequence of the plant pathogen Xylella fastidiosa* (SIMPSON *et al.*, 2000). Este trabalho contém toda a sequência do genoma da bactéria *Xylella fastidiosa*, causadora de enfermidades em citros. Por ser o primeiro trabalho desta natureza produzido no Brasil, é referido como o primeiro trabalho em bioinformática em território nacional. Uma importante característica desta pesquisa, que auxiliou a expansão da bioinformática no país, é seu caráter não centrista. Na pesquisa participaram mais de vinte profissionais distribuídos por diversos laboratórios do país, alguns, a saber: Instituto Ludwig de Pesquisas sobre o Câncer, a Universidade Estadual de Campinas (UNICAMP), a Universidade Estadual Paulista (UNESP), Universidade Federal de São Paulo (UNIFESP), Instituto Biológico (IB), Instituto Agrônomo de Campinas (IAC), Instituto Butantan, Universidade de Mogi das Cruzes (UMC) e Universidade de Ribeirão Preto (UNAERP) (BONGIOLO, 2006).

Em pouco tempo outros trabalhos importantes produzidos no Brasil foram publicados, como o estudo do genoma de *Mycoplasma hyopneumoni*, uma bactéria que causa pneumonia em porcos. (VASCONCELOS *et al.* 2005). Hoje, alguns trabalhos brasileiros estão voltados ao sequenciamento do genoma da onça-pintada, *Panthera onca* (HEIDTMANN, 2014). Contudo, estas não são as únicas fontes de pesquisa de Bioinformática no Brasil.

O Ministério da Ciência e Tecnologia e Inovação (MCTIC) por meio do Conselho Nacional de Desenvolvimento Científico e tecnológico (CNPq) apoiou a implantação de redes regionais com o objetivo de realizar estudos genômicos de organismos que possuam interesse social, regional ou econômico. No ano de dois mil e seis, nove campos de pesquisa distribuídos por todo o território brasileiro estavam com suas pesquisas voltadas ao genoma de organismos importantes de suas respectivas regiões. (BONGIOLO, 2006).

Na Região Sul, o Instituto de Biologia Molecular do Paraná e o Programa Genoma do Paraná (GenoPar) implementaram um programa para a busca sistêmica de genes expressos pelo protozoário *Trypanosoma cruzii* e determinar o genoma da bactéria *Herbaspirillum seropedicae*, respectivamente (PEDROSA *et al.*, 2011). Na região também há o Programa de Investigação de Genomas do Sul, que é focalizada em saúde animal, especialmente suínos, e busca decifrar o genoma da bactéria *Mycoplasma synoviae* (VASCONCELOS *et al.*, 2005).

Na Região Sudeste, a Rede Genoma Minas Gerais e a Rede Genômica do Estado do Rio de Janeiro objetivam construir um retrato da expressão gênica do inseto *Anopheles darlingi* (MARINOTTI *et al.*, 2013) em diferentes situações e sequenciar o genoma da bactéria *Gluconacetobacter diazotrophicus*, que causa danos à cana-de-acúcar (BERTALAN *et al.*, 2009).

No Nordeste, por sua vez, a Rede Genômica do Nordeste tem por objetivo estudar os genes expressos pelo protozoário *Leishmania chagasi*, na busca de entender o genoma do organismo causador da Leishmaniose (DINIZ; OLIVEIRA, 2001). Na Rede Genômica do Centro-Oeste os trabalhos estão focados em mapear o genoma funcional do organismo *Paracoccidioides brasiliensis*, o fungo responsável pelas denominadas micoses. (FELIPE *et al.*, 2005). Por sua vez, a Rede da Amazônia Legal de Pesquisa Genômica (Realgene) tem seu foco voltado para o genoma do guaranazeiro (*Paullinia cupana*). (ÂNGELO, *et al.*, 2007; FREITAS *et al.*, 2007).

Para mais, ressalta-se o fato de que países da América Latina como a Argentina (BASSI; GONZÁLEZ; PARISI, 2007), Colômbia (BENITEZ-PAEZ ; CARDENAS-BRITO, 2010) e Cuba (PONS; MONTERO; FEBLES, 2007) buscaram descrever como a bioinformática é utilizada em cada região, com suas dificuldades e avanços. No Brasil, pela mesma época, o trabalho de Neshich (2007) é o maior expoente do estado da arte em que se encontra a pesquisa em Biologia Computacional no País.

## 1.5 BIOINFORMÁTICA: UM INÍCIO PARA O IFSULDEMINAS – Campus INCONFIDENTES

Mesmo em face ao grande crescimento da bioinformática nos últimos anos, Barata (2003) ressalva que a capacitação ainda é o grande obstáculo para a bioinformática em nosso país. O autor comenta que os Estados Unidos da América se mantem em primeiro lugar em pesquisas de biotecnologia, principalmente pelo Centro Nacional de Informação Biotecnológica (NCBI), seguido pelos países Europeus, como a Inglaterra, França, Alemanha e Suíça, após encontra-se o Japão com seu Banco de dados de DNA (DDBJ). Nos países em desenvolvimento, Brasil e Índia são os países que mais investem em pesquisa na área.

Alguns autores como Júnior *et al.* (2011), Farias, Chacon e Silva (2012) criaram cursos introdutórios às ferramentas de bioinformática, pois ao que descrevem, a bioinformática ainda é desconhecida de muitos estudantes do ensino superior, mesmo com todo o acesso tecnológico que temos. No caso de Júnior *et al.* (2011), o autor expõe que há um

grande desconhecimento da área no Nordeste, principalmente porque não há trabalhos a nível de graduação que coloquem os alunos em contato direto com a área. Há somente encontros e trabalhos em nível de pós-graduações, em mestrados e doutorados. Mesmo assim, o contato só ocorre em seminários e eventos com palestrantes que provém de outras regiões. Em suas considerações finais, ambos os autores discorrem sobre a importância da aplicação e trabalho com conceitos novos e emergentes, contribuindo para um melhor conhecimento dos alunos nas áreas que trabalham diretamente com o genoma, buscando fazer uma conexão direta entre o conhecimento *a priori* concebido através da biologia e o trabalho que se pode realizar com ferramentas disponíveis para a análise de dados gerados a todo o instante, seja na genômica, transcriptoma ou proteômica.

Assim, devido à escassez de profissionais atuantes no Brasil, a inovação prestada por um curso introdutório à bioinformática em um Instituto Federal de Ensino Superior pode promover uma nova cultura de estudos para diversos alunos. A bioinformática ainda é grande desconhecida de muitos graduandos e é muitas vezes tratada de maneira secundária no debate acadêmico.

Sabbatini (1999) diz que o profissional bioinformata no Brasil tem sua formação de maneira autodidata e que, antes de tudo, o profissional para atuar nesse ramo precisa ter uma mentalidade que aproxime a biologia, matemática, informática e outros ramos de pesquisa necessários à área. Além disso, o autor completa que a bioinformática é uma profissão promissora para futuro e que a universidade precisa urgentemente se movimentar para formar esse profissional que será crucial no futuro das ciências. Ainda, segundo Araújo *et al.* (2008) a bioinformática têm sido alvo de grande interesse nos tempos atuais, principalmente nos aspectos tangentes ao âmbito biomédico, sendo caracterizada como uma das mais importantes ferramentas para essa área da saúde.

Outro ramo da pesquisa que tem muito se desenvolvido com o apoio das ferramentas de bioinformática é o *Structure-based drug design* (Planejamento de fármacos baseado na estrutura do receptor, em português. Sigla SBDD). Nesse contexto, salienta Guido, Andricopulo e Oliva (2010), que o auxílio prestado pela bioinformática possibilita uma melhor organização, gerenciamento, visualização e interpretação da informação resultante do estudo das estruturas tridimensionais, aumentando a integralidade dos dados biológicos em um sistema.

Seguindo esses aspectos, torna-se necessária a abertura da Universidade ao estudo da bioinformática. Com o curso, pretendeu-se promover aos alunos participantes o conhecimento primário em um ramo de pesquisa que se mostra cada vez mais promissora para o desenvolvimento científico. Assim, pode-se dizer que a aplicação de um curso com características inovadoras no ramo da ciência acrescenta em muito não só para os alunos que cursam a graduação, mas para o próprio IFSULEMINAS- Campus Inconfidentes que passa a fornecer suporte àqueles que buscam doar seu tempo a novas áreas de ensino e pesquisa.

Há ainda de se ressaltar que, conforme Padilha *et al.* (2008), a bioinformática tem crescido ao passo em que novas descobertas científicas são geradas, contudo a falta de recursos humanos dificulta o aprendizado e difusão da bioinformática no Brasil. Uma das maneiras que torna possível levar esse conhecimento aos estudantes é através de cursos de introdutórios à área.

Neste trabalho, buscou-se levar a uma parcela dos alunos licenciandos em ciências biológicas a participação em um curso introdutório às ferramentas utilizadas em bioinformática. Ante a expansão da bioinformática nos últimos anos, e segundo Farias, Chacon e Silva (2012) torna-se imprescindível a oferta de cursos que ofereçam capacitação para a compreensão de problemas biológicos e utilização de ferramentas de bioinformática.

A escolha de trabalhar este tema está baseado na necessidade da abertura da universidade às novas possibilidades de conteúdos a serem trabalhados na Licenciatura em Ciências Biológicas, em suas áreas correlatas e as áreas ligadas à computação. Este é um passo adiante em um período que tem se voltado ao trabalho acadêmico que busca unir, por meio de ferramentas computacionais, os mais diferentes ramos das pesquisas em busca de objetivos e fins únicos.

A Bioinformática tem tomado espaço cada vez maior no que tange a construção e armazenamento de conteúdos em genômica, proteômica, e outras áreas relacionadas com o código genético. “assim como há mais de 20 anos a Biologia Molecular passou a fazer parte da rotina de quase todos os laboratórios de biologia do mundo, o mesmo vai ocorrer com a bioinformática no futuro próximo” (GERAQUE, 2003).

Por tudo isso, mostra-se necessária a abertura do espaço da educação superior para a bioinformática, para que cada vez mais alunos se interessem pela área e busquem auxiliar no desenvolvimento biotecnológico, não só do país, como da própria ciência, que necessita de novos conhecimentos e métodos para desenvolver-se permanentemente.

## 2. METODOLOGIA

Para este trabalho, a metodologia está dividida em duas partes, a saber: o Pré-curso e o Curso.

### 2.1. PRÉ-CURSO

Para informar os licenciandos acerca do curso e obter informação sobre a disponibilidade e o interesse em participação destes no projeto, nos dias 04/03/2016 e 07/03/2016 o autor deste trabalho apresentou-se em todas as turmas convidando a todos para participação e informando-os de como será trabalhado o curso. Para captar as informações necessárias para dar início ao curso houve a distribuição de um questionário anônimo (anexo I) para mensurar a quantidade de alunos disponíveis para participação. Em seguida, durante os dias 15/03/2016 e 16/03/2016 um documento (anexo II) foi distribuído aos estudantes com o intuito de recolher os nomes e e-mails dos interessados em participar do evento. Os dados obtidos com os questionários serão discutidos adiante, no capítulo sobre os Resultados e Discussões.

Para a seleção dos estudantes foi levado em consideração o critério de possuir *notebook*. Ao final do primeiro critério, vinte e cinco alunos possuíam os requisitos necessários para a participação. Para selecionar apenas vinte estudantes, houve a realização de um sorteio simples, com o nome de todos os vinte e cinco possíveis participantes. Por motivos de ética o nome dos estudantes não foram expostos em momento algum deste trabalho. Sempre que necessário, foram substituídos por Aluno 1, Aluno 2, e assim por diante.

## 2.2 CURSO

O curso foi realizado no Laboratório de Física Experimental do Instituto Federal do Sul de Minas Gerais – Campus Inconfidentes. A aplicação do curso ocorreu em um período de três meses, com encontros semanais às quartas-feiras, entre as 17h15min e 18h50min totalizando 20hs de curso. Os quadros I e II mostram como foram dispostos os conteúdos teóricos e práticos do curso. É importante ressaltar que todos os participantes finalistas receberam um certificado de participação no curso com 20 horas de Atividades Acadêmicas extracurriculares.

Aulas Teóricas		
Aula	Data	Conteúdos
1	23/03	Apresentação
2	31/03	Origem, história e ramos da bioinformática
3	06/04	Somente prática
4	13/04	Somente prática
5	20/04	Dogma central da Biologia Molecular
6	27/04	Proteínas
7	04/05	Avaliação 1
8	11/05	Alinhamento de proteínas
9	18/05	Cristalografia/ Modelagem por homologia
10	25/05	Somente prática
11	01/06	Revisão Alinhamento e Modelagem por homologia
12	15/06	Avaliação 2

**Quadro I:** Aulas teóricas. **Fonte:** Autor do trabalho.

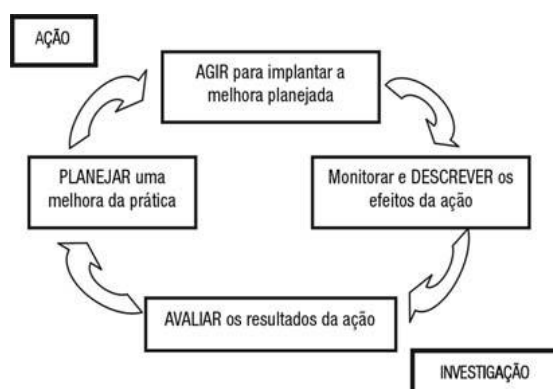
Aulas Práticas		
Prática	Datas	Conteúdos
1	30/03	Banco de dados PDB
2	06/04	Visualização de aminoácidos
3	13/04	Visualização e manipulação de estrutura secundária de proteína
4	20/04 e 27/04	Estrutura terciária e quaternária de proteínas/ Alinhamento estrutural
5	11/05 e 18/05	Alinhamento de proteínas
6	25/05 e 01/06	Modelagem por homologia

**Quadro II:** Aulas práticas. **Fonte:** Autor do trabalho.

Em conformidade com a necessidade de se trabalhar os conteúdos de bioinformática e poder, ao final, construir uma ementa para a disciplina de Introdução à bioinformática Estrutural, a pesquisa proposta aqui possuiu um cunho de pesquisa-ação. Segundo Novaes e Gil (2009) A pesquisa-ação é uma pesquisa de caráter humanista e crítica, de cunho

interpretativista, ou seja, não propõe corroborar com assertivas fixas e imutáveis, mas, sim, com resoluções norteadoras de interpretação da realidade. Este tipo de pesquisa foi também escolhido, pois segundo os mesmos autores, é de grande valia quando se deseja promover mudanças organizacionais. Tripp (2005) ressalta que a pesquisa-ação é somente um de vários tipos de investigação-ação, sendo este um termo para a pesquisa que aprimora a prática a partir do agir no campo prático e a investigação posterior. O diagrama I demonstra como ocorre a pesquisa ação.

Partes teóricas, principalmente com temática relacionada às áreas de Biologia Molecular, Bioquímica e Evolução serão contempladas. Faz-se necessário o trabalho teórico dentro desse curso, uma vez que conhecimentos como o fluxo de informações biológicas na célula (Dogma Central da Biologia), os processos de duplicação, transcrição e tradução, e, do mesmo modo, a função dos aminoácidos, peptídeos e proteínas estão subjetivamente ligados à manipulação computacional.



**Diagrama I:** As fases da investigação ação. Note que mesmo não estando representado, o pesquisador é parte central da pesquisa planejando, agindo, descrevendo e avaliando todo o trabalho. **Fonte:** Tripp (2005)

## 2.3 CONSTRUÇÃO DAS AULAS TEÓRICAS E AULAS PRÁTICAS.

A princípio, a construção e aplicação das aulas práticas e teóricas se dariam em momentos distintos ao decorrer do curso. As aulas de cunho teórico seriam trabalhadas nas primeiras seis aulas do curso, postergando para as aulas finais os conteúdos práticos. Todavia, para não tornar o ensino tedioso e monótono – ainda mais em um conteúdo cujos estudantes ainda não possuem conhecimento prévio – ficou claro a necessidade de mesclar as aulas teóricas e as aulas práticas, trabalhando ambas em conjunto, com explicações teóricas sempre a frente das práticas, constituindo em aulas teórico-práticas.

As aulas teóricas tiveram cunho expositiva e buscavam passar aos participantes os conhecimentos necessários para um bom entendimento das práticas, estas essenciais para o aprendizado da manipulação dos programas de bioinformática. Para auxílio nas aulas, foram utilizados *Data Show* e quadro negro.

Para conduzir as aulas, os livros textos utilizados foram Verli (2014), intitulado *Bioinformática: da Biologia à flexibilidade Molecular*. O *ebook* pode ser baixado livre e gratuitamente no seguinte link: (<http://www.ufrgs.br/bioinfo/ebook/>) e Alberts (2010) *Biologia Molecular da Célula*, disponível na biblioteca desta instituição de ensino.

As aulas práticas seguiram roteiros previamente preparados e eram efetuadas após as aulas teóricas. O intuito das aulas práticas é familiarizar os alunos e seus conhecimentos teóricos através de programas disponíveis livremente na *web*. Por livremente, entenda-se que todos os *softwares* utilizados durante este trabalho estão disponíveis gratuitamente sobre a licença GPL GNU (Licença Pública Geral). Durante o curso foram utilizados diversos programas para auxiliar no processo de aprendizagem, sendo eles:

### 2.3.1 VMD

VMD (*Visual Molecular Dynamics*) é projetado para modelagem, visualização e análise de sistemas biológicos, tais como proteínas e ácidos nucleicos. Por ler arquivos em formato PDB, o programa pode visualizar qualquer tipo de molécula que esteja gravada sob estes aspectos. O VMD fornece uma ampla variedade de métodos para renderização e colorir uma molécula: pontos simples, linhas, esferas, entre outros. VMD pode ser usado para analisar a trajetória de uma simulação de dinâmica molecular (DM). Em particular, VMD pode atuar como uma interface gráfica de um programa externo de DM, exibindo uma



molécula submetido a simulação em computador remoto ou *software* remoto. (HUMPHREY; DALKE; SCHULTEN, 1996)

### 2.3.2. PyMOL

PyMOL é uma ferramenta utilizada para visualizar moléculas biológicas. O *software* é compatível com Windows, Linux e MacOS. PyMOL tem excelentes capacidades na criação de imagens de alta qualidade a partir de estruturas 3D, possui funções bem desenvolvidas para manipulação de estruturas e algumas funções básicas para analisar suas propriedades químicas. As possibilidades para escrever scripts e plugins, bem como para incorporar PyMOL em software personalizado são vastas e superior à maioria dos outros programas (ROTHER, 2005).

### 2.3.3 Swiss Model

O Swiss Model é um sistema automatizado para a modelagem de estruturas 3D a partir da sequência de aminoácidos. Swiss Model é o primeiro servidor completamente automatizado para modelar estruturas proteicas homólogas e é sucessivamente melhorado por seus idealizadores. Possui uma interface de fácil manuseio, o que permite que pessoas que não sejam especialistas na área possam modelar suas proteínas de interesse. O Swiss Model é um dos mais acessados servidores da web para modelagem proteica no mundo, contando com quase 1 milhão de acessos anuais. (BIASINI *et al.*, 2014)

É necessário deixar registrado que o servidor Swiss Model não estava programado para ser usado durante o curso por possuir qualidade inferior ao servidor *I-TASSER* (ZHANG, 2008), contudo a utilização deste servidor se mostrou inviável uma vez que para sua utilização é necessário que os alunos possuam e-mail acadêmico, entretanto os alunos relataram durante as aulas não possuírem qualquer e-mail sem ser o pessoal, o que invalida a utilização do *I-TASSER*. Mesmo assim, o fato de o servidor utilizado durante o curso possuir qualidade inferior não é capaz de produzir quaisquer danos, uma vez que a utilização de tal ferramenta foi pensada apenas para fins didáticos.

## 2.4 COLETA DE DADOS

Para melhor averiguação do curso ministrado, quatro tipos de documentos foram produzidos durante os três meses de pesquisa. Os documentos criados foram dois diários de campo, um para as aulas práticas e um para as aulas teóricas. Ademais, os alunos finalizantes do curso responderam a duas provas cada um e construíram seis práticas. Todos estes documentos serão explanados no capítulo de resultados e discussões.

### Os diários de campo

...são um instrumento magnífico para identificar quais questões são dilemas para cada professor e como ele vai enfrentá-los. Lendo os diários, vê-se, algumas vezes com clareza e outras vezes nas entrelinhas, quais são os dilemas que mais preocupam esse professor ou essa professora, qual o tipo de situações da dinâmica de sua aula que se transformam em momentos dilemáticos e como raciocina (e vive) a resolução desses momentos. (ZABALZA, 2003)

Assim sendo, os diários foram escolhidos, pois, segundo o mesmo autor, auxiliam o pesquisador a melhor compreender seus dilemas e a natureza complexa da atividade de lecionar, criando questões muito próximas à realidade. Com isso, objetiva-se com os diários, e os dilemas criados a partir deste, melhor interpretar as questões auxiliares à construção da ementa disciplinar. Os diários eram escrito pré-aula e pós-aula. Antes do início das aulas, era anotado como a esta foi construída e como deveria transcorrer. Ao final, a segunda parte consistia em descrever como ocorreram as atividades e quais semelhanças e diferenças houve entre o que foi escrito anteriormente e o escrito posteriormente.

Por parte dos alunos, a construção das aulas práticas ocorria durante as aulas ou em casa. Para construir as práticas, os alunos seguiam os roteiros e respondiam às questões neles contidas. Durante o curso houve dois exames. O primeiro era constituído de quatro questões e possuía caráter diagnóstico, buscavam identificar no aluno quais eram seus apontamentos sobre o curso, de maneira a fazer o curso atender às pretensões dos alunos. O último exame possuiu caráter somativo, com o intuito de mensurar estatisticamente se os conhecimentos trabalhados durante o curso foram compreendidos pelos estudantes. O último exame foi composto de vinte questões ao todo. Todas as discussões sobre estes temas estarão presentes nos Resultados e Discussões.

## 2.5 CONTEÚDOS

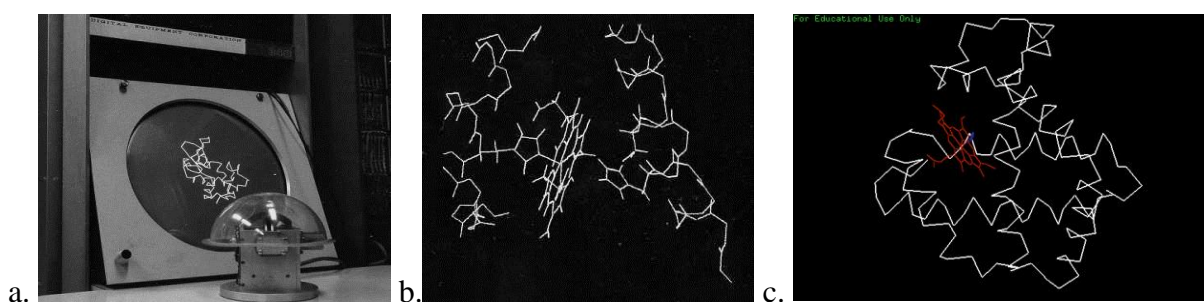
Nesta seção serão descritos teoricamente todos os conteúdos trabalhados durante o curso em ordem cronológica.

### 2.5.1 A história da Bioinformática. Do seu início até o primeiro banco de dados.

O momento chave para o início da bioinformática pode ser traçado na década de 1950, quando Watson e Crick (1953) publicaram o trabalho “*Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid*” sobre a estrutura em hélice do DNA (ácido desoxirribonucleico) (VERLI, 2014).

Ao contrário do que se possa presumir, as ferramentas computacionais começaram a ser aplicadas à Biologia Molecular muito antes do começo da internet e dos processos de sequenciamento genômico. Já na década de 1960 os cientistas tiveram a necessidade de combinar técnicas de biologia molecular, matemática e computação para defrontar os desafios que a biologia possuía devido à grande quantidade de dados acerca da química de proteínas (FRANCO; CEDIEL; PAYÁN, 2008).

O primeiro trabalho a utilizar ferramentas computacionais para visualizar moléculas biológicas foi publicado somente treze anos após o trabalho de Watson e Crick. Levinthal (1966) desenvolveu um construtor de modelos proteicos que permitia a visualização e o estudo da interação de átomos em pequenas moléculas. A visualização ocorria em um osciloscópio monocromático (Figura 1).



**Figura 1:** (a) Osciloscópio utilizado por Levinthal em seus experimentos. Detalhe para a proteína exposta na tela. (b) Detalhes da estrutura da mioglobina mostrando o grupo heme com dois segmentos da cadeia de polipeptídeo que o rodeia. **Disponível em:** <http://www.umass.edu/molvis/francoeur/levinthal/lev-index.html>. **Acesso em:** 22/07/2016 (c) representação da proteína mioglobina (PDB ID 1MBO) no *software* PyMOL (The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC.). Em destaque na cor vermelho está representado o grupo Heme da proteína.

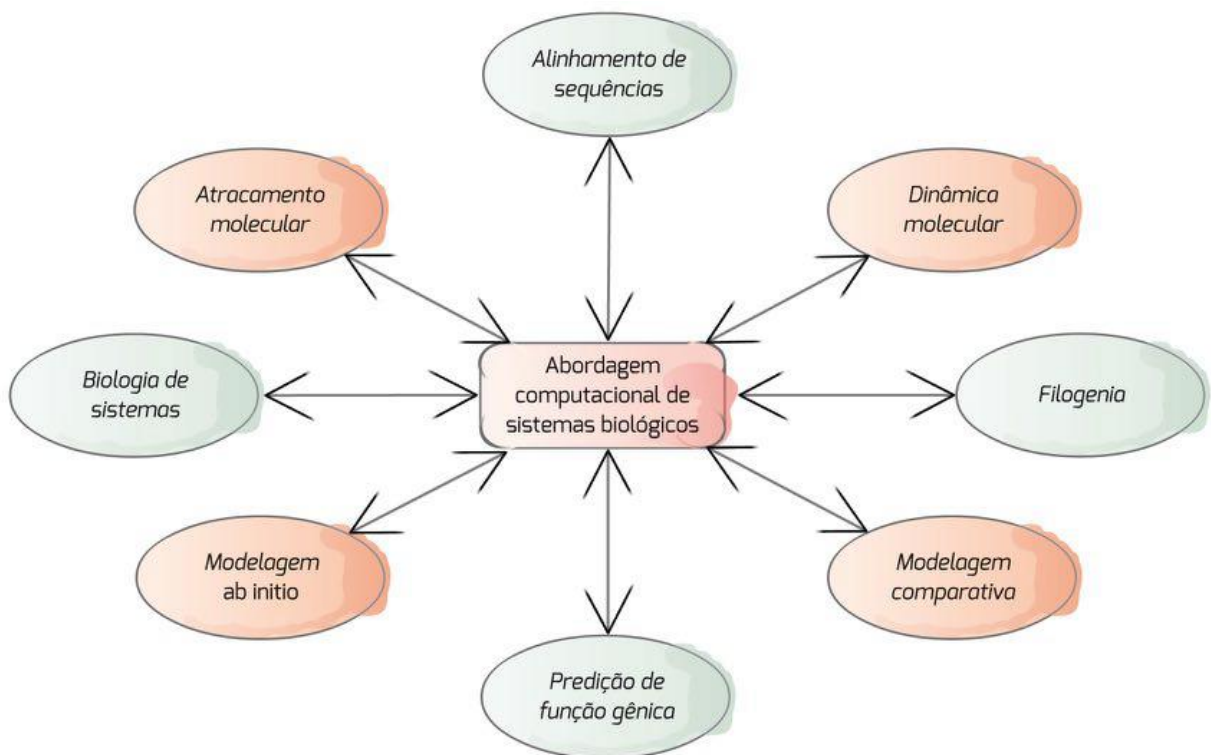
Com o desenvolvimento de métodos de sequenciamento (SANGER; COULSON, 1975) foi possível obter os dados de diversas das mais comuns famílias proteicas, como os

citocromos de diversos organismos. Por este período, a Dr. Margaret Dayhoff e seus colaboradores na *National Biomedical Research Foundation* (Fundação Nacional de Pesquisa Biomédica, em português. Sigla NBRF) desenvolveram métodos computacionais que permitiram comparar sequências de proteínas a partir de alinhamentos e investigar as relações entre eles e, portanto, a história evolutiva entre os diferentes Reinos, Filos e Taxa biológicos. Sua obra final contendo todas as 65 proteínas estudadas por ela foi armazenado em um livro intitulado *Atlas of protein sequence and structure*, publicado no ano de 1965 (FRANCO; CEDIEL; PAYÁN, 2008). Durante esse período o centro de coleta desses pesquisadores tornou-se conhecido pelo nome de *Protein Identification Research* (PIR). Este banco de dados foi mantido pelo NBRF a partir de 1984 e em 1988 foi criado o PIR – *Protein Information Resource* (<http://pir.georgetown.edu/>), em colaboração com o *Munich Information Center for Protein Sequences* (Centro para Sequências Proteicas de Munique, em português. Sigla: MIPS) e o *Japanese International Protein Sequence Database* (Banco de Dados Internacional de Proteínas do Japão, em português. Sigla: JIPID). Dra. Dayhoff também classificava as proteínas em famílias e superfamílias baseado no grau de similaridade sequencial. O trabalho conduzido por Dayhoff possui grande valor para a bioinformática dado que a partir deles surgem os primeiros bancos de dados de sequências proteicas. (MOUNT, 2001).

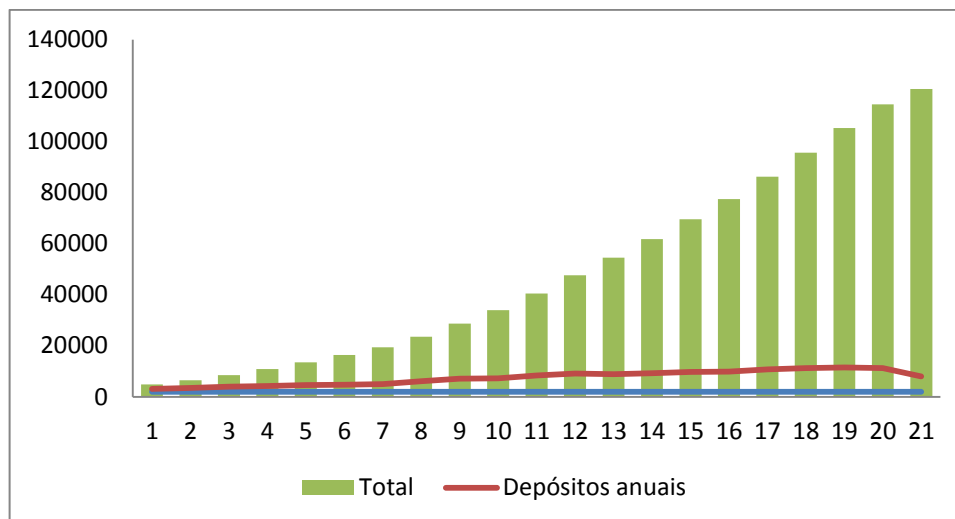
Segundo Santos (2010) a primeira vez que a palavra bioinformática foi utilizada para dar nome a um novo ramo da pesquisa ocorreu no ano de 1979, quando a Dra. Paulien Hogeweg publicou um artigo intitulado *Bioinformatica: een werkconcept* (Bioinformática: um conceito de trabalho, em português). Durante os primeiros anos da bioinformática seu principal uso se dá nas áreas de genética e genômica, especialmente para manuseio de grande quantidade de informação advinda do sequenciamento de ácido desoxiribonucleico (DNA), ácido ribonucleico (RNA) e aminoácidos. Hoje, a bioinformática é utilizada em diversas áreas e contribui para a criação de bancos de dados, nas análises de sequências genômicas, na análise e busca por genes e na predição de funções de proteínas, além de auxiliar no entendimento e resolução de estruturas proteicas tridimensionais, filogenia, construção de modelos evolutivos e bibliotecas genômicas. Hoje podemos dividir a bioinformática, conforme Verli (2014) em duas grandes áreas. A primeira é chamada de Bioinformática tradicional, por vezes chamada de clássica, que trabalha com a informação biológica obtida a partir de nucleotídeos e aminoácidos e proteínas. A segunda é conhecida como Bioinformática estrutural, que se preocupa com questões biológicas a partir do ponto de vista tridimensional (Figura 2).

### 2.5.2 Banco de dados

Banco de dados são coleções organizadas que estruturam e organizam informações para facilitar a consulta ao acervo e a sua renovação constante. A necessidade de organizar as informações biológicas para facilitar o acesso e a pesquisa tornou indispensável a construção de banco de dados para armazenar a informação crescente. O gráfico abaixo (Gráfico 1) retrata o crescimento do banco de dados internacional *Protein Data Bank* (PDB) que entre 1996 e 2016 saltou de 4985 estruturas para 120624 estruturas em 2016. Isso representa um crescimento médio de 5781 estruturas ao ano, aproximadamente. (BERMAN *et al.*, 2016).



**Figura 2:** Representação das principais áreas da Bioinformática. Em verde visualizamos as áreas relacionadas à bioinformática tradicional. Em laranja visualizamos as áreas relacionadas à bioinformática estrutural. **Fonte:** (VERLI, 2014).



**Gráfico 1:** Crescimento do Banco de Dados Protein Data Bank (PDB) entre os anos de 1996 (1) e 2016 (21). **Fonte:** <http://www.rcsb.org/pdb/statistics/contentGrowthChart.do?content=total&seqid=100>. **Acesso em:** 07/07/2016

Os bancos de dados podem ser classificados em primários e secundários. Banco de dados primários apresentam resultados de pesquisas experimentais, onde não há análises cuidadosas dos dados com relação aos dados depositados anteriormente. O GenBank, Sayers *et al.* (2009) e o PDB, Berman *et al.* (2016) são exemplos de bancos primários. Os bancos de dados secundários são aqueles onde os dados fornecidos são compilados e interpretados por vários cientistas, de forma a obter dados mais representativos e interessantes. São exemplos o *Reference Sequence* (RefSeq) (PRUITT; TATUSOVA; MAGLOTT, 2004) e o Universal Protein Resource (UniProt) (The UniProt Consortium, 2014).

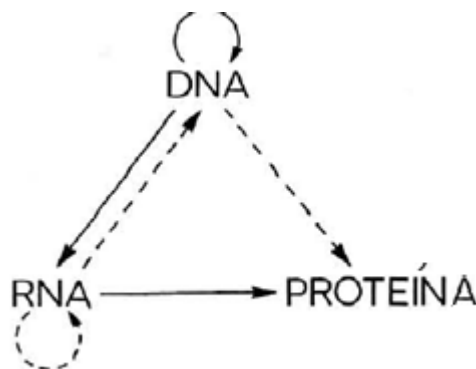
Nos últimos anos a quantidade de bancos de dados criados é muito grande. São tantos que todos os anos a revista internacional *Nucleic Acids Research* (2016) reserva duas publicações anuais apenas para apresentar os bancos de dados criados nos últimos seis meses ou as melhorias ocorridas em bancos de dados já existentes.

### 2.5.3 O dogma central da Biologia. Aspectos de duplicação, transcrição e tradução.

“O dogma central da biologia molecular lida com a transferência detalhada, resíduo por resíduo, da informação sequencial. Determina que tal informação não possa ser transferida de proteína para outra proteína, nem de proteína para ácidos nucleicos.” (CRICK, 1970). Em suma, o dogma central da biologia é um fluxo de informação em sistemas biológicos, tendo

como ponto de partida a molécula de DNA até a produção da proteína, sendo os mecanismos de duplicação, transcrição e tradução necessários (VERLI, 2014).

A partir do conhecimento das regras gerais acerca do dogma da transferência de informações sequenciais de um polímero para outro, foi possível estabelecer uma orientação por onde a informação pode ser transmitida e recebida (Figura 3). Para além disso, análises mostraram que a transferência possa ser dividida em três grupos distintos: Grupo de transferência geral, grupo de transferência especial e grupo de transferência desconhecida. Os grupos de transferência geral comportam a transferência de informações genéticas de DNA para DNA, DNA para RNA e RNA para proteína. Esse grupo recebe esse nome dado que pode ocorrer em qualquer célula. Grupo de transferência especial podem não ocorrer na maioria das células, todavia pode ocorrer em determinadas circunstâncias. São elas RNA para DNA, RNA para RNA e DNA para proteínas. Para terminar este quadro, há ainda o grupo de transferências não conhecidas, postuladas pelo dogma como de ocorrência não existente. São elas, proteína para proteína, proteína para RNA e proteína para DNA (CRICK, 1970).

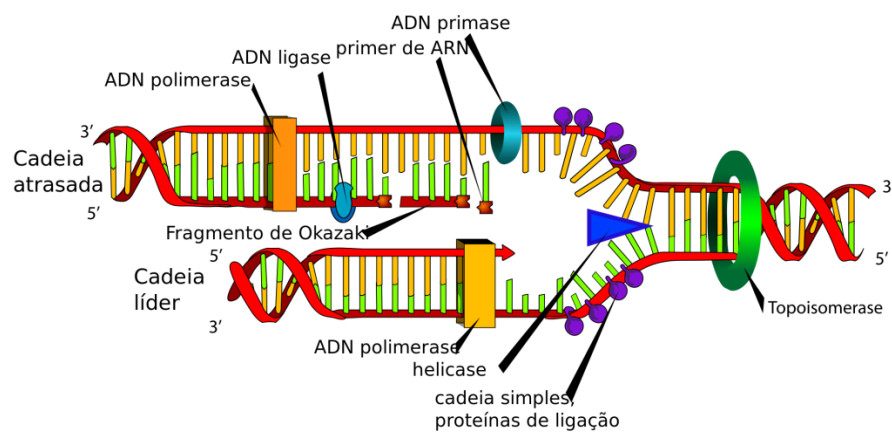


**Figura 3:** Representação do Dogma Central da Biologia. Em linhas contínuas observamos o grupo de transferência geral. Em linhas tracejadas, observamos os grupos de transcrição especial. **Adaptado de:** (CRICK, 1970)

Para manter toda a informação e um alto grau de organização, é necessária a duplicação de grandes quantias de DNA de maneira exata, e, embora divergências e recombinações garantam uma melhor sobrevivência de uma determinada espécie, a manutenção da vida do organismo depende de estabilidade genética. Alterações no DNA são conhecidas como mutações, podendo em alguns casos levar à morte o organismo mutante (ALBERTS *et al.*, 2010).

A proposta de dupla hélice para o DNA como uma estrutura que contem duas cadeias em forma de hélices girando sobre o mesmo eixo (WATSON; CRICK, 1953) abriu espaço

para um grande número de hipóteses de como o DNA poderia se replicar. No ano de 1958 o experimento de Menselson provou que o DNA se duplica de forma semiconservativa, ou seja, a partir de um DNA molde, dois novos DNA's são produzidos, sendo que cada novo DNA contém uma cadeia antiga e uma nova cadeia, sendo esta produzida a partir da cadeia antiga (MESELSON; STAHL, 1958). Como acima mencionado, o DNA é composto por fitas duplas, sendo uma conhecida pelo nome de fita-líder, ou cadeia líder, e a outra pelo nome de fita descontínua, ou cadeia descontínua. A identificação das fitas se dá pela disposição de transcrição 5' 3' contínua na cadeia líder (*leading strand*, em inglês), e a presença de descontinuidades (fragmentos de Okasaki) 5' 3' na cadeia atrasada (*lagging strand*, em inglês).



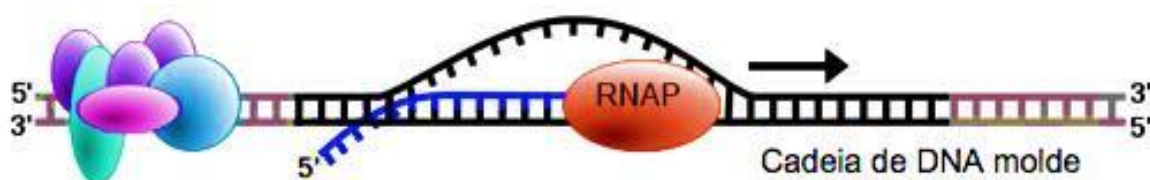
**Figura 4:** Esquema representativo da duplicação do DNA. Fonte: [https://www.lookfordiagnosis.com/mesh\\_info.php?term=replica%C3%A7%C3%A3o+do+dna&lang=3](https://www.lookfordiagnosis.com/mesh_info.php?term=replica%C3%A7%C3%A3o+do+dna&lang=3). Acesso em: 07/07/2016

De maneira sucinta, para que a duplicação seja efetiva, vários fatores são necessários. Elencando os principais de maneira cronológica na duplicação, é necessária a presença de uma origem de replicação, DNA's helicases, DNA's-primases, DNA's polimerases, DNA's-ligases (no caso, para estabelecer ligação entre os fragmentos de Okasaki), proteínas desestabilizadoras de hélices, um cinta reguladora que mantém a DNA-polimerase ligada à fita em duplicação, DNA's topoisomerases e a atuação da telomerase nas extremidades de cada telômero (nos eucariotos). Além de mecanismos que auxiliam na recuperação e reparo de lesões sofridas pelo DNA durante replicação, como é o caso do reparo por excisão de bases e o reparo por excisão de nucleotídeos e muitos outros (ALBERTS *et al.*, 2010).

No passo a frente, o RNA é transcrito a partir do DNA por um processo chamado transcrição. Este é o passo intermediário para transformar o conteúdo sequencial presente no



DNA em macromoléculas biológicas para que estas atuem em diversas áreas e de diversas maneiras por entre todo um organismo vivo. A transcrição produz um RNA que é complementar a uma das fitas que compõem o DNA, e as enzimas que realizam a transcrição são chamadas de RNA-polimerases. Para a transcrição é necessário uma subunidade destacável conhecida como fator sigma, que forma a holoenzima RNA-polimerase quando em contato com o cerne da polimerase. Em seguida o fator sigma liga-se a região promotora do DNA, começando o processo de criação de RNA pela RNA-polimerase. A transcrição ocorre até que se encontre uma região no DNA conhecida como região terminadora, finalizando o processo de confecção do RNA.

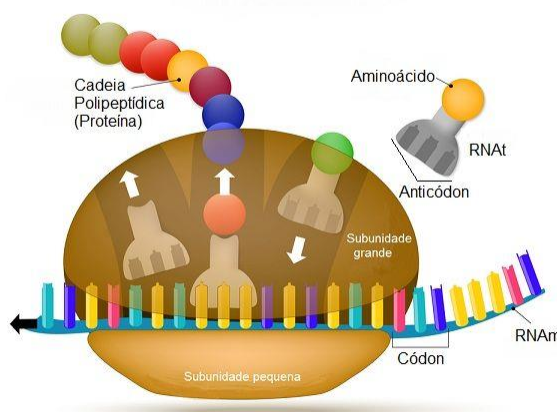


**Figura 5:** Representação dos passos para a transcrição de um DNA em RNA. **Fonte:** <http://wikiciencias.casadasciencias.org/wiki/index.php/Transcri%C3%A7%C3%A3o>. **Acesso em:** 07/07/2016

O RNA produzido é chamado de pré-RNA, pois ainda contém em si partes não codificantes de proteínas. Nos pré-RNA encontramos éxons, que são partes codificantes de proteínas e íntrons, que são partes não codificantes. Há dois mecanismos para a retirada dessas regiões não codificadoras. O primeiro deles é chamado de *splicing*. Este é um mecanismo coordenado por snRNA's (*small nuclear* RNA. Pequeno RNA nuclear, em português) que consiste em rearranjos estruturais e posterior retirada de íntrons. Além desse mecanismo ocorre em organismos inferiores o chamado *auto-splicing*, onde o próprio pré-RNA cumpre o papel da retirada dos íntrons. Acredita-se que o *splicing* é uma característica apomórfica sobre o *auto-splicing*. Ao final desse processo os RNA's já são considerados maduros e são seletivamente exportados do núcleo para o citosol, onde podem codificar as proteínas. Os principais RNA's produzidos nas células são: mRNAs (RNA mensageiro, que codificam proteínas), rRNAs (RNA ribossômicos, que formam a estrutura básica do ribossomo), tRNAs (RNA transportador, essenciais na síntese proteica), snRNAs (*small nuclear* RNA. Auxiliam no *splicing*), snoRNAs (*small nucleolar* RNA. Atua no processamento e modificação química de rRNAs), scaRNAs (*small cajal* RNA. Modificam

snoRNAs e snRNAs), miRNAs (micro RNA. Regulam a expressão gênica por bloqueio da tradução de mRNAs), siRNAs (*small interfering RNA*. Degradam mRNAs, auxiliando na expressão gênica) e outros RNAs não-codificantes (diversos processos celulares) (ALBERTS *et al.*, 2010).

O último estágio é conhecido como tradução. Este processo é intermediado por tRNAs que transportam aminoácidos para os códons (sequências de três nucleotídeos) do mRNA. O tRNA possui uma região chamada de anti-códon, cuja a função é ligar-se ao códon responsável por aquele aminoácido. Por exemplo, o anticódon UAC liga-se ao códon AUG, responsável pelo aminoácido metionina. O anticódon UUU liga-se ao códon AAA, sendo responsável pelo aminoácido lisina e assim por diante. Todas essas ligações são intermediadas pelos ribossomos. Os ribossomos por sua vez contêm três regiões distintas, em sequência: sítio A, sítio P e sítio E. No sítio A, liga-se o tRNA contendo o aminoácido, dando prosseguimento, o tRNA é levado em direção à região P onde ocorre reações químicas que ligam o aminoácido a outros aminoácidos formando a cadeia proteica. Quando terminada a ligação o tRNA é levado para o sítio E onde é excluído de dentro do ribossomo, estando livre para se ligar novamente a outros aminoácidos. O primeiro aminoácido a ser colocado em uma cadeia proteica em eucariotos é a metionina e o último aminoácido depositado na cadeia é seguido pelos códons de terminação (UAA, UAG ou UGA). Chegando nesse estágio a transcrição está terminada. Em passos mais alongados, outros fatores são determinantes para a futura operação das proteínas, como chaperonas, chaperoninas, vias de ubiquitina e proteossomas. Todos esses atuam regulando a expressão de genes em diferentes fases pós-transcrição (ALBERTS *et al.*, 2010).



**Figura 6:** Representação do esquema de tradução do RNA em DNA. Fonte: <https://www.todamateria.com.br/sintese-proteica/>. Acesso em: 07/07/2016.

## 2.5.4 Proteínas: Níveis de informação

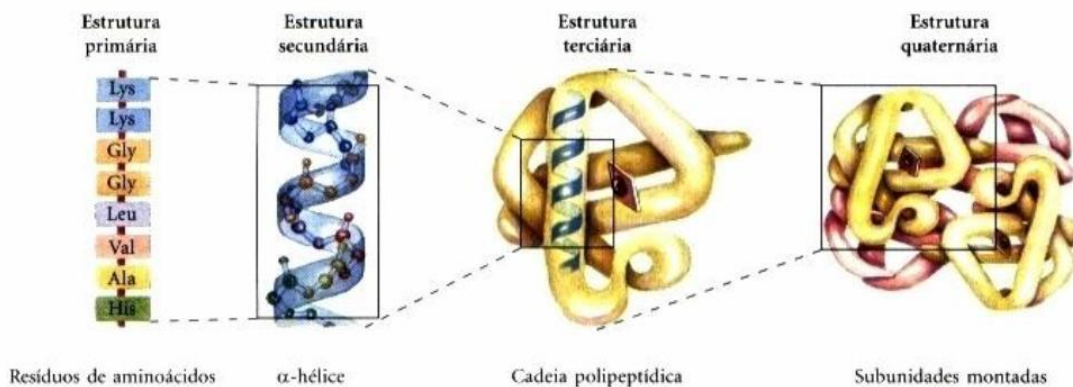
Proteínas são, por definição, macromoléculas biológicas que desempenham funções vitais em um organismo (NELSON; COX, 2011). A função que uma proteína exerce está intrinsecamente ligada à sua conformação estrutural. Para entender sua estrutura, é importante frisar que uma proteína é normalmente dividida em quatro níveis diferentes de informação:

I) A estrutura primária, formada pela sequência de aminoácidos, onde cada aminoácido está unido a outro por ligações peptídicas.

II) A estrutura secundária, formada pelas conformações estruturais locais (Hélices-  $\alpha$  e folhas-  $\beta$ )

III) A estrutura terciária, formada pelo enovelamento da cadeia peptídica, formando uma região hidrofílica na superfície e no núcleo uma região hidrofóbica.

IV) A estrutura quaternária, complexos formados quando estruturas terciárias se unem. Não ocorre em todas as proteínas.



**Figura 7:** Níveis de estruturas nas proteínas. **Fonte:** <http://docplayer.com.br/5057458-Determinacao-da-estrutura-tridimensional-de-proteinas-por-difracao-de-raios-x.html>. **Acesso em:** 07/07/2016

### 2.5.5 Alinhamento

Alinhamento de sequências é o procedimento de comparar duas (alinhamento simples) ou mais (alinhamento múltiplo) sequências biológicas, seja sequências de DNA, RNA ou de proteínas, buscando uma série de caracteres individuais ou padrões de caracteres que estão na mesma ordem nas sequências. Duas sequências são alinhadas transcrevendo-as através de uma página em duas fileiras. Caracteres idênticos (*match*) são colocados na mesma coluna, e caracteres não idênticos (*mismatch*) podem ser colocados na mesma coluna como uma incompatibilidade ou em frente a uma lacuna (*gap*) na outra sequência. Em um alinhamento ótimo, caracteres não idênticos e as lacunas são dispostos de maneira a resultar no maior número de caracteres idênticos ou semelhantes possíveis. As sequências que podem ser facilmente alinhadas deste modo são referidas como sendo semelhantes.

Existem dois tipos de alinhamento de sequências, o alinhamento global e o alinhamento local (Figura 8). As sequências que possuem maiores níveis de semelhança e aproximadamente do mesmo comprimento são candidatos adequados para alinhamento global. No alinhamento local, parte de uma dada sequência com a maior densidade de caracteres idênticos (*match*) são alinhados. Alinhamentos locais são mais adequados para o alinhamento de sequências que são semelhantes em pontos isolados, partilhando uma região conservada ou um domínio idêntico, mas que diferem em outras características, como o comprimento (MOUNT, 2001). Os alinhamentos são produzidos para se buscar regiões conservadas em proteínas, auxiliando no entendimento evolutivo, funcional e dimensional da proteína, uma vez que proteínas evolutivamente próximas, pertencentes ou não a mesma família de proteínas, possuem regiões muito semelhantes nas suas sequências de aminoácidos, e por conseguinte, regiões semelhantes estruturalmente.

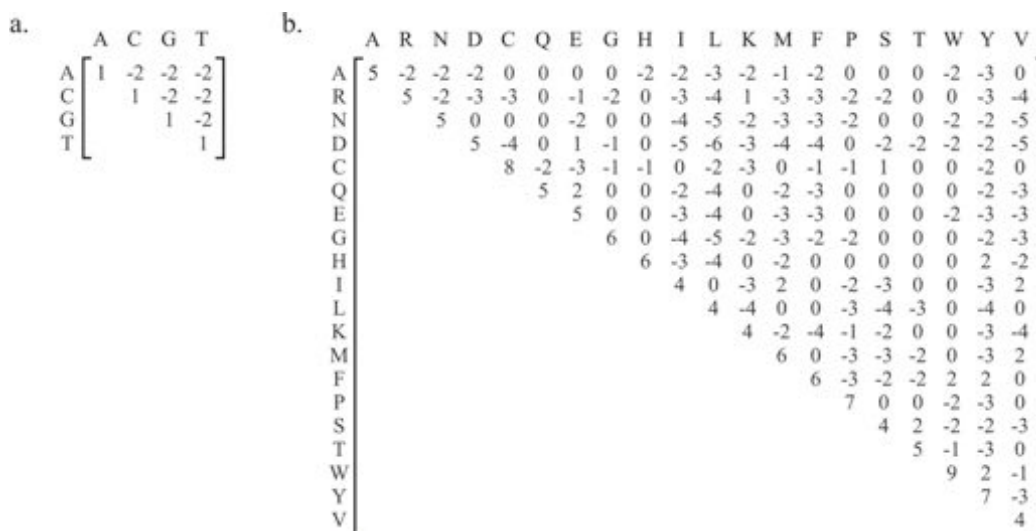


**Figura 8:** Distinção entre alinhamento local e alinhamento global. Adaptado de (MOUNT, 2001).

### 2.5.5.1 Matrizes de custo e Blosum 62

Biologicamente, as alterações nos nucleotídeos ou aminoácidos não ocorrem com a mesma probabilidade, sendo assim é possível empregar diferentes valores a cada substituição que ocorra em diferentes sequências. A família de matriz BLOSUM (*Blocks Substitution Matrix*, em inglês) (figura 9) foi desenvolvida para empregar pontuações diferentes para qualquer substituição que ocorra nos nucleotídeos ou em sequências de aminoácidos em um alinhamento. As matrizes são baseadas em um banco de dados conhecido por BLOCKS, o banco de dados consiste em alinhamentos múltiplos sem existência de lacunas (*gaps*) que representam as regiões mais conservadas de famílias proteicas (HENIKOFF; HENIKOFF, 1996).

Além disso, é necessário que as lacunas (*gaps*) também recebam diferentes pontuações, uma vez que são comumente encontradas em alinhamentos de sequências biológicas. Por convenção a abertura de lacuna é pontuada como -4 (menos quatro pontos) e a extensão da lacuna é pontuada em -3 (menos três pontos). (VERLI, 2014)



**Figura 9:** Matrizes de alinhamento. A. Matriz de substituição utilizada para alinhamentos de bases nitrogenadas. B. Matriz BLOSUM 62, utilizada para a substituição de aminoácidos. **Fonte:** (VERLI, 2014).

### 2.5.5.2 BLAST

O BLAST (*Basic Local Alignment Search Tool*. Ferramenta de busca por alinhamento, em português) é um algoritmo capaz de localizar regiões de similaridade entre sequências biológicas. O programa compara nucleotídeos ou sequências de proteínas em um banco de dados e dá significância estatística entre alinhamentos.

Para funcionar, o programa somente necessita de uma sequência de busca (*query*) e sequências alvos, estas comumente já se encontram em um banco de dados próprio do algoritmo. A ferramenta pode ser acessada em (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>).

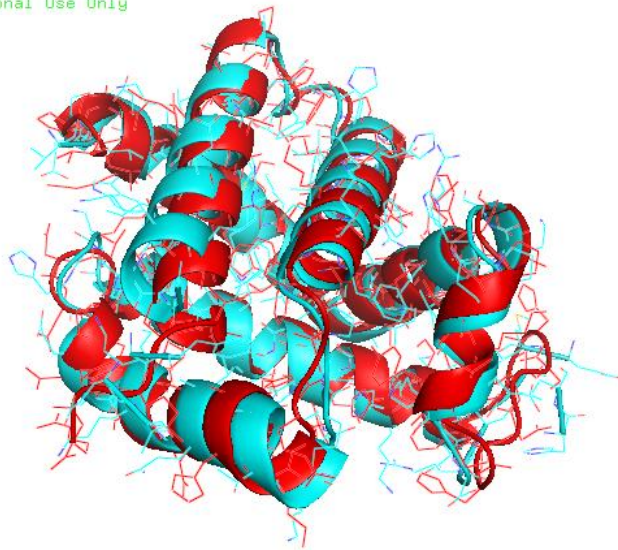
### 2.5.5.3 Alinhamento estrutural

Alguns aspectos da análise da sequência levam diretamente para análise estrutural. Como no caso de sequências, uma questão fundamental na análise de estruturas é conceber e calcular uma medida de semelhança entre duas sequências. Se duas moléculas têm estruturas idênticas ou muito semelhantes, podemos sobrepô-las de modo que cada átomo correspondente esteja tão próximo quanto possível em dada sequência. Assim, para mensurar a distância média entre átomos correspondentes utiliza-se uma medida de semelhança estrutural. Na prática, é convencional utilizar uma métrica chamada de RMSD (*Root Mean Square Deviation*) (Figura 10) que calcula a distância média entre os C $\alpha$  da cadeia principal de cada proteína. (LESK, 2002). Quando menor o valor do cálculo, menor as distâncias médias entre C $\alpha$  das sequências sobrepostas, o que necessariamente implica em maior similaridade estrutural.

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N \delta_i^2}$$

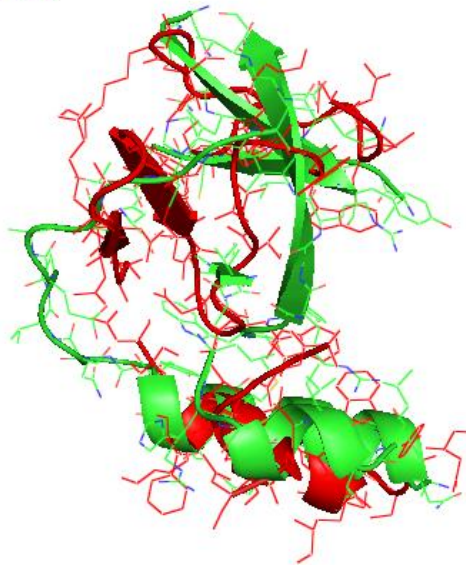
**Figura 10:** *Root mean square deviation*. Onde  $\delta_i$  é a distância entre o átomo  $i$  e um átomo de referência em uma estrutura superposta na mesma posição  $N$ . O RMSD somente considera os átomos da cadeia principal de cada proteína ou somente os C $\alpha$  de cada estrutura.

For Educational Use Only



**Figura 11:** Alinhamento estrutural de proteínas similares: Uma cadeia  $\alpha$  da Hemoglobina (PDB ID 1BBB, em vermelho) e Mioglobina (PDB ID 1MBO, em azul). O Valor calculado do RMSD é de 1.76 Å. Calculado por (ZHANG, 2005)

For Educational Use Only



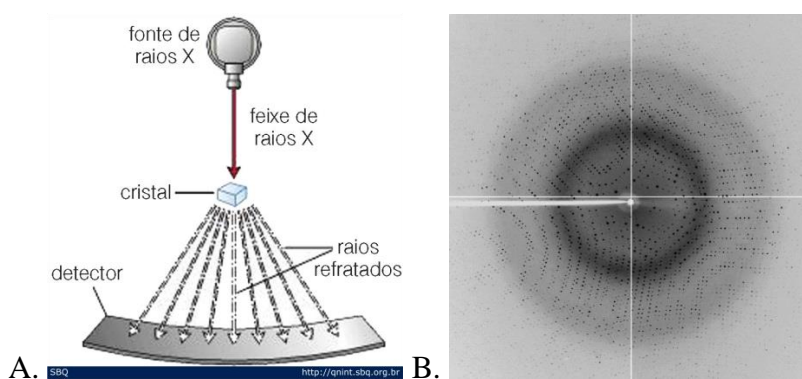
**Figura 12:** Alinhamento estrutural de proteínas não similares: TDRD2 humana (PDB ID 3FDR, em vermelho) e lisozima presente no ovo de galinha (PDB ID 6LYZ, em verde). O RMSD calculado é de 4.5 Å. Calculado por (ZHANG, 2005)

## 2.5.6 Cristalografia de Raio-X

A cristalografia Raio-X dedica-se ao estudo das estruturas molecular e cristalina, assim como a relação entre estas. Sua utilização para elucidação da estrutura tridimensional de proteínas deu origem à área hoje conhecida por cristalografia de proteínas, determinando a criação de uma nova área também na biologia, denominada Biologia Estrutural, que tem foco na investigação da estrutura e função de macromoléculas. Para a bioinformática os estudos cristalográficos são fundamentais em diversas áreas, como a modelagem molecular e a construção de fármacos a partir da técnica de Planejamento de fármacos baseado na estrutura dos receptores (*structure-based drug design*). A determinação de estruturas 3D proteicas a partir do método cristalográfico é o primeiro passo para modelos moleculares que tencionam resolver a estrutura e função de proteínas homólogas ou o planejamento de novas moléculas bioativas (VERLI, 2014).

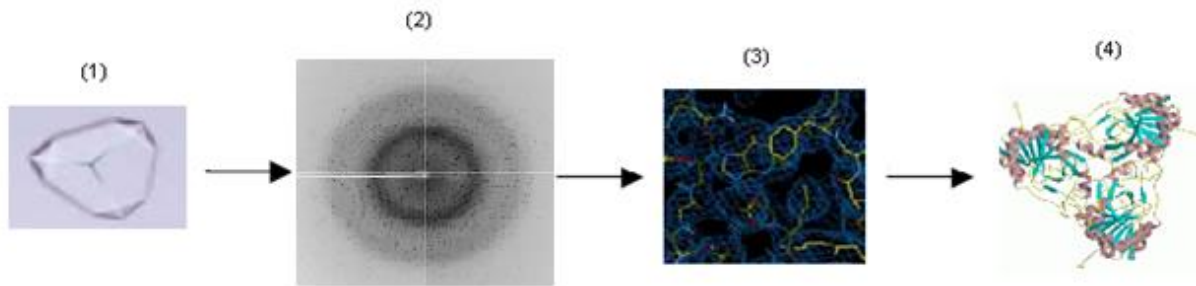
Para a construção da imagem tridimensional de uma proteína, um cristal supersaturado de proteínas é exposto a um feixe de Raio-X (figura 13a). Os Raios-X utilizados nos estudos estruturais são produzidos por geradores em laboratórios, ou com mais frequência ultimamente, por um síncrotron. O padrão de difração resultante que se forma após o cristal ter sido atingido pelos Raios-X demonstra o mapa de densidade eletrônica da proteína estudada (figura 13b) (VOET; VOET; PRATT, 2008).

De maneira a construir a estrutura tridimensional final, o mapa de densidade eletrônica é trabalhado a partir de técnicas de refinamento contínuas até que haja convergência de dados e não seja observado modificações significativas entre a estrutura criada e o mapa de densidade eletrônica.



**Figura 13** : Cristalografia de Raio-X. A. Esquema representativo da captação da densidade eletrônica em cristalografia de raio x. **Fonte:** [http://qnint.s bq.org.br/qni/popup\\_visualizarConceito.php?idConceito=73&semFrame=1](http://qnint.s bq.org.br/qni/popup_visualizarConceito.php?idConceito=73&semFrame=1). **Acesso em:** 22/07/2016. B. Fotografia de difração de raios X. A intensidade de cada difração é em função da densidade eletrônica do cristal. **Fonte:** <http://www.diabeticool.com/wp-content/uploads/2013/01/cristalografia-raios-x-diabetes.jpg>. **Acesso em:** 22/07/2016





**Figura 14:** Representação resumida do processo de cristalografia. 1. Cristal proteico. 2. Mapa de densidade eletrônica. 3. Refinamento. 4. Estrutura final. **Fonte:** <http://s3.amazonaws.com/magoo/ABAAAaczAG-0.jpg>. **Acesso em:** 22/07/2016

### 2.5.7 Modelagem de proteínas por homologia

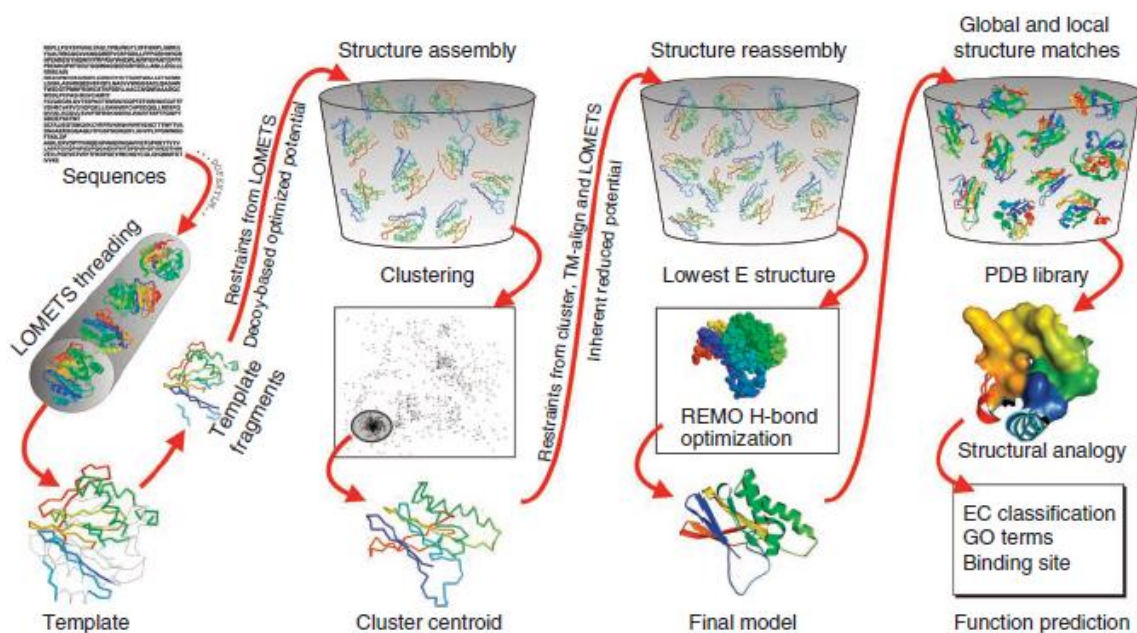
A modelagem de proteínas por homologia é a ferramenta mais bem sucedida para a predição de estrutura terciária de proteínas (FILHO; ALENCASTRO, 2003). Zhang (2008) define a predição de estrutura proteica como o esforço em gerar modelos tridimensionais de proteínas a partir da sequência de aminoácidos utilizando algoritmos computacionais como auxílio.

É necessário construir modelos tridimensionais de proteínas, uma vez que o paradigma atual indica que a função exercida pelas proteínas está diretamente relacionada à sua estrutura, sendo que a perda desta caracteriza-se em posterior perda da função proteica (REIS, 2013). Roy, Kucukural e Zhang (2010) apontam para o fato de que a porcentagem total de proteínas depositadas no banco de dados PDB que possuem estruturas tridimensionais resolvidas são de 2%, 1,2% e 0,6% nos anos de 2004, 2007 e 2009, respectivamente.

Vários servidores são utilizados ao redor do globo por diversos pesquisadores com o objetivo de construir modelos tridimensionais. Para avaliar qual a capacidade de cada servidor em modelar proteínas corretamente, há bienalmente, desde o ano de 1994, o *Critical Assessment of Structure Prediction* (Avaliação crítica de predição estrutural, em português. Sigla CASP) que tem por objetivo auxiliar no desenvolvimento dos métodos utilizados para a modelagem proteica. O servidor *I-TASSER* (ZHANG, 2008) tem constantemente sido ranqueado como o melhor método para predição de estruturas terciárias na competição CASP (YANG *et al.*, 2014).

O servidor *I-TASSER* se divide em quatro estágios até o modelo final da proteína (ROY; KUCUKURAL; ZHANG, 2010). O primeiro estágio utilizado pelo servidor é

construir a estrutura secundária da proteína a partir da sequência de aminoácidos por *machine learning* (aprendizagem automática). Em seguida, a sequência e a estrutura secundária são alinhadas com estruturas modeladas e armazenadas anteriormente em banco de dados para identificar os melhores moldes possíveis. Esse procedimento é chamado de *threading*. No segundo estágio, após o *threading*, o servidor fragmenta os moldes em pequenas regiões baseado nas sequências de alinhamento dos moldes e, após, remonta os fragmentos a partir da técnica de repetições sucessivas, chamada técnica de Monte Carlo, para melhor moldar regiões que possuem maiores dificuldades de modelagem, como regiões de alça, que não possuem estrutura secundária definida. No terceiro estágio, os modelos então criados no estágio dois novamente passam pelo mesmo processo anterior. O objetivo deste passo é aperfeiçoar as redes de ligações de hidrogênio e refinar a topologia global da proteína. O último estágio é identificar a função biológica que a proteína moldada desempenha. Para isso, as proteínas são novamente alinhadas com proteínas que possuam funções já caracterizadas. A figura quinze representa os estágios que o *I-TASSER* perpassa a fim de modelar proteínas.

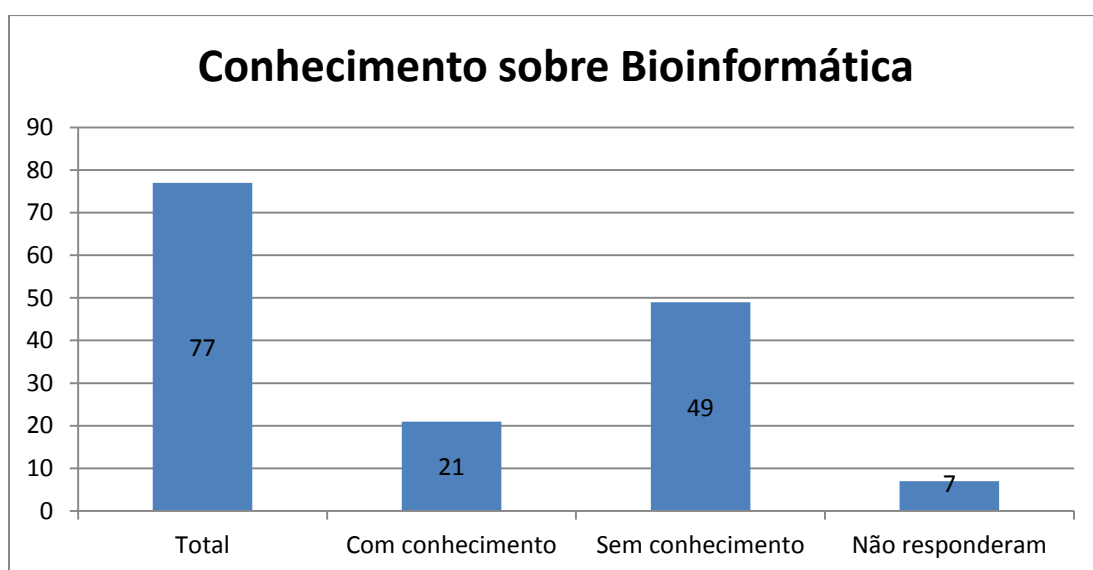


**Figura 15:** Caminho completo utilizado pelo servidor I-TASSER para modelagem proteica. Cada coluna representa um estágio. **Fonte:** (ROY;KUCUKURAL;ZHANG, 2010).

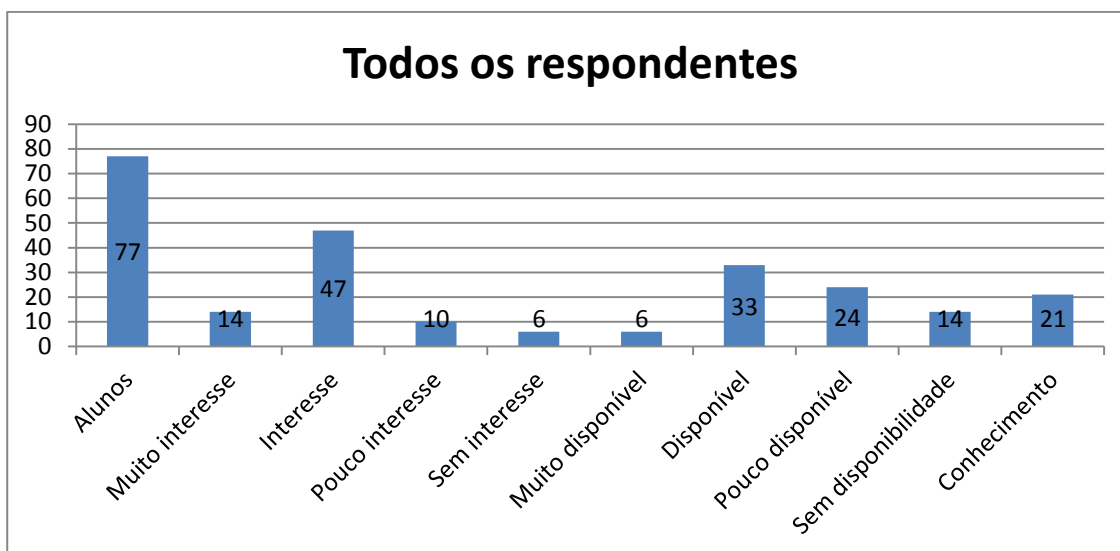
### 3. RESULTADOS E DISCUSSÕES

Os gráficos expostos abaixo se relacionam diretamente às respostas dos alunos ao Anexo I deste trabalho. As variáveis observadas nestes gráficos demonstram que do total de setenta e sete (77) estudantes do curso de Licenciatura em Ciências Biológicas que responderam ao questionário, quarenta e nove (49), ou a porcentagem de 63, 64% dos estudantes responderam que não possuíam conhecimento sobre o que seja bioinformática. Estes dados reforçam o que foi anteriormente discutido em outros trabalhos, como Júnior *et. al* (2011); Farias, Chacon e Silva (2012), sobre o desconhecimento por parte de graduandos sobre esta nova área das ciências da vida.

Outra informação relevante para que fosse dado início ao curso sobre bioinformática era a necessidade de haver mais do que vinte (20) estudantes muito interessados/interessados e muito disponíveis/disponíveis. Com os dados obtidos, observou-se que trinta e cinco (35) estudantes mostravam-se aptos para se inscreverem no curso.



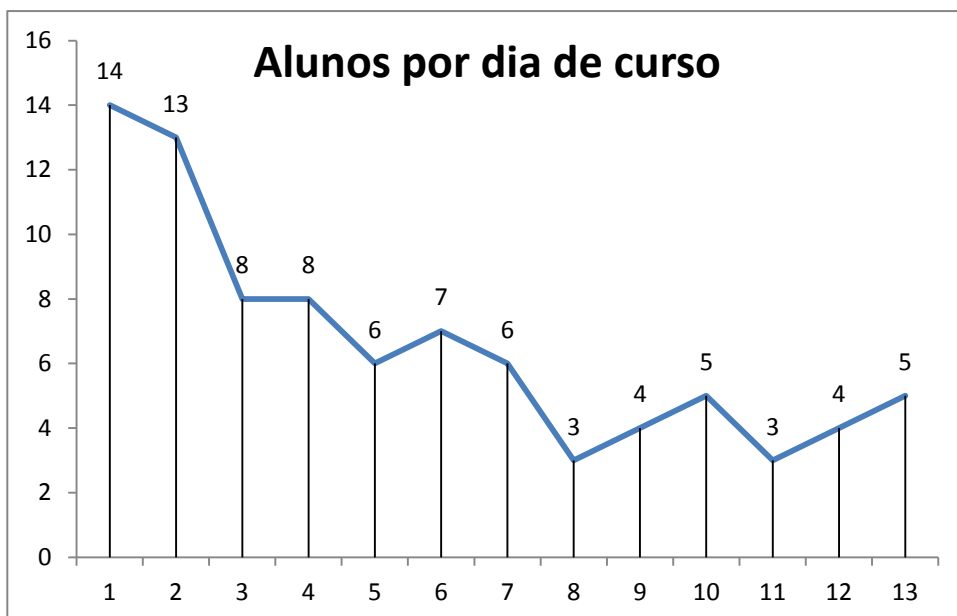
A.



B.

**Gráfico 2:** A. Conhecimento dos alunos do IFSULDEMINAS- Campus Inconfidentes sobre o campo da Bioinformática.. B. Disponibilidade e interesse dos alunos graduandos em ciências biológicas para participar do mini-curso em bioinformática. Dados relativos ao primeiro semestre de 2016 no IFSULDEMINAS- Campus Inconfidentes.

O gráfico abaixo (Gráfico 3) demonstra a quantidade de alunos por aula. Dos vinte (20) estudantes inscritos (100%), no primeiro dia compareceram quatorze (14) (70%) e cinco (5) estudantes (25%) completaram o curso e receberam a certificação. Contudo, esses dados, em comparação com dados de cursos MOOC (Cursos online abertos e massivos) demonstram que o índice de evasão para este curso está abaixo da média para cursos daquela categoria, como relata Jordan (2015), a média de concluintes em cursos MOOC é de 15%. Ding *et al.* (2014) avaliou cursos MOOC em bioinformática e seus dados mostram que a média de todos os alunos inscritos nos cursos que assistiram a pelo menos uma aula é de 65% e apenas 1,55% dos alunos inicialmente inscritos concluíram seus estudos. Isso posto, mostra-se uma superioridade na quantidade de concluintes neste curso presencial em bioinformática do que em virtuais.



**Gráfico 3:** Número de estudantes por aula. No primeiro dia havia 14 estudantes presentes. Houve cinco alunos finalizantes. **FONTE:** Autor do trabalho.

### 3.1 PRÁTICAS

Durante o curso foi realizado por parte dos alunos seis práticas e duas avaliações. As práticas neste trabalho se mostraram de grande importância para a construção do conhecimento, adicionando ao conhecimento teórico trabalhado as capacidades práticas necessárias. As práticas, principalmente aquelas produzidas mais ao final (em especial a prática número VI) construíram uma boa forma de auxiliar na equalização do ensino, não por parte do professor, mas por parte do próprio aluno. Foi possível visualizar e anotar os comportamentos ocorrentes ao final do curso que mostraram que nas práticas as diferenças nas habilidades (por habilidade entenda-se o manuseio dos computadores e das ferramentas) não tornaram o curso desequilibrado, mas construíram uma cadeia de auxílio mútuo onde os alunos que se destacavam mais (dois dos alunos) auxiliavam os outros três alunos que porventura sentiam algumas dificuldades no percorrer das práticas.

Nas práticas, que se seguiram da visualização de aminoácidos até a modelagem de proteínas por homologia, seguiu-se uma tendência de, prática a prática, ser agregado aos alunos novas informações. Com isso, a construção das práticas evidenciou uma característica

de construir novos conhecimentos a partir de conhecimentos obtidos previamente noutras práticas e em aulas teóricas. Em correlação a isso Martins (2009) relata que entre as vantagens das atividades práticas consistem em “estabelecer relação entre os conhecimentos prévios dos alunos e as informações novas que irão descobrindo no estudo do objeto ou do fato.”. Com isso, percebe-se a relação necessária entre a prática e o pensar pedagógico relacionando os conteúdos de maneira a determinar áreas de aprendizagem sem que durante a trajetória do estudante haja lapsos na relação entre os conteúdos teóricos e práticos.

Outras características relacionadas às práticas durante o curso demonstraram que a confecção destas, para além de necessitar diretamente de um esforço por parte do aluno, seja esse esforço particular ou em grupo, mostrou-se como desafios às suas capacidades. Dessa forma, o aluno necessitou colocar-se como agente direto do seu próprio conhecimento. Nesse caso, o professor por instantes deixou de ser o centro das atenções e tornou-se somente facilitador da aprendizagem. Essas constatações vão de encontro às novas perspectivas docentes que implicam em “uma nova postura do professor em sala de aula, de forma a imergir os alunos em práticas significativas.” (CORREIA, 2016. p. 260) Por outro lado, as práticas também evidenciaram questões negativas. Como relatado nos diários de campo, foi percebido aos primeiros dias de curso que alguns alunos possuíam dificuldades básicas no manuseio do próprio computador. Como relatado em meu próprio diário de campo:

“Devo abrir um espaço para dizer que a princípio não imaginava tanto desconhecimento de funcionalidade do próprio computador. Alguns dos primeiros passos eram completamente intuitivos e, mesmo assim, alguns alunos já encontraram dificuldades. Esses passos eram abrir uma aba de navegação, acessar um *site* e buscar informações.”.

Não há somente a questão de dificuldades no manuseio da máquina durante as práticas. Outro fato que ficou presente durante as primeiras práticas foi certa passividade, ou incerteza, dos alunos perante a máquina. Conforme Damásio (2007) esse fato está relacionado a um padrão já existente onde a vasta maioria (cerca de 85%) dos indivíduos somente adotam uma nova tecnologia depois de algum tempo desde seu primeiro contato com o novo objeto. Segundo o autor, esse grupo de pessoas pode ser descritos como céticos, renitentes, conservadores e cautelosos no que concerne ao aprendizado e utilização de novas propostas tecnológicas, dificultando de início a apropriação dos estudantes das tecnologias.

Não somente isso, mas a renitência de os estudantes adaptarem-se às tecnologias apresentadas no decorrer dos primeiros dias de curso e a falta de competência apresentada pelos estudantes no manuseio de seus próprios computadores levou a novas questões. A

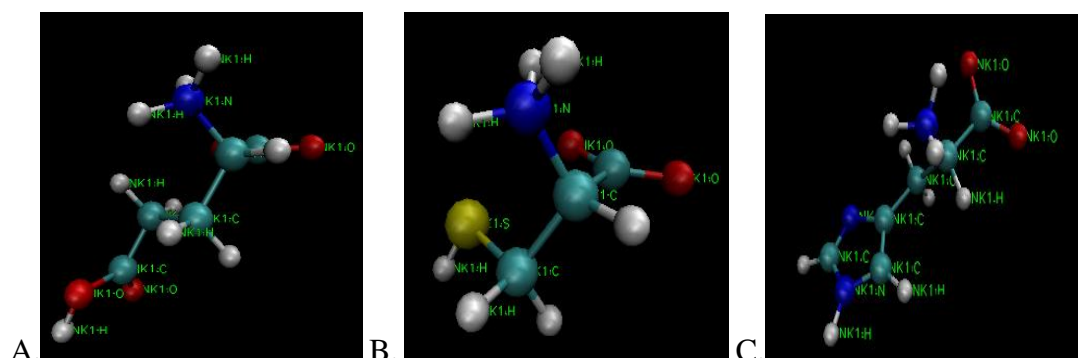
princípio, o questionamento era como ingressar licenciandos em Ciências Biológicas na Bioinformática, contudo mostra-se também necessário a adequação do estudante ao nível de apropriação exigido para o manuseio das ferramentas propostas. Tal constatação é necessária para a confecção da ementa para o curso de Bioinformática que deverá conter aspectos relacionados a computadores, naquilo que tange aos conhecimentos sobre a manipulação correta da máquina anteriormente às partes relacionadas diretamente ao conteúdo de Bioinformática.

Para além de novas questões, as dificuldades apresentadas levaram a necessidade de construir e reconstruir práticas, explicitando a necessidade de o planejamento ser flexível. Essa remodelação põe em contraposição aquilo que estava suposto ser trabalhado e aquilo que realmente foi trabalhado decorrer deste curso. Conteúdos necessitaram ser inseridos, da mesma maneira que alguns necessitaram ser retirados. Dessa maneira, a reflexão sobre a prática pedagógica é um ponto fundamental para o referencial que será utilizado para construir a ementa. Mostra-se necessário a experiência primária do pesquisador com o tema antes da efetiva construção da ementa, dado que a construção pura e simples de um documento que albergue os conteúdos a serem trabalhados durante um semestre não se mostrará errado do ponto de vista conteudista, contudo pecará pela falta de vivência prática e reflexão sobre a ação de ensinar. Planejar é dar sentido a ação (PASSOS, 2006). Sem ação o planejar se torna vago de referência e sem adequação ao contexto onde está inserido. Gama e Figueiredo (2009) consideram o planejamento um organismo vivo, sendo assim, o bom planejamento deve levar em consideração a realidade sócio-cultural em que o aluno está inserido, não somente os conteúdos a serem trabalhados.

As práticas se mostraram, tanto nos seus pontos positivos quanto em seus pontos negativos, o fio condutor que moveu o curso realizado. Ficou evidente o caráter de anexar conteúdos que as práticas possuíam. As práticas, junto com os dois exames, mais do que as aulas teóricas, auxiliaram de forma mais direta a construção final da ementa para a disciplina. Os roteiros de cada prática estarão anexados ao final deste trabalho. O intuito da análise das práticas não é somente trazer para os resultados e discussões aqui debatidos as respostas dos alunos, mas trazer as questões mais relevantes ao debate daquilo que se mostrou mais motivador e aquilo que se mostrou mais laboroso na introdução das práticas e das ferramentas utilizadas. Essa análise das práticas é necessária uma vez que a reflexão sobre a própria atividade conscientiza sobre a própria atuação, segundo Zabala (1998, p. 45) “Como também sabemos, não basta repetir um exercício sem mais nem menos. Para poder melhorá-lo

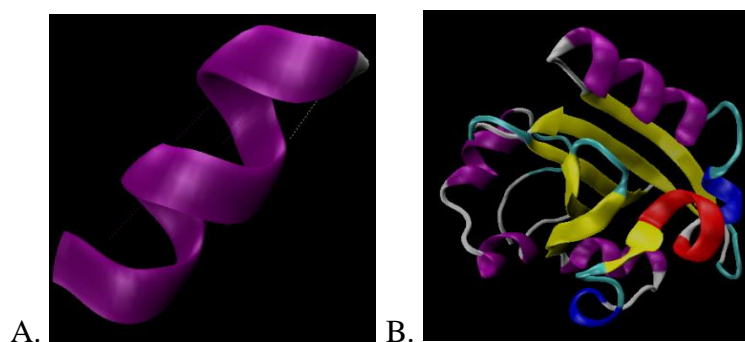
devemos ser capazes de refletir sobre a maneira de realizá-lo e sob quais são as condições ideais de seu uso.” O enfoque das reflexões aqui apresentadas está no processo, não no produto das práticas. Para o objetivo final deste trabalho, a construção de uma ementa, o processo mostra-se muito mais necessário de debate do que o produto, pois aquele necessita de contínuo aperfeiçoamento e reflexão, este é somente a mensuração do processo. Segundo Luckesi (2011) o processo é o conjunto de obras que utilizamos para alcançarmos um resultado. O produto é por sua vez o resultado final ao qual se chegou.

Abaixo estão sucintamente relatado os produtos de cada prática. A primeira prática (30/03) consistiu em iniciar os estudantes em bancos de dados e obter as informações provenientes deste. O banco de dados utilizado nessa prática foi o *Protein Data Bank* (PDB) um dos maiores repositórios mundiais de sequências biológicas. A segunda prática (06/04) consistiu em observar e retirar informações dos aminoácidos através do *software* VMD (Figura 16).



**Figura 16:** Exemplos de aminoácidos visualizados na primeira prática. A, Ácido glutâmico. B, Cisteína. C, Histidina. Observar que todos os átomos estão nomeados conforme o elemento químico. **Fonte:** Autor do trabalho.

A terceira prática (13/04) consistiu em observar e retirar informações de ligações peptídicas e estruturas secundárias de proteínas. Figura abaixo.



**Figura 17:** Estruturas secundárias. A. Representação da Alfa-Hélice de uma polialanina. B. Representação da proteína Amarela Fotoativa. As cores nas imagens representam estruturas secundárias diversas: Amarelo: folha-Beta; Roxo: Alfa-Hélice; Azul: Hélice  $3_{10}$ ; Vermelho: Hélice Pi. **Fonte:** Autor do trabalho.



A prática de número quatro (20/04 e 27/04) trabalhou questões sobre estruturas terciárias e quaternárias de proteínas e alinhamento estrutural. Figura abaixo.



**Figura 18:** Alinhamento estrutural das proteínas hemoglobina (em verde) e mioglobina (em rosa). O alinhamento estrutural busca comparar a estrutura tridimensional de proteínas para busca de similaridades entre as estruturas. **Fonte:** Autor do trabalho.

A quinta prática (11/05 e 18/05) consistiu na análise matemática de dois alinhamentos sequenciais para averiguar qual dos alinhamentos se mostrava mais parcimonioso. Para essa análise foi utilizada a matriz de substituição BLOSUM62 e o alinhamento foi produzido através da ferramenta BLAST. A análise resultou na superioridade matemática do segundo alinhamento sobre o primeiro. A figura dezenove representa os dois alinhamentos averiguados. A última prática (25/05 e 01/06) consistiu na modelagem por homologia e construção do gráfico de Ramachandran, utilizando as ferramentas Swiss Model e VMD, da enzima Manganês Peroxidase produzida pelo fungo *Trametes versicolor*. A sequência a ser modelada, o modelo resultante e o gráfico de Ramachandran estão disponíveis na figura vinte.

Download ▾ GenPept Graphics

PREDICTED: peroxidase 2 [Cucumis sativus]  
Sequence ID: [ref|XP\\_004142149.1](#) Length: 318 Number of Matches: 1

Range 1: 1 to 318 [GenPept](#) [Graphics](#) ▾ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
411 bits(1057)	6e-142	Compositional matrix adjust.	209/323(65%)	248/323(76%)	6/323(1%)
Query 1	MAASSKIVSVLVLCLMMAVSVRSQLSSTFYDTTCPNVSSIVHGVMQALQSDDRAGAKII				60
Sbjct 1	MASLFRVAFFLFLGLMVRAS-QAQLCPTFYDESCPDVSNIVRRVVQALVSDERAGARLI				59
Query 61	RLHFHDCFVGDGSGVLLLEDQDGITSELGAPNGGITGFNIVNDIKTAVENVC PGVVSCA				120
Sbjct 60	RLHFHDCFV+GCDGSVLLEDQ G+ SEL APGN ITGFNIVN+IK AVE CPGVVSCA				119
Query 121	DILALGSRDAVTLASGQGWTVQLGRRDSRTANLQGARDRLPSPFESLSNIQGI FRDVGLN				180
Sbjct 120	DILA+ S ++V LA G W VOLGRRDSR ANLQGA D LPSPFE+++ ++ F V L				178
Query 181	DNTDLVALSGAHTFGRSRCMFFSGRLN-NPNADDSPIDSTYASQLNQTCSGSGTFVDL				239
Sbjct 179	D+TDLVALSGAHTFG+SRC FF RLN +NP DS ++ YA QL Q C SG TFV+L				235
Query 240	DPTTPNFD+NYITNLQ+N GLL SDQVL STPG T+ VN A+S++ F ++F QSMI				299
Sbjct 236	DPTTPNKFQDKNYITNLQSNIGLLSDQVLHSTPGEDTVKIVNLFASQNFESFGQSMI				295
Query 300	RMGNLDPKTGTTGEIRTCRRLN		322		
Sbjct 296	MGN+ P TG GEIR+NCRRLN		318		

Download ▾ GenPept Graphics

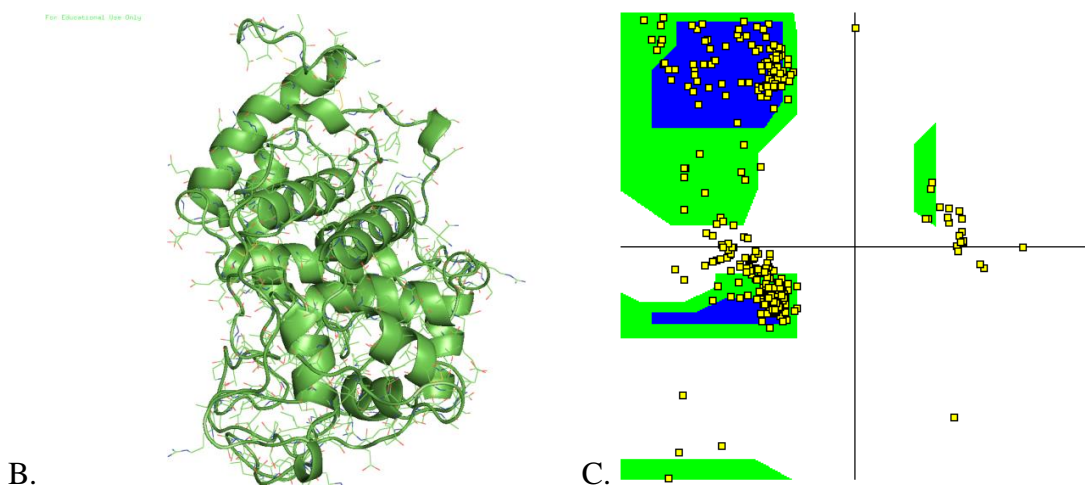
PREDICTED: peroxidase 2 [Cucumis melo]  
Sequence ID: [ref|XP\\_008449778.1](#) Length: 318 Number of Matches: 1

Range 1: 1 to 318 [GenPept](#) [Graphics](#) ▾ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
410 bits(1054)	2e-141	Compositional matrix adjust.	210/323(65%)	247/323(76%)	6/323(1%)
Query 1	MAASSKIVSVLVLCLMMAVSVRSQLSSTFYDTTCPNVSSIVHGVMQALQSDDRAGAKII				60
Sbjct 1	MASLFKVAFFLFLGLMVRAS-QAQLCPTFYDESCPDVSNIVRGVVKQALAFDERAGARLI				59
Query 61	RLHFHDCFVGDGSGVLLLEDQDGITSELGAPNGGITGFNIVNDIKTAVENVC PGVVSCA				120
Sbjct 60	RLHFHDCFVNGCDGSVLLEDQPGVSELGAPGNANITGFNIVNIIKAAVEKACPGVVSCA				119
Query 121	DILALGSRDAVTLASGQGWTVQLGRRDSRTANLQGARDRLPSPFESLSNIQGI FRDVGLN				180
Sbjct 120	DILA+ S +AV LA G W VOLGRRDSR ANLQGA D LPSPFE ++ ++ F VGL				178
Query 181	DNTDLVALSGAHTFGRSRCMFFSGRLN-NPNADDSPIDSTYASQLNQTCSGSGTFVDL				239
Sbjct 179	D+TDLVALSGAHTFG+SRC FF RLN +NP D+ +D+ +A QL + C S TFV+L				235
Query 240	DPTTPNFD+NYITNLQ+N GLL SDQVL STPG T+ VN A S+ F D+F QSMI				299
Sbjct 236	DPTTPNQFQDKNYITNLQSNIGLLSDQVLHSTPGEDTVDIVNVFAVSQQFFD+FGQSMI				295
Query 300	RMGNLDPKTGTTGEIRTCRRLN		322		
Sbjct 296	MGNIRPLTGNQGEIRSNCRRLN		318		

**Figura 19:** Alinhamentos sequenciais realizados para a proteína Manganês Peroxidase de *Trametes versicolor*. O alinhamento sequencial busca encontrar proteínas que possuam a maior correlação entre seus aminoácidos. **Fonte:** Autor do trabalho.

A. MAFKTLASLLSVLVTIQVAGGALTRRVACPDGVNTATNAACCQLFAVRDDIQQLNFDGGECGEEV  
 HESLRLTFHDAIGISPSIASRGQFGGGADGSIALFEDIETNFHANLGVDEIIDEQRPFARHNLTAD  
 FIQFAGAIGVSNCPGAPQLDVFIGRPDATQPAPDLTVPEPFDTVDSIIERFSDAGGFTPAEIVALLVSH  
 TIAAADHVDPSIPGTPFDSTPEEFDTQFFIETQLRGTLFPGTGGNQGEVESPLRGELRLQSDSELARD  
 SRTACEWQSFVNNQAKLQSAFKA AFRKMTVLGHDESLLIECSELVPTPPPATSVAHFPAGLSNADV  
 EQACAETPFPTLPTDGPVTTVAPVPPS



**Figura 20:** Modelagem por homologia. A. Sequência de 365 aminoácidos da enzima Manganês Peroxidase sem estrutura terciária resolvida. B. Estrutura terciária da enzima Manganês Peroxidase produzida pelo Swiss Model. C. Gráfico de Ramachandram para o modelo tridimensional criado, produzido pelo *software* VMD. O gráfico de Ramachandram demonstra as regiões energeticamente favoráveis (regiões em azul e verde) para os ângulos diedros  $\phi$  e  $\psi$ , ao redor do  $C\alpha$ . O gráfico auxilia na validação da qualidade de modelos construídos *in silico*, pois nos oferece a capacidade de observar quais aminoácidos estão em regiões energeticamente favoráveis e quais aqueles que encontram-se fora dessas regiões. **Fonte:** Autor do trabalho.

### 3.2 AVALIAÇÕES

Quanto às avaliações, a primeira consistiu em um questionário de opinião qualitativa que objetivava receber do aluno todas as informações, positivas ou negativas, sobre o curso. A análise dos dados desta primeira avaliação mostrou alguns pontos positivos que os alunos mais focaram em suas respostas, sendo eles:

1. O manuseio dos programas VMD E PyMOL.
2. Os roteiros práticos que auxiliam na aprendizagem e no manuseio correto dos programas.
3. A metodologia utilizada no curso, unindo os conteúdos teóricos e práticos.
4. Visão mais ampla da Bioquímica e Biologia Molecular.
5. Necessidade de tecnologias para melhor compreender moléculas e funções biológicas.

Quanto aos pontos negativos:

1. Os programas utilizados são configurados em inglês somente.
2. A maneira como as práticas são construídas sem a participação do professor ministrante.
3. O conhecimento dos alunos, uma vez que estes possuem conhecimentos diferentes por pertencerem a períodos diferentes no curso.

Observando os pontos positivos (pontos 1, 2, 3 e 5) e negativos (1 e 2) pode-se relacionar a opinião dos alunos à necessidade da compreensão e utilização de novas tecnologias. Nesse ponto, em conformidade com Damásio (2007) é possível relacionar a necessidade em compreender e utilizar as tecnologias como um novo panorama na aprendizagem. A necessidade desta informação tecnológica transforma o aprendiz, outrora passivo, em um sujeito ativo da própria aprendizagem, com capacidade de atribuir sentido àquilo que trabalha e posteriormente aprende, auxiliando ainda na maneira com que o aprendiz recebe e estrutura o próprio conhecimento. Dessa forma, é possível dizer que o curso auxiliou os alunos participantes a buscarem a informação de maneira mais ativa em seu aprendizado por auxiliá-los na apropriação e manipulação de tecnologias utilizadas atualmente dentro da Bioinformática.

As frases abaixo foram retiradas da primeira avaliação do curso, produzida pelos estudantes. Os erros de digitação e concordância são mantidos para evitar alterações nas respostas dos estudantes.

“O curso superou minhas expectativas, foi legal e produtivo aprender a mexer em tais programas como o PYMOL e VMD, ver as proteínas em 3d foi muito interessante, podendo girar a molécula e ver elas de outros ângulos”. Aluno 1.

“Tive um pouco de dificuldade nos programas por serem com configurações em inglês, com o auxílio de outras pessoas foi facilitando a execução das práticas”. Aluno 1.

“ As guias para as práticas direcionam muito bem a forma com que devemos trabalhar, facilitando muito a utilização e compreensão dos programas”. Aluno 2.

“O conteúdo tem sido satisfatório, muita coisa já havia aprendido nas aulas da faculdade, mas muita coisa nova que ainda vou usar foi acrescentada a meu conhecimento”. Aluno 3.

“Quanto às praticas, são bem eficientes para entender como funciona o programa utilizado, além de um pequeno tutorial de como fazer os passos da questão no programa. Depois a parte dissertativa ajuda a fixar melhor o objetivo. Acho muito boa a metodologia porque ajuda a fixar de forma muito eficiente o que foi ensinado.” Aluno 3

“Vejo que o curso está proporcionando uma visão mais ampla no universo da biologia molecular e da bioquímica sendo possível aprender determinados conceitos relacionados a essas duas áreas”. Aluno 4.

“Gostaria de ressaltar que através desses encontros semanais estou compreendendo melhor a importância de usufruir das tecnologias para aperfeiçoar e aprimorar meus conhecimentos relacionados a estruturas biológicas e suas funções”. Aluno 4.

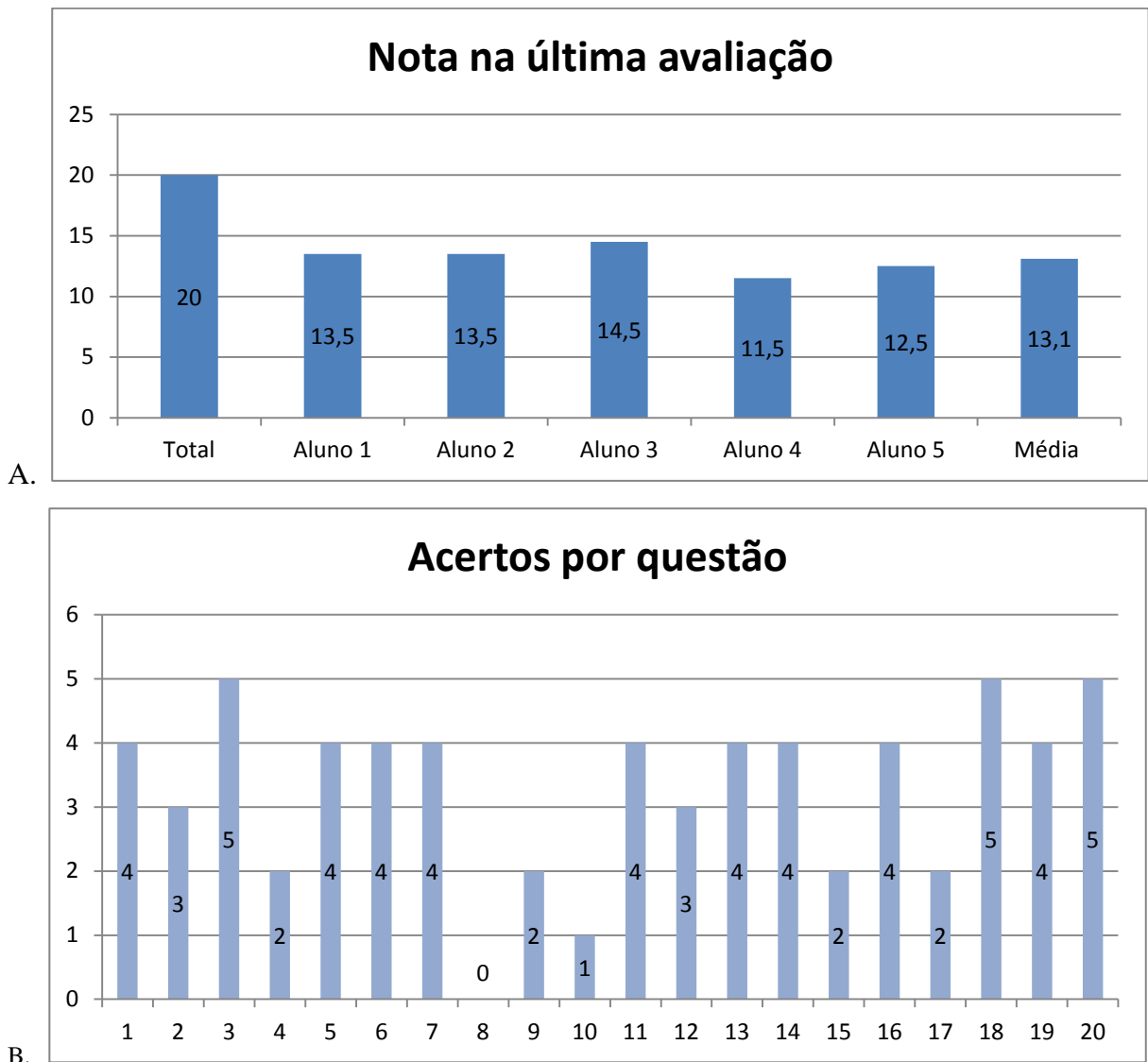
“Aprendemos muito sobre os programas que para mim, foi uma excelente maneira fazer uma folha alto explicativa dos passos que devemos realizar para conseguir visualizar o que se pede. Porque caso faltamos, em casa conseguimos acompanhar o que foi ministrado.” Aluno 5 .

“Só sugiro que se entregasse a folha da pratica aos alunos para todos seguirem como se faz, no entanto, realizasse no data show junto com a gente para acompanharmos o que é para fazer, para render mais a aula, porque senão fica um a um chamando na mesa. E também fazendo no data show dá para ir explicando para que serve o que você está visualizando e fazendo”. Aluno 5.

“As práticas têm possibilitado um aprendizado inédito e motivante sobre o assunto. Uma vez que os programas utilizados são de fáceis manuseios, e entendimento. Com isso, nos possibilita uma visão mais ampla e pratica sobre o conteúdo de uma forma menos abstrata. A forma como todas as práticas foram elaboradas tem ajudado muito a reforça o entendimento de uma forma geral do que foi trabalhado até o momento.” Aluno 6.

Os alunos de forma geral aprovaram a maneira como o curso foi ministrado. Os alunos mostraram-se simpáticos às práticas e exporam que elas auxiliam significativamente na aprendizagem. Além disso, os alunos constataram conhecimentos únicos em bioinformática e outras áreas, como a Bioquímica e Biologia Molecular, dado ao caráter interdisciplinar que a bioinformática possui. Nesse caso, adota-se a visão de interdisciplinaridade utilizado por Zabala (1998) como a interação entre duas ou mais disciplinas, com graus diferentes de ligação, indo desde o simples encontro de ideias até a utilização recíproca de conceitos fundamentais em ambas as partes.

A segunda avaliação, ministrada no último dia de minicurso, possuiu caráter somativo. Mesmo em face à avaliação como um processo, não somente produto (SANTOS, 2016) há a necessidade de obter médias aritméticas sobre o processo para melhor mensurar o nível de apropriação dos conceitos trabalhados em sala. As questões de 0-6 tratavam de temas referentes a conceitos básicos sobre Bioinformática, trabalhado nos primeiros dias de curso; As questões de 7-10 consistiam em questões sobre Biologia Molecular e Bioquímica, áreas correlatas da bioinformática; As questões de 11-19 consistiam em questões sobre o os conteúdos próprios de bioinformática trabalhados durante o curso. A última questão dava aos estudantes a oportunidade de avaliar novamente o curso. O gráfico abaixo demonstra os acertos de cada aluno na avaliação e sua média final.



**Gráfico 4:** A. Notas médias dos alunos na avaliação somativa final. B. Acertos por questão. **Fonte:** Autor do trabalho.

A última avaliação permite observar que todos os participantes obtiveram média superior a 50% do valor da prova. A única questão que nenhum aluno obteve acerto foi a questão de número 8 da avaliação. A questão faz parte do grupo de questões relacionadas à Biologia Molecular e Bioquímica. Há duas causas para a falta de acertos na questão tratada, o primeiro como consequência de os alunos terem tido contato com esse tema somente durante o mini-curso, por alguns deles ainda não terem participado da disciplina Biologia Molecular, onde esse conteúdo é comumente trabalhado. Outra possível causa pode estar no fato de que os conteúdos sobre duplicação, transcrição e replicação não possuíam caráter prático, somente

teórico. O tema é abordado pela necessidade de o estudante ter contato com a área teórica que engloba o estudo *in silico* de moléculas biológicas. A questão é descrita abaixo. A resposta correta está sublinhada:

8. O que se entende por *splicing*?

a. processo que retira os *introns* e *éxons* do RNA, mantendo somente a parte necessária para posterior tradução do RNA em proteína.

b. processo que retira os *introns* do RNA, mantendo somente os *éxons*, a parte necessária para posterior tradução do RNA em proteína.

c. processo que retira os *éxons* do RNA, mantendo somente os *introns*, a parte necessária para posterior tradução do RNA em proteína.

d. processo que retira os *introns* e *éxons* do DNA, mantendo somente a parte necessária para posterior transcrição do DNA em RNA.

Abaixo está um quadro com a respectiva nota de cada estudante durante o curso. Para dar valor final ao curso foi levado em consideração as práticas e a última avaliação. Pelo caráter prático que este curso proporcionou, as atividades práticas possuem valor maior sobre o resultado, sendo destinadas a esta sessenta por cento do valor final. Sendo assim, para cada prática entregue é dado o valor de um ponto. Para a última avaliação está destinado o valor de quarenta por cento do valor final, para adequar o valor da avaliação ao valor final, todos os valores fora divididos por cinco. A primeira avaliação, por não possuir características qualitativas, somente quantitativas, não foi adicionada à nota final.

	<b>Nota práticas</b>	<b>Nota avaliação</b>	<b>Nota final</b>
<b>Aluno 1</b>	5	2,7	7,7
<b>Aluno 2</b>	5	2,7	7,7
<b>Aluno 3</b>	6	2,9	8,9
<b>Aluno 4</b>	6	2,3	8,3
<b>Aluno 5</b>	6	2,5	8,5

**Quadro 3:** Nota final dos alunos concluintes do curso de Bioinformática. A menor nota obtida é de 7,7 dos alunos 1 e 2 e a maior nota, 8,9, pertence ao Aluno 3.

### 3.3 EMENTA.

Com todas as reflexões possíveis durante o curso, modificando as aulas para atender algumas demandas, está disponível abaixo o produto final deste trabalho, uma ementa para uma disciplina de Introdução à Bioinformática Estrutural. Junto com a ementa, deixo também um plano de ensino construído durante o curso. O intuito do plano de ensino é dar bases para aqueles que queiram estudar bioinformática. Assim, terão um material inicial para estudos.

<b>EMENTA</b>
Noções Básicas de Informática. Aspectos de Biologia Molecular e Bioquímica. O que é Bioinformática, origem e relação com o projeto genoma. Banco de dados. Alinhamento sequencial, <i>threading</i> e alinhamento estrutural de moléculas biológicas. Modelagem molecular por homologia.
<b>BIBLIOGRAFIA BÁSICA</b>
<ul style="list-style-type: none"><li>• ALBERTS, B. et al. <b>A biologia molecular da célula</b>. 5. ed. Porto Alegre: Artmed, 2010. 1396 p.</li><li>• MANZANO, A. N. G.; MANZANO, M. I. N. G. <b>Informática Básica</b>. 7. Ed. São Paulo. Érica, 2007. 250 p.</li><li>• VERLI, H. (Org.). <b>Bioinformática: da Biologia à flexibilidade molecular</b>. São Paulo: Sbbq, 2014. 282 p. Disponível em: &lt;<a href="http://www.ufrgs.br/bioinfo/ebook/">http://www.ufrgs.br/bioinfo/ebook/</a>&gt;. Acesso em: 24 fev. 2016.</li></ul>
<b>BIBLIOGRAFIA COMPLEMENTAR</b>
<ul style="list-style-type: none"><li>• ARAÚJO, N. D. et al. A era da Bioinformática: Seu potencial e suas implicações para as ciências da saúde. <b>Estudos de Biologia</b>, Curitiba, v. 70, n. 30, p.143-148, 2008. Disponível em: &lt;<a href="http://www2.pucpr.br/reol/index.php/BS?dd1=4621&amp;dd99=pdf">www2.pucpr.br/reol/index.php/BS?dd1=4621&amp;dd99=pdf</a>&gt;.</li><li>• CLAVERIE, J. M.; NOTREDAME, C. <b>Bioinformatics for Dummies</b>. 2. ed. Indianápolis: Wiley Publishing, Inc., 2007. 436 p.</li><li>• FILHO, O. A. A.; ALENCASTRO, R. B. Modelagem de proteínas por homologia. <b>Química Nova</b>, São Paulo, v. 26, n. 2, p.253-259, set. 2003</li><li>• LESK, A.M. <b>Introduction to Bioinformatics</b>. New York: Oxford University Press Inc, 2002. 255 p.</li><li>• NESHICH, G. Computational Biology in Brazil. <b>Plos Computational Biology</b>, São Francisco, v. 3, n. 10, p.1845-1848, out. 2007. Disponível em: &lt;<a href="http://dx.doi.org/10.1371/journal.pcbi.0030185">http://dx.doi.org/10.1371/journal.pcbi.0030185</a>&gt;. Acesso em: 12 fev. 2016.</li><li>• PROSDOCIMI, F. et al. Bioinformática: Manual do usuário. <b>Biociência e Desenvolvimento</b>, v. 29, n. 5, p.11-25, jan. 2002. Disponível em: &lt;<a href="http://www.biotecnologia.com.br/revista/bio29/bio29.pdf">http://www.biotecnologia.com.br/revista/bio29/bio29.pdf</a>&gt;. Acesso em: 24 fev. 2016.</li></ul>



PLANO DE ENSINO	
CURSO	DISCIPLINA
Licenciatura em Ciências Biológicas	Introdução à Bioinformática Estrutural
Semestre	ANO
Sexto	2016
CARGA HORÁRIA (SEMESTRAL)	AULAS PREVISTAS (SEMESTRAL)
33h20m	18
<b>PROFESSOR:</b>	

EMENTA
Noções Básicas de Informática. Rudimentos de Biologia Molecular e Bioquímica. O que é Bioinformática, origem e relação com o projeto genoma. Banco de dados. Alinhamento sequencial, <i>threading</i> e alinhamento estrutural de moléculas biológicas. Modelagem molecular por homologia.

OBJETIVOS DA DISCIPLINA
<p><b>GERAL:</b></p> <ul style="list-style-type: none"> <li>• Compreender o que é bioinformática e quais são suas áreas de atuação</li> <li>• Conhecer as ferramentas utilizadas por bioinformatas.</li> </ul> <p><b>ESPECÍFICO:</b></p> <ul style="list-style-type: none"> <li>• Ser capaz de relacionar as diversas áreas afins, tais como biologia molecular, evolução, genética e bioquímica com a bioinformática.</li> <li>• Compreender a importância do projeto genoma e o crescimento dos bancos de dados para a bioinformática</li> <li>• Estabelecer relações entre estrutura e função de proteínas</li> <li>• Operar análises <i>in silico</i> em <i>softwares</i> livres para construir alinhamentos e modelar proteínas.</li> </ul>

CONTEÚDO PROGRAMÁTICO
<p><b>Noções Básicas de Informática, Aspectos de Biologia Molecular e Bioquímica</b></p> <ol style="list-style-type: none"> <li>1. Noções de <i>hardware</i> e <i>software</i>.</li> <li>2. Noções básicas de sistemas operacionais: manipulação de arquivos e diretórios; configurações básicas de desktop.</li> <li>3. Duplicação, transcrição e tradução.</li> <li>4. Proteínas, enzimas, cristalografia.</li> </ol> <p><b>O que é Bioinformática, origem e relação com o projeto genoma. Banco de dados.</b></p> <ol style="list-style-type: none"> <li>1. Origem: de Watson e Crick até os primeiros bancos de dados.</li> <li>2. Primeiros trabalhos em bioinformática.</li> </ol>

3. Projeto Genoma.
4. Bancos de Dados.
5. Vertentes dentro da bioinformática.
6. Principais problemas alvos da Bioinformática atualmente.

### **Alinhamento sequencial, *threading* e alinhamento estrutural de moléculas biológicas.**

1. Qual a necessidade de alinhar moléculas biológicas?
2. Tipos de Alinhamento.
3. Matrizes de Alinhamento. BLOSUM 62.
4. Alinhamento sequencial, *threading*, alinhamento estrutural.
5. BLAST.

### **Modelagem molecular por homologia**

1. Identificação e seleção de molde.
2. Alinhamento das sequências.
3. Construção das coordenadas atômicas.
4. Validação do modelo.

### **ESTRATÉGIAS DE ENSINO**

- Aulas expositivas em quadro branco e pincel com uso de *datashow*;
- Confecção de práticas;
- Simulações com pacotes computacionais livres (licença livre);
- Atendimento ao aluno em horários extra-classe.

### **SISTEMA DE AVALIAÇÃO**

<b>Instrumentos</b>	<b>Valores</b>
1. Avaliação 1 (P1)	20%
2. Avaliação 2 (P2)	20%
3. Trabalhos Práticos (individuais) (T)	60%

Nota final (NF) é dada por:

$NF = 0,20 \cdot P1 + 0,20 \cdot P2 + 0,60 \cdot T$ , onde as notas  $P_i$  e T estão na escala [0-10].

Para aprovação, a NF deve ser superior ou igual à 6,0 e o aluno deve obter 75% de presença nas aulas, conforme artigo 20, Capítulo VI da Portaria n.71 de 25 de novembro de 2013, disponível em:

[http://www.ifs.ifsuldeminas.edu.br/images/arquivos\\_noticias/2015/04\\_abril/resolucao.071.pdf](http://www.ifs.ifsuldeminas.edu.br/images/arquivos_noticias/2015/04_abril/resolucao.071.pdf)

Terão direito à prova de exame final os alunos que obtiverem ao menos 75% de presença, nota final inferior a 6,0 e superior ou igual a 4,0. Para esses discentes a nova Nota Final será calculada pela média ponderada entre a nota do exame (NE) e a nota obtida durante as avaliações do semestre (NS), sendo que a prova de exame terá peso dois enquanto a nota semestral terá peso um. Assim:

$$NF = ((NE * 2) + NS) / 3$$

#### BIBLIOGRAFIA BÁSICA

- ALBERTS, B. et al. **A biologia molecular da célula**. 5. ed. Porto Alegre: Artmed, 2010. 1396 p.
- MANZANO, A. N. G.; MANZANO, M. I. N. G. **Informática Básica**. 7. Ed. São Paulo. Érica, 2007. 250 p.
- VERLI, H. et al (Org.). **Bioinformática: da Biologia à flexibilidade molecular**. São Paulo: Sbbq, 2014. 282 p. Disponível em: <<http://www.ufrgs.br/bioinfo/ebook/>>. Acesso em: 24 fev. 2016.

#### BIBLIOGRAFIA COMPLEMENTAR

- ARAÚJO, N. D. et al. A era da Bioinformática: Seu potencial e suas implicações para as ciências da saúde. **Estudos de Biologia**, Curitiba, v. 70, n. 30, p.143-148, 2008. Disponível em: <[www2.pucpr.br/reol/index.php/BS?ddl=4621&dd99=pdf](http://www2.pucpr.br/reol/index.php/BS?ddl=4621&dd99=pdf)>.
- CLAVERIE, J. M.; NOTREDAME, C. **Bioinformatics for Dummies**. 2. ed. Indianápolis: Wiley Publishing, Inc., 2007. 436 p.
- FILHO, O. A. A.; ALENCASTRO, R. B. Modelagem de proteínas por homologia. **Química Nova**, São Paulo, v. 26, n. 2, p.253-259, set. 2003
- LESK, A.M. **Introduction to Bioinformatics**. New York: Oxford University Press Inc, 2002. 255 p.
- NESHICH, G. Computational Biology in Brazil. **Plos Computational Biology**, São Francisco, v. 3, n. 10, p.1845-1848, out. 2007. Disponível em: <<http://dx.doi.org/10.1371/journal.pcbi.0030185>>. Acesso em: 12 fev. 2016.
- PROSDOCIMI, F. et al. Bioinformática: Manual do usuário. **Biotecnologia Ciência e Desenvolvimento**, v. 29, n. 5, p.11-25, jan. 2002. Disponível em: <<http://www.biotecnologia.com.br/revista/bio29/bio29.pdf>>. Acesso em: 24 fev. 2016.

#### 4. CONCLUSÃO

Durante o curso, alguns alunos demonstraram grandes dificuldades em relação à utilização de tecnologias, necessitando auxílio constante para a manipulação correta de seus próprios computadores pessoais. Infelizmente, como trago à luz por Damásio (2007) muitos não estão dispostos a mudança, preferindo não deixar suas zonas de conforto e manterem-se alheios às novas tecnologias.

Mesmo com essa carência, o curso obteve sucesso em apresentar a uma parcela dos estudantes licenciandos em ciências biológicas algumas ferramentas utilizadas em bioinformática e os conteúdos relacionados à Bioinformática. O quadro 3 demonstra isso. As práticas durante o curso foram importantes para auxiliar na relação estudante-conteúdo, tornando o estudante mais ativo em relação ao seu aprendizado. Sendo a bioinformática uma ciência puramente prática, é necessário reforço constante sobre as práticas para o bom entendimento desta. Além do mais, o curso proporcionou aos participantes momentos de auxílio mútuo, dando às práticas uma característica democratizadora de ensino, no sentido que auxilia na diminuição das diferenças de habilidades entre as pessoas. Outro aspecto importante a ser ressaltado é papel do professor durante o curso, que foi tornando-se cada vez mais secundário a cada dia, conforme os estudantes iam melhor adaptando-se às ferramentas utilizadas, demonstrando no curso um caráter em crescente debate no campo educacional dos últimos anos, do professor como um facilitador do ensino em detrimento do professor como o dono do saber.

## 5. REFERÊNCIAS BIBLIOGRÁFICAS

ALBERTS, B. et al. Mecanismos genéticos básicos. In: ALBERTS, B. et al. **Biologia Molecular da Célula**. 5. ed. Porto Alegre: Artmed, 2010. Cap. 5 e 6. p. 263-410.

ÂNGELO, P. C. S. et al. Guarana (*Paullinia cupana* var. *sorbilis*), an anciently consumed stimulant from the Amazon rain forest: the seeded-fruit transcriptome. **Plant Cell Reports**, v. 27, n. 1, p.117-124, 5 out. 2007. Disponível em: <<http://dx.doi.org/10.1007/s00299-007-0456-y>>. Acesso em: 07 jun. 2016.

ARAÚJO, N. D. et al. A era da Bioinformática: Seu potencial e suas implicações para as ciências da saúde. **Estudos de Biologia**, Curitiba, v. 70, n. 30, p.143-148, 2008. Disponível em: <[www2.pucpr.br/reol/index.php/BS?ddl=4621&dd99=pdf](http://www2.pucpr.br/reol/index.php/BS?ddl=4621&dd99=pdf)>. Acesso em: 02 mar. 2016.

ARBEX, W.; COSTA, V. M. M. S.; SILVA, M. V. G. B. Bioinformática como ferramenta nas pesquisas atuais. In: III ENCONTRO DE GENÉTICA E MELHORAMENTO DE UNIVERSIDADE FEDERAL DE VIÇOSA, Anais do III Encontro de Genética e Melhoramento da Universidade Federal de Viçosa. Viçosa: Universidade Federal de Viçosa, 2006.

BARATA, G. **Capacitação ainda é o principal obstáculo para a área**. 2003. Disponível em: <<http://www.comciencia.br/reportagens/bioinformatica/bio08.shtml>>. Acesso em: 17 mar. 2016.

BIASINI, M. et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. **Nucleic Acids Research**, Oxford, v. 42, n. 1, p.252-258, 2014. Disponível em: <<http://dx.doi.org/10.1093/nar/gku340>>. Acesso em: 07 jun. 2016.

BASSI, S.; GONZÁLEZ, V.; PARISI, G. Computational Biology in Argentina. **Plos Computational Biology**, São Francisco, v. 3, n. 12, p.2426-2431, 2007. Disponível em: <<http://dx.doi.org/10.1371/journal.pcbi.0030257>>. Acesso em 07 jun. 2016.

BENÍTEZ-PÁEZ, A.; CÁRDENAS-BRITO, S. Bioinformática en Colombia: presente y futuro de la investigación biocomputacional. **Biomédica**, Bogotá, v. 30, n. 30, p.170-177, 2010.

BERMAN, H. M. et al. **RCSB PDB**. 2016. Disponível em:< <http://www.rcsb.org/pdb/home/home.do>> Acesso em: 14 mar. 2016.

BERTALAN, M. et al. Complete genome sequence of the sugarcane nitrogen-fixing endophyte *Gluconacetobacter diazotrophicus* Pal5. **Bmc Genomics**, v. 10, n. 1, p.1-17, set. 2009. Disponível em: <<http://dx.doi.org/10.1186/1471-2164-10-450>>. Acesso em: 07 jun. 2016

BONGIOLO, E. **Análise Panorâmica da Bioinformática no Brasil: Propostas da gestão de pessoas para os laboratórios de pesquisa**. 2006. 155 p. Monografia (Especialização em Gestão de Pessoas) - Universidade do Extremo Sul Catarinense.

CLAVERIE, J. M.; NOTREDAME, C. Finding out what bioinformatics can do for you: What is Bioinformatics. In: CLAVERIE, J. M.; NOTREDAME, C. **Bioinformatics for Dummies**. 2. ed. Indianapolis: Wiley Publishing, Inc., 2007. Cap. 1. p. 7-28.

CORREIA, K.. Projetos de Letramento no Ensino Médio: novas perspectivas e desafios. **Educação e realidade**, Porto Alegre, v. 41, n. 1, p.259-277, mar. 2016. Disponível em:<[dx.doi.org/10.1590/2175-623653625](http://dx.doi.org/10.1590/2175-623653625)>. Acesso em: 07 jun. 2016

CRICK, F. Central Dogma of Molecular Biology. **Nature**, Cambridge, v. 227, n. 5258, p.561-563, 1970. Disponível em: <<http://www.nature.com/nature/journal/v227/n5258/abs/227561a0.html>>. Acesso em: 31 nov. 2015.

DAMÁSIO, M. Quadro de referência: Modelo de evolução das tecnologias da informação e da comunicação. In: DAMÁSIO, Manuel José. **Tecnologia e Educação: As tecnologias da informação e da comunicação e o processo educativo**. Lisboa: Vegas, 2007. Cap. 1. p. 36-66

DINIZ, N. B.; OLIVEIRA, D. M. Integração de aplicativos e algoritmos da plataforma WEBPHYLIP na análise transcriptômica e proteômica de *Leishmania chagasi*. In: VI SEMANA UNIVERSITÁRIA DA UECE. 2001, Fortaleza. **Anais**. Disponível em: <[http://www.uece.br/propgpq/semana\\_universitaria/anais/anais2001/Iniciacao\\_Cientifica/biologicas/ic/biol087.htm](http://www.uece.br/propgpq/semana_universitaria/anais/anais2001/Iniciacao_Cientifica/biologicas/ic/biol087.htm)>. Acesso em: 07 jul. 2016.

DING, Y. et al. Bioinformatics: Introduction and Methods, a Bilingual massive open online course (MOOC) as a new example for global bioinformatics education. **Plos Computational Biology**, São Francisco, v. 10, n. 12, p.1-10, 2014. Disponível em: <<http://dx.doi.org/10.1371/journal.pcbi.1003955.t001>>. Acesso em: 07 jun. 2016

FARIAS, A. Q. P; CHACON, P. F. S; SILVA, N. R. R. A bioinformática como ferramenta de formação de recursos humanos no IFRN. **Holos**, Natal, v. 6, p.113-123, 29 jan. 2013. Disponível em: <<http://www2.ifrn.edu.br/ojs/index.php/HOLOS/article/view/688>>. Acesso em: 10 mar. 2016.

FELIPE, M. S. S. et al. Transcriptional Profiles of the Human Pathogenic Fungus *Paracoccidioides brasiliensis* in Mycelium and Yeast Cells. **Journal Of Biological Chemistry**, v. 280, n. 26, p.24706-24714. 2005. Disponível em: <<http://dx.doi.org/10.1074/jbc.m500625200>>. Acesso em: 07 jun. 2016.

FILHO, C. F. **A história da Computação: O caminho do pensamento e da tecnologia**. Porto Alegre: Edipucrs, 2007. 205 p.

FILHO, O. A. S.; ALENCASTRO, R. B. Modelagem de proteínas por homologia. **Química Nova**, São Paulo, v. 26, n. 2, p.253-259, 2003.

FRANCO, M. L.; CEDIEL, J. F.; PAYÁN, C. Breve história de la bioinformática. **Colombia Médica**, Bogotá, v. 39, n. 1, p.117-120, 2008.

FREITAS, D. V. et al. Karyotype with 210 chromosomes in guaraná (*Paullinia cupana* 'Sorbilis'). **Journal Of Plant Research**, v. 120, n. 3, p.399-404, 2007. Disponível em: <<http://dx.doi.org/10.1007/s10265-007-0073-4>>. Acesso em: 07 jun. 2016

GAMA, A. S.; FIGUEIREDO, S. A. O planejamento no contexto escolar. **Discursividade**, Campo Grande, n. 4, p.1-13, 2009.

GERAQUE, Eduardo. A revolução da Bioinformática. **Pesquisa Fapesp**, São Paulo, p.34-36, 2003

GUIDO, R.V. C.; ANDRICOPULO, A. D.; OLIVA, G.. Planejamento de fármacos, biotecnologia e química medicinal: aplicações em doenças infecciosas. **Estudos Avançados**, São Paulo, v. 24, n. 70, p.81-98, 2010. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0103-40142010000300006](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-40142010000300006)>. Acesso em: 17 mar. 2016.

HEIDTMANN, Laura Moretti. **Caracterização do genoma mitocondrial de onça-pintada (*Panthera onca*) e elucidação da filogenia mitogenômica do gênero *Panthera***. 2014. 40 p. Dissertação (Mestrado em Zoologia) - Pontifícia Universidade Católica do Rio Grande do Sul.

HENIKOFF, J. G.; HENIKOFF, S. Blocks database and its applications. **Methods In Enzymology**, v. 266, p.88-105, 1996. Disponível em: <[http://dx.doi.org/10.1016/s0076-6879\(96\)66008-x](http://dx.doi.org/10.1016/s0076-6879(96)66008-x)>. Acesso em: 07 jun. 2016.

HUMPHREY, W.; DALKE, A.; SCHULTEN, K. VMD: Visual Molecular Dynamics. **Journal Of Molecular Graphics**, Nova York, v. 14, n. 14, p.33-38, 1996. Disponível em: <<http://www.ks.uiuc.edu/Publications/Papers/PDF/HUMP96/HUMP96.pdf>>. Acesso em: 13 mar. 2016

JORDAN, Katy. **MOOC Completion Rates: The Data**. Disponível em: <<http://www.katyjordan.com/MOOCproject.html>>. Acesso em: 25 de jul. 2016.

JUNIOR, H. L. R. et al. Abordagem prática de bioinformática em evento Acadêmico-Científico na cidade de Fortaleza-Ce. **Revista Brasileira em Ensino de Ciência e Tecnologia**, Curitiba, v. 4, n. 1, p.79-91, abr. 2011. Disponível em: <[https://www.researchgate.net/profile/Howard\\_Ribeiro\\_Junior/publication/266594530\\_Abordagem\\_Pratica\\_de\\_Bioinformtica\\_em\\_Evento\\_Acadmico-Cientfico\\_na\\_Cidade\\_de\\_Fortaleza-Ce/links/5436cf9c0cf2bf1f1f2d40c3.pdf](https://www.researchgate.net/profile/Howard_Ribeiro_Junior/publication/266594530_Abordagem_Pratica_de_Bioinformtica_em_Evento_Acadmico-Cientfico_na_Cidade_de_Fortaleza-Ce/links/5436cf9c0cf2bf1f1f2d40c3.pdf)>. Acesso em: 08 abr. 2016

LEINER, B. M. et al. **Brief History of the Internet**. 2012. Disponível em: <[http://www.internetsociety.org/sites/default/files/Brief\\_History\\_of\\_the\\_Internet.pdf](http://www.internetsociety.org/sites/default/files/Brief_History_of_the_Internet.pdf)>. Acesso em: 07 jul. 2016.

LENGAUER, T. Computational Biology at the Beginning of the Post-genomic Era. In: R. WILHELM (Ed.). **Informatics: 10 Years Back - 10 Years Ahead**. Berlim: Springer-Verlag, 2001. p. 341-355.

LESK, A. M. **Introduction to Bioinformatics**. New York: Oxford University Press Inc, 2002. 255 p.

LEVINTHAL, C. (Ed.). Molecular Model-building by Computer. **Scientific American**, v. 6, n. 214, p.42-52, 1966. Disponível em: <<http://www.scientificamerican.com/article/molecular-model-building-by-compute/>>. Acesso em: 12 fev. 2016.

LUCKESI, C. C. Avaliação da Aprendizagem... mais uma vez. In: LUCKESI, C. C. **Avaliação da aprendizagem escolar: Estudos e proposições**. 22. ed. São Paulo: Cortez, 2011. Cap. 4. p. 61-66.

MARINOTTI, O. et al. The Genome of *Anopheles darlingi*, the main neotropical malaria vector. **Nucleic Acids Research**, Oxford, v. 41, n. 15, p.7387-7400, 12 jun. 2013. Disponível em: <<http://dx.doi.org/10.1093/nar/gkt484>>. Acesso em: 07 jun. 2016.

MARTINS, J.S. **Situações Práticas de Ensino: aprendizagem significativa**. Campinas: Autores Associados, 2009. 144 p.

MESELSON, M.; STAHL, F. W. The replication of DNA in *Escherichia coli*. **Proceedings Of The National Academy Of Sciences**, v. 44, n. 7, p.671-682, 1958. Disponível em: <<http://www.pnas.org/content/44/7/671.full?sid=3ab1fded-d028-4335-b62e-c235857a5523>>. Acesso em: 07 jun. 2016

MOUNT, D.W. **Bioinformatics: Sequence and genome analysis**. New York: Cold Spring Harbor Laboratory Press, 2000. 564 p.

NELSON, D. L.; COX, M. M. **Princípios de Bioquímica de Leninger**. 5. ed. Porto Alegre: Artmed, 2011. 1274 p.

NESHICH, G. Computational Biology in Brazil. **Plos Computational Biology**, São Francisco, v. 3, n. 10, p.1845-1848, out. 2007. Disponível em: <<http://dx.doi.org/10.1371/journal.pcbi.0030185>>. Acesso em: 12 fev. 2016.

NEVES, B. C.; SANTOS, R. R.; GOMES, H. F. A condição estruturante das tecnologias nas relações sociais: potencialidades na interlocução entre biblioteca e usuário. **Informação & Sociedade: Estudos**, João Pessoa, v. 3, n. 22, p.25-31, dez. 2012.

NOVAES, M. B. C.; GIL, A. C. A pesquisa ação participante como estratégia metodológica para o estudo do empreendedorismo social em administração de empresas. **Revista de Administração Mackenzie**, São Paulo, v. 10, n. 1, p.134-160, 2009.

**NUCLEIC ACIDS RESEARCH**. Oxford: Oxford University Press, 2016.

PADILHA, Itácio Queiroz de Mello et al. A Bioinformática como instrumento de inserção digital e de difusão da biotecnologia. **Revista Eletrônica Extensão Cidadã**, v. 5, n. 111, p.1-5, jun. 2008. Disponível em: <<http://periodicos.ufpb.br/ojs/index.php/extensaocidada/article/view/2073/1831>>. Acesso em: 02 fev. 2016.

PASSOS, C. M. B. Planejamento: para além do burocratismo. Fortaleza. 2006

PEDROSA, F. O. et al. Genome of *Herbaspirillum seropedicae* Strain SmR1, a Specialized Diazotrophic Endophyte of Tropical Grasses. **Plos Genetics**, v. 7, n. 5, p.1-10, 12 maio 2011. Disponível em: <http://dx.doi.org/10.1371/journal.pgen.1002064>. Acesso em: 07 jun. 2016

POLSSON, K. **Chronology of Apple Computer Personal Computers**. 2009. Disponível em: <<http://www.islandnet.com/~kpolsson/applehis/>>. Acesso em: 07 jul. 2016.

PONS, T.; MONTERO, L.A.; FEBLES, J.P. Computational Biology in Cuba: An Opportunity to Promote Science in a Developing Country. **Plos Computational Biology**, São Francisco, v. 3, n. 11, p.2047-2051, 2007. Disponível em: <<http://dx.doi.org/10.1371/journal.pcbi.0030227>>. Acesso em: 07 jun. 2016



PROSDOCIMI, F. et al. Bioinformática: Manual do usuário. **Biotecnologia Ciência e Desenvolvimento**, Brasília, v. 29, n. 5, p.11-25, jan. 2002. Disponível em: <<http://www.biotecnologia.com.br/revista/bio29/bio29.pdf>>. Acesso em: 24 fev. 2016.

PROSDOCIMI, F.; SANTOS, F. R. Sobre bioinformática, genoma e ciência. **Ciência Hoje**, v. 35, n. 209, p.54-57, out. 2004.

PRUITT, K. D.; TATUSOVA, T.; MAGLOTT, D. R.. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. **Nucleic Acids Research, Oxford**, v. 33, n. 0, p.501-504, 17 dez. 2004. Disponível em: <<http://dx.doi.org/10.1093/nar/gki025>>. Acesso em: 07 jun. 2016.

REIS, M. A. **Aplicação das técnicas de espalhamento de Raio-X na caracterização estrutural de proteínas e modelagem computacional utilizando vínculos experimentais obtidos por SAXS**. 2013. 124 p. Tese (Doutorado em Ciências) – Unicamp.

RICOY, M. C.; COUTO, M. J. V. S. As boas práticas com TIC e a utilidade atribuída pelos alunos recém-integrados na universidade. **Educação Pesquisa**, [s.l.], p.897-912, 2014. Disponível em: <<http://dx.doi.org/10.1590/s1517-97022014005000005>>. Acesso em: 07 jun 2016.

ROTHER, Kristian. **Introduction to PyMOL: What is PyMOL?**. 2005. Disponível em: <[http://pages.jh.edu/pfleming/bioinform/files/PyMOL\\_Tutorial.pdf](http://pages.jh.edu/pfleming/bioinform/files/PyMOL_Tutorial.pdf)>. Acesso em: 15 ago. 2016.

ROY, A.; KUCUKURAL, A.; ZHANG, Y. I-TASSER: a unified platform for automated protein structure and function prediction. **Nature Protocols**, v. 5, n. 4, p.725-738, 25 mar. 2010. Disponível em: <<http://dx.doi.org/10.1038/nprot.2010.5>>. Acesso em: 07 jun. 2016

SABBATINI, R. Bioinformática, a profissão do futuro. **Correio Popular**. Campinas, p. 1-2. 1999. Disponível em: <<http://www.sabbatini.com/renato/correio/cp990430.htm>>. Acesso em: 16 mar. 2016.

SANGER, F.; COULSON, A.R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. **Journal Of Molecular Biology**, v. 94, n. 3, p.441-448, 1975. Disponível em: <[http://dx.doi.org/10.1016/0022-2836\(75\)90213-2](http://dx.doi.org/10.1016/0022-2836(75)90213-2)>. Acesso em: 10 mar. 2016.

SANTOS, L. A articulação entre a avaliação somativa e a formativa, na prática pedagógica: uma impossibilidade ou um desafio?. **Ensaio: avaliação e políticas públicas em educação**, Rio de Janeiro, v. 24, n. 92, p. 637-669, 2016. Disponível em <<http://dx.doi.org/10.1590/S0104-40362016000300006>>. Acesso em 22 ago. 2016.

SANTOS, L.A. **Uso de ferramentas de bioinformática para estudos de epidemiologia molecular, filogeografia e filodinâmica viral**. 2010. 139 p. Dissertação (Mestrado em Biotecnologia) - Fundação Oswaldo Cruz Centro de Pesquisas.

SAYERS, E. W. et al. Database resources of the National Center for Biotechnology Information. **Nucleic Acids Research**, [s.l.], v. 37, n. 0, p.5-15, 1 jan. 2009. Disponível em: <<http://dx.doi.org/10.1093/nar/gkn741>>. Acesso em 07 jun 2016.

SIMPSON, A. J. G. et al. The genome sequence of the plant pathogen *Xylella fastidiosa*. **Nature**, v. 406, n. 6792, p. 151-157, 2000.

SPUDEIT, D. **Elaboração do Plano de ensino e do Plano de aula**. 2014. Disponível em: <<http://www2.unirio.br/unirio/cchs/eb/ELABORAODOPLANODEENSINOEDOPLANODEAULA.pdf>>. Acesso em: 24 out. 2016.

THE UNIPROT CONSORTIUM (Ed.). UniProt: A hub for protein information. **Nucleic Acids Research**, Oxford, v. 43, n. 11, p.204-212, 2014. Disponível em: <<http://dx.doi.org/10.1093/nar/gku989>>. Acesso em 07 jun. 2016.

TISDALL, J. What is Bioinformatics. In: TISDALL, James. **Beginning Perl for Bioinformatics**. Sebastopol: O'reilly, 2001. p. 4-5.

TRIPP, D. Pesquisa-ação: uma introdução metodológica. **Educação e Pesquisa**, São Paulo, v. 3, n. 31, p.443-466, dez. 2005.

VASCONCELOS, A. T. R. et al. Swine and Poultry Pathogens: the Complete Genome Sequences of Two Strains of *Mycoplasma hyopneumoniae* and a Strain of *Mycoplasma synoviae*. **Journal Of Bacteriology**, v. 187, n. 16, p.5568-5577, 2005.

VERLI, H. (Org.). **Bioinformática: da Biologia à Flexibilidade molecular**. São Paulo: Sbbq, 2014. 282 p. Disponível em: <<http://www.ufrgs.br/bioinfo/ebook/>>. Acesso em: 30 nov. 2015.

VOET, D.; VOET, J. G.; PRATT, C.W. **Fundamentos de Bioquímica: A vida em nível molecular**. 2. ed. Porto Alegre: Artmed, 2008. 1264 p.

WATSON, J. D.; CRICK, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. **Nature**, v. 171, n. 4356, p.737-738, 1953. Disponível em: <<http://dx.doi.org/10.1038/171737a0>>. Acesso em: 07 jun. 2016

YANG, J. et al. The I-TASSER Suite: protein structure and function prediction. **Nature Methods**, v. 12, n. 1, p.7-8, 2014. Disponível em: <<http://dx.doi.org/10.1038/nmeth.3213>>. Acesso em: 07 jun. 2016

ZABALA, A. A prática educativa: como ensinar. São Paulo: Artmed, 1998. 224 p.

ZABALZA, M. Os dilemas práticos dos professores. **Revista Pátio**, n. 27, 8-11. 2003.

ZHANG, Y. I-TASSER server for protein 3D structure prediction. **Bmc Bioinformatics**, v. 9, n. 1, p.40-48, 2008. Disponível em: <<http://dx.doi.org/10.1186/1471-2105-9-40>>.

ZHANG, Y. TM-align: a protein structure alignment algorithm based on the TM-score. **Nucleic Acids Research**, Oxford, v. 33, n. 7, p.2302-2309, 2005. Disponível em: <<http://dx.doi.org/10.1093/nar/gki5>>. Acesso em: 07 jun. 2016.

## ANEXO I



Projeto de TCC: Tópicos em bioinformática estrutural: a concepção de um currículo.

Aluno: Samuel Messias de Oliveira

Orientador: Marcelo Augusto dos Reis

Curso: Licenciatura em Ciências Biológicas

e-mail: samueloliveira30@hotmail.com

Dia e horário: As quartas-feiras, das 17h às 18h30min

Bioinformática: Em suma, a bioinformática pode ser descrita como a aplicação de tecnologias computacionais para o entendimento e o uso de informações biológicas, permitindo que pesquisadores possam ampliar seus conhecimentos de sistemas biológicos complexos *in silico*.

Proposta: Participação no desenvolvimento e aplicação de um curso relacionado à área de bioinformática estrutural.

O curso: Deverá ter início em 23/03/2016 e finalizado juntamente com o semestre letivo.

Objetivos gerais: Compreender o que é bioinformática e quais são suas áreas de atuação. Conhecer as ferramentas utilizadas por bioinformatas.

Para os ingressantes: Além do ganho de novos conhecimentos, a participação no curso poderá render horas de AACC.

### Questionário

Marque um x em frente cada resposta

1. Qual seu período de estudos nessa IES?

Primeiro: Terceiro: Quinto: Sétimo: Outro:

2. Qual seu nível de interesse em participar do curso?

Bastante interessado: Interessado: Pouco interessado:

Não-interessado:

3. Qual sua disponibilidade em participar do curso?

Bastante disponível: Disponível: Pouco disponível:

Não-disponível:

4. Você tinha ciência desta área de pesquisa (bioinformática) antes desta apresentação? Caso tenha, onde passou a conhecer?

## ANEXO II



Tópicos em Bioinformática: A concepção de um currículo	
Aluno: Samuel Messias de Oliveira	Curso: Licenciatura em Ciências Biológicas
Orientador: Dr. Marcelo Augusto dos Reis	Dia e hora: As Quartas-feiras, entre as 17h15min e 18h45min
Bioinformática: Em suma, a bioinformática pode ser descrita como a aplicação de tecnologias computacionais para o entendimento e o uso de informações biológicas, permitindo que pesquisadores possam ampliar seus conhecimentos de sistemas biológicos complexos <i>in silico</i> .	
Local do curso: Laboratório de Física. No prédio principal, próximo ao auditório.	Número de vagas: 20
Proposta: Participação no desenvolvimento e aplicação de um curso relacionado à área de bioinformática estrutural.	
O curso: Deverá ter início em 23/03/2016 e finalizado juntamente com o semestre letivo.	
Objetivos gerais: Compreender o que é bioinformática e quais são suas áreas de atuação. Conhecer as ferramentas utilizadas por bioinformatas.	
Para os ingressantes: Além do ganho de novos conhecimentos, a participação no curso poderá render horas de AACC.	
Para os interessados: É favor deixar o nome, e-mail e período na tabela abaixo. Manteremos contato por e-mail.	
Dúvidas: samueloliveira30@hotmail.com	

Nome	E-mail	Período	Possui notebook?

## ANEXO III

### Prática I: Banco de dados de Proteínas

**Objetivos:** Familiarizar-se com o PDB. Conhecer arquivos PDB e FASTA.

**Materiais para uso:** Computador pessoal e acesso à Internet.

1. Com o *browser* acessar o *site* do PDB: [www.rcsb.org/pdb](http://www.rcsb.org/pdb).
2. Baixar as coordenadas atômicas da proteína RUBISCO, proveniente de folhas de espinafre, código de acesso 1RCX.
3. Localizar no arquivo PDB ou *site* e anotar as informações sobre autores, fonte da proteína, quantidade de cadeias e técnica utilizada para resolução da proteína.
4. No *site* do PDB identificar e anotar a massa molecular da proteína (Total Structure Weight).
5. Baixar a sequência no formato FASTA.
6. Identificar e anotar o número de resíduos de aminoácido.
7. Baixar as coordenadas atômicas da mioglobina, código de acesso: 1MBO
8. Localizar no arquivo PDB ou *site* e anotar as informações sobre autores, fonte da proteína, quantidade de cadeias e técnica utilizada para resolução da proteína.
9. No *site* do PDB identificar e anotar a massa molecular da proteína.
10. Baixar a sequência no formato FASTA.
11. Identificar e anotar o número de resíduos de aminoácido.
12. Baixar as coordenadas atômicas da lisozima, código de acesso: 6LYZ.
13. Localizar no arquivo PDB ou *site* e anotar as informações sobre autores, fonte da proteína, quantidade de cadeias e técnica utilizada para resolução da proteína.
14. No *site* do PDB identificar e anotar a massa molecular da proteína.
15. Baixar a sequência no formato FASTA.
16. Identificar e anotar o número de resíduos de aminoácido.
17. Qual comparação pode ser feita entre o formato PDB e o formato FASTA?

## ANEXO IV

### Prática II: Visualização de aminoácidos

Objetivos: Visualizar e familiarizar-se com a estrutura tridimensional de aminoácidos através do programa VMD. Identificação dos átomos que fazem parte dos aminoácidos.

Materiais: Computador pessoal, arquivo com coordenadas atômicas de aminoácidos e programa VMD.

#### Procedimento:

1. Carregar as coordenadas atômicas da alanina (ala.pdb) da seguinte maneira. No programa VMD clique em *File* e em *New Molecule*, em seguida clique em *Browse*. Selecione o arquivo PDB correspondente a partir do diretório (pasta) onde salvou os arquivos enviados por e-mail. Clique em abrir e *Load*.
2. Observe a tela gráfica. As linhas representam ligações covalentes entre os átomos. As cores representam cada átomo. Azul, nitrogênio; Branco, hidrogênio; Vermelho, oxigênio; Azul claro, carbono.
3. Identificação dos átomos. Clique VMD *Main* → *Mouse* → *Label* → *Atoms*. Clique sobre cada átomo. Desta maneira, identifique cada átomo que compõe o aminoácido alanina. Identificar o carbono alfa, nitrogênio, carbono e oxigênio da ligação carbonila, carbono beta (carbono da cadeia lateral). Representar a fórmula molecular no quadro de fórmula molecular.
4. Visualizar os aminoácidos a partir de outros esquemas gráficos. *Graphical Representations* → *Drawing Method* → *VDW* (raio de Van der Waals). Qual a principal característica dessa representação?  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_
5. *Graphical Representations* → *Drawing Method* → *CPK*. Qual a principal característica dessa representação?  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_
6. Representação dos aminoácidos em figuras. A partir da representação gráfica CPK, gere as figuras em formato bmp, procedimento a seguir: VMD *Main* → *File* → *Render* → *render using: snapshot*. Selecione um diretório (pasta) e um nome para a imagem a ser gerada. Clique em *start rendering*. Cole a imagem gerada no quadro de imagens.
7. Selecione outros dez aminoácidos e refaça o mesmo procedimento.

Quadro de fórmula Molecular

Aminoácido	Código de três letras (Código de uma letra)	Arquivo PDB	Fórmula molecular
Alanina	ALA (A)	ala.pdb	C <sub>3</sub> H <sub>7</sub> NO <sub>2</sub>

Tabela de Imagens

Aminoácido	Imagem
Alanina	


#### **Referência**

AZEVEDO JUNIOR, Walter Filgueira de. **Visualização de aminoácidos**. 2007. Disponível em: <[http://www.biocristalografia.df.ibilce.unesp.br/cursos/curso\\_bioinfo/roteiros.pdf](http://www.biocristalografia.df.ibilce.unesp.br/cursos/curso_bioinfo/roteiros.pdf)>. Acesso em: 15 fev. 2016.



## ANEXO V

### Prática III: Visualização de Hélices

**Objetivos:** Visualizar a estrutura tridimensional de hélices a partir de programas computacionais. Saber distinguir Hélice-alfa, hélice<sub>310</sub> e hélice pi. Visualizar as ligações de hidrogênio que estabilizam as hélices.

**Materiais:** Computador pessoal, programa VMD e arquivos com coordenadas atômicas.

**Procedimento:**

1. Visualização de polialanina (poliala.pdb). Para a visualização, seguir passos 1 e 2 da prática II.
2. Que estrutura secundária esse peptídeo representa?  
\_\_\_\_\_
3. Visualização de ligações de hidrogênio. Clique VMD *Main* → *Graphics* → *Representations*. Clique no Botão *Create rep.* A seguir selecione *Drawing Method* e *HBond*. Aumente o *Distance Cutoff* para 3.4. Veja o Resultado gerado. Cada nova linha tracejada representa uma ligação de hidrogênio.
4. Quantas ligações de hidrogênio possui o peptídeo?  
\_\_\_\_\_
5. Montar uma tabela com as ligações de hidrogênio. Para isso, precisamos dar nome aos átomos presentes no peptídeo. VMD *Main* → *Mouse* → *Labels* → *Atom*. Agora, clicando em cada átomo participante da ligação de hidrogênio é possível saber qual é. As ligações de hidrogênio possuem sempre um grupo doador da ligação (NH) e um grupo receptor da ligação (carbonila).
6. Clique em todos os átomos doadores e receptores de hidrogênio e complete a tabela abaixo.

Ligação de Hidrogênio	Atomo Aceitador	Atomo Receptor
1	ALA 1:O	ALA 5:N
2		
3		
4		
5		
6		

7. Há um padrão nas ligações de hidrogênio? Se sim, Qual?  
\_\_\_\_\_  
\_\_\_\_\_
8. Determinação do comprimento do peptídeo. A cada resíduo de aminoácido temos um padrão de tamanho de 1,5 Å. Uma hélice alfa possui 3,6 resíduos de aminoácidos,

obtendo um comprimento de 5,4 Å. A equação para descobrir o comprimento de uma hélice-alfa pode ser descrita como  $L_H = N_a \cdot 1,5 \text{ Å}$ . Onde  $L_H$  representa o tamanho da hélice e  $N_a$  o número de resíduos de aminoácidos. Qual o comprimento do peptídeo Poliala?

9. Observar o peptídeo em outros esquemas gráficos. Para isso seguir o passo 4 da prática II. Visualizar em VDW, CPK e *New Cartoon*.

10. Qual a principal característica da representação *New Cartoon*?

11. Observe a tabela abaixo, ela representa o parâmetro de hélices. Alfa hélices, Hélices  $3_{10}$  e Hélice pi.

Hélice	N (resíduos por volta)	D (Å)	P(A)	Nº de átomos entre ligações de hidrogênio.	Representação no VMD
$3_{10}$	3	2,0	6,0	10	G(Rosa)
Alfa	3,6	1,5	5,4	13	H(Roxo)
Pi	4,4	1,1	4,8	16	I (Vermelho)

Onde D(A) é o comprimento em Angstroms de cada resíduo de aminoácidos e P(A) é o tamanho completo de uma hélice.  $P(A) = N \cdot D(A)$ .

12. Hélices  $3_{10}$  e Pi são muito raras em proteínas devido ao estado conformacional menos estável que a hélice alfa. Entretanto algumas proteínas apresentam essas conformações. A estrutura da proteína amarela fotoativa (2phy.pdb) apresenta um pequeno trecho em hélice pi. Visualizaremos sua estrutura.

13. Carregue as coordenadas atômicas da proteína seguindo o procedimento 1. Visualize na representação *New Cartoon*. VMD  
Main → Graphics → Representations → Graphical representations → New Cartoon.

14. Nesse menu, selecione *Coloring Method* → *Structure*

15. Qual o início e o final da hélice pi dessa estrutura?

## Referência

AZEVEDO JUNIOR, Walter Filgueira de. Visualização de hélices. 2007. Disponível em: <[http://www.biocristalografia.df.ibilce.unesp.br/cursos/curso\\_bioinfo/roteiros.pdf](http://www.biocristalografia.df.ibilce.unesp.br/cursos/curso_bioinfo/roteiros.pdf)>. Acesso em: 15 fev. 2016.

## ANEXO VI

### Prática IV: Estrutura tridimensional de proteínas

Objetivos: Visualizar estruturas proteicas. Compreender características inerentes ao funcionamento e uma proteína. Alinhamento estrutural.

Materiais: PC, conexão à internet, arquivos PDB, software PyMol.

#### Parte I

Procedimento:

1. Abrir o programa "PyMol". Perceba que ao abrir o programa duas telas são mostradas. Disponha-as de tal modo que seja possível visualizar ambas as páginas.
2. Para inserir a proteína vá em *Plugin* → *PDB loader service* e digite o código PDB da proteína mioglobina (1MBO) e aperte *OK*. Ir em *Action (A)* → *Remove Waters*.
3. Vá em *Display* e depois *Sequence* para visualizar a sequência de aminoácidos da proteína.
4. Alterar a forma de apresentação da proteína. Vá em *Show (S)* → *as*. Altere as visualizações entre *Lines*, *Sticks*, *Ribbon*, *Cartoon*, *Dots*, *Spheres*, *Mesh* e *Surface*. No final volte à forma de visualização *Cartoon*.
5. Clique com o botão esquerdo do *mouse* sobre a proteína para poder girar a proteína para os lados em que desejar. Para aumentar ou diminuir o zoom sobre a proteína, aperte *Ctrl* e mantenha, em seguida gire a bolinha do *mouse* para frente ou para trás.
6. Vá em *S* → *as* → *Surface*. Qual a principal característica que podemos perceber na proteína a partir dessa visualização?

---

7. *S* → *as* → *nonbonded* (A intenção é deixar a tela sem qualquer imagem. Caso sobre alguns pontos vermelhos (oxigênios não ligados covalentemente à proteína) vá em *A* → *remove Waters*.) Em seguida *S* → *side chain* → *sticks*. *S* → *main chain* → *lines*. *Main chain* (cadeia principal) se refere ao conjunto de átomos diretamente ligados ao *C $\alpha$*  de cada resíduo de aminoácido e *Side chain* (cadeia lateral) se refere aos átomos presentes nas cadeias laterais de cada resíduo de aminoácido.

8. *S* → *cartoon*. Observando a molécula seguindo estes parâmetros e compreendendo a diferença entre *main* e *side chain*, Responda: A visualização *cartoon* representa estruturalmente qual tipo de cadeia?

---

9. *S* → *as* → *Surface*. Note que a proteína é formada por apenas uma cadeia de aminoácidos. Ao final da sequência há um grupo prostético (heme, HEM), selecione-o clicando sobre ele. Vá em (sele) → *S* → *as* → *lines*. Em seguida selecione o oxigênio na posição 555 (OXY) clicando sobre ele. Vá em (sele) → *S* → *as* → *lines*, novamente.

10. Qual a relação entre a resposta do procedimento número 6 com o que descobriram no procedimento número 9?

---

---

---

11. Volte a visualização por linhas. *All →S →as →lines*. Para melhor compreender a interação do grupo prostético com a proteína selecione os aminoácidos H64, H93, V67 e F43. Selecione também o OXY e HEM e vá em *(sele) →S →as →lines*. Em seguida, selecione somente os aminoácidos e vá em *(sele) →S →as →spheres*.

#### Parte II – Alinhamento estrutural.

1. Insira no PyMol, o código da hemoglobina: 1BBB, como feito como feito no procedimento 2 da parte I. Em seguida insira também a mioglobina 1MBO.

2. Coloque ambas as estruturas no formato *cartoon*, identifique e coloque em evidência (*spheres*) os quatro grupos heme da hemoglobina e o heme da mioglobina

3. Vá em *1BBB →A →align →to molecule →1MBO*.

#### Questões

1. Identifique quais os tipos de estruturas secundárias presentes na hemoglobina. Quantos tipos de estruturas secundárias formam essa proteína? E quantas subunidades?

2. Quais são as principais semelhanças e diferenças entre as estruturas da mioglobina e da hemoglobina?

#### Referência:

OLIVEIRA, Marcos Túlio. *Visualização estrutura tridimensional de proteínas*. 2015. Disponível em: <<http://www.fcav.unesp.br/Home/departamentos/tecnologia/marcostuliooliveira/roteiro-aula-pratica--pymol.pdf>>. Acesso em: 28 fev. 2016.

# ANEXO VII

## Prática V: Alinhamento de proteínas

Objetivos: Calcular qual o melhor alinhamento para (C) utilizando-se de (A) e (B).

Materiais: Computador pessoal, arquivo com coordenadas atômicas de aminoácidos e programa VMD.

### A. Matrizes de substituição de nucleotídeos (a) e aminoácidos BLOSUM 62 (b)

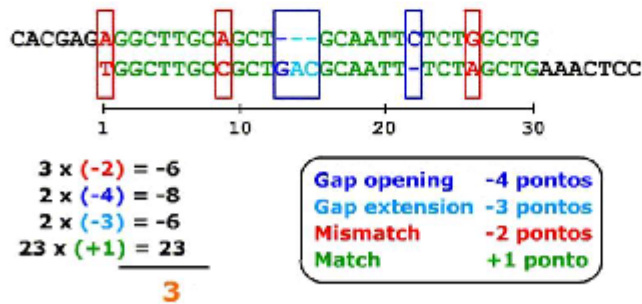
a.

	A	C	G	T
A	1	-2	-2	-2
C		1	-2	-2
G			1	-2
T				1

b.

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	
A	5	-2	-2	-2	0	0	0	-2	-2	-3	-2	-1	-2	0	0	0	-2	-3	0		
R		5	-2	-3	-3	0	-1	-2	0	-3	-4	1	-3	-3	-2	-2	0	0	-3	-4	
N			5	0	0	0	-2	0	0	-4	-5	-2	-3	-3	-2	0	0	-2	-2	-5	
D				5	-4	0	1	-1	0	-5	6	3	-4	-4	0	-2	-2	2	2	-5	
C					8	-2	-3	-1	-1	0	-2	-3	0	-1	-1	1	0	0	-2	0	
Q						5	2	0	0	-2	-4	0	-2	-3	0	0	0	0	-2	-3	
E							5	0	0	-3	-4	0	-3	-3	0	0	0	-2	-3	-3	
G								6	0	-4	-5	-2	-3	-2	-2	0	0	0	-2	-3	
H									6	-3	-4	0	-2	0	0	0	0	0	2	-2	
I										4	0	-5	2	0	-2	-3	0	0	-5	7	
L											4	0	0	-3	-4	-3	0	-4	0		
K												4	-2	-4	-1	-2	0	0	-3	-4	
M													6	0	-3	-3	-2	0	-3	2	
F														6	-3	-2	-2	2	2	0	
P															7	0	0	-2	-3	0	
S																4	2	-2	2	-3	
T																	5	-1	-3	0	
W																		9	2	-1	
Y																				7	-3
V																					4

### B. Exemplo:



C: Dispostos abaixo estão dois alinhamentos produzidos através do *software* BLAST. Utilizando-se do conhecimento adquirido nesta aula, qual dos dois alinhamentos possui um valor de pontuação melhor? Considere somente os primeiros trinta aminoácidos de cada alinhamento (Metionina (M) 1 até Tirosina (Y) 30)

[Download](#) [GenPept](#) [Graphics](#)

PREDICTED: peroxidase 2 [Cucumis sativus]  
 Sequence ID: [ref|XP\\_004142149.1](#) Length: 318 Number of Matches: 1

Range 1: 1 to 318 [GenPept](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
411 bits(1057)	6e-142	Compositional matrix adjust.	209/323(65%)	248/323(76%)	6/323(1%)
Query 1	MAASSKIVVSLVLCMPAVSVRSQLSSTFYDTTCPIWSSIVHGVMOALQSDRAGAKII	60			
Sbjct 1	MALFKVAVFFLLGLMVRAS-QAQLCPTFYDESCPDVSHIVRRVVKQALAFDERAGARLI	59			
Query 61	RLHFHDCFVDCDGSVLLLEDQDITSELGAPNGGITGFNIVNDIKTAVENWCPGVVSCA	120			
Sbjct 60	RLHFHDCFV+GCDGSVLLLEDQ G+ SEL APGN ITGFNIVN+IK AVE CPGVVSCA	119			
Query 121	DILALGSRDAVTLASGQGTAVLQGRDSTRANLQGARLPSPFESLNIQGFROVGLN	180			
Sbjct 120	DILAIASVESVAVLGGPCHEVQLGRRDSRRANLQGAIDGLPSPFEDVTQLKRFGRVGL-	178			
Query 181	DNTDLVALSGAHTFGSRCHFFSGRLN-MNPMADDSPIDSTYASQLNQTCSGSGTFVDL	239			
Sbjct 179	DSTDLVALSGAHTFGKSRQCFDRLRLVNSNP---DSTLNPYRQQLRQACSSGSDTFVNL	235			
Query 240	DPTTPNFDRIYYTINLQNIQGLLRSDQVLFSTPGASTIATVNSLASSAFADAFQSHI	299			
Sbjct 236	DPTTPNFDRIYYTINLQNIQGLLRSDQVLFSTPGASTIATVNSLASSAFADAFQSHI	295			
Query 300	RNGNLDPKTGTGTEIRTKRRLN 322				
Sbjct 296	RNGNIRPLTGNQGEIRSKRRLN 318				

[Download](#) [GenPept](#) [Graphics](#)

PREDICTED: peroxidase 2 [Cucumis melo]  
 Sequence ID: [ref|XP\\_008449778.1](#) Length: 318 Number of Matches: 1

Range 1: 1 to 318 [GenPept](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
410 bits(1054)	2e-141	Compositional matrix adjust.	210/323(65%)	247/323(76%)	6/323(1%)
Query 1	MAASSKIVVSLVLCMPAVSVRSQLSSTFYDTTCPIWSSIVHGVMOALQSDRAGAKII	60			
Sbjct 1	MALFKVAVFFLLGLMVRAS-QAQLCPTFYDESCPDVSHIVRGVVKQALAFDERAGARLI	59			
Query 61	RLHFHDCFVDCDGSVLLLEDQDITSELGAPNGGITGFNIVNDIKTAVENWCPGVVSCA	120			
Sbjct 60	RLHFHDCFV+GCDGSVLLLEDQ G+ SELGAPGN ITGFNIVN+IK AVE CPGVVSCA	119			
Query 121	DILALGSRDAVTLASGQGTAVLQGRDSTRANLQGARLPSPFESLNIQGFROVGLN	180			
Sbjct 120	DILAIASVEAVLGGPCHEVQLGRRDSRRANLQGAIDGLPSPFEDVTQLKRFGRVGL-	178			
Query 181	DNTDLVALSGAHTFGSRCHFFSGRLN-MNPMADDSPIDSTYASQLNQTCSGSGTFVDL	239			
Sbjct 179	DSTDLVALSGAHTFGKSRQCFDRLRLVNSNP---DNTLDARFAQLRRACSSSDTFVNL	235			
Query 240	DPTTPNFDRIYYTINLQNIQGLLRSDQVLFSTPGASTIATVNSLASSAFADAFQSHI	299			
Sbjct 236	DPTTPNFDRIYYTINLQNIQGLLRSDQVLFSTPGASTIATVNSLASSAFADAFQSHI	295			
Query 300	RNGNLDPKTGTGTEIRTKRRLN 322				
Sbjct 296	RNGNIRPLTGNQGEIRSKRRLN 318				

## ANEXO VIII

### Prática VI: Predição estrutural de proteínas

Objetivos: Utilizar-se de ferramentas computacionais para predição de estruturas proteicas por meio de enovelamento proteico (threading). Criação do gráfico de Ramachandram.

Materiais: Computador pessoal, acesso à internet e *softwares* PYMOL e VMD.

1. Abrir O Swiss Model: <http://swissmodel.expasy.org/> . ;
- 2 Ir em Crie Account e fazer o que se pede.
3. Fazer Log In.
4. Entrar no site do NCBI: <http://www.ncbi.nlm.nih.gov/>
5. Modificar o campo “All databases” para “protein” e pesquisar por “manganese peroxidase”
6. Escolher “manganese peroxidase [Trametes versicolor]” e clicar em “FASTA”. Copie toda a sequência.
7. Voltar para o Swiss Model. Modelling→Automated Model.
8. Colar a informação em “paste your sequence here” e clique em “Search for template”
9. Quando o site dispuser os resultados, clicar em “identity” de modo a deixar todos os resultados que possuem maior identidade a frente. Selecione os três primeiros resultados.
10. Clique em seguida em “Build Models”
11. Quando o site mostrar os resultados, selecione o primeiro clicando na seta apontada para baixo logo abaixo da imagem. Clique em seguida em PDB file. Quando a página abrir, clique com o botão direito sobre a página e vá em salvar como. Salve o documento no formato PDB em algum diretório (pasta) em seu computador. Refaça o mesmo com os outros dois resultados.
12. Abrir o VMD e inserir os três resultados da maneira em que foi ensinado na prática II.
13. Clique sobre o primeiro modelo criado e vá em Extensions→analysis→Ramachandran Plot
14. Clique no quadro ao lado da palavra “Molecule” e selecione a molécula de número zero. Verá que em seguida aparecerá uma tela com um gráfico dividido em quatro partes com algumas cores e pontos amarelos. Você sabe o que é cada uma dessas cores e pra que elas são úteis? Para melhor compreender um gráfico de Ramachandran, é interessante ler:

<http://www1.ci.uc.pt/pessoal/manolo/ramachandran.html> (É possível deixar essa leitura para após o tempo de curso, para não diminuir o ritmo da prática.)

15. Clique em Create 3D histogram. Vá à interface gráfica do programa (onde é possível visualizar as proteínas.) e observe o histograma representando o gráfico de Ramachandran. Em seguida, salve o histograma como imagem, como ensinado na prática III. Também salvar a imagem da própria proteína.

16. Faça o mesmo com os outros dois resultados.

17. Adicione as imagens do gráfico no Quadro 1.

18. Abrir o PyMol. Adicionar o resultado número 1 e o resultado número 2. Para isso, vá em Fileopen e insira um arquivo PDB por vez. Em seguida, coloque as duas moléculas para a visualização cartoon. Em seguida, vá em Aalignto molecule. Em seguida, anotar no quadro 2 o valor do RMS. Salvar a imagem do alinhamento e coloca-la no quadro 3.

19. Exclua a molécula 2. Para isso Adelete object. E adicione a molécula 3. Refaça o procedimento acima.

20. Exclua a molécula 1 e adicione a molécula 2. Refaça o procedimento acima.

21. Conceitue Gráfico de Ramachandram e discorra sobre sua importância para a avaliação de predição proteica.



**Quadro 1-Gráfico de Ramachandran**

<b>Nome e imagem da proteína produzida pelo Swiss Model</b>	<b>Imagem gráfico de Ramachandran</b>

**Quadro 2 – Valor RMS**

<b>Alinhamento entre as moléculas</b>	<b>Valor de RMS</b>

**Quadro 3 - Alinhamentos**

Alinhamento entre as moléculas	Imagem