



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Avaliação de Interação Humano-Computador: um estudo de caso para bioinformática

Gabriella de Oliveira Esteves

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

Orientadora

Prof. Dr.^a Maria Emília Machado Telles Walter

Brasília
2016

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

Coordenador: Prof. Dr. Rodrigo Bonifácio de Almeida

Banca examinadora composta por:

Prof. Dr.^a Maria Emília Machado Telles Walter (Orientadora) — CIC/UnB
Prof. Dr.^a Fernanda Lima — CIC/UnB
Prof. Dr. Professor II — CIC/UnB

CIP — Catalogação Internacional na Publicação

Esteves, Gabriella de Oliveira.

Avaliação de Interação Humano-Computador: um estudo de caso para bioinformática / Gabriella de Oliveira Esteves. Brasília : UnB, 2016.

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

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

1. Bioinformática, 2. Redes Metabólicas, 3. NoSQL, 4. Grafo,
5. Interação Humano-Computador, 6. Método de Análise de
Computabilidade

CDU 004.4

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



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Avaliação de Interação Humano-Computador: um estudo de caso para bioinformática

Gabriella de Oliveira Esteves

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

Prof. Dr.^a Maria Emília Machado Telles Walter (Orientadora)
CIC/UnB

Prof. Dr.^a Fernanda Lima Prof. Dr. Professor II
CIC/UnB CIC/UnB

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

Brasília, 08 de Julho de 2016

Dedicatória

Dedicatória

Agradecimentos

Agradecimento

Resumo

Resumo em português

Palavras-chave: Bioinformática, Redes Metabólicas, NoSQL, Grafo, Interação Humano-Computador, Método de Análise de Computabilidade

Abstract

Abstract in english

Keywords: Bioinformatics, Metabolic Networks, NoSQL, Graph, Human-Computer Interaction, Method for Evaluating Software Communicability

Sumário

1	Introdução	1
1.1	Justificativa	2
1.2	Problema	2
1.3	Objetivo	2
1.4	Descrição dos Capítulos	2
2	Avaliação em Interação Humano-Computador	4
2.1	Conceitos Básicos de IHC	4
2.2	Avaliação Através de Observação	6
2.2.1	Método de Avaliação de Comunicabilidade	7
3	Redes Metabólicas	12
3.1	Conceitos Básicos de Biologia Molecular	12
3.1.1	Ácidos Nucleicos	12
3.1.2	Proteínas	13
3.1.3	Síntese de Proteína	15
3.2	Conceitos Básicos de Metabolismo	17
3.2.1	Metabolismo Primário	17
3.2.2	Metabolismo Secundário	18
3.3	Visualização de Redes Metabólicas	19
3.3.1	Reactome, KEGG e BioCyc	20
3.3.2	2Path	27
4	Projeto de Interface	29
4.1	Comunicação usuário-sistema	29
4.1.1	Design de Interação	29
4.1.2	Mapa de Objetivos	29
4.1.3	Tratamento de Rupturas na Comunicação	29
4.1.4	Modelagem da Interação com Linguagem MoLIC	29
4.2	Interface do Sistema	29
4.3	Detalhes de Implementação	30
4.3.1	Banco de Dados OrientDB	30
4.3.2	Desafios	30
5	Resultados	31
5.1	Interface	31
5.2	Avaliação e Discussão	31

6	Conclusão e Trabalhos Futuros	32
6.1	Conclusão	32
6.2	Trabalhos Futuros	32
7	Cronograma	33
	Referências	34

Lista de Figuras

2.1	Exemplo de tipos de caixa de texto que representam o mesmo campo (Nome) porém com níveis de <i>affordance</i> diferentes, sendo que o último apresenta melhor o objetivo da caixa e, portanto, ela é reconhecida mais facilmente. Adaptado de [17].	5
3.1	Nucleotídeo com grupo fosfato (P), pentose abaixo à esquerda e base nitrogenada G (no caso a Guanina). Adaptado de [20].	13
3.2	Representação das duas fitas de ADN, que se ligam por pontes de hidrogênio entre os nucleotídeos. Adaptado de [20].	14
3.3	Quatro estruturas de proteínas. As unidades esféricas na estrutura primária representa os aminoácidos da proteína. Adaptado de [5].	15
3.4	Representação da etapa final da síntese de proteína, que sempre finaliza com o códon UGA. Adaptado de [3].	16
3.5	Exemplo de via metabólica retirado do site MetaCyC. Esta representa o processo de produção da bioluminescência de bactérias [10].	18
3.6	A célula que se divide por meio de mitose gera duas outras células idênticas à original. A célula que se divide por meio de meiose gera quatro células, cada uma com metade do material genético da célula original. Adaptado de [6].	18
3.7	Um exemplo de elemento químico para cada classe, ou grupo, de metabolismo secundário.	19
3.8	Visão geral da rede metabólica do <i>Homo sapiens</i> do Reactome, com destaque na via metabólica de remoção do Orc1 da cromatina do processo de regulação da replicação do ADN. A interface pode ser dividida nas seções [1], [2], [3] e [4].	21
3.9	Via metabólica de remoção do Orc1 da cromatina apresentada pela ferramenta de visualização do Reactome.	22
3.10	Visão geral da rede metabólica de referência do KEGG, com destaque na via metabólica de biossíntese do <i>backbone</i> de terpenóides. A interface pode ser dividida nas seções [1], [2] e [3].	24
3.11	Via metabólica de biossíntese do <i>backbone</i> de terpenóides apresentada pela ferramenta de visualização do KEGG.	26
3.12	Representação modificada da Figura 3.5 com maior nível de detalhes apresentada pelo MetaCyc.	28

Lista de Tabelas

2.1	Descrição das etiquetas, de acordo com a falha de comunicação que representam [7].	9
3.1	Código Genético que mapeia cada códon à um dos 20 aminoácidos, representados de maneira abreviada.	16
7.1	Cronograma	33

Capítulo 1

Introdução

Desde a descoberta da estrutura helicoidal do DNA, por Watson e Crick em 1953 [21], várias linhas de pesquisas em Biologia Molecular foram desencadeadas. O estudo do conjunto de processos celulares que ocorrem envolvendo o DNA e o RNA, denominado Dogma Central [21], permitiu o sequenciamento dessas macromoléculas, bem como o sequenciamento de estruturas maiores produzidas por elas, como as proteínas. Essa tarefa, porém, não é simples. Para sequenciar os genes humanos, por exemplo, foi necessário criar o Projeto Genoma Humano, financiado pelo Departamento de Energia dos Estados Unidos e pelo Instituto Nacional de Saúde dos Estados Unidos. Mesmo contando com colaboração de laboratórios internacionais, o projeto levou aproximadamente 17 anos para ser concluído [19].

Atualmente, pesquisadores já podem sequenciar cadeias de DNA, RNA e proteína [19] de vários organismos mais rapidamente. Esses dados, denominados dados ômicos, geram uma quantidade de informação tão extensa e complexa que apenas ferramentas de Big Data podem ser usadas para análise, visualização, busca, etc, para um tratamento eficiente [8]. Existem grandes áreas da Biologia Molecular voltadas para estudo desse dados, tais como genoma (conjunto de genes), proteoma (conjunto de proteínas) e metaboloma (conjunto de metabólitos) [8].

Existe também um campo da Biologia Molecular específico para visualização de informação, chamado interactoma, uma vez que os pesquisadores precisam de uma ferramenta para analisar os dados ômicos. Essa visualização pode ser disponível de maneira dinâmica, permitindo que o pesquisador manipule com os dados de maneira interativa.

Na área de redes metabólicas, existem diversas ferramentas de visualização das vias, cada uma provendo sua própria perspectiva dos dados para o pesquisador. Neste trabalho serão apresentados o estado da arte de quatro ferramentas de visualização de vias metabólicas: *Reactome Pathway Browser*, *KEGG Pathway*, *ByoCyc* e *2Path*. Devido à complexidade dos dados, o objetivo dessas ferramentas é apresentar um conteúdo de maneira confortável ao usuário, de maneira que ele não fique sobrecarregado com a quantidade de informação inerentemente extensa dos dados ômicos.

Em computação existe uma área voltada para o estudo de interação entre usuários e sistemas computacionais, chamado Interação Humano-Computador (IHC). Existem diversas técnicas e métodos que especificam a maneira com que uma interface deve ser estruturada para aumentar o nível de qualidade de um sistema [7]. Além disso, existem também vários métodos de avaliação que verificam se um sistema já pronto ou ainda

em produção satisfaz os critérios de qualidade [7] desejados (usabilidade, experiência de usuário, acessibilidade e comunicabilidade).

Neste trabalho será formulizada e desenvolvida a ferramenta de visualização de vias metabólicas *2Path* seguindo os padrões de projeto de interface do campo de IHC. Para verificar se o sistema satisfaz os critérios básicos de qualidades, este projeto utiliza um método chamado Método de Avaliação de Comunicabilidade, que fornece uma sequência de atividades a serem aplicadas para um grupo de pesquisadores biólogos e, ao final, coleta dados relativos à reação dos mesmos ao navegar no sistema.

1.1 Justificativa

Atualmente, a quantidade de dados ômicos estudados pelos pesquisadores é extensa e complexa. Uma maneira de amenizar o esforço feito para analisá-los e compreendê-los é oferecer uma ferramenta que aproxime o usuário (pesquisador) e os dados e a maneira mais natural é representar tais dados em forma de grafo (redes metabólicas). Esta ferramenta deverá permitir que o usuário facilmente visualize e interaja com os dados dinamicamente.

1.2 Problema

Não há uma visualização interativa de redes metabólicas armazenadas no *2Path* que permita ao pesquisador explorar os aspectos biológicos do organismo estudado, pelos padrões de interface sugeridos pela área de Interface Humano-Computador.

1.3 Objetivo

Construir uma interface que acesse redes metabólicas armazenadas no *2Path* e gere uma visualização interativa. Além disso, verificar se o sistema satisfaz os critérios de qualidade de acordo com as especificações de Interação Humano-Computador.

1.4 Descrição dos Capítulos

No Capítulo 2 serão apresentados os conceitos básicos de Interação Humano-Computador e alguns métodos de avaliação da qualidade de sistemas computacionais, com foco no Método de Avaliação de Comunicabilidade utilizado neste trabalho.

No Capítulo 3 serão descritos os conceitos básicos de Biologia Molecular, metabolismo primário e, principalmente, metabolismo secundário. Serão descritas também as ferramentas de visualização de redes metabólicas mais conhecidas, que são o *Reactome Pathway Browser*, *KEGG Pathway* e *BioCyc*, bem como a ferramenta implementada neste trabalho, o *2Path*.

O Capítulo 4 aborda o método de elaboração da interface auto-explicativa e consistente do sistema desde sua concepção, incluindo certos detalhes sobre a implementação.

Já o Capítulo 5 apresenta os resultados da aplicação do Método de Avaliação de Comunicabilidade definido no Capítulo 2 sobre o *2Path*.

O Capítulo 6 conclui o projeto e oferece sugestões para trabalhos futuros.

Por fim, o Capítulo 7 descreve as atividades realizadas neste trabalho através de um cronograma.

Capítulo 2

Avaliação em Interação Humano-Computador

Atualmente a tecnologia está presente mais do que nunca em grande parte das atividades das pessoas, modernizando as casas, ruas, escolas, diferentes ambientes de trabalho, integrando-se na vida pessoal de cada um, se difundindo cada vez mais rápido. Essas tecnologias são cada uma um sistema computacional distinto que interage a sua maneira com os usuário. Elas podem interferir direta e/ou indiretamente nos processos de comunicação e informação dos indivíduos [12].

Para satisfazer certos critérios de qualidade, estes sistemas devem possuir algumas características que facilitem seu uso para seus usuários específicos. Ainda, vários métodos de avaliação podem ser aplicados sobre eles, cada um com um foco diferente a ser analisado [7]. Em Ciência da Computação, a área de Interação Humano-Computador (IHC) é responsável por verificar a qualidade destes sistema mensurando certas propriedades e avaliar o impacto dos mesmos na vida dos usuários.

Na Seção 2.1 serão descritos os conceitos básicos de IHC bem como os critérios de qualidade avaliados nos sistemas computacionais. Na Seção 2.2 será apresentado um modelo de avaliação de sistemas computacionais, chamado Avaliação Através de Observação. Um método que utiliza este modelo, chamado Método de Avaliação de Comunicabilidade, será descrito detalhadamente também nessa seção e será utilizado neste projeto.

2.1 Conceitos Básicos de IHC

Esta seção apresenta os componentes básicos utilizados na matéria interdisciplinar Interação Humano-Computador, envolvidos na interação entre usuários e algum sistema computacional. A **interação** engloba todo o contato realizado pelo usuário com o objetivo de exercer uma tarefa através do sistema. Segundo John Kammersgaard, 1988, existem quatro perspectivas de interação usuário-sistema [7]:

- 1 **Sistema:** Usuário conhece linguagem específica voltada para o sistema e seu objetivo é a transmissão correta dos dados da maneira mais rápida possível;
- 2 **Parceiro de discurso:** Usuário interage com o sistema através de uma inteligência artificial que personifica um interlocutor especialista naquilo que usuário procura;

- 3 **Mídia:** Usuários interagem entre si por meio do sistema, que pode oferecer diversos recursos como jogos e portal de notícias, porém é focado na qualidade da comunicação entre pessoas;
- 4 **Ferramenta:** Usuário utiliza o sistema de maneira automática como instrumento de propósito geral para realizar suas tarefas.

Em um sistema interativo, a **interface** é responsável por manter o contato motor (como a *webcam* e o teclado), perceptivo (como o monitor) e conceitual (interpretação do contato físico) do usuário [7].

O conjunto dos elementos da interface que expõem seu funcionamento de maneira implícita é chamado de ***affordance*** [7]. Esta palavra não possui tradução para o português, mas entende-se que significa "reconhecimento", pois uma vez que o usuário é levado a realizar os passos corretos para cumprir seus objetivos em uma interface, mesmo sem nunca ter interagido com ela antes, considera-se que ele já conhecia algumas características que o sistema oferece e obteve respostas já esperadas. Nesse sentido, a *affordance* é geralmente uma característica muito desejável nos sistemas. Na Figura 2.1 é possível perceber três graus diferentes de *affordance* na 2.1.

The figure shows three examples of text input fields for the name field, arranged vertically. Each example starts with the label 'Nome:'. The first example has a simple rectangular box. The second example has a rectangular box with a thin border. The third example has a rectangular box with a thin border and the placeholder text 'type your name here' inside.

Figura 2.1: Exemplo de tipos de caixa de texto que representam o mesmo campo (Nome) porém com níveis de *affordance* diferentes, sendo que o último apresenta melhor o objetivo da caixa e, portanto, ela é reconhecida mais facilmente. Adaptado de [17].

Em IHC, a qualidade de um certo sistema está fortemente relacionada à sua interface e interação [7], pois para que os usuários aproveitem o sistema por completo, eles devem estar confortáveis com o ambiente. As características da interação e interface que quantificam a qualidade de um sistema são chamadas de critérios de qualidade, apresentados a seguir:

- 1 **Usabilidade:** Medida de complexidade no aprendizado do uso da interface para atingir objetivos de maneira eficiente (no menor tempo possível, utilizando o menor número de recursos do sistema) e eficaz (de modo que o usuário execute as tarefas o mais automático possível, sem apresentar dúvidas a respeito da interface);
- 2 **Experiência do usuário:** Medida de satisfação do usuário em relação ao sistema;
- 3 **Acessibilidade:** Medida da flexibilidade do sistema, ou seja, da capacidade de usuários interagirem com o mesmo;
- 4 **Comunicabilidade:** Medida da capacidade de transmissão das intenções do projetista do sistema para o usuário por meio da interface.

Ao avaliar a qualidade de uso de um sistema, é necessário escolher um tipo de avaliação apropriado para o objeto estudado e para o critério de qualidade em foco, se existir. A avaliação através de inspeção, por exemplo, tenta conhecer as possíveis respostas do usuário em certas situações no sistema ainda em produção. Essa avaliação possui três métodos distintos, cada um com um objetivo diferente. O método pode ser heurístico (focada em problemas de usabilidade durante um processo iterativo), cognitivo (focada em avaliar a facilidade de aprendizado de um sistema interativo) ou inspeção semiótica (focada em avaliar a quantidade de informação transmitida do projetista do sistema para o usuário).

Outro tipo de avaliação bastante utilizado é o de avaliação por observação, o qual será utilizado neste trabalho. Esse tipo de avaliação considera problemas reais que os usuários (ou participantes com o mesmo perfil dos usuários) enfrenta ao utilizar o sistema. Existem três métodos de avaliação baseados em avaliação por observação: teste de usabilidade (focado em avaliar a usabilidade através de registros de performance e opiniões dos usuários), método de avaliação de comunicabilidade (focado em avaliar a quantidade de informação que chega ao usuário através do sistema) e prototipação em papel (focado em avaliar a usabilidade através de simulações do sistema em papel).

2.2 Avaliação Através de Observação

Para manter os critérios básicos de qualidades, é importante que exista algum tipo de avaliação do sistema antes de sua entrega ao(s) usuário(s). Na perspectiva de quem desenvolve o sistema, a avaliação deve verificar se o sistema recebe as entradas, processa os dados e gera o resultado na saída corretamente. Por outro lado, na perspectiva do usuário, a avaliação deve verificar como o comportamento da interface afeta a experiência de uso do sistema [7], recaindo nos quatro critérios de qualidade descritos na Seção 2.1. A etapa de avaliação pode ocorrer durante (**avaliação formativa**) ou após (**avaliação somativa**) o processo de desenvolvimento do sistema. Na primeira opção, o foco é a identificação de possíveis problemas, com a vantagem do baixo custo de correção. Na segunda opção, o foco é a verificação dos níveis de qualidade do protótipo de escopo definido [7].

Os tópicos mais avaliados em IHC são os problemas na interação e na interface do sistema, que são classificados de acordo com a frequência com que ocorrem, com sua gravidade ou com os quatro critérios de qualidade. Um exemplo de questionário associado a esse tópico de avaliação encontra-se na parte de consolidação dos resultados, como será descrito a seguir nessa seção.

A fundamentação dos métodos de avaliação é dada pela teoria da Engenharia Semiótica, focada em dois tipos de comunicações: usuário-sistema e projetista-usuário através do sistema. Esse último tipo de comunicação recebe o nome de **metacomunicação**, uma vez que é feita indiretamente através da interface criada pelo projetista para o usuário. Segundo esta teoria, toda aplicação computacional é um artefato de metacomunicação por onde o projetista, como interlocutor, comunica-se com o usuário.

A **metamensagem** é a mensagem transmitida via metacomunicação. Em IHC, a metamensagem possui um padrão fixo e generalizado, que pode ser utilizado para facilitar o entendimento dos problemas do usuário, o que pode ser feito para solucioná-los e como

fazê-lo. O texto a seguir representa esta mensagem e os campos entre os símbolos “<” e “>” representam um espaço a ser ocupado de acordo com o sistema desenvolvido:

Este é o meu entendimento, como <projetista do sistema>, de quem você, <usuário específico / grupo de usuários>, é, do que aprendi que você quer ou precisa <fazer / executar / visualizar / se comunicar>, de que maneiras <prefere / é mais natural> fazer, e <por quê / por quê não outras>. O <nome do sistema>, portanto, é o sistema que projetei para você, e <esta navegação específica> é a forma como você pode ou deve utilizá-lo para alcançar uma gama de objetivos que se encaixam nesta visão.

O projetista de um sistema cria signos para se comunicar com o usuário através da interface. Estes signos podem ser estáticos (como rótulos, imagens, botões, cores), dinâmicos (como ações, transições de tela, notificações) ou metaliguísticos (informações referentes à outros signos). A Engenharia Semiótica também estuda o processo de significação dos signos, denominado semiose. A semiose abrange tudo o que é percebido pelos seres humanos e produz sensação, na mente dos intérpretes. Esse é um processo que transforma os fenômenos perceptíveis através dos sentidos em experiências dos indivíduos [13].

Em IHC, a semiose é utilizada para compreender as ações tomadas, bem como as reações de resposta, pelos usuários que interagem com sistemas cobertos de signos. Assim, é possível elaborar um perfil semiótico do sistema para retratar os problemas de comunicabilidade, e, para isso, a metamensagem pode ser utilizada como base.

Existem dois métodos de avaliação baseados em Engenharia Semiótica. Quando o objetivo é avaliar a quantidade de emissão de metacomunicação, utiliza-se o Método de Inspeção Semiótica [7]. Quando é avaliar a quantidade de recepção de metacomunicação, utiliza-se o Método de Avaliação de Comunicabilidade [7]. O primeiro método, porém, tem como objetivo antever possíveis problemas causados por decisões do projetista, logo não envolve obstáculos reais dos usuários. Já o segundo método visa compreender as dificuldades enfrentadas pelos usuários enquanto navegam de fato pelo sistema. Nesse sentido, o Método de Avaliação de Comunicabilidade será utilizado neste trabalho.

2.2.1 Método de Avaliação de Comunicabilidade

O Método de Avaliação de Comunicabilidade (MAC) é um método qualitativo de avaliação somativa, cujo objeto são os problemas na interação e interface do sistema, com foco na percepção do usuário. Ele consiste em uma análise de gravações em vídeo de pessoas utilizando o sistema com base em 13 etiquetas (expressões linguísticas que caracterizam ruptura na comunicação) e é composto por cinco etapas[7]: preparação, coleta de dados, interpretação, consolidação dos resultados e relato dos resultados. As pessoas filmadas não são necessariamente os usuários do sistema, elas podem ser representantes de usuários que possuem o mesmo perfil daqueles para qual o sistema é direcionado.

2.2.1.1 Preparação

Nesta etapa, o avaliador é responsável por definir o perfil dos usuários, selecionar os representantes dos mesmos e elaborar o ambiente e tarefa realizada por eles. O material

da gravação é preparado e verificado fazendo-se um teste piloto. Um termo de consentimento da avaliação deve ser redigido para assinarem o avaliador e cada representante. Devem ser elaborados dois questionários: um pré-teste, para coletar informações de cada representante tais como conhecimento sobre o domínio do sistema e um pós-teste, para coletar informações referentes à opinião dos participantes sobre suas experiências com o sistema.

2.2.1.2 Coleta de Dados

Aqui o avaliador deve receber os representantes de usuários, explicar o procedimento da tarefa realizada, entregar uma via do termo de consentimento para cada representante assinar, já com a assinatura do avaliador. Os questionários pré-teste são entregues e cabe ao avaliador decidir se é necessário ler em voz alta as perguntas, caso alguma possa conter mais de uma interpretação. Não é especificado tempo mínimo nem máximo para esta tarefa. Ao final do questionário, o avaliador posiciona os equipamentos de gravação perto de cada participante, de maneira que a imagem gravada seja a tela, o *mouse/mousepad*, o teclado e as mãos dos mesmos, e o som captado seja apenas a voz dos mesmos.

Uma técnica muito conhecida de coleta de dados na gravação é chamada *think aloud*, onde os participantes devem relatar em voz alta tudo aquilo que estão pensando em relação ao sistema, como execução, planejamento de execução, reação às respostas da interface, etc. O resultado final facilita bastante a análise do avaliador, porém o ato de falar enquanto desempenha a tarefa pode descentralizar o pensamento dos participantes. Uma vez que estes sintam-se seguros em realizar as tarefas ao mesmo tempo em que se expressam verbalmente, os avaliadores terão muito mais dados (etiquetas) para interpretar na fase seguinte.

Terminada a gravação, ocorre uma entrevista para coletar a opinião de todos os participantes em relação ao sistema, bem como tirar dúvidas sobre seus desempenhos.

2.2.1.3 Interpretação

Nesta etapa o avaliador assiste todos os vídeos contabilizando certas expressões linguísticas chamadas de etiquetas apresentadas na Tabela 2.1. Deve-se levar em consideração os perfis dos usuários traçados pelo questionário pré-teste, bem como as entrevistas pós-teste, uma vez que as experiências passadas de cada um pode afetar seu comportamento no sistema.

As expressões linguísticas possuem na maioria conotação ruim com respeito ao sistema, tais como indagação sobre seu comportamento e aversão às respostas. Observando os vídeos, o avaliador pode perceber as barreiras de comunicabilidade e gargalos do sistema do ponto de vista do usuário, chamados de **falha de comunicabilidade**. Quando o número de participantes é muito baixo, o avaliador pode fazer a interpretação ao mesmo tempo em que coleta dados, sem precisar de gravação.

Após a análise dos dados, o avaliador deve redefinir um protótipo da próxima versão do sistema já com a solução para os problemas mais simples e listar todos os obstáculos na interface que podem ser corrigidos.

Tabela 2.1: Descrição das etiquetas, de acordo com a falha de comunicação que representam [7].

Falhas de comunicação completas: efeito obtido é inconsistente com a intenção comunicativa do usuário		
aspecto semiótico	característica específica	etiqueta
O usuário termina uma semiose malsucedida, mas não inicia outra para obter o resultado esperado	porque, mesmo percebendo que não obteve o resultado esperado, não possui mais recursos, capacidade ou vontade de continuar tentando	Desisto
	porque não percebe que não obteve o resultado esperado	Para mim está bom...
Falhas de comunicação parciais: o efeito obtido é somente parte do efeito pretendido de acordo com a intenção do usuário		
aspecto semiótico	característica específica	etiqueta
O usuário abandona uma semiose antes de obter o resultado esperado, e inicia outra com o mesmo propósito	porque, embora entenda a solução de IHC proposta, prefere seguir por outro caminho no momento	Não, obrigado
	porque não entende a solução de IHC proposta	Vai de outro jeito
Falhas de comunicação temporárias: o efeito parcial do processo de interpretação (semiose) e de comunicação (interação) do usuário é inconsistente e incoerente com sua intenção de comunicação		
aspecto semiótico	característica específica	etiqueta
O usuário interrompe temporariamente sua semiose	porque não encontra uma expressão apropriada para sua intenção de comunicação	Cadê?
	porque não percebe ou não entende a expressão do sistema (preposto do projetista)	Ué, o que houve?
	porque não consegue formular sua próxima intenção de comunicação	E agora?
O usuário percebe que seu ato comunicativo não foi bem-sucedido	porque percebeu que havia "falado" algo no contexto errado	Onde estou?
	porque percebeu que havia "falado" algo errado	Epa!
	porque não obteve o resultado esperado depois de conversar com o sistema (preposto do projetista) por algum tempo, alternando vários turnos de fala com ele	Assim não dá.
O usuário procura compreender o ato comunicativo do sistema (preposto do projetista)	através da metacomunicação implícita	O que é isto?
	através da metacomunicação explícita	Socorro!
	testando várias hipóteses sobre o significado do que o sistema comunicou	Por que não funciona?

2.2.1.4 Consolidação dos Resultados

Este é o momento em que o avaliador busca diferenciar as características do grupo das características individuais observando a recorrência das etiquetas em certos pontos da navegação no sistema. Pode-se atribuir significado às etiquetas de acordo com frequência de uso em um certo contexto, com sequência de uso e com o nível de problemas, por exemplo. Além disso, como o foco desta avaliação é verificar os problemas relacionados à interação e interface, nesta etapa devem ser respondidas as seguintes questões:

- O usuário consegue operar o sistema?
- Ele atinge seu objetivo? Com quanta eficiência? Em quanto tempo? Após cometer quantos erros?
- Que parte da interface e da interação o deixa insatisfeito?
- Que parte da interface o desmotiva a explorar novas funcionalidades?
- Ele entende o que significa e para que serve cada elemento da interface?
- Ele vai entender o que deve fazer em seguida?
- Que problemas de IHC dificultam ou impedem o usuário de alcançar seus objetivos?
- Onde esses problemas se manifestam? Com que frequência tendem a ocorrer? Qual é a gravidade desses problemas?
- Quais barreiras o usuário encontra para atingir seus objetivos?
- Ele tem acesso a todas as informações oferecidas pelo sistema?

Respondidas as perguntas, o avaliador pode criar então um perfil semiótico para o sistema com base na metamensagem a seguir:

Este é o meu entendimento, como projetista, de **quem você, usuário, é** (1), do que aprendi que você **quer ou precisa fazer** (2), de **que maneiras prefere fazer** (3), e **por quê** (4). Este, portanto, é o sistema que projetei para você, e esta é **a forma como você pode ou deve utilizá-lo** (5) para alcançar uma gama de objetivos que se encaixam nesta visão.

Cada item numerado na meta mensagem deve fornecer as seguintes informações buscadas:

- (1) Qual é o perfil dos usuários?
- (2) Quais são seus desejos e o que a metacomunicação realiza para satisfazê-los?
- (3) Quais as maneiras de realizar seus desejos e de que maneira eles preferem fazer?
- (4) O que os leva a ter esta preferência?
- (5) Quão bem o conteúdo da metacomunicação é transmitidos aos usuários?

2.2.1.5 Relato dos Resultados

Ao final, o avaliador deverá possuir material suficiente para apresentar aos *stakeholders* responsáveis pelo desenvolvimento do sistema os seguintes resultados:

- Os objetos da avaliação, tais como interface do sistema e perfil dos usuários;
- Breve descrição do Método de Avaliação de Comunicabilidade;
- Quantidade e perfil dos avaliadores, bem como dos usuários ou dos participantes que os representam;
- Descrição detalhada de todas as tarefas executadas pelos participantes;
- Resultado das etiquetagens, em geral contabilizando as etiquetas por usuário e tarefa;
- Lista de problemas de comunicabilidade encontrados;
- Sugestões de melhoria para cada falha de comunicação;
- Perfil semiótico do sistema, de acordo com a metamensagem padrão.

Capítulo 3

Redes Metabólicas

Neste capítulo serão descritos conceitos básicos da Biologia Molecular, em particular de metabolismo nos organismos além de bancos de dados específicos para redes metabólicas. A Seção 3.1 detalha as principais envolvidas no metabolismo, tais como DNA e enzima. A Seção 3.2 descreve como ocorre o processo do metabolismo. Por fim, a Seção 3.3 apresenta os principais bancos de dados voltados para redes metabólicas.

3.1 Conceitos Básicos de Biologia Molecular

Nesta seção, inicialmente descrevemos os ácidos nucleicos (ADN e ARN), e em seguida proteínas e o Dogma Central.

3.1.1 Ácidos Nucleicos

Os ácidos nucleicos são biomoléculas responsáveis pelo armazenamento, transmissão e tradução das informações genéticas dos seres vivos. Isto é possível devido ao processo de síntese de proteínas que constitui a base da herança biológica. Os ácidos nucleicos são polímeros, macromoléculas formadas por estruturas menores chamadas monômeros¹, que nesse caso são nucleotídeos.

Nucleotídeos são compostos de três elementos: um radical fosfato², uma pentose (um monossacarídeo³ formado por cinco átomos de carbono), e uma base nitrogenada. Existem cinco tipos de bases nitrogenadas que podem compor um nucleotídeo: Adenina (A), Timina (T), Citosina (C), Guanina (G) e Uracila (U).

Na Figura 3.1, observa-se que no nucleotídeo existe uma numeração de 1' à 5', que representam os carbonos presentes na pentose. Para a criação de uma fita de ácido nucleico, existe uma ligação entre o carbono da posição 5' de um nucleotídeo e o carbono de posição 3' de outro [21]. Por definição o sentido da leitura de uma fita de ácido nucleico é 5' → 3', o que deve ser levado em consideração ao se fazer interpretação de dados do material genético.

Dois tipos de ácidos nucleicos são encontrados nos seres vivos: ácido desoxirribonucleico (ADN) e ácido ribonucleico (ARN). Eles diferenciam-se na estrutura das bases

¹Pequenas moléculas que ligam-se entre si formando moléculas maiores.

²Grupo que confere características ácidas às moléculas.

³Unidade básica dos carboidratos.

nitrogenadas e em suas funções. O ADN é formado por uma desoxirribose, indicada na figura 3.2. Cada nucleotídeo de uma fita liga-se a partir de suas bases nitrogenadas com um nucleotídeo de outra fita, formando assim um eixo helicoidal tridimensional chamada de dupla hélice [21]. Esta ligação é feita entre grupos de purinas⁴, Adenina (A) e Guanina (G), e pirimidinas⁵, Timina (T) e Citosina (C). Já o ARN possui apenas uma fita, onde a Timina (T) é substituída pela Uracila (U).

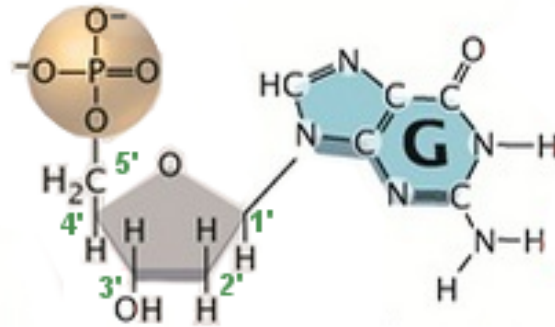


Figura 3.1: Nucleotídeo com grupo fosfato (P), pentose abaixo à esquerda e base nitrogenada G (no caso a Guanina). Adaptado de [20].

O ADN é uma biomolécula que armazena as informações referentes ao funcionamento de todas as células dos seres vivos. Uma fita de ADN pode conter centenas de milhões de nucleotídeos. A representação do ADN, seja nos livros ou computacionalmente, é dada por um par em paralelo de strings de letras A, T, G e C. Como explicado no início dessa seção, o sentido padrão da leitura de uma fita é de $5' \rightarrow 3'$, e no caso do ADN, as hélices são dispostas de maneira antiparalela, ou seja, uma é lida de $5' \rightarrow 3'$ e a outra, de $3' \rightarrow 5'$. Observa-se que a partir de uma hélice, pode-se inferir a sequência de sua hélice complementar. Por exemplo, seja uma hélice H1 igual a AGTAAGC; então H2 em seu sentido oposto é H2' igual a TCATTCG, e no sentido biológico, igual a GCTTACT. A Figura 3.2 apresenta a estrutura do ADN como explicada nesta seção.

O ARN é uma biomolécula que possui diversas funções nas células dos organismos. Existem três tipos de ARNs presentes no citoplasma⁶. Cada um possui funções específicas que serão detalhadas na Seção 3.1.3. De maneira resumida, o ARN mensageiro (mARN) é responsável pela transferência de informação do ADN. Em seguida o ARN ribossômico (rARN) será combinado com o ARN transportador (tARN) para realizar a síntese de proteína.

3.1.2 Proteínas

As proteínas são biomoléculas com diversas funcionalidades nas células dos seres vivos. As proteínas fibrosas, como o colágeno, compõem a estrutura do corpo e para isso precisam ser resistentes e insolúveis em água. As proteínas globulares, como a hemoglobina, possuem formato esférico e são compactas, o que as permite realizar processos dinâmico pelo corpo

⁴Composto orgânico que possui um anel duplo de carbono.

⁵Composto orgânico que possui um anel simples de carbono.

⁶Espaço entre a membrana plasmática e o núcleo da célula.

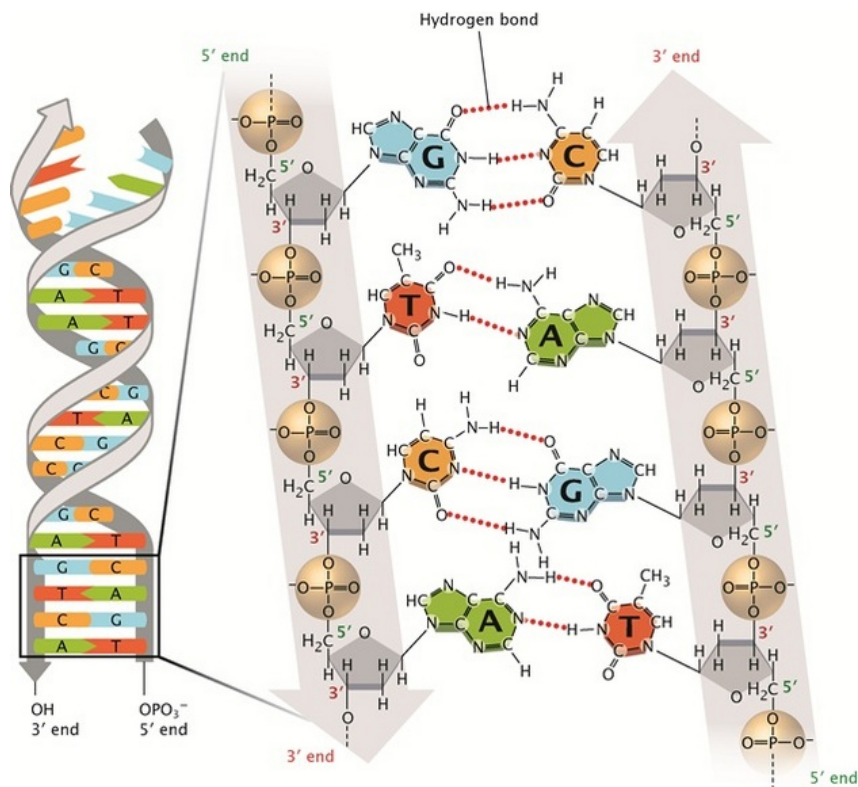


Figura 3.2: Representação das duas fitas de ADN, que se ligam por pontes de hidrogênio entre os nucleotídeos. Adaptado de [20].

[16]. Cada tarefa é realizada por uma proteína com uma estrutura específica e adaptada pra tal.

Assim como os ácidos nucleicos, as proteínas são polímeros, macromoléculas cujos monômeros são aminoácidos. Aminoácidos são moléculas que possuem cinco componentes: amina (NH_2), carbono (C), hidrogênio (H), ácido carboxílico (COOH) e uma cadeia lateral que funciona como identificador de cada um dos 20 tipos de aminoácidos presentes nos seres vivos. A maneira como eles são criados será explicada com mais detalhes em seguida, pois envolve um processo complexo de síntese de proteína. A ligação, ou polimerização, de dois aminoácidos é feita unindo o grupo amina de um com o ácido carboxílico do outro, liberando uma molécula de água (H_2O) e formando uma cadeia chamada de dipeptídeo. Como houve liberação de água na ligação, o dipeptídeo não é formado por aminoácidos, mas sim resíduos dos mesmos. Nesse sentido, cadeias peptídicas de 100 à 5000 diferentes resíduos aminoácidos, ou cadeia polipeptídicas, constituem a proteína.

Existem quatro estruturas para caracterização de uma proteína [21], apresentadas na Figura 3.3. A mais simples é chamada de estrutura primária e é composta por uma sequência linear de resíduos aminoácidos. A estrutura secundária é tridimensional e estabiliza-se por meio de ligações de hidrogênio na cadeia principal, chamada de *backbone*. Dependendo da disposição dos resíduos de aminoácidos, esta cadeia pode se dar forma de hélice (α -Helix) ou em forma de folha (β -Helix). A estrutura terciária é dada pela união de várias estruturas secundárias e, por fim, a estrutura quaternária é composta de múltiplas estruturas terciárias [5].

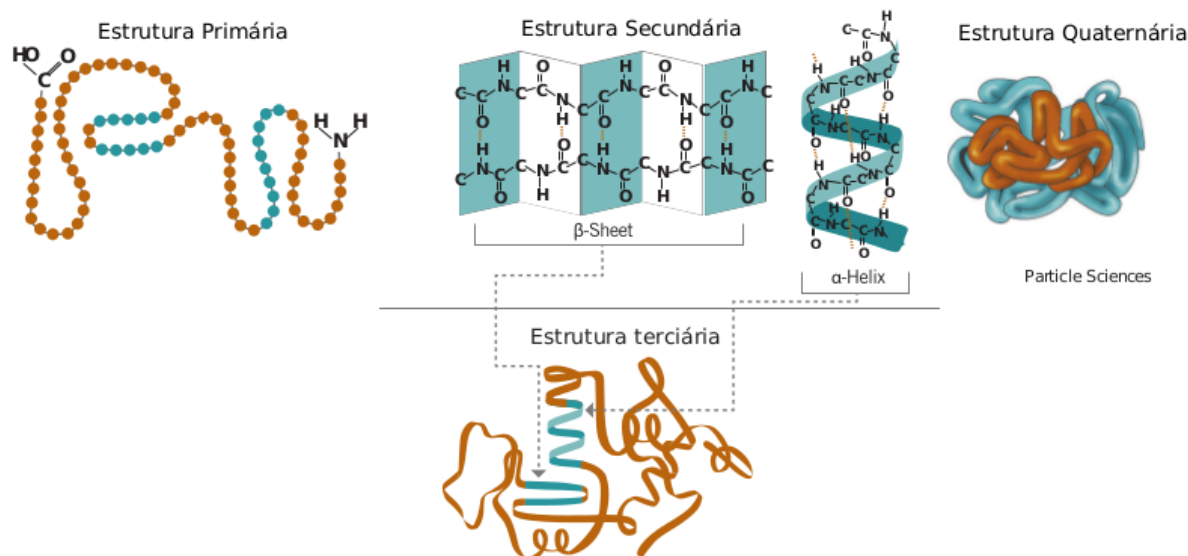


Figura 3.3: Quatro estruturas de proteínas. As unidades esféricas na estrutura primária representa os aminoácidos da proteína. Adaptado de [5].

3.1.3 Síntese de Proteína

A **transcrição** é o processo de produção de mARN a partir do ADN e ele ocorre da seguinte forma: O início de cada gene possui um identificador em uma das fitas para indicar o local da codificação e, a partir dali, uma cópia inversa (A, T, C, G são traduzidos para U, A, G, C respectivamente) do mesmo é feita sob forma de molécula de mARN que, por consequência, obterá a mesma sequência que a cadeia codificadora (a qual não possui o identificador), porém trocando o U por T.

O mARN deixa, então, o núcleo celular e inicia a **tradução** no citoplasma. O processo ocorre no interior de uma organela celular chamada de ribossomo, constituído de proteínas e rARN e cuja função é construir a molécula de proteína a partir de duas entradas, o mARN e tARN. A estrutura do tARN é tal que de um lado se encaixa exatamente um códon⁷ e no oposto, seu aminoácido correspondente, conforme ilustrado na Figura 3.4. O processo de tradução se dá da seguinte forma: à medida em que o mARN passa pelo interior do ribossomo, este atrai quaisquer tARNs das proximidades cujos códons sejam correspondentes ao da subsequência corrente do mARN. No momento em que o códon do tARN conecta-se com um dos códons do mARN, a molécula de proteína em desenvolvimento é liberada e, com o auxílio da catálise de uma enzima, agregada no aminoácido que estava fixado naquele tARN. A tabela contendo a tradução de códon para aminoácido é fixa e chama-se código genético, apresentado na Tabela 3.1. A tradução é finalmente completa quando o mARN apresenta um códon de parada, pois nenhum tARN possui correspondência para tal [21]. Uma proteína simples é, então, formada.

⁷Sequência de três nucleotídeos.

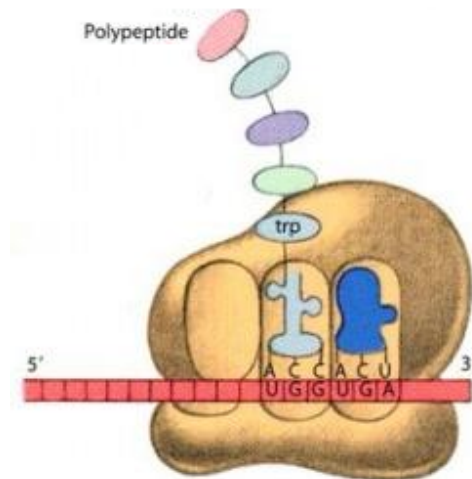


Figura 3.4: Representação da etapa final da síntese de proteína, que sempre finaliza com o códon UGA. Adaptado de [3].

Tabela 3.1: Código Genético que mapeia cada códon à um dos 20 aminoácidos, representados de maneira abreviada.

Primeira Posição	Segunda Posição				Terceira Posição
	G	A	C	U	
G	Gly	Glu	Ala	Val	G
	Gly	Glu	Ala	Val	A
	Gly	Asp	Ala	Val	C
	Gly	Asp	Ala	Val	U
A	Arg	Lys	Thr	Met	G
	Arg	Lys	Thr	Ile	A
	Ser	Asn	Thr	Ile	C
	Ser	Asn	Thr	Ile	U
C	Arg	Gln	Pro	Leu	G
	Arg	Gln	Pro	Leu	A
	Arg	His	Pro	Leu	C
	Arg	His	Pro	Leu	U
U	Trp	FIM	Ser	Leu	G
	FIM	FIM	Ser	Leu	A
	Cys	Tyr	Ser	Phe	C
	Cys	Tyr	Ser	Phe	U

3.2 Conceitos Básicos de Metabolismo

As proteínas podem ser enzimas, macromoléculas responsáveis por auxiliar a realização de biossíntese (construção) e biodegradação de moléculas no metabolismo. As enzimas têm o propósito de catalisar, ou acelerar, reações bioquímicas que naturalmente levariam muito mais tempo para serem realizadas no organismo [21]. Geralmente as enzimas catalisam uma reação específica e seus nomes correspondem à tarefa que elas executam com adição do sufixo "ase". Por exemplo, a Sintetase realiza biossíntese, a Desidrogenases realiza oxirredução (transferência de elétrons) e a Quinase transporta elementos químicos de uma molécula para outra [11].

As reações bioquímicas são alterações químicas que fornecem um ou mais produtos a partir de um ou mais substratos [15]. Esses produtos e substratos são compostos químicos chamados de metabólitos. A presença de um metabólito em um certo local de um organismo depende do tipo de célula e do tipo de compartimento em que ele se encontra dentro da célula. Normalmente as reações ocorrem em apenas uma direção, ou seja, os produtos não pode gerar os substratos de uma mesma reação [15].

Ao catalisar uma reação, uma enzima possui comportar um ou mais substratos em um local pré-determinado em formato côncavo chamado de sítio ativo. Se ela comporta apenas um substrato, a estrutura que se forma com o preenchimento do sítio ativo é um complexo enzima-substrato. Porém se ela comporta mais de um substrato, a estrutura é chamada de complexo ternário intermediário [18]. Duas enzimas ainda podem possuir a mesma atividade enzimática porém apresentar estruturas físicas diferentes. Essas são chamadas isoenzimas [18].

Dentro das células existem também pequenas moléculas que regulam (aumentam ou diminuem) as atividades enzimáticas, chamados cofatores. Eles são componentes químicos não-proteicos e podem ser orgânicos ou inorgânicos [15]. Cofatores podem ser coenzimas, associadas momentaneamente às enzimas, ou grupos prostéticos, associados firmemente à elas [18].

Uma sequência de reações bioquímicas é chamada de via metabólica. As vias metabólicas podem ser anabólicas (realizam síntese de moléculas complexas gastando energia) ou catabólicas (realizam a quebra de moléculas complexas produzindo energia) [15]. Geralmente a energia liberada pelas reações catabólicas é usada para impulsionar as reações anabólicas [9]. A Figura 3.5 apresenta um exemplo de via metabólica, a bioluminescência de uma bactéria. Ela possui oito reações, quatro produtos intermediários e um produto final, que é um photon (luz).

O conjunto de todas as vias metabólicas de um determinado organismo é chamado de rede metabólica. Pesquisadores de Biologia Molecular podem analisar tanto as vias metabólicas separadamente quanto em conjunto, avaliando as interações entre elas dentro de uma rede metabólica [15]. Uma das vantagens de se estudar as vias em conjunto é poder explorar vias alternativas para um mesmo fim biológico [15].

3.2.1 Metabolismo Primário

O metabolismo, além de ser subdividido em catabolismo e anabolismo, também pode ser classificado em relação à função que exerce no organismo. Se a função realizada é fundamental no organismo, como crescimento, desenvolvimento e reprodução, ele é

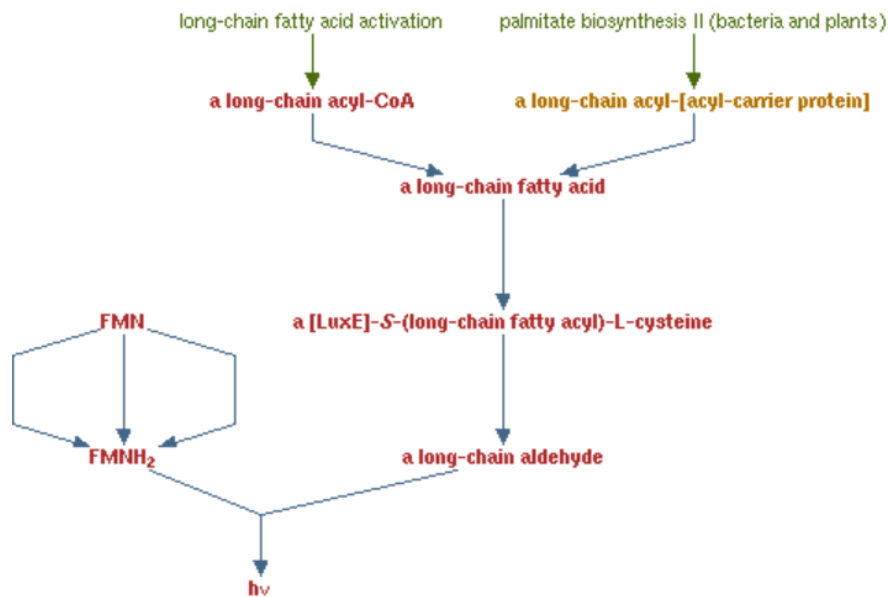


Figura 3.5: Exemplo de via metabólica retirado do site MetaCyC. Esta representa o processo de produção da bioluminescência de bactérias [10].

denominado metabolismo primário [22]. São exemplos de metabolismos primários os processos de divisão celular mitose e meiose, a primeira para crescimento e a segunda para reprodução, conforme indicado na Figura 3.6.

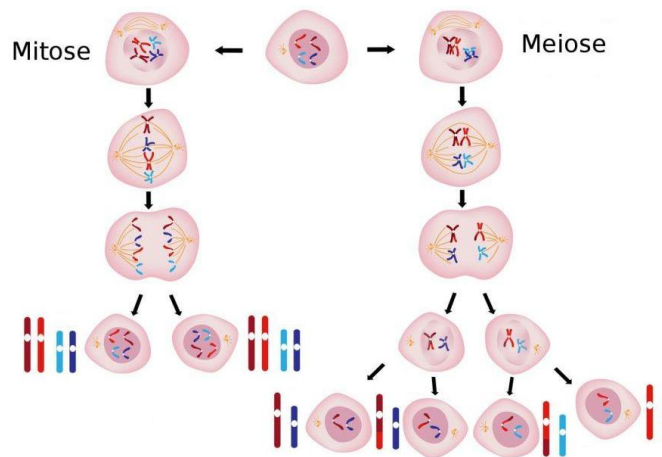


Figura 3.6: A célula que se divide por meio de mitose gera duas outras células idênticas à original. A célula que se divide por meio de meiose gera quatro células, cada uma com metade do material genético da célula original. Adaptado de [6].

3.2.2 Metabolismo Secundário

No caso dos metabolismos que não realizam função essencial no organismo, eles são classificados como metabolismos secundários. Esses são caracterizados pela vasta diversidade química e, desta forma, são responsáveis pela sobrevivência do organismo em

diferentes meio ambientes de acordo com os fatores bióticos (elementos causados pela interação entre organismos como, por exemplo, cadeia alimentar) e abióticos (elementos naturais independente de organismos como, por exemplo, luz e temperatura) [22].

Enquanto 20% dos metabólitos secundário são encontrados em bactérias, fungos e organismos sésseis⁸ marinhos, os outros 80% encontram-se em plantas vasculares⁹ [22] e esses podem ser subdivididos em três classes: terpenóides, alcaloides e fenólicos [14]. A seguir está uma breve explicação de cada uma dessas classe, acompanhada da Figura 3.7 que apresenta um exemplo de composto químico para cada grupo.

- Os **terpenóides** constituem no grupo mais abundante de produtos naturais, apresentando uma grande variedade estrutural e funcional, principalmente no Reino *Plantae*. No metabolismo secundário, possuem funções como produção de óleos, esteroides, cera, resinas e borracha natural, produção de compostos usados para defesa contra herbívoros ou aromas usados para atrair polinizadores [22];
- Os **alcaloides** são majoritariamente tóxicos a outros organismos diferentes daquele que os produz. Nesse sentido, eles possuem nas plantas função de defesa contra herbívoros e podem ser encontrados principalmente nos locais mais propícios à ataques, por exemplo, nas sementes, flores e tecidos periféricos em crescimento. Para o consumo dos seres humanos, são usados na fabricação de estimulantes, como cafeína e nicotina, e drogas, como morfina [22]. Por apresentarem alta diversidade estrutural, é difícil classificá-los. A tentativa mais recente baseia-se na semelhanças entre os esqueletos carbônicos [14];
- Os **fenólicos** são caracterizados por suas propriedades anti-oxidantes, anti-inflamatórias e anti-cancerígenas e muitos deles são bactericidas, antissépticos e vermífugos. Eles estão presentes em praticamente todas as plantas e são utilizados na Química, Biologia, Agricultura e Medicina [14].

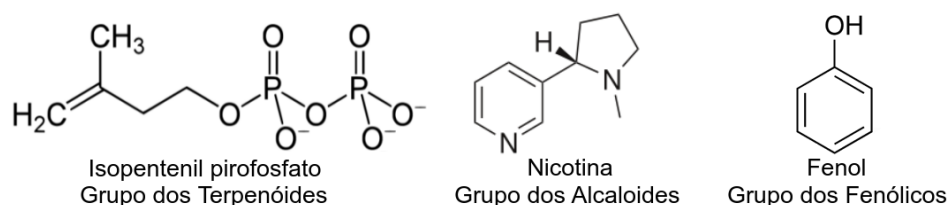


Figura 3.7: Um exemplo de elemento químico para cada classe, ou grupo, de metabolismo secundário.

3.3 Visualização de Redes Metabólicas

Desde o descobrimento da estrutura do ADN por Crick e Watson, o número de sequências de proteínas descobertas cresceu, aumentando também a necessidade de serem criados bancos de dados para armazená-las. A físico-química norte-americana Margaret Dayhoff,

⁸Que vivem fixos, sem capacidade de locomoção.

⁹Plantas com tecidos especializados em transporte de água e seiva.

com colaboração de alguns membros do *National Biomedical Research Foundation* em Washington, foi a primeira a construir um banco de dados com este propósito em um tipo de atlas de proteínas na década de 60 [19]. Somente em 1984 esta coleção foi intitulada de *Protein Information Resource* [19]. Os dados eram organizados de acordo com o grau de similaridade das sequências, onde o agrupamento das mesmas era dado em forma de árvore filogenética representando famílias e superfamílias de proteínas. Caso a semelhança fosse alta, é provável que elas teriam as mesmas funções bioquímicas e estrutura tridimensional.

A partir da árvore gerada, foi possível calcular as mutações que ocorreram nos aminoácidos durante a evolução genética e, então, produzir uma tabela utilizada até hoje, chamada PAM (*Percent Acept Mutation*), que apresenta tais dados¹⁰. Hoje em dia também utiliza-se uma outra matrix, chamada BLOSUM (*Block Substitution Matrix*), que, ao contrário da PAM que identifica as semelhanças entre sequências de proteínas, identifica a divergência evolucionária entre sequências de proteínas.

Outro banco de dados de grande porte e bastante utilizado nos dias de hoje é o *GenBank*, criado em 1982 por Walter Goad e demais colaboradores com o objetivo de catalogar sequências genéticas e coleções de anotações de todos os ADNs públicos, agora, com o patrocínio do *National Center for Biotechnology Information* (NCBI). Os dois bancos são públicos e continuam crescendo exponencialmente [19].

Nos dias de hoje, a quantidade de dados é tão grande e que os biólogos enfrentam dificuldades em tarefas como análise, busca, armazenamento, visualização e atualização de dados. Nesse sentido, eles utilizam agora ferramentas de Big Data em suas pesquisas. Existem grandes áreas da Biologia Molecular voltadas para estudo desse dados (chamados de dados ômicos), tais como, por exemplo, genoma¹¹, transcrito¹², proteoma¹³, metaboloma¹⁴ e interactoma¹⁵.

Nesta seção serão apresentados quatro grandes bancos de dados utilizados no estudo do metaboloma (Reactome, KEGG, MetaCyc e 2Path), bem como seus sistemas de visualização de dados. O foco será dado ao 2Path, sistema desenvolvido neste projeto.

3.3.1 Reactome, KEGG e BioCyc

Reactome

Reactome¹⁶ é um banco de dados de reações de mudança de estado, ou seja, além de reações bioquímicas, ele também abrange reações de ativação, de degradação e de ligação, por exemplo [4]. Ele faz uma ligação sistemática entre as proteínas de um certo organismo e as funções moleculares do mesmo, fornecendo uma base de funções que pode ser utilizada para pesquisas sobre expressão de genes ou mutações somáticas. O Reactome disponibiliza o *Pathway Browser*, uma rede geral para cada organismo, que representa os vários seus sistemas, como reprodução e metabolismo, por exemplo. Algumas sub-redes estão

¹⁰1 PAM é uma medida de tempo para representar 1 mutação para cada 100 aminoácidos.

¹¹Material genético de um organismo.

¹²Conjunto dos transcritos mARN, tARN, rARN e microARN.

¹³Conjunto de proteínas e suas variantes em um organismo.

¹⁴Conjunto de metabólitos de um organismo.

¹⁵Conjunto de interações moleculares em um organismo.

¹⁶Disponível pela web, em <http://www.reactome.org>.

conectadas (por exemplo, replicação de ADN e ciclo de célula), outra não (por exemplo, contração muscular e reprodução).

Nesta rede, cada nó representa uma via cujo número de entidade se reflete no raio do nó, e cada aresta representa a relação entre estas vias. A página ainda possui uma ferramenta de análise de dados baseada nas correspondências entre as reações na redes dos organismos comparados.

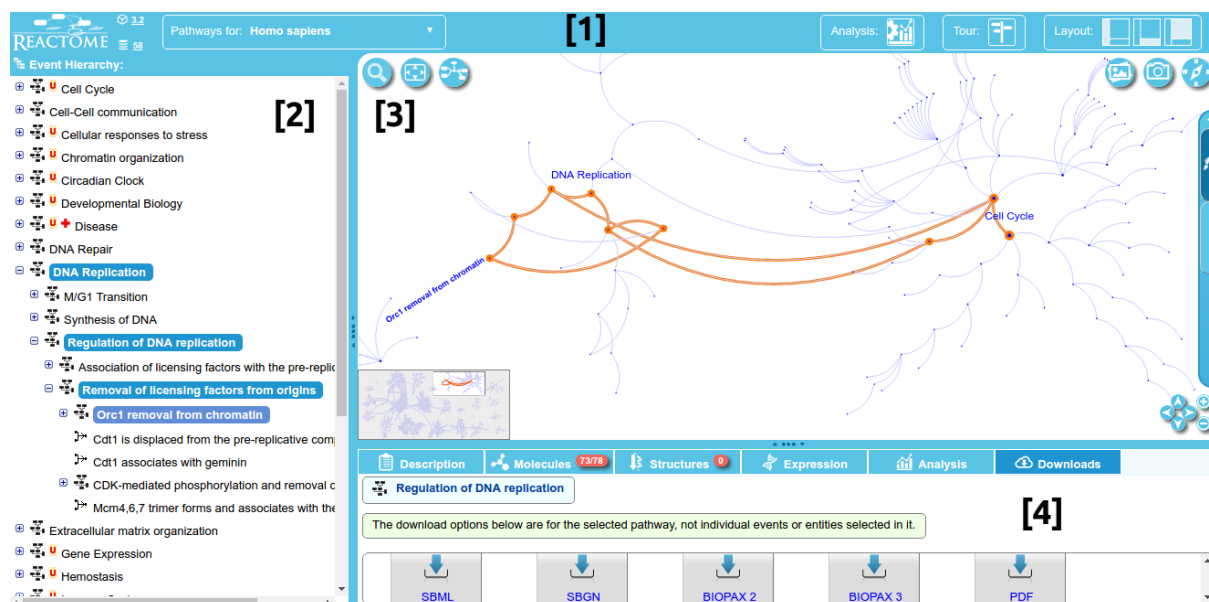


Figura 3.8: Visão geral da rede metabólica do *Homo sapiens* do Reactome, com destaque na via metabólica de remoção do Orc1 da cromatina do processo de regulação da replicação do ADN. A interface pode ser dividida nas seções [1], [2], [3] e [4].

Para acessar sua ferramenta de visualização, basta navegar a partir da *home page*:

<http://www.reactome.org/> ⇒

Browse Pathways

Caso o usuário queira visualizar a(s) via(s) metabólica detalhadamente, basta clicar nos nós do mapa, cujo tamanho está diretamente relacionado ao número de entidades que compõem a via. A imagem do grafo desaparece e então um conjunto dos elementos que o compõe aparece em seu lugar. A Figura 3.9 apresenta a via da remoção do Orc1 da cromatina do processo de regulação da replicação do ADN.

Segue abaixo a descrição de cada elemento da interface da ferramenta de visualização do Reactome, subdivididos conforme indicado na Figura 3.8.

- [1] Menu de navegação direta e indiretamente relacionado às vias metabólicas. À esquerda tem-se a versão do *Pathway Browser* e de seu banco de dados e um menu *dropdown* para se selecionar o organismo mostrado na interface. À direita tem-se um *link* (*Analysis*) para uma página modal onde é possível analisar os dados do usuário e fazer comparação entre espécies, um *link* (*Tour*) para um vídeo explicando como *Pathway Browser* funciona e um conjunto de *layouts* que alteram a visualização geral da página;

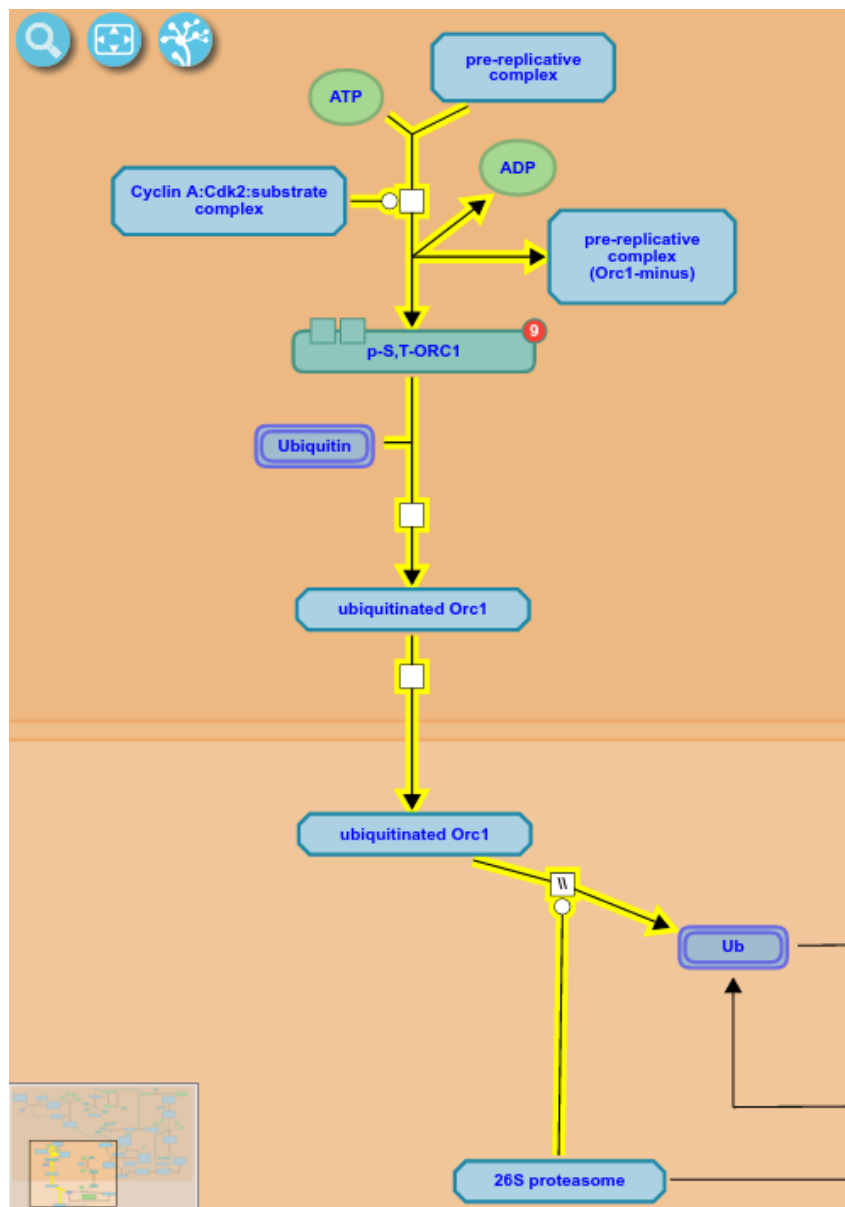


Figura 3.9: Via metabólica de remoção do Orc1 da cromatina apresentada pela ferramenta de visualização do Reactome.

[2] Lista de redes metabólicas ordenadas alfabeticamente;

[3] Mapa gerado em forma de grafo interativo. Quando uma rede metabólica da lista [2] é selecionada, ela se destaca no mapa. No topo, à esquerda é possível buscar por palavras chaves no mapa e visualizá-lo em tela cheia e à direita é possível visualizar imagens detalhadas das redes metabólicas, capturar uma imagem do mapa gerado e ler sobre a notação utilizada na página. Abaixo, à esquerda tem-se um mini-map por onde o usuário pode se localizar no mapa geral e à direita tem-se botões com o mesmo propósito e mais dois: (+) *Zoom in* e (-) *Zoom out*. Por fim, à direita do mapa existe um painel que se abre da direita para a esquerda que permite que o usuário edite as cores do mapa e obtenha informação sobre o mesmo.

- [4] Menu de operação e informações sobre as redes metabólicas selecionadas em [2]. O *Description* contém seus detalhes tais como compartimento celular e referências externas. O *Molecules* contém uma lista de compostos químicos e proteínas associadas às vias. O *Structures* oferece uma lista de imagens que representam a via e seus componentes. O *Expression* apresenta detalhes sobre o local no organismo onde ocorrem as reações das vias selecionadas. O *Analysis* apresenta o resultado das análises feitas em [1]. Finalmente o botão Downloads oferece arquivos das vias para baixar em diversos formatos tais como SBML e PDF.

KEGG

O KEGG¹⁷ (*Kyoto Encyclopedia of Genes and Genomes*) é uma base de informações sobre sistemas biológicos em nível molecular, sobretudo sobre conjuntos de dados em larga escala gerados por sequenciamento de genoma [2].

As informações sobre os sistemas podem ser dadas em forma de módulos, unidades funcionais com identificação otimizada para análise dos dados, em forma de *brite*, coleção de arquivos estruturados hierarquicamente sobre as funções das entidades biológicas, ou em forma de vias, mapa de interações moleculares e reações químicas.

Dado que o metabolismo é um conjunto de reações e transformações químicas, a maneira natural de representá-lo é por meio de uma rede de interações, ou seja, em forma de vias. O KEGG oferece uma ferramenta de busca de vias metabólicas sobre várias redes metabólicas, dos vários organismos que constituem o banco de dados.

Para acessar sua ferramenta de visualização, basta navegar a partir da *home page*:

<http://www.kegg.jp/> ⇒

Data-oriented entry points ⇨ KEGG PATHWAY ⇒

Pathway Maps ⇨ 1. Metabolism ⇒

1.0 Global and overview maps ⇨ Metabolic pathways

Segue abaixo a descrição de cada elemento da interface da ferramenta de visualização do KEGG, subdivididos conforme indicado na Figura 3.10.

- [1] Menu de navegação direta e indiretamente relacionado às vias metabólicas. O *Pathway menu* redireciona para uma página que contém uma lista *links* para todos os diagramas mostrados no mapa geral. O *Organism menu* redireciona para uma página com uma lista de todos os organismos conhecidos pelo KEGG (346 eucariotas e 4159 procariotas¹⁸). O *Pathway entry* redireciona para a descrição da via metabólica selecionada. O *Hide module list*, quando clicado, oculta a área [2] da página. O *User data mapping* abre uma nova janela que solicita um objeto seguido de uma cor escolhida pelo usuário para dar sua própria coloração às vias. Por fim, o *Image (png) file* oferece a imagem do mapa para *download*. Abaixo é possível escolher o organismo apresentado no mapa, sua resolução e buscar por palavras chaves;

¹⁷Disponível pela *web*, em <http://www.kegg.jp/>.

¹⁸Visitado em 2016-10-06

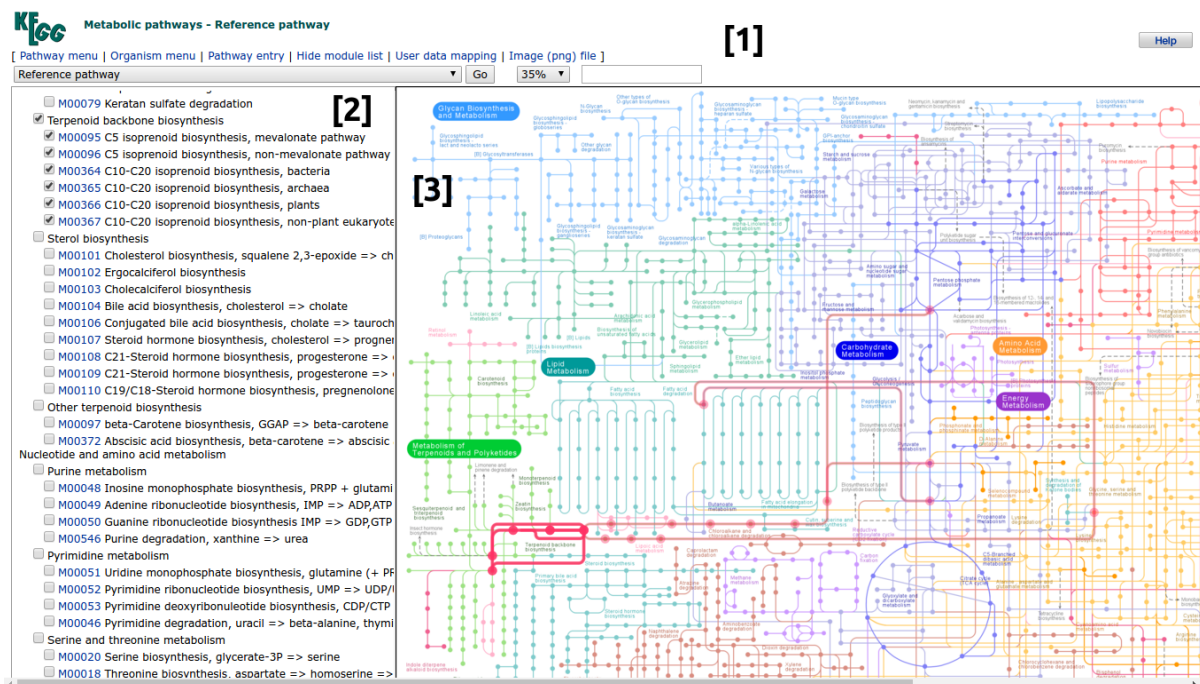


Figura 3.10: Visão geral da rede metabólica de referência do KEGG, com destaque na via metabólica de biossíntese do *backbone* de terpenóides. A interface pode ser dividida nas seções [1], [2] e [3].

[2] Lista de redes metabólicas organizadas de acordo com suas funções;

[3] Mapa gerada em forma de grafo interativo. Quando uma rede metabólica da lista [2] é selecionada, ela se destaca no mapa.

Caso o usuário queira visualizar a(s) via(s) metabólica detalhadamente, o KEGG oferece uma lista de diagramas¹⁹ desenhados manualmente em formato de grafo interativo cujos nós, quando clicados, redirecionam para uma página contendo a descrição do objeto, que pode ser enzima, composto químico ou outra rede metabólica, por exemplo. Continuando com o exemplo da biossíntese do *backbone* de terpenóides, para acessar seu diagrama, apresentado na figura 3.11, basta navegar a partir da *home page*:

<http://www.kegg.jp/> ⇒

Data-oriented entry points ⇔ KEGG PATHWAY ⇒

Pathway Maps ⇔ 1. Metabolism ⇒

1.9 Metabolism of terpenoids and polyketides ⇔ Terpenoid backbone biosynthesis

¹⁹Detalhes da notação dos diagramas: http://www.genome.jp/kegg/document/help_pathway.html

TERPENOID BACKBONE BIOSYNTHESIS

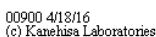


Figura 3.11: Via metabólica de biossíntese do *backbone* de terpenóides apresentada pela ferramenta de visualização do KEGG.

BioCyc

O BioCyc²⁰ é um sistema de coleção de aproximadamente 7 mil bancos de dados chamados PGDBs (*Pathway/Genome Databases*) pois possuem duas maneiras diferentes de representar as informações: modelo de vias metabólicas, que enfatiza as sequências de reações, substratos e produtos de múltiplos organismos, ou modelo de sequência genômica, que destaca a localização e descrição dos genes de cada organismo específico [1].

Os bancos PGDBs são organizado em três camadas de acordo com a frequência de atualizações/refinações e da maneira com que os dados foram obtidos. O BioCyc possui um banco de dados específico para redes metabólicas determinadas experimentalmente, chamado MetaCyc²¹. Este é o único banco de dados multi-organismos do grupo BioCyc e ele é referência na ferramenta gratuita *Pathway Tools* desenvolvida pelo instituto de pesquisa *SRI International*.

Diferente das ferramentas do KEGG e Reactome, o Pathway Tools não possui um mapa global de visualização das vias metabólicas. Para acessá-las, o usuário deve buscar por alguma palavra chave conhecida ou navegar pelo sumário de vias²² classificado hierarquicamente com base em suas funções biológicas e na classe dos metabólitos produzidos/consumidos. O acesso à via da Figura 3.5, por exemplo, se deu pela navegação a partir da *home page*:

<http://metacyc.org/> ⇒

Metabolism ⇨ Browse Pathway Ontology ⇒

Pathways ⇨ Bioluminescence (1 instances) ⇨ bacterial bioluminescence

Cada nó das vias metabólicas quando clicado redireciona o usuário para uma página com mais detalhes sobre o objeto, que pode ser outra via, um composto químico ou uma enzima, por exemplo. A Figura 3.12 é uma modificação da Figura 3.5 apresentando diretamente o conjunto destes elementos.

3.3.2 2Path

Conteúdo estará disponível quando site estiver pronto.

²⁰Disponível pela *web*, em <http://biocyc.org>.

²¹Disponível pela *web* através do site <http://metacyc.org/>.

²²Na última visita, em 2016-10-06, a página possuía 3.461 vias metabólicas.

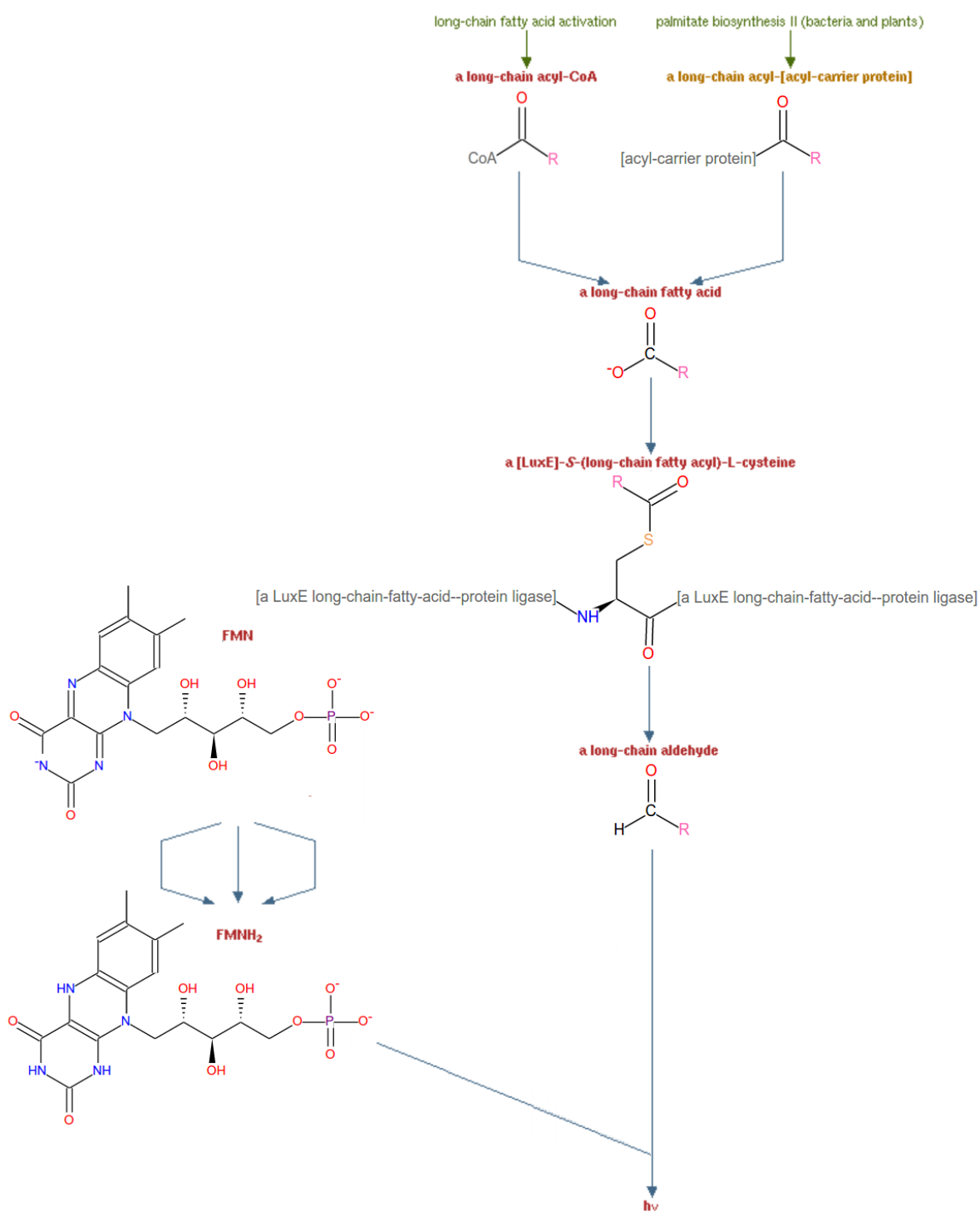


Figura 3.12: Representação modificada da Figura 3.5 com maior nível de detalhes apresentada pelo MetaCyc.

Capítulo 4

Projeto de Interface

4.1 Comunicação usuário-sistema

Página 212.

Modelo do livro: Criar tabela de conversação entre usuário e sistema que explicita os signos (elementos da interface) e o que o usuário pensa (representado como fala) a respeito deles. Vantagens: Designer aprende o que o user quer fazer: **são especificados as entradas e saídas esperadas**; Site pequeno \Rightarrow **Apenas um cenário** de conversação para alcance do objetivo do user. Desvantagem: Só tem no livro

4.1.1 Design de Interação

Página 213. TABELA DE COMUNICAÇÃO USER-SISTEMA AO LONGO DO SITE TODO.

4.1.2 Mapa de Objetivos

Página 216. DIAGRAMA DE ELEMENTOS DO SITE.

4.1.3 Tratamento de Rupturas na Comunicação

Página 223. EXPLICAÇÃO DOS TIPOS DE TRATAMENTOS.

4.1.4 Modelagem da Interação com Linguagem MoLIC

Página 229. DIAGRAMA DE TAREFAS.

4.2 Interface do Sistema

Página 243. ESTILO DE INTERAÇÃO: FERRAMENTA

4.3 Detalhes de Implementação

O sistema desenvolvido para este projeto é uma aplicação web chamada *2Path*. O usuário deve se cadastrar no *website* para ter acesso às redes metabólicas do banco de dados do sistema, bem como pesquisar por palavras chaves no mesmo. Nesta seção serão apresentadas as linguagens e ferramentas utilizadas no desenvolvimento do *website*, as características, funcionalidades e limites do sistema e, por fim, as dificuldades enfrentadas na implementação do projeto.

O sistema foi desenvolvido no ambiente de desenvolvimento integrado *open source* Eclipse Java EE - *Java Platform, Enterprise Edition*, versão Mars 4.5.2. Para simplificar a obtenção das dependências do projeto, ou seja, pacotes de arquivos java (extensão .jar), foi utilizada o Apache Maven¹. Este *software* opera sobre o arquivo *pom.xml*, onde POM significa *Project Object Model* e contém as especificações de cada projeto que se tornará dependência do sistema em desenvolvimento, além de outros aspectos do código. O servidor selecionado para hospedagem local, *localhost* porta 8080, do sistema foi o Apache TomCat versão 7.0.

As páginas da aplicação foram desenvolvidas na linguagem de marcação XHTML, *Extensible Hypertext Markup Language*, e a estilização em CSS, *Cascading Style Sheets*. Para conexão entre *front-end* e *back-end* foram utilizados JSF², Primefaces³, e AngularJS, para a construção do grafo de vias metabólicas.

JSF: https://projetos.inf.ufsc.br/arquivos_projetos/projeto_1214/Ferramenta%20de%20mapeamento%20de%20UIDs%20para%20JSF.pdf

CORES: <http://www.hermancerrato.com/graphic-design/images/color-images/the-meaning-of-colors-book.pdf>, <http://www.awwwards.com/trendy-web-color-palettes-and-html>, <https://medium.com/wdstack/how-the-bootstrap-grid-really-works-471d7a089cfc#.k7w39z2uh>

4.3.1 Banco de Dados OrientDB

O banco de dados escolhido para armazenar as enzimas, compostos, reações e demais elementos foi o OrientDB.

CITAR: (1) Graph database, (2) SQL, (3) Commercial Friendly License, (4) Low Total Cost of Ownership (TCO) Community Edition, (5) Open Source. <http://orientdb.com/why-orientdb/>, <http://orientdb.com/docs/last/Java-API.html>

4.3.2 Desafios

O que foi o trabalho. Decrever todo o ambiente usado Neste capítulo serão apresentados os primeiros resultados experimentais obtidos.

¹*Software* de gerenciamento de projeto e ferramenta de compreensão de programa.

²Especificação Java para criação de componentes de interface de aplicação *web*.

³Biblioteca com uma coleção de componentes de interface voltadas para JSF.

Capítulo 5

Resultados

Aplicação do método da avaliação da comunicabilidade

5.1 Interface

5.2 Avaliação e Discussão

Capítulo 6

Conclusão e Trabalhos Futuros

6.1 Conclusão

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

6.2 Trabalhos Futuros

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

- x

Capítulo 7

Cronograma

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

Tabela 7.1: Cronograma

Atividades	2016					
	Jul	Ago	Set	Out	Nov	Dez
Revisão bibliográfica	X	X				
Familiaridade com ambiente de desenvolvimento		X	X			
Implementação da aplicação		X	X	X	X	
Aplicação do MAC sobre o 2Path					X	
Interpretação dos resultado						X
Defesa						X

Referências

- [1] Introduction to biocyc. <http://biocyc.org/intro.shtml>, visitado em 2016-10-01. 27
- [2] Kegg overview. <http://www.kegg.jp/kegg/kegg1a.html>, visitado em 2016-10-01. 24
- [3] Protein syntesis steps. <http://www.proteinsynthesis.org/protein-synthesis-steps/>, visitado em 2016-08-19. vii, 16
- [4] Usersguide. <http://wiki.reactome.org/index.php/Usersguide>, visitado em 2016-10-01. 20
- [5] Protein structure, 2009. <http://www.particlesciences.com/news/technical-briefs/2009/protein-structure.html>, visitado em 2016-01-02. vii, 14, 15
- [6] Márcio Santos Aleixo. Divisão celular. <http://www.infoescola.com/citologia/divisao-celular/>, visitado em 2016-11-19. vii, 18
- [7] Simone Diniz Junqueira Barbosa and Bruno Santana da Silva. *Interação Humano-computador*. Elsevier Editora Ltda., 2010. viii, 1, 2, 4, 5, 6, 7, 9
- [8] Bonnie Berger, Jian Peng, and Mona Singh. Computational solutions for omics data. *Nature reviews. Genetics*, 14(5):333–346, May 2013. 1
- [9] Perry Carter. Catabolic and anabolic reactions. <http://classes.midlandstech.edu/carterp/courses/bio225/chap05/lecture1.htm>, visitado em 2016-10-12. 17
- [10] Ron Caspi. Metacyc pathway: bacterial bioluminescence. <http://metacyc.org/META/new-image?object=PWY-7723>, visitado em 2016-10-13. vii, 18
- [11] Krukemberghe Divino Kirk da Fonseca Ribeiro. Enzimas. <http://mundoeducacao.bol.uol.com.br/biologia/enzimas.htm>, visitado em 2016-11-20. 17
- [12] Aline Marques da Silva Almeida. Vivendo uma nova era: a tecnologia e o homem, ambos integrantes de uma sociedade que progride rumo ao desenvolvimento, 2014. <http://www.seduc.mt.gov.br/Paginas/Vivendo-uma-nova-era-a-tecnologia-e-o-homem,-ambos-integrantes-de-uma-sociedade-que-progride-rumo-ao-desenvolvimento.aspx>, visitado em 2016-11-20. 4

- [13] Claudio Manoel de Carvalho Correia. Semiose e desenvolvimento cognitivo: Estudo sobre as estratégias de construção dos processos sígnicos em seqüências lógicas, 2001. <http://www.filologia.org.br/vicnlf/anais/semiose.html>, visitado em 2016-10-06. 7
- [14] Justin N. Kabera, Edmond Semana, Ally R. Mussa, and Xin He. Plant secondary metabolites: Biosynthesis, classification, function and pharmacological properties. *Journal of Pharmacy and Pharmacology*, 2:377–392, 2014. 19
- [15] Vincent Lacroix, Ludovic Cottret, Patricia Thébault, and Marie-France Sagot. An introduction to metabolic networks and their structural analysis. *IEEE/ACM Trans. Comput. Biology Bioinform.*, 5(4):594–617, 2008. 17
- [16] José G. Sampedro y Hugo Nájera Laura González-Torres, Alfredo Téllez-Valencia. Las proteínas en la nutrición. *Respyn - Revista Salud Pública y Nutrición*, 8, 2007. 14
- [17] Nicole Legault. Affordance: What does it mean for e-learning? <https://community.articulate.com/articles/affordance-what-does-it-mean-for-e-learning>, visitado em 2016-10-06. vii, 5
- [18] Gerhard Michal and Dietmar Schomburg. *The Cell and Its Contents*, pages 14–36. John Wiley & Sons, Inc., 2012. 17
- [19] David W. Mount. *Bioinformatics : sequence and genome analysis*. Cold Spring Harbor, N.Y. Cold Spring Harbor Laboratory Press, 2001. 1, 20
- [20] Leslie A. Pray. Discovery of dna structure and function: Watson and crick, 2008. <http://www.nature.com/scitable/topicpage/discovery-of-dna-structure-and-function-watson-397>, visitado em 2016-01-15. vii, 13, 14
- [21] João Carlos Setubal and João Meidanis. *Introduction to computational molecular biology*. PWS Publishing Company, 1997. 1, 12, 13, 14, 15, 17
- [22] Röbbbe Wünschiers, Martina Jahn, Dieter Jahn, Ida Schomburg, Susanne Peifer, Elmar Heinzle, Helmut Burtscher, Julia Garbe, Annika Steen, Max Schobert, Dieter Oesterheld, Josef Wachtveitl, and Antje Chang. *Metabolism*, pages 37–209. John Wiley & Sons, Inc., 2012. 18, 19