



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
Dr.^a Tainá Raiol Alencar — IB/UnB
Prof. Dr. Rodrigo Bonifacio de Almeida — 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.

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

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

1. Bioinformática, 2. Redes Metabólicas, 3. Grafo, 4. Interação Humano-Computador, 5. 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 Emilia Machado Telles Walter (Orientadora)
CIC/UnB

Dr.^a Tainá Raiol Alencar Prof. Dr. Rodrigo Bonifacio de Almeida
IB/UnB CIC/UnB

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

Brasília, 08 de Julho de 2016

Agradecimentos

Agradeço à professora Maria Emilia pelo e ao pesquisador Waldeyr Mendes, pela paciência, apoio e orientação. Agradeço à minha família pelo suporte e dedicação em me ajudar nessa fase final da graduação.

Resumo

O *2Path* é um banco de dados de redes metabólicas. Ele possui um sistema que recebe um arquivo FASTA de múltiplas sequências e verifica se as sequências de enzimas presentes no banco correspondem às sequências presentes no FASTA. Caso alguma corresponda, então essa enzima está presente no organismo em questão. O objetivo deste trabalho é apresentar um projeto de interface para o banco *2Path* em que seja possível fazer *upload* de arquivos FASTA que representam organismos, além de pesquisar por enzimas e vias metabólicas entre dois compostos no mesmo. Para isso, serão analisadas as interfaces dos bancos de dados de redes metabólicas KEGG, BioCyc e Reactome. A interface proposta será testada e aperfeiçoada de acordo com o Método de Avaliação de Comunicabilidade, da área de Interação Humano-Computador.

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

Abstract

The 2Path is a metabolic network database. It has a system that receives as an input a multiple sequence FASTA file and verifies if the enzymes on the database can be produced by the uploaded organism represented by the FASTA file. If so, then the organism is linked to them on the database. The goal of this project is to propose an interface to the 2Path database such that users can upload a FASTA file and search for enzymes and metabolic pathways between two compounds in the database. To achieve this goal, the interface of the most known metabolic database (KEGG, BioCyc and Reactome) will be analysed. The proposed interface will be tested and improved according to the Communicability Evaluation Method, from the Human-Computer Interaction field.

Keywords: Bioinformatics, Metabolic Networks, Graph, Human-Computer Interaction, Communicability Evaluation Method

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
2.3	Projeto de Interface Centrado na Comunicação	11
2.3.1	Tabela de Interações	11
2.3.2	Mapa de Objetivos	12
2.3.3	Modelagem de Tarefas	13
2.3.4	Tratamento de Rupturas na Comunicação	14
3	Redes Metabólicas	16
3.1	Conceitos Básicos de Biologia Molecular	16
3.1.1	Ácidos Nucleicos	16
3.1.2	Proteínas	17
3.1.3	Síntese de Proteína	19
3.2	Conceitos Básicos de Metabolismo	21
3.2.1	Metabolismo Primário	22
3.2.2	Metabolismo Secundário	22
3.3	Visualização de Redes Metabólicas	24
3.3.1	Reactome, KEGG e BioCyc	25
3.3.2	2Path	33
4	Interface do 2Path	35
4.1	Projeto de Interfaces	35
4.1.1	Tabela de Interações	36
4.1.2	Mapa de Objetivos	36
4.1.3	Modelagem de Tarefas	36
4.1.4	Tratamento de Rupturas na Comunicação	41
4.2	Implementação da Interface	41
4.2.1	Detalhes de Implementação	41
4.2.2	Interface do Sistema	43

4.3 Projeto de Avaliação	47
5 Resultados dos testes	48
5.1 Avaliação da Interface	48
5.2 Nova Interface do Sistema	51
6 Conclusão e Trabalhos Futuros	55
Referências	56
A Questionários de dados pessoais dos biólogos	58
B Tarefas realizadas pelos biólogos no sistema 2Path	60
C Questionários de interface conforme o Método de Avaliação de Comunicabilidade	62

Listas de Figuras

2.1	ExemploS 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 [18].	5
2.2	Representação generalizada do mapa de objetivos.	13
2.3	Estruturas de tarefas (a) sequencial, (b) independente de ordem, (c) alternativa, (d) iterativa, (e) ubíqua e (f) opcional.	14
3.1	Nucleotídeo com grupo fosfato (P), pentose abaixo à esquerda e base nitrogenada G (no caso a Guanina). Adaptado de [22].	17
3.2	Representação das duas fitas de ADN, que se ligam por pontes de hidrogênio entre os nucleotídeos. Adaptado de [22].	18
3.3	Quatro estruturas de proteínas. As unidades esféricas na estrutura primária representam os aminoácidos da proteína. Adaptado de [5].	19
3.4	Representação da etapa final da síntese de proteína, que sempre finaliza com o códon UGA, UAA ou UAG. Adaptado de [3].	20
3.5	Representação de uma reação bioquímica catalisada por uma enzima. Adaptado de [14].	21
3.6	Exemplo de via metabólica retirado do MetaCyc [10], que representa o processo de produção da bioluminescência de bactérias.	22
3.7	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].	23
3.8	Um exemplo de elemento químico para cada classe, ou grupo, de metabolismo secundário.	24
3.9	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].	26
3.10	Via metabólica de remoção do Orc1 da cromatina apresentada pela ferramenta de visualização do Reactome.	27
3.11	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].	28
3.12	Via metabólica de biossíntese do <i>backbone</i> de terpenóides apresentada pela ferramenta de visualização do KEGG.	30

3.13 Representação modificada da Figura 3.6 com maior nível de detalhes apresentada pelo MetaCyc.	32
3.14 Modelo do banco de dados 2Path.	34
4.1 Mapa de objetivos finais e instrumentais.	39
4.2 Objetivo de pesquisar enzima no banco de dados completo do 2Path.	39
4.3 Objetivo de pesquisar enzima no banco de dados específico para a rede metabólica construída a partir do organismo do usuário.	40
4.4 Objetivo de pesquisar via metabólica entre substrato e produto no banco de dados completo do 2Path.	40
4.5 Objetivo de pesquisar via metabólica entre substrato e produto no banco de dados específico para a rede metabólica construída a partir do organismo do usuário.	41
4.6 Diagrama geral de uma via metabólica na interface do 2Path.	43
4.7 Página inicial do 2Path.	44
4.8 Página principal de busca no 2Path.	44
4.9 Página de busca por enzima no banco completo	45
4.10 Página de apresentação da reação bioquímica catalisada pela enzima buscada.	45
4.11 Página de busca por via metabólica no banco de dados formado para o organismo <i>Arabidopsis thaliana</i>	46
4.12 Página de apresentação da via metabólica de reações que ocorrem no organismo <i>Arabidopsis thaliana</i>	46
5.1	52
5.2	52
5.3	53
5.4	53
5.5	53
5.6 PROBLEMA: ENZIMA PODE CATALIZAR MAIS DE UMA REACAO	54

Lista de Tabelas

2.1	Descrição das etiquetas, de acordo com a falha de comunicação que representam [7]	9
2.2	Modelo geral de representação da interação entre usuário (U) e projetista (P). Os signos representam o foco de cada conversa.	12
3.1	Código Genético que mapeia cada códon à um dos 20 aminoácidos, representados de maneira abreviada.	20
4.1	Representação da interação entre usuário (U) e projetista (P). Os signos representam o foco de cada conversa.	37
4.2	Campos de entrada e manipulação dos usuários do sistema 2Path	42
5.1	Contabilização das etiquetas para cada usuário. Cada etiqueta é registrada no tempo do vídeo em que ela ocorreu.	49

Capítulo 1

Introdução

Desde a descoberta da estrutura helicoidal do DNA, por Watson e Crick [23] em 1953, 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 [23], 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 [20].

Atualmente, pesquisadores já podem sequenciar cadeias de DNA, RNA e proteína [20] 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 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 de quatro bancos de dados de redes metabólicas: *Reactome*, *KEGG* e *MetaCyc* e *2Path*, bem como as ferramentas de visualização de dados dos três primeiros. 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 inherentemente 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 desenvolvida uma ferramenta de visualização de vias metabólicas, a ser acoplada ao *2Path*, seguindo os padrões de projeto de interface do campo de IHC. Para verificar se o sistema satisfaz os critérios básicos de qualidade, 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 o *2Path* não possui uma interface para que os pesquisadores biólogos façam consulta no banco de dados. Nesse sentido, a única maneira de interagir com o banco é por meio de *queries* em CYPHER, linguagem específica do banco de dados não relacional em grafos, Neo4j, sobre o qual o *2Path* foi modelado. A ferramenta desenvolvida neste projeto deverá permitir que o usuário facilmente pesquise, visualize e interaja com os dados dinamicamente.

1.2 Problema

Não há uma visualização de redes metabólicas armazenadas no *2Path* que facilite ao pesquisador explorar os aspectos biológicos do organismo estudado.

1.3 Objetivo

O objetivo principal do projeto é construir uma interface que acesse redes metabólicas armazenadas no *2Path* e gere uma visualização interativa dessas informações. Essa interface deverá ser avaliada de acordo com os critérios de qualidade especificados em Interação Humano-Computador.

Os objetivos específicos são:

- 2 : Propor uma interface adequada para o *2Path*;
- 3 : Propor avaliação de usabilidade, experiência do usuário, acessibilidade e comunicabilidade da interface para a interface proposta;
- 4 : Avaliar a interface desenvolvida com três biólogos;
- 5 : Analisar a usabilidade e a comunicabilidade de modo a aprimorar a interface.

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, bem como a ferramenta implementada neste trabalho, o *2Path*.

O Capítulo 4 descreve o método de elaboração da interface auto-explicativa e consistente do sistema desde sua concepção, incluindo detalhes da sua 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.

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, difundindo-se cada vez mais rápido. Essas tecnologias são, cada uma, um sistema computacional distinto, que interage a sua maneira com os usuários. Elas podem interferir direta e/ou indiretamente nos processos de comunicação e informação dos indivíduos [12].

Para satisfazer certos critérios de qualidade, estes sistemas computacionais devem possuir algumas características que facilitem seu uso para 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 sistemas [7], mensurando certas propriedades e avaliando 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, pois será utilizado neste projeto.

2.1 Conceitos Básicos de IHC

Esta seção apresenta os componentes básicos utilizados na matéria interdisciplinar IHC, 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 Kammersgaard [7], existem quatro perspectivas de interação usuário-sistema:

- 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*.

The figure shows three examples of text input fields, each preceded by the label "Nome:". The first two are simple rectangular boxes. The third is a box with the placeholder text "type your name here".

Figura 2.1: Exemplos 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 [18].

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 deve 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 (focado em problemas de usabilidade durante um processo interativo), cognitivo (focado em avaliar a facilidade de aprendizado de um sistema interativo) ou inspeção semiótica (focado em avaliar a quantidade de informação transmitida do projetista do sistema para o usuário).

Outro tipo 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) enfrentam 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 qualidade, 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 essa 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 **semióse**. A semióse 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 semióse 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 5.1. 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 **falta 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.

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

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?

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.” [7, pag. 78]

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.

2.3 Projeto de Interface Centrado na Comunicação

Para construir um sistema com foco na comunicabilidade, o projetista deve elaborar um modelo conceitual de elementos do sistema computacional sendo desenvolvido. Esse modelo deve descrever detalhadamente como será realizada a interação usuário-sistema, os objetivos dos usuários e todas as possíveis tarefas que um usuário pode realizar. Além disso, para certos elementos da interface facilmente mal interpretados pelo usuário no sistema, o projetista deve definir os métodos de recuperação e/ou prevenção de rupturas comunicativas [7]. Esses e outros conceitos serão explicados com detalhes a seguir.

O objetivo desta seção é apresentar a tabela de interações usuário-sistema, o mapa de objetivos de usuário, o diagrama de tarefas do sistema e a tabela de prevenção e recuperação de rupturas comunicativas. Esses quatro modelos serão utilizadas neste trabalho para o desenvolvimento da interface do sistema *2Path*.

De maneira geral, a tabela de interações tem o objetivo de representar a interação entre usuário e sistema de uma maneira fácil de entender, para que fiquem claros o que o sistema espera do usuário e vice-e-versa. O mapa de objetivos apresenta a lista de todos os motivos pelos quais o usuário está utilizando o sistema, indicando também os tipos diferentes de objetivos e usuários. O diagrama de tarefas apresenta a navegação completa no sistema para atingir todos os objetivos, elaborados previamente. Por fim, a tabela de prevenção e recuperação de rupturas aponta o tratamento realizado após um erro já cometido ou prestes a ser cometido pelo usuário.

2.3.1 Tabela de Interações

Na Engenharia Semiótica, cabe ao projetista elaborar a metacomunicação do sistema, minimizando ao máximo qualquer tipo de falha de comunicação que pode levar a possíveis

rupturas comunicativas. Para simplificar essa tarefa, pode-se representar todas as interações entre usuário e sistema em forma de conversação [7]. Nesse sentido, o projetista pode criar uma tabela nos padrões da Tabela 2.2, onde o usuário parece estar se comunicando com o projetista através de suas ações e as respostas obtidas.

Cada conversa possui um tópico geral, como “Realizar compra” ou “Cadastrar membro”. Essas ações não serão executadas de fato no sistema, mas possuem o propósito de engajar o projetista a iniciar uma conversa mais específica. Os subtópicos que se estendem a partir de um tópico geral são chamados diálogos. Esses últimos representam ações reais que podem precisar de alguma entrada do usuário (através de um formulário, por exemplo) e produzem respostas para dadas ações (mensagens de sucesso ou fracasso, por exemplo).

Tabela 2.2: Modelo geral de representação da interação entre usuário (U) e projetista (P). Os signos representam o foco de cada conversa.

Tópico -> Subtópico (diálogo)	Falas e Signos U: Usuário e P: Projetista
Ação com objetivo generalizado	U: Fala de usuário que precisa realizar um objetivo generalizado.
> Ação com objetivo específico	U: Detalhes da requisição P: Resposta do sistema para o usuário.

2.3.2 Mapa de Objetivos

Um dos primeiros passos para construir um projeto de interação é a definição dos objetivos dos usuários. Os objetivos podem ser finais ou instrumentais [7]. Os objetivos finais são o motivo pelo qual o usuário está utilizando o sistema. Em um *site* de venda de roupas, por exemplo, o objetivo final pode ser “Comprar camiseta”. Os objetivos instrumentais não são naturalmente percebidos pelos usuários. Eles servem para simplificar o objetivo final, realizando-o passo a passo. No mesmo exemplo, um objetivo instrumental pode ser “Adicionar camiseta no carrinho de compras”.

Existem dois tipos de objetivos instrumentais: direto e indireto [7]. Os objetivos instrumentais diretos facilitam a realização do objetivo final de maneira imediata. No exemplo acima, a ação “Adicionar camiseta no carrinho de compras” é um objetivo instrumental direto. Já o indireto prepara atividades planejadas para o futuro. A ação “Adicionar cartão de crédito” pode ser considerada um objetivo instrumental indireto, pois não envolve diretamente a aquisição da camiseta, porém será uma informação necessária para a finalização da compra.

A escolha dos objetivos já envolve tomar decisões de interação usuário-sistema [7]. Nesse sentido, ao criar o mapa de objetivos, representado de maneira generalizada na Figura 2.2, o projetista já consegue imaginar qual deverá ser o foco da implementação do sistema para melhor atender as necessidades dos usuários. Esses usuários podem não possuir o mesmo perfil, e consequentemente, apresentar papéis diferentes. Assim, também cabe ao projetista definir as permissões de acesso de certos elementos em uma mesma páginas, de acordo com o perfil dos usuários e seus objetivos.

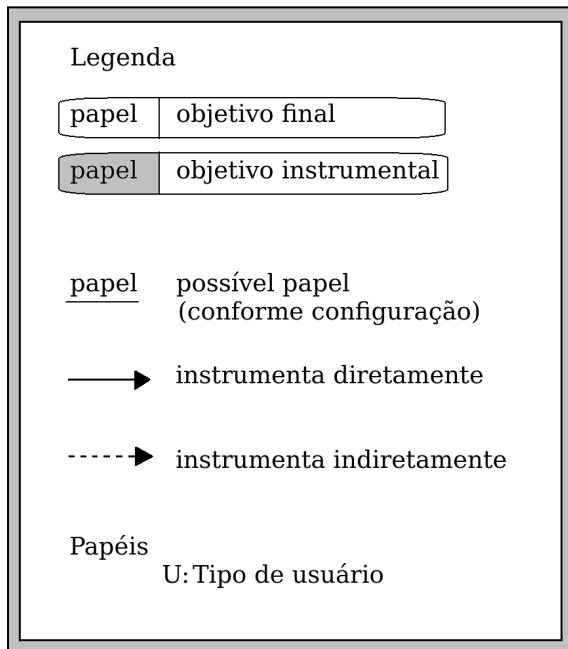


Figura 2.2: Representação generalizada do mapa de objetivos.

2.3.3 Modelagem de Tarefas

Conhecendo os objetivos dos usuários, o projetista pode formular uma sequência de tarefas a serem realizadas no sistema para cada um. Independente da plataforma, o diagrama de modelagem de tarefas descrito a seguir deverá permitir que o projetista tenha uma visualização hierarquizada de todas as tarefas que podem ser realizadas dado um objetivo de usuário.

Diferente do mapa de objetivos, a modelagem de tarefas não é focada na definição das permissões de usuários nem nos tipos de objetivos (finais ou instrumentais). Ela tem a finalidade de apresentar em formato de árvore as navegações que um usuário genérico pode realizar no sistema a partir de um objetivo [7]. Em um sistema de grande porte, podem existir diversos diagramas de tarefas, um para cada objetivo de usuário.

Ao elaborar esse modelo, o projetista deve fazer uma distinção entre tarefa, objetivo e operador. Objetivos são representados de maneira geral, em formato retangular, e não são executáveis de verdade no sistema. A raiz (elemento inicial) do diagrama de tarefas é um objetivo. As tarefas compõem um objetivo, e elas também são representadas por retângulos. As tarefas são os nós (elementos) internos do diagrama. Quando uma ação interage de fato com a interface, ou seja, quando a ação deixa de ser abstrata, ela é dita operador. Operadores são as folhas (elementos finais) do diagrama de tarefas.

A Figura 2.3 apresenta os tipos de estruturas que compõem o diagrama. Elas são:

- Estruturas sequenciais representam tarefas que devem ser executadas em ordem seguindo a numeração no canto superior esquerdo do retângulo;
- Estruturas independentes de ordem são auto-explicativas. Geralmente o projetista sugere a ordem que considera a mais otimizada para se atingir o objetivo, indicada na

numeração seguida de um ponto de interrogação. Porém o usuário possui autonomia para decidir qual a ordem o deixa mais confortável;

- (c) Estruturas alternativas oferecem mais de um caminho, ou navegação, que o usuário pode seguir. Ele, portanto, deve escolher apenas uma opção dentre as oferecidas na subárvore. Cada nó da subárvore é representado com letras maiúsculas no canto superior esquerdo dos retângulos e um círculo em branco no canto superior direito;
- (d) Estruturas iterativas podem ser realizadas mais de uma vez e os retângulos possuem um asterisco no canto superior direito. Para definir quantidade mínima (m) e máxima (n) de execução, pode-se alinhar à direita a expressão [m...n];
- (e) Estrutura ubíqua é utilizada quando existe uma tarefa que pode ser realizada a qualquer momento e em qualquer ordem. Ela é representada com um círculo preenchido no canto superior direito;
- (f) Estruturas opcionais, cuja borda é tracejada, representam tarefas opcionais.

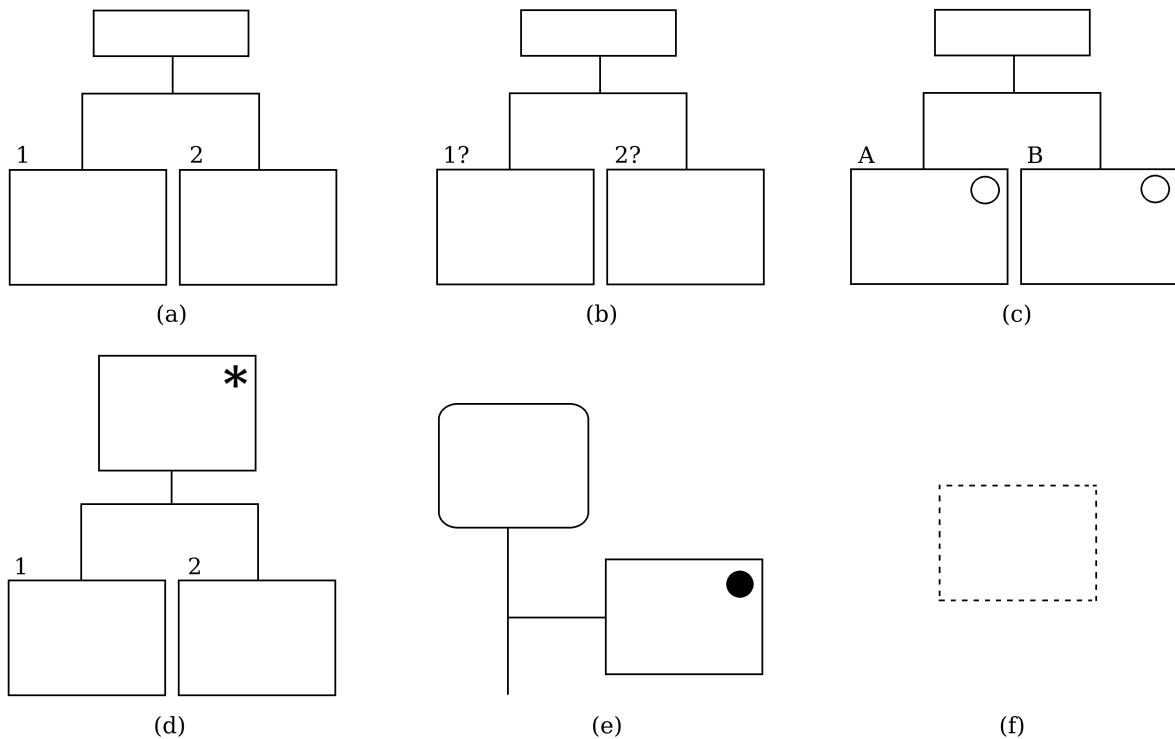


Figura 2.3: Estruturas de tarefas (a) sequencial, (b) independente de ordem, (c) alternativa, (d) iterativa, (e) ubíqua e (f) opcional.

2.3.4 Tratamento de Rupturas na Comunicação

Na Engenharia Semiótica, as rupturas na comunicação devem ser fortemente evitadas. Rupturas são consequências de falhas na comunicação e ocorrem durante a interação. Elas podem ser, por exemplo, cliques indevidos, ações executadas erroneamente ou erros de

interpretação. O projetista deve identificar os elementos que podem causar rupturas na comunicação e implementar ou uma prevenção, ou uma recuperação da mesma. Existem cinco categorias de prevenção e recuperação de rupturas [7], conforme descrito abaixo.

Mais de uma categoria pode ser utilizada para prevenir e/ou recuperar uma tarefa, interação ou signo. O Projetista deve criar uma tabela com a lista dos principais signos da interface, cada um com zero ou mais métodos de prevenção e zero ou mais métodos de recuperação.

- **Prevenção passiva:** O sistema *tenta evitar* que o usuário execute uma ação de maneira errada, porém sem impedi-lo de fazê-lo. Para isso, ele pode apresentar textos explicativos sobre o elemento específico (um campo em um formulário, por exemplo). Ainda, ele pode utilizar imagens e símbolos para guiar o usuário na execução de todas as ações de maneira correta;
- **Prevenção ativa:** O sistema *deve interferir* diretamente na interface do usuário, impedindo-o de executar certas ações que podem não ser seguras. Computacionalmente isso significa desabilitar botões, campos de entrada, *links*, restringir entrada de dados apresentando uma lista fechada de opções em vez de espaço em branco para preenchimento livre, entre outros;
- **Prevenção apoiada:** Nesse caso, o projetista identificou que um certo ponto da interface pode muito facilmente causar uma ruptura, portanto o sistema desenvolvido *deve descrever* com detalhes as consequências da interação com o conjunto de elementos naquele ponto em questão. Normalmente o projetista cria uma janela *pop-up* para chamar a atenção do usuário e solicitar a confirmação de sua decisão. Por exemplo, o usuário pode tentar sobrescrever um arquivo já existente, pensando que está criando um novo. O sistema, então, deverá informá-lo das consequências de sua ação;
- **Recuperação apoiada:** Caso uma ruptura já tenha acontecido, e o usuário tenha a chance de voltar atrás em sua decisão, o sistema deve fornecer um meio para que ele possa voltar. A recuperação apoiada expõe ao usuário seu erro cometido e oferece um meio para que ele *corrija*. Um exemplo disso são os campos obrigatórios de um formulário que ficam destacados em vermelho se o usuário tenta enviar sem preenchê-los;
- **Captura de erro:** Caso uma ruptura já tenha acontecido e não exista possibilidade de recuperação, o sistema *deve descrever* uma outra maneira de o usuário retomar suas atividades novamente. Por exemplo, caso o usuário envie um formulário com uma informação incorreta, ele deve poder editá-la, seja no mesmo sistema ou enviando um email para alguém com mais permissões que ele.

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 de 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 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 [23]. Por definição, o sentido da leitura de uma fita de ácido nucleico é 5' → 3', o que deve ser levado em consideração ao se interpretar de dados do material genético gerado em sequenciadores.

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, pentose possuindo um átomo de hidrogênio ligado ao carbono 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 [23], conforme indicado na Figura 3.2. 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).

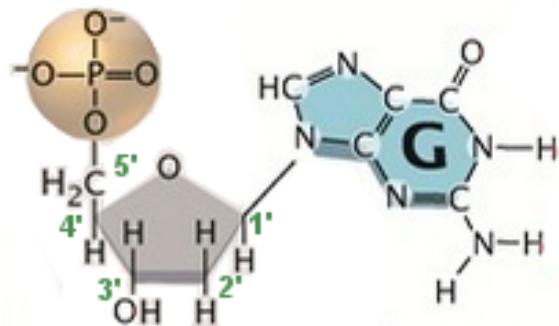


Figura 3.1: Nucleotídeo com grupo fosfato (P), pentose abaixo à esquerda e base nitrogenada G (no caso a Guanina). Adaptado de [22].

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 antes, o sentido padrão da leitura de uma fita é de 5' → 3', e no caso do ADN, as hélices são dispostas de maneira antiparalela, ou seja, uma é lida de 5' → 3' e a outra, de 3' → 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 TCATTTCG, e no sentido biológico, igual a GCTTACT.

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, posuem formato esférico e são compactas, o que as permite realizar processos dinâmicos pelo

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.

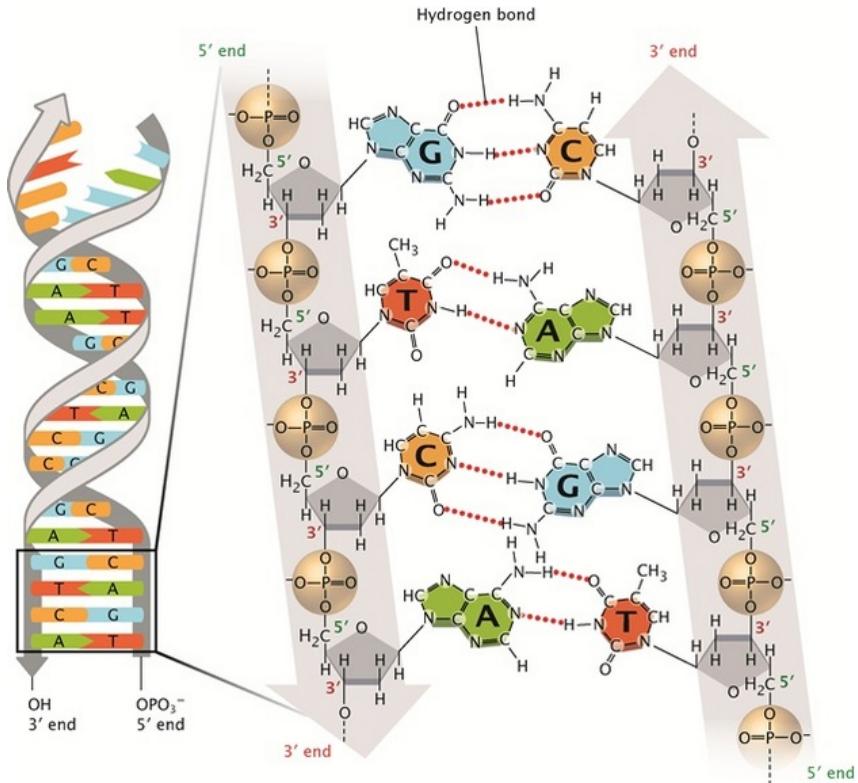


Figura 3.2: Representação das duas fitas de ADN, que se ligam por pontes de hidrogênio entre os nucleotídeos. Adaptado de [22].

corpo [15]. Cada tarefa é realizada por uma proteína com uma estrutura específica e adaptada para tal.

Assim como os ácidos nucleicos, as proteínas são polímeros, macromoléculas cujos monômeros são aminoácidos. Aminoácidos são moléculas que possuem cinco componentes: grupo 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 os aminoácidos são produzidos 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 amino 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 5.000 diferentes resíduos de aminoácidos, ou cadeia polipeptídica, constituem a proteína.

Existem quatro estruturas para caracterização de uma proteína [23], 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 na forma de hélice (α -Helix) ou na 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].

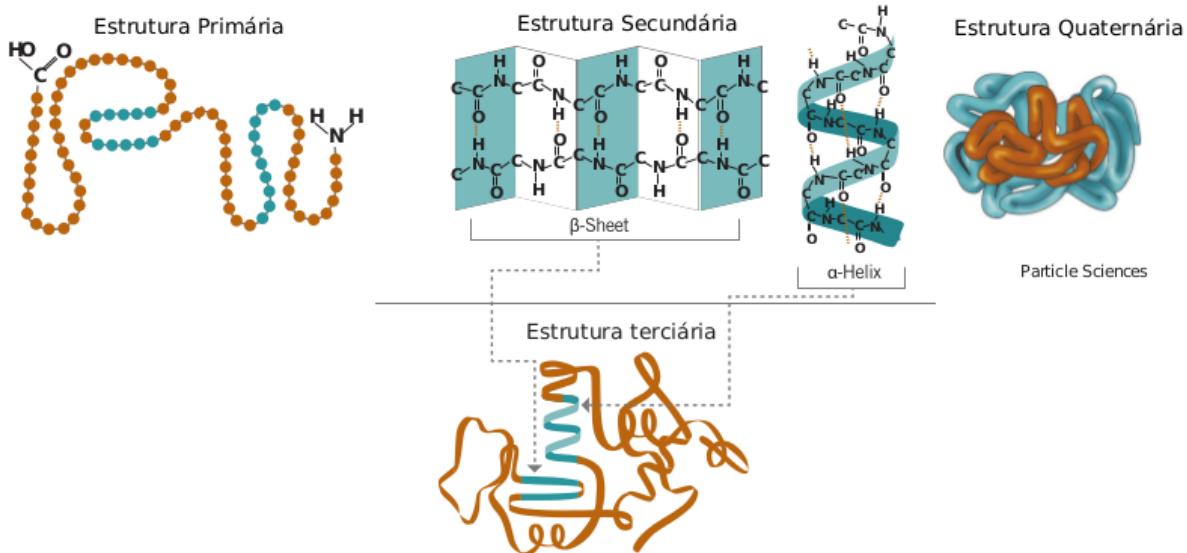


Figura 3.3: Quatro estruturas de proteínas. As unidades esféricas na estrutura primária representam os aminoácidos da proteína. Adaptado de [5].

3.1.3 Síntese de Proteína

O Dogma Central da Biologia Molecular é exatamente uma descrição da síntese de proteínas, e ocorre basicamente em duas grandes etapas, a transcrição e a tradução.

A **transcrição** é o processo de produção de mARN a partir do ADN. 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. O mARN VAI TER 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, 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 ao 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 completada quando o mARN apresenta um códon de parada, pois nenhum tARN possui correspondência para tal [23]. Uma proteína simples é, então, formada.

⁷Sequência de três nucleotídeos.

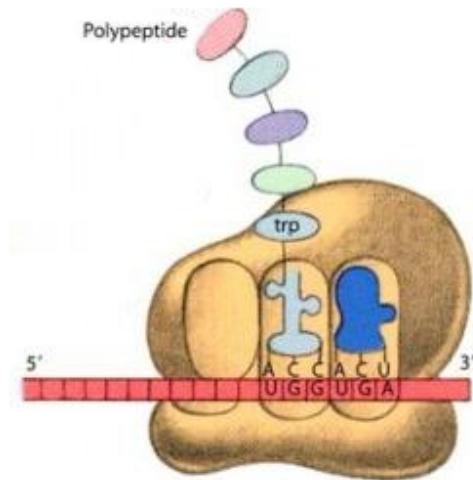


Figura 3.4: Representação da etapa final da síntese de proteína, que sempre finaliza com o códon UGA, UAA ou UAG. Adaptado de [3].

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 [23]. 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 oxi-redução (transferência de elétrons) e a Quinase transporta elementos químicos de uma molécula para outra [11].

As reações bioquímicas são alterações químicas que produzem um ou mais produtos a partir de um ou mais substratos [17]. 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 geram os substratos de uma mesma reação [17].

Ao catalisar uma reação, uma enzima pode receber um ou mais substratos em seus sítios ativos. O sítio ativo é um local pré-determinado na enzima em formato côncavo onde ocorre uma reação bioquímica, conforme indicado na Figura ???. Se ela recebe apenas um substrato, a estrutura que se forma com o preenchimento do sítio ativo é um complexo enzima-substrato. Porém se ela recebe mais de um substrato, a estrutura é chamada de complexo ternário intermediário [19]. Duas enzimas ainda podem possuir a mesma atividade enzimática, porém apresentar estruturas físicas diferentes. Essas são chamadas isoenzimas [19].

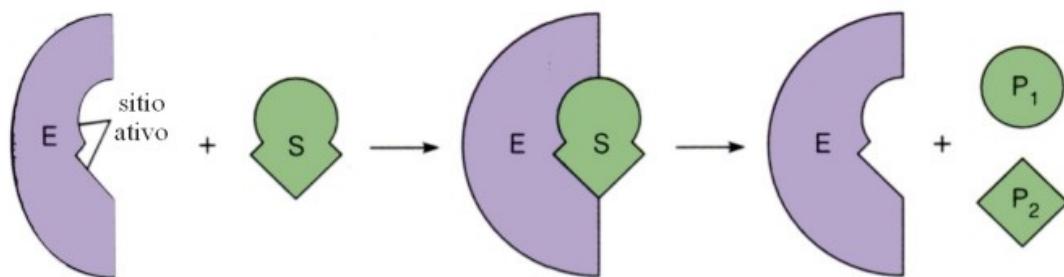


Figura 3.5: Representação de uma reação bioquímica catalisada por uma enzima. Adaptado de [14].

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 [17]. Cofatores podem ser coenzimas, associadas momentaneamente às enzima, ou grupos prostéticos, associados firmemente às enzimas [19].

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) [17]. Geralmente a energia liberada pelas reações catabólicas é usada para impulsionar as reações

anabólicas [9]. A Figura 3.6 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).

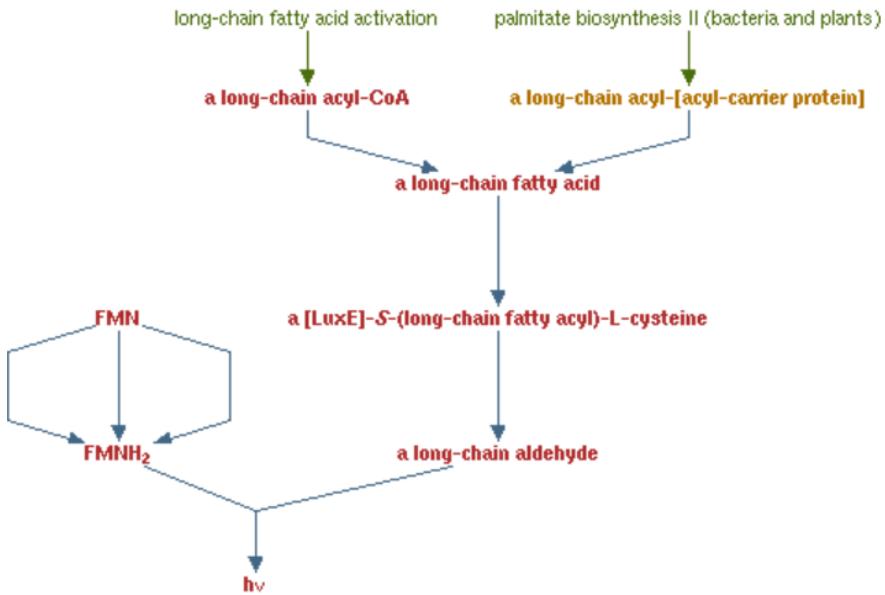


Figura 3.6: Exemplo de via metabólica retirado do MetaCyc [10], que representa o processo de produção da bioluminescência de bactérias.

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 [17]. Uma das vantagens de se estudar as vias em conjunto é poder explorar vias alternativas para um mesmo fim biológico [17].

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 denominado metabolismo primário [25]. São exemplos de metabolismos primários os processos de divisão celular mitose e meiose, a primeira para crescimento e a segunda para reprodução, conforme mostrado na Figura 3.7.

3.2.2 Metabolismo Secundário

Os metabolismos que não realizam função essencial no organismo, são classificados como metabolismos secundários. Esses são caracterizados por vasta diversidade química e, desta forma, são responsáveis pela sobrevivência do organismo em diferentes ambientes, de acordo com os fatores bióticos (elementos causados pela interação entre organismos, por exemplo, cadeia alimentar) e abióticos (elementos naturais independente de organismos, por exemplo, luz e temperatura) [25].

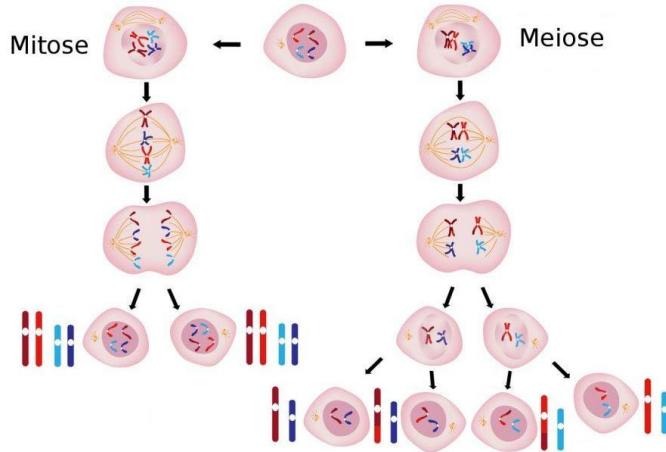


Figura 3.7: A célula que se divide por meio de mitose gera duas outras células idênticas à original. A célula que se divide por meio de meiose gera quatro células, cada uma com metade do material genético da célula original. Adaptado de [6].

Enquanto 20% dos metabólitos secundário são encontrados em bactérias, fungos e organismos sésseis⁸ marinhos, os outros 80% encontram-se em plantas vasculares⁹ [25], que podem ser subdivididos em três classes: terpenóides, alcaloides e fenólicos [16]. A seguir está uma breve explicação de cada uma dessas classes. A Figura 3.8 que apresenta um exemplo de composto químico para cada grupo:

- Os **terpenóides** constituem o 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 [25];
- 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 [25]. Por apresentarem alta diversidade estrutural, é difícil classificá-los. A tentativa mais recente baseia-se na semelhanças entre os esqueletos carbônicos [16];
- Os **fenólicos** são caracterizados pelas propriedades anti-oxidantes, anti-cancerígenas e anti-inflamatórias 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 [16].

⁸Que vivem fixos, sem capacidade de locomoção.

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

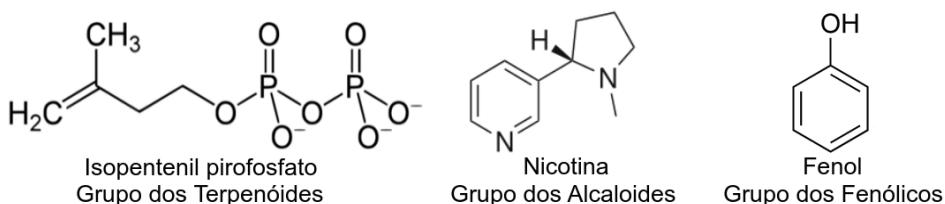


Figura 3.8: Um exemplo de elemento químico para cada classe, ou grupo, de metabolismo secundário.

3.3 Visualização de Redes Metabólicas

Desde o descobrimento da estrutura do ADN por Watson e Crick, 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, 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 [20]. Somente em 1984 esta coleção foi intitulada de *Protein Information Resource* [20]. 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 Accept Mutation*), que apresenta tais dados¹⁰. Hoje em dia também utiliza-se uma outra matrix, chamada BLOSUM (*Block Substitution Matrix*). Enquanto a PAM identifica as semelhanças entre sequências de proteínas, A BLOSUM 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, contam com o patrocínio do *National Center for Biotechnology Information* (NCBI). Os dois bancos são públicos e continuam crescendo exponencialmente [20].

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, transcriptoma¹², proteoma¹³, metaboloma¹⁴ e interactoma¹⁵.

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.

Um rede metabólica pode ser gerada computacionalmente a partir de um conjunto de dados ômicos [21]. Por exemplo, se um pesquisador detém um arquivo em formato FASTA¹⁶ contendo um sequência de nucleotídeos do genoma de um certo organismo, então ele pode verificar se aquela sequência gera uma certa enzima. Se sim, então naquele organismo pode existir a reação catalisada por essa enzima. Um arquivo FASTA de um organismo pode possuir várias sequências de nucleotídeos, que representam várias enzimas, que por sua vez catalisam reações em sequência. Nesse caso é possível construir uma via metabólica compreendendo essas reações em cadeia. Portanto, com um maior número de dados, é possível construir uma rede metabólica completa para um organismo, ou generalizada, para mais de um organismo.

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

Reactome¹⁷ 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 seus sistemas, como reprodução e metabolismo, por exemplo. Algumas sub-redes estão 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.

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 3.10 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 3.9:

¹⁶Sequências de nucleotídeos ou de aminoácidos são geralmente armazenadas em um arquivo em formato FASTA em texto ASCII.

¹⁷Disponível pela web, em <http://www.reactome.org>.

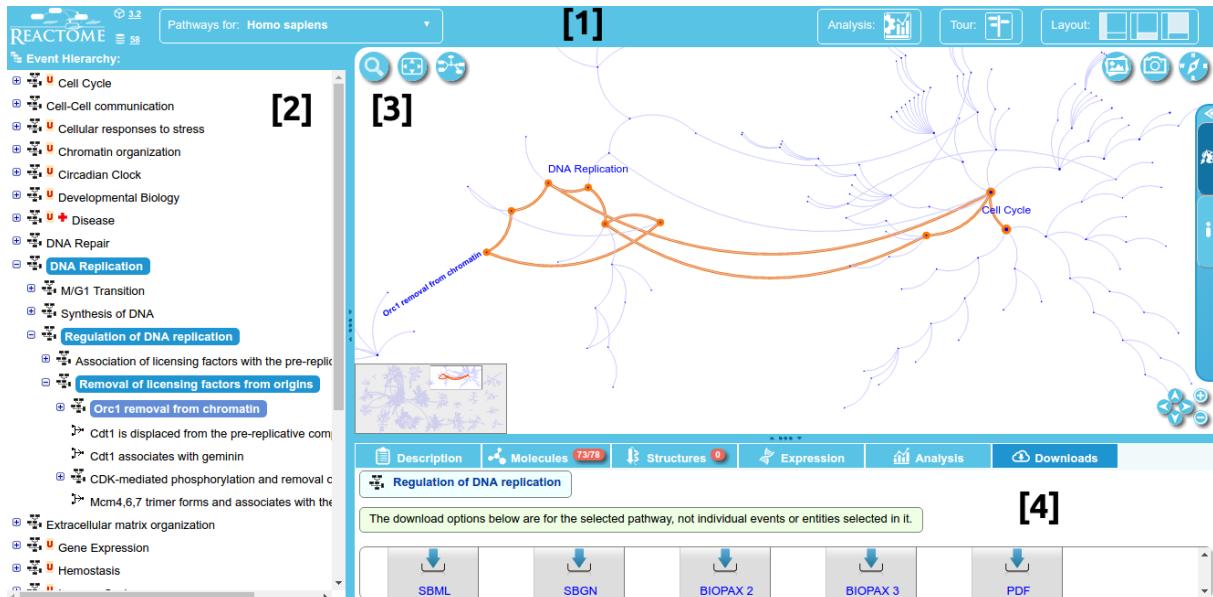


Figura 3.9: Visão geral da rede metabólica do *Homo sapiens* do Reactome, com destaque na via metabólica de remoção do Orc1 da cromatina do processo de regulação da replicação do ADN. A interface pode ser dividida nas seções [1], [2], [3] e [4].

- [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 a 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;
- [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-mapa 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

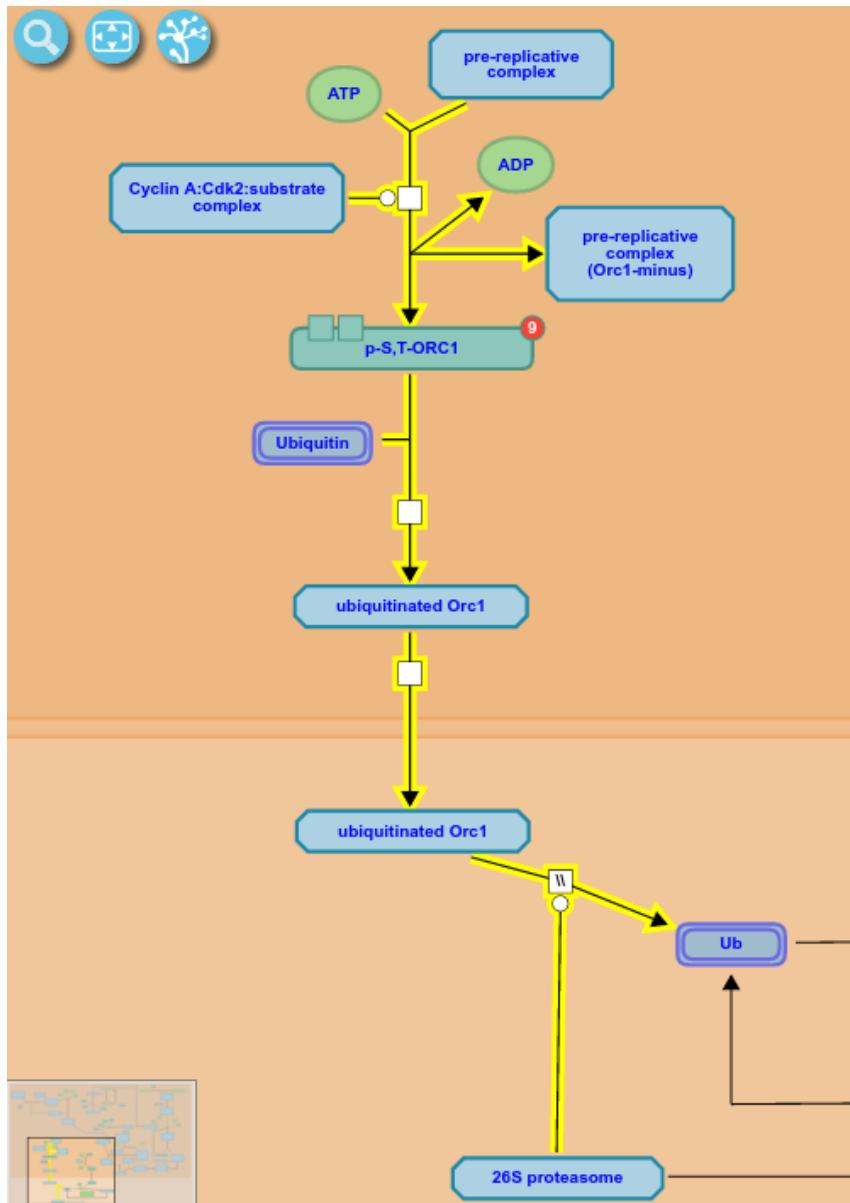


Figura 3.10: Via metabólica de remoção do Orc1 da cromatina apresentada pela ferramenta de visualização do Reactome.

das análise feita em [1]. Finalmente o botão Downloads oferece arquivos das vias para baixar em diversos formatos tais como SBML e PDF.

KEGG

O KEGG¹⁸ (*Kyoto Encyclopedia of Genes and Genomes*) é uma base de informações sobre sistemas biológicos em nível molecular, sobretudo sobre conjuntos de dados em

¹⁸Disponível pela web, em <http://www.kegg.jp/>.

larga escala gerados por sequenciamento de genoma [2]. As informações sobre os sistemas podem ser dadas em forma de módulos¹⁹, em forma de *brite*²⁰ ou em forma de vias²¹.

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.

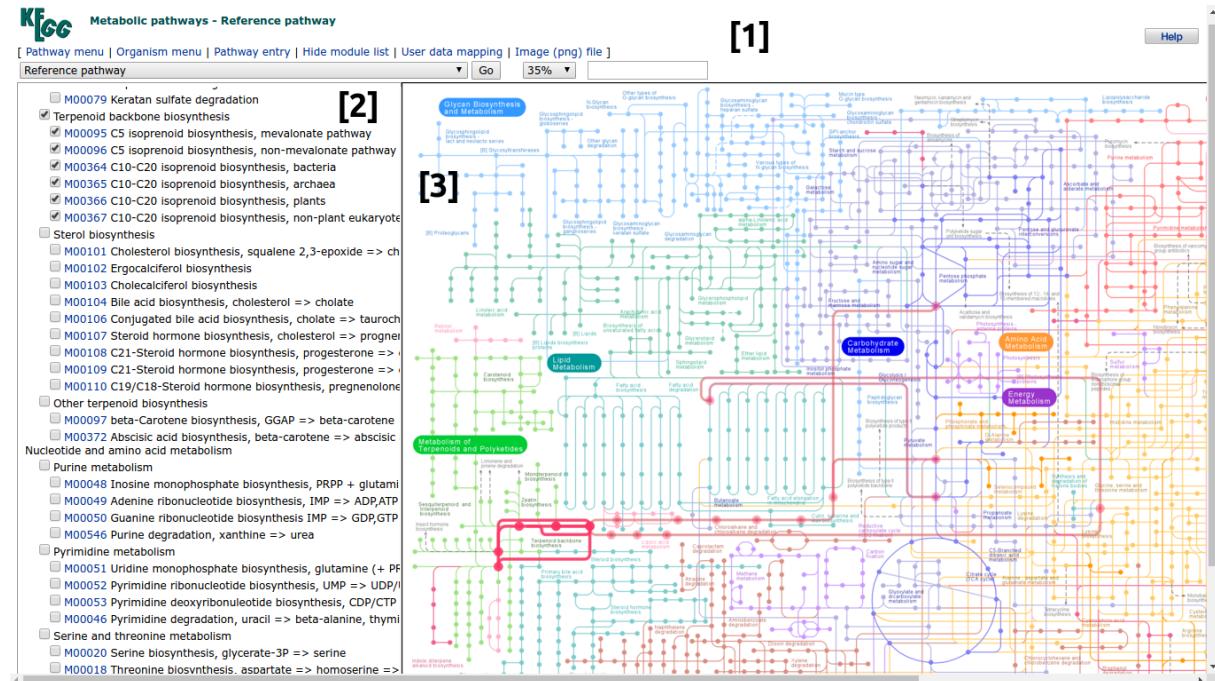


Figura 3.11: Visão geral da rede metabólica de referência do KEGG, com destaque na via metabólica de biossíntese do *backbone* de terpenóides. A interface pode ser dividida nas seções [1], [2] e [3].

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 3.11:

[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 à uma

¹⁹Unidades funcionais com identificação otimizada para análise dos dados.

²⁰Coleção de arquivos estruturados hierarquicamente sobre as funções das entidades biológicos.

²¹Mapa de interações moleculares e reações químicas.

página com uma lista de todos os organismos conhecidos pelo KEGG (346 eucariontes e 4.159 procariotas²²). 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;

- [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.

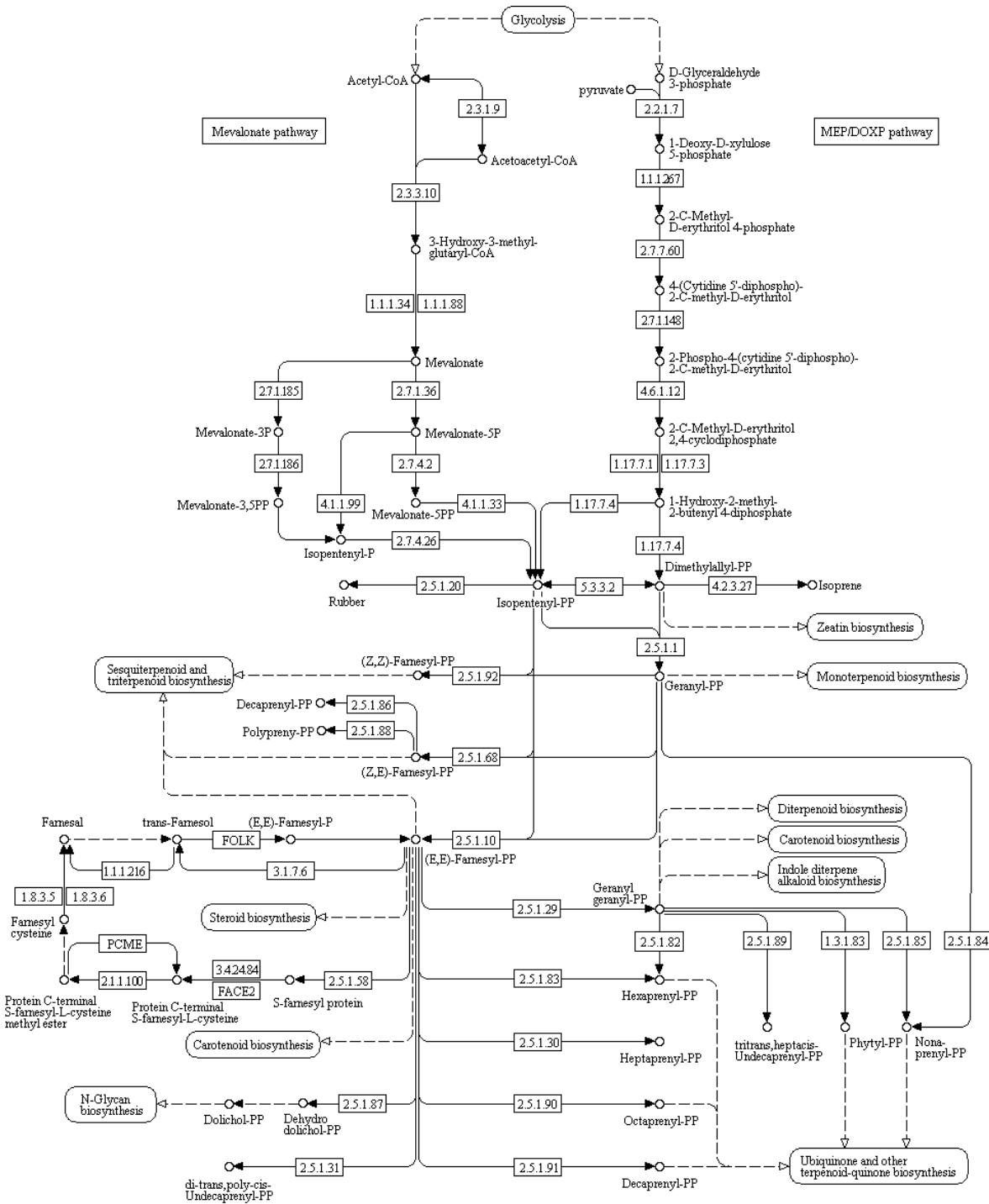
Caso o usuário queira visualizar a(s) via(s) metabólica detalhadamente, o KEGG oferece uma lista de diagramas²³ desenhados manualmente em formato de grafo interativo cujos nós, quando clicados, redirecionam para uma página contendo a descrição do objeto, que pode ser enzima, composto químico ou outra rede metabólica, por exemplo. Continuando com o exemplo da biossíntese do *backbone* de terpenóides, para acessar seu diagrama, apresentado na Figura 3.12, 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

²²Visitado em 2016-10-06

²³Detalhes da notação dos diagramas: http://www.genome.jp/kegg/document/help_pathway.html

TERPENOID BACKBONE BIOSYNTHESIS



00900 4/18/16
(c) Kanehisa Laboratories

Figura 3.12: Via metabólica de biossíntese do *backbone* de terpenóides apresentada pela ferramenta de visualização do KEGG.

BioCyc

O BioCyc²⁴ é um sistema de coleção de aproximadamente 7 mil bancos de dados chamados PGDBs (*Pathway/Genome Databases*), que 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 organizados 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 MetaCyc²⁵. 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 vias²⁶ classificado hierarquicamente com base em suas funções biológicas e na classe dos metabólitos produzidos consumidos. O acesso à via da Figura 3.6, 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 3.13 é uma modificação da Figura 3.6 apresentando diretamente o conjunto destes elementos.

²⁴Disponível pela web, em <http://biocyc.org>.

²⁵Disponível pela web através do site <http://metacyc.org/>.

²⁶Na última visita, em 2016-10-06, a página possuía 3.461 vias metabólicas.

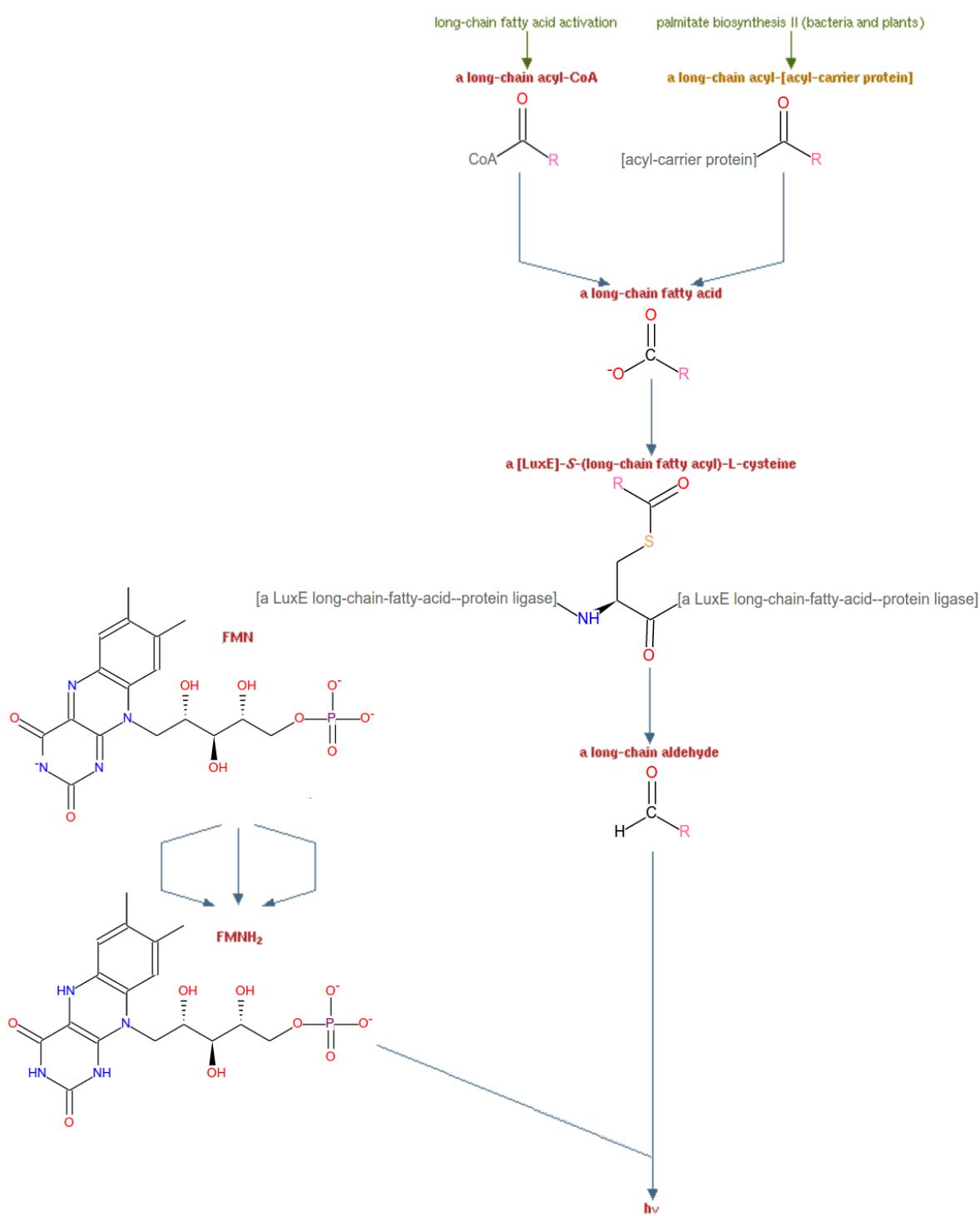


Figura 3.13: Representação modificada da Figura 3.6 com maior nível de detalhes apresentada pelo MetaCyc.

3.3.2 2Path

O *2Path* é um banco de dados de redes metabólicas de terpenoides. Ele possui milhares de reações de metabolismos secundários, preservando, assim, as principais características de biossíntese dos terpenos [24]. O 2Path foi modelado em uma banco de dados não-relacional baseado em grafos, onde os nós são enzimas, reações, cofatores, compostos, compartimento celular e organismos. A relação entre os nós é apresentada na Figura 3.14.

O *2Path* foi constituído a partir de dados obtidos de outros bancos. Dessa forma, a maioria dos nós e arestas possuem propriedades que fazem referências aos mesmos. A enzima, por exemplo, possui um nome, um número EC²⁷ e referências para seus dados nos bancos *Universal Protein Resource* (Uniprot) e Brenda - *Comprehensive Enzyme Information system*.

A modelagem em grafo do 2Path foi feita de tal forma a permitir a obtenção de respostas das seguintes questões biológicas [24]:

- Dado x_1, x_2, \dots, x_n metabólitos, existe alguma via metabólica que os ligue à y_1, y_2, \dots, y_m produtos?
- Para qual organismo?
- Baseado na similaridade, qual a localização celular onde ocorre uma reação específica?
- Em qual tecido um metabólito particular é armazenado?

A *feature* mais nova do *2Path*, que ainda está em fase de teste, é um sistema que permite a construção de uma rede metabólica de terpenoide a partir do genoma de um organismo. Para isso, os usuários do sistema deveriam fazer *upload* de seus arquivos FASTA, cada um contendo sequências de nucleotídeos do genoma um organismo. Então, o banco retornaria uma rede metabólica para cada organismo, construídos de acordo com as reações que as enzimas mapeadas pelas sequências catalisam.

Entretanto, o *2Path* ainda não possui uma interface para lidar com os usuários desse sistema que está sendo desenvolvido. Este trabalho propõe uma interface que permita que os usuários obtenham as respostas das questões biológicas que o banco de dados 2Path é capaz de responder.

²⁷O número *Enzyme Commission* (EC) é um esquema de classificação de enzimas.

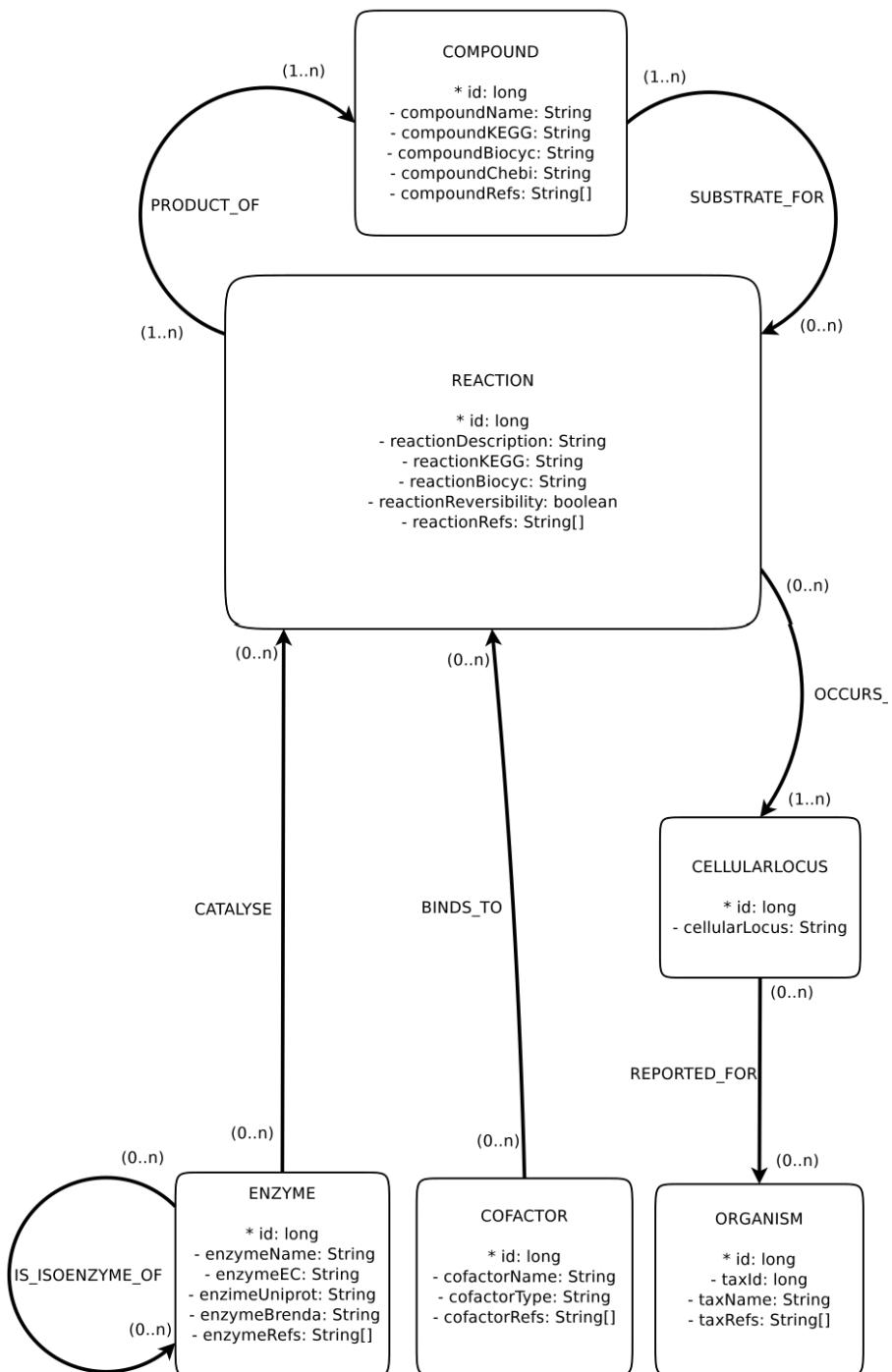


Figura 3.14: Modelo do banco de dados 2Path.

Capítulo 4

Interface do 2Path

Este capítulo apresenta uma interface para o *2Path*, utilizando as informações dos capítulos 2 e 3. São descritos quatro elementos de IHC utilizados para definir o escopo do trabalho. A tabela de interações representa de forma simples a metacomunicação. A partir dela é possível estabelecer os objetivos finais e instrumentais, organizados em um diagrama chamado mapa de objetivos. Uma vez conhecendo os motivos de os usuários utilizarem o sistema, o projetista utiliza os diagramas de modelagem de tarefas para representar de forma abstrata a navegação pelo sistema do ponto de vista de cada tipo de usuário. Por fim, a tabela de tratamento de rupturas na comunicação tem o objetivos de documentar os métodos utilizados para impedir / prevenir / recuperar erros cometidos pelo usuário ao interagir com o sistema.

Na Seção 4.1, são descritos de maneira geral todas as ferramentas, softwares e linguagens utilizadas na elaboração de uma interface para o *2Path*. De maneira breve, a interface do 2Path foi desenvolvida na IDE Eclipse utilizando a linguagem Java para *Back-end* e XHTML, JSF e Primefaces para *Front-end*. O banco de dados utilizado foi o Neo4j [24], um banco não relacional direcionado a grafos. A visualização das vias metabólicas foi implementada em javascript, com uma biblioteca de grafos e gráficos chamada D3 (Data-Driven Documents);

Na Seção 4.2, são descritos o ambiente de desenvolvimento e os detalhes de implementação do *back-end*, e *front-end* do sistema.

Na Seção 4.3, é detalhada a coleta de dados do Método de Avaliação de Comunicabilidade com os três usuários biólogos.

4.1 Projeto de Interfaces

O conjunto de diagramas do projeto de interface apresentam um modelo de elementos do sistema a ser seguido na implementação de fato das páginas *web*. Esses elementos devem ter o propósito de permitir que o usuário atinja seu objetivo, conservando os quatro critérios de qualidade: usabilidade, experiência do usuário, acessibilidade e comunicabilidade.

4.1.1 Tabela de Interações

De acordo com as especificações do capítulo 2, uma tabela de interações deve conter uma simulação de conversa entre o projetista da interface e o usuário que irá interagir com ela. A Tabela 4.1 apresenta este conteúdo de maneira simples, de forma a facilitar o entendimento do que se é requisitado pelo usuário.

Após a criação a tabela, ficou claro quais eram os campos de entrada e saída esperados pelo usuário. Assim, a partir dessa informação foi possível construir mais facilmente os demais diagramas.

4.1.2 Mapa de Objetivos

O 2Path foi desenvolvido considerando que um usuário já fez *login* no sistema e já possui pelo menos um organismo em seu banco de dados particular. Nesse sentido, o usuário possui dois objetivos finais:

- 1 : Verificar se existe uma via metabólica entre dois compostos no banco de dados público do 2Path e/ou em algum de seus organismos privados;
- 2 : Verificar se existe uma enzima específica no banco de dados público do 2Path e/ou em algum dos seus organismos privados.

Para atingir tais objetivos, ele precisa realizar certos objetivos instrumentais diretos e indiretos. Observe que a manipulação do grafo na visualização da via metabólica envolve clicar, mover e passar o mouse por cima nos nós e arestas.

Direto : Informar dois elementos: substrato e produto;

Direto : Informar enzima;

Direto : Manipular grafo para obter informações sobre os nós e arestas;

Indireto : Fazer *login* no sistema (Considerado já feito);

Indireto : Fazer *upload* de arquivos FASTA de organismos (Considerado já feito).

A Figura 4.1 apresenta o Mapa de Objetivos do sistema 2Path construído com base nos objetivos finais e instrumentais (diretos e indiretos) dos usuários.

4.1.3 Modelagem de Tarefas

A partir do mapa de objetivos, têm-se uma ideia geral da navegação do site, uma vez que foi feito um mapeamento generalizado das sequências de passos para atingi-los. O propósito da modelagem de tarefas é, portanto, formularizar essa sequências de passos de maneira específica para cada objetivo final.

No caso, foram elaborados dois objetivos finais: Verificar via metabólica entre dois metabólitos no banco de dados; Verificar existência de enzima no banco de dados. Entretanto a pesquisa pode ser feita tanto no banco de dados completo do 2Path, quanto na rede metabólica construída para um organismo específico. Assim, esses objetivos finais

Tabela 4.1: Representação da interação entre usuário (U) e projetista (P). Os signos representam o foco de cada conversa.

Tópico >Subtópico (diálogo)	Falas e Signos U: Usuário e P: Projetista
Pesquisar enzima	U: Quero procurar uma <i>enzima</i> no banco de dados 2Path.
> Informar dados da enzima	<p>P: Qual o número EC (<i>Enzyme Commission</i>) da enzima?</p> <p>U: O número EC é (...).</p> <p>P: OK. A enzima está no banco de dados e ela catalisa as reações (...) cujos substratos são (...) e os produtos são (...).</p>
Pesquisar enzima em organismo	U: Quero saber se o genoma de um dos meus organismos possui sequencia que produz certa enzima.
> Informar organismo	<p>P: Em qual dos seus organismos você quer buscar essa enzima?</p> <p>U: O organismo é (...).</p>
> Informar dados da enzima	<p>P: Qual o número EC da enzima?</p> <p>U: O número EC é (...).</p> <p>P: OK. A está no banco de dados e ela catalisa as reações (...) cujos substratos são (...) e os produtos são (...). O organismo (...) possui as sequências (...) que produziram tal enzima.</p>
Procurar caminho entre dois metabólitos	U: Quero saber se um certo metabólito é substrato de alguma via metabólica no 2Path cujo produto é um outro certo metabólito.
> Informar dados dos metabólitos	<p>P: Qual o substrato? Qual o produto?</p> <p>U: O substrato é (...) e o produto é (...).</p> <p>P: OK. Existe uma via que liga estes dois metabólitos. As reações (...) e os compostos (...) estão entre eles.</p>
Procurar caminho entre dois metabólitos em organismo	U: Agora quero verificar se há uma via metabólica entre dois metabólitos em um dos meus organismos.
> Informar organismo	<p>P: Em qual dos seus organismos você quer buscar essa enzima?</p> <p>U: O organismo é (...).</p>
> Informar dados da metabólitos	<p>P: Qual o substrato? Qual o produto?</p> <p>U: O substrato é (...) e o produto é (...).</p> <p>P: OK. Existe uma via que liga estes dois metabólitos. O organismo possui as sequências (...) que geram as enzimas (...) que, por sua vez, catalisam as reações. Estas são as reações que compõem a via metabólica e os compostos (...) são seus substratos e produtos.</p>

foram subdivididos em dois na modelagem de tarefas: Verificar se enzima está no banco de dados público do 2Path (Figura 4.2); Verificar se enzima está no banco de dados privado (Figura 4.3); Verificar se há via entre dois metabólitos no banco de dados público (Figura 4.4); Verificar se há via metabólica no banco de dados privado (Figura 4.5);

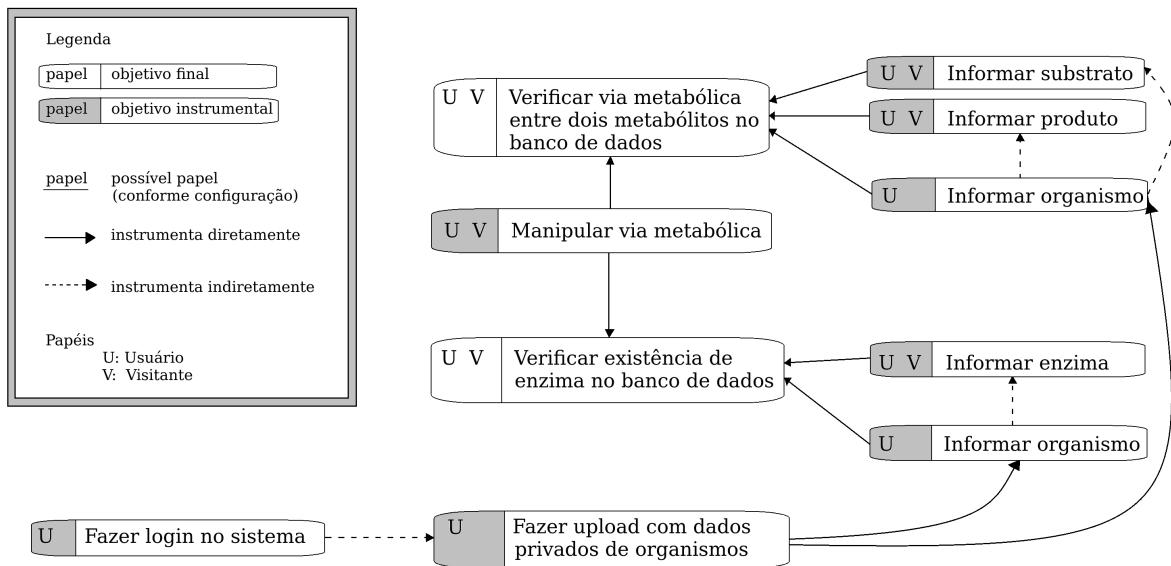


Figura 4.1: Mapa de objetivos finais e intrumentais.

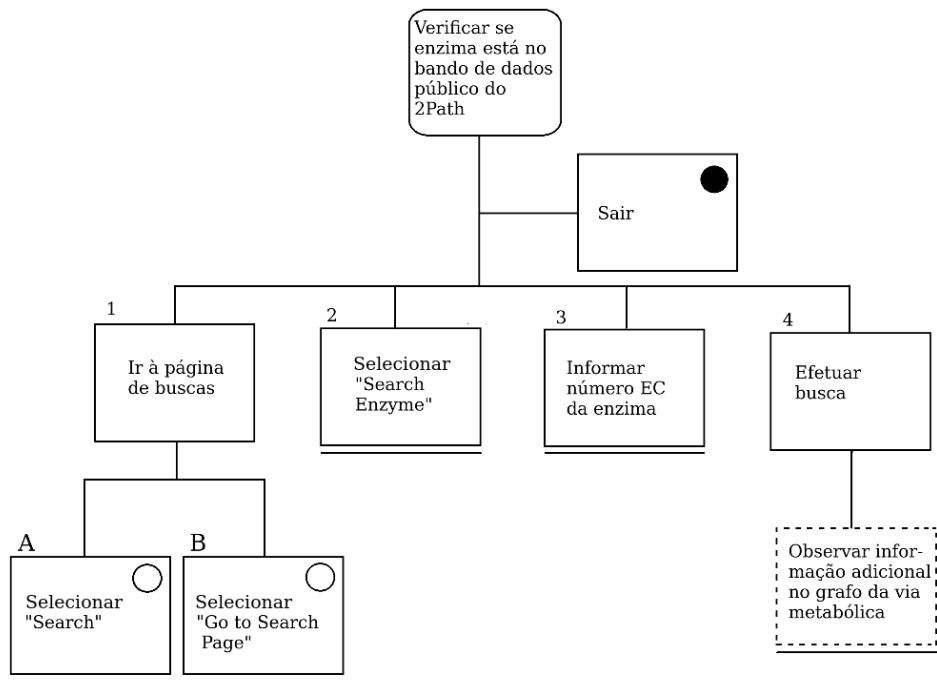


Figura 4.2: Objetivo de pesquisar enzima no banco de dados completo do 2Path.

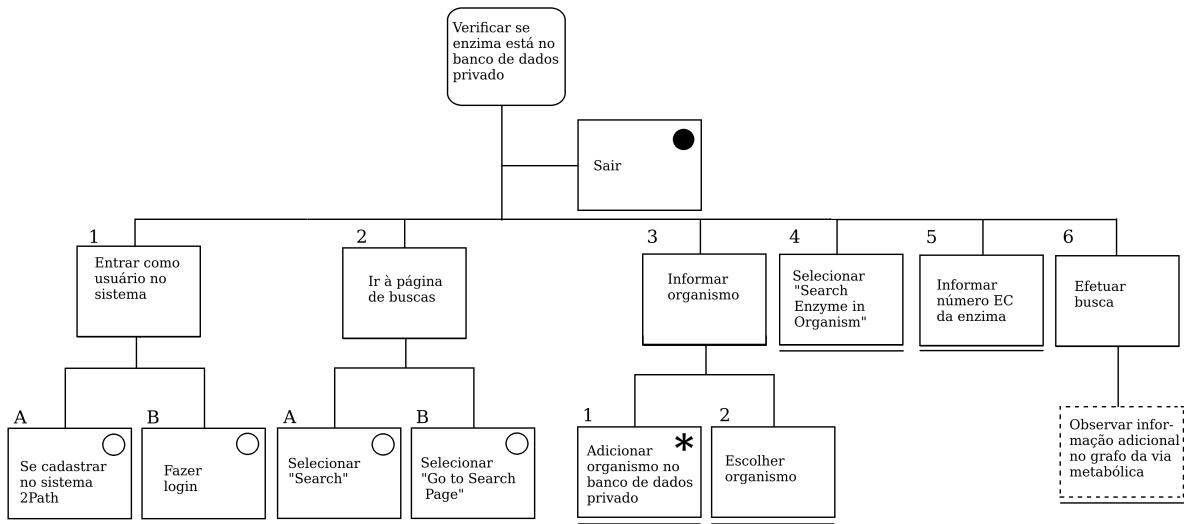


Figura 4.3: Objetivo de pesquisar enzima no banco de dados específico para a rede metabólica construída a partir do organismo do usuário.

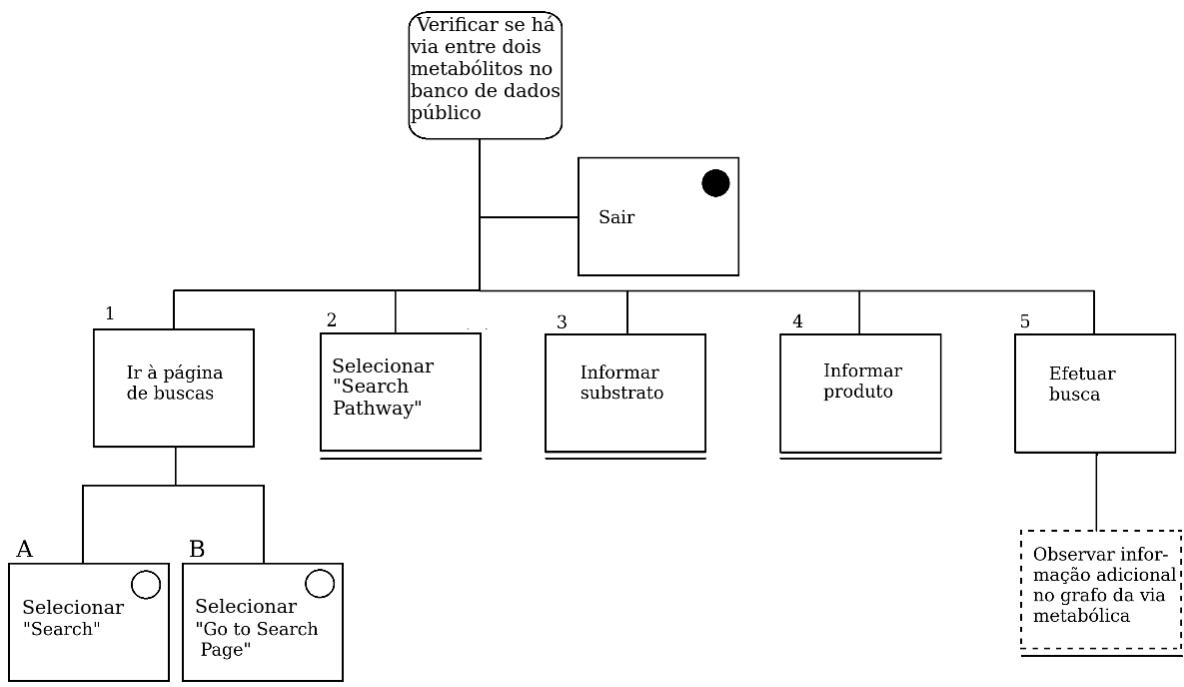


Figura 4.4: Objetivo de pesquisar via metabólica entre substrato e produto no banco de dados completo do 2Path.

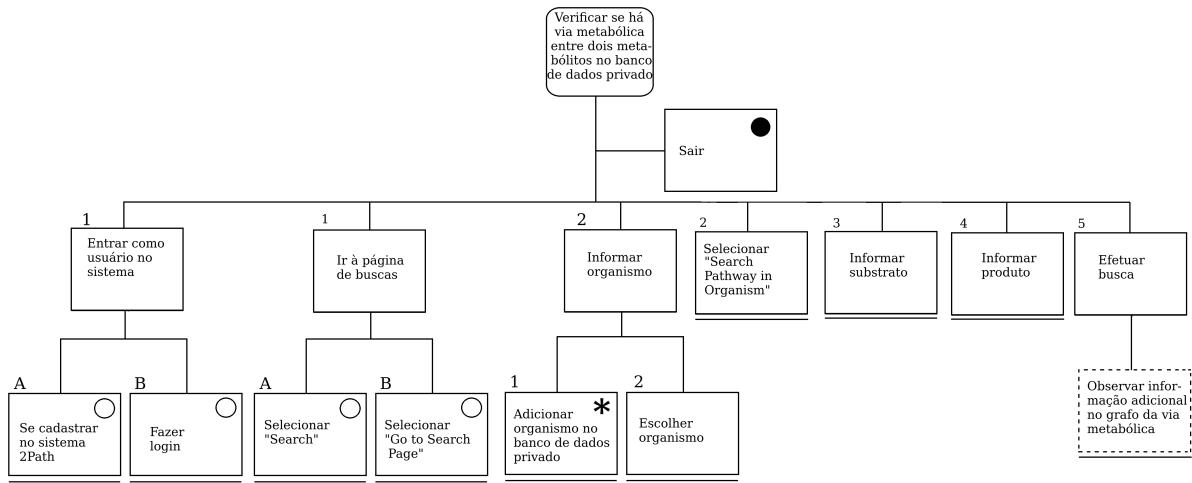


Figura 4.5: Objetivo de pesquisar via metabólica entre substrato e produto no banco de dados específico para a rede metabólica construída a partir do organismo do usuário.

4.1.4 Tratamento de Rupturas na Comunicação

A Tabela 4.2 apresenta as medidas tomadas para prevenção e recuperação de falhas de comunicação na interface do 2Path, assim como especificado no Capítulo 2. Nela, PP significa a prevenção passiva, PA significa prevenção ativa e RA significa recuperação apoiada.

Não houve nenhum tipo de prevenção apoiada, pois não foi observado tipo de interação que o usuário possa fazer que cause dano grave como, por exemplo, excluir um arquivo. Quando este projeto evoluir, esse tipo de prevenção pode ser aplicada, por exemplo, ao se tratar o *upload* de um arquivo FASTA, onde o projetista deverá verificar se existe um organismo com o mesmo nome já no banco de dados privado do usuário.

Também não foi constatado a necessidade de implementar captura de erro, uma vez que todas as possíveis rupturas observadas são tratáveis por meio de mensagens de erro.

4.2 Implementação da Interface

O sistema desenvolvido para este projeto é uma aplicação web chamada *EnzymeGraph*. 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.

4.2.1 Detalhes de Implementação

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),

Tabela 4.2: Campos de entrada e manipulação dos usuários do sistema 2Path

Signo	Prevenção	Recuperação
Seleção de organismo	PA: Campo obrigatório com indicador “*” para formulários “ <i>Search Enzyme in Organism</i> ” e “ <i>Search Pathway in Organism</i> ”. Caso o usuário esqueça de selecionar este campo, o organismo escolhido será o primeiro e ordem alfabética.	-
Seleção de enzima	PP: Campo obrigatório com indicador “*”; PP: Campo possui função <i>auto-complete</i> , que fornece ao usuário todos os números EC existentes	RA: Mensagem de texto em vermelho aparece acima do formulário informando o usuário de que o campo obrigatório não foi preenchido: “ <i>Enzyme: Validation Error: Value is required</i> ”.
Seleção de substrato	PP: Campo obrigatório com indicador “*”; PP: Campo possui função <i>auto-complete</i> , que fornece ao usuário todos os compostos existentes no 2Path	RA: Mensagem de texto em vermelho aparece acima do formulário informando o usuário de que o campo obrigatório não foi preenchido: “ <i>From: Validation Error: Value is required</i> .”. O “From” significa que a via tem início neste metabólito
Seleção de produto	PP: Campo obrigatório com indicador “*”; PP: Campo possui função <i>auto-complete</i> , que fornece ao usuário todos os compostos existentes no 2Path	RA: Mensagem de texto em vermelho aparece acima do formulário informando o usuário de que o campo obrigatório não foi preenchido: “ <i>To: Validation Error: Value is required</i> .”. O “To” significa que a via finaliza neste metabólito
Manipulação da via metabólica	-	-

foi utilizada o Apache Maven¹. Este *software* opera sobre o arquivo *pom.xml* (*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 Javascript, para a construção do grafo de vias metabólicas.

Em javascript, foi utilizada a biblioteca D3 na apresentação das vias metabólicas na tela. Essa biblioteca permite que a construção de um grafo a partir de um arquivo ou *string* no formato JSON com os vetores *nodes* para os nós e *links* para as arestas. A figura 4.6 apresenta as cores escolhidas para cada nó do grafo. Não houve padrão definido para a escolha das cores.

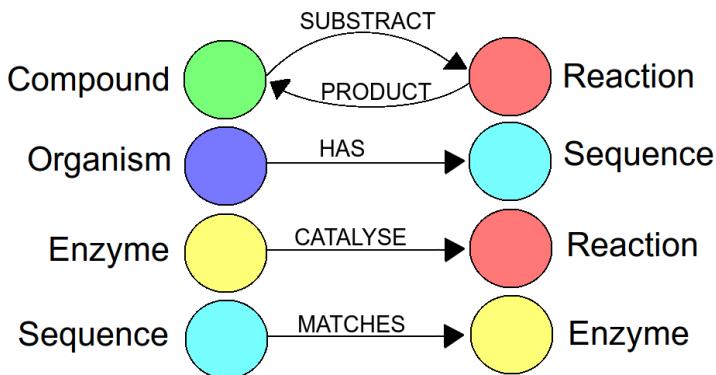


Figura 4.6: Diagrama geral de uma via metabólica na interface do 2Path.

O banco de dados em grafo Neo4j foi utilizado no sistema, pois o próprio 2Path foi desenvolvido no mesmo. Devido à facilidade de implementação do mesmo em Java e da alta usabilidade de sua interface, foi relativamente simples implementar as *queries* de busca e fazer a conexão em Java. As *queries* foram escritas em *CYPHER*, linguagem nativa do Neo4j. Como o banco não é relacional, essa linguagem não tem a estrutura do SQL, porém foi elaborada para facilitar buscas que realizam travessia em grafos.

O Neo4j, porém, retorna em suas transações uma string em formato JSON que não é compatível com o especificado pelo D3. Nesse caso, foi feito um parser para a construção de uma nova string compatível. Os dados passados de uma página para a outra, inclusive essa string em JSON, são todos armazenados em uma sessão, cujo ciclo de vida é determinado pelo JSF.

4.2.2 Interface do Sistema

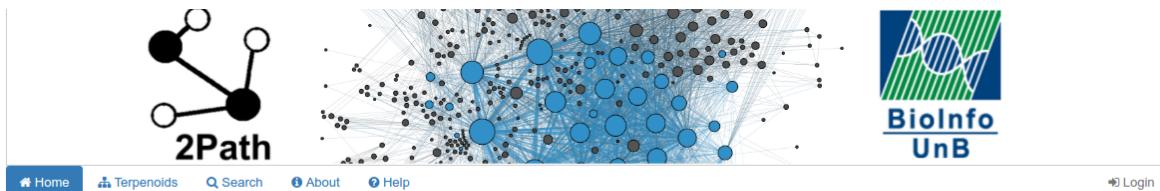
A implementação do *front-end* resultou nas figuras a seguir. Todos os elementos estão em inglês, pois o sistema criado pode ser utilizado mundialmente. A Figura 4.7 mostra a

¹Software 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.

página inicial, que possui uma breve explicação sobre terpenoides, sobre o sistema 2Path e sobre as possíveis buscas que o usuário pode realizar. O menu ao topo está em todas as páginas do sistema. Nele é possível navegar até a página inicial (Home), página sobre terpenoides (Terpenoids), página de busca (Search), página sobre o 2Path (About) e, por fim, página de ajuda (Help). A Figura 4.8 apresenta a página de busca principal, onde a partir dela é possível fazer as quatro tipos de buscas: por enzima no 2Path; por enzima no organismo; por via metabólica no 2Path e por via metabólica no organismo.



About 2Path

2Path is a **terpenoid** metabolic network modeled and stored using a graph database, which aims to preserve important terpenoid biosynthesis characteristics and mitigate challenges in genome-scale storage metabolic networks.

[View details »](#)

Terpenoid

Terpenoids have critical ecological, industrial and commercial relevance. Interactions as signaling for communication intra/inter species, signal molecules to attract pollinating insects, or defense against herbivores and microbes, are protagonized by terpenoids. Due to their chemical composition, many terpenoids possess vast pharmacological applicability in medicine and biotechnology. The biosynthesis of terpenes has been widely studied over the years and today it is known that they can be synthesized from two metabolic pathways: Mevalonate pathway (MVA) and Non-mevalonate pathway (MEP).

[View details »](#)

2Path System

1. Go to the Search Page
2. Search for

Public

- A keyword in our database
- A pathway between two components in our database

Private

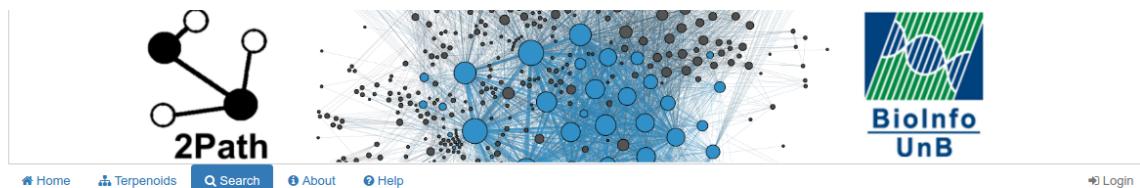
- A keyword in your organisms merged with our database
- A pathway between two components in your organism merged with our database

3. Visualize the data with the 2Path graph

[Go to Search Page](#)

[See Full Example](#)

Figura 4.7: Página inicial do 2Path.



Search in the 2Path database

When you don't select an organisms, you will search for all the enzymes from the 2Path database.

[Search Enzyme](#)

When you don't select an organism, you will search for all the possible pathways from one compound (substract) to another (product).

[Search Pathway](#)

Search in one Organism merged with 2Path database

Organism: *

Arabidopsis thaliana

Discover if the enzyme you are looking for is on your Organism by giving its EC number. Find out the reactions it can catalyze on terpenoids. The reactions come with their description, Kegg ID and link and type.

Visualize a pathway from a compound (substract) to another (product). If there is a pathway, then the genome of organism selected has the sequence for all the enzymes envolved on the reactions.

[Search Enzyme in Organisms](#)

[Search Pathway in Organism](#)

Figura 4.8: Página principal de busca no 2Path.

Em relação às buscas específicas, decidiu-se apresentar o grafo das vias metabólicas em uma página separada da busca, pois ele poderia ficar extenso e complexo. Nesse

sentido, a busca de fato é realizada em uma página intermediária, que apenas responde se a enzima/via foi encontrada ou não. Caso o usuário queira saber detalhes sobre a mesma, ele deve selecionar o botão *View Interactive Data*.

A Figura 4.9 apresenta um exemplo de busca pela enzima *2.7.7.60* no banco de dados completo do 2Path. Esta enzima foi encontrada e o grafo da Figura 4.10 apresenta o substrato (*2-C-Methyl-D-erythritol 4-phosphate*) que ela recebe e o produto (*4-(Cytidine 5'-diphospho)-2-C-methyl-D-erythritol*) que ela produz.

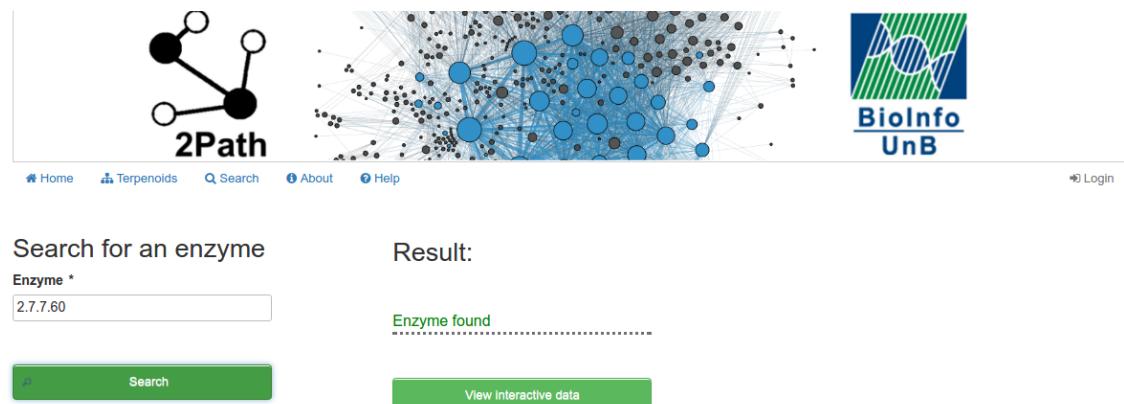


Figura 4.9: Página de busca por enzima no banco completo

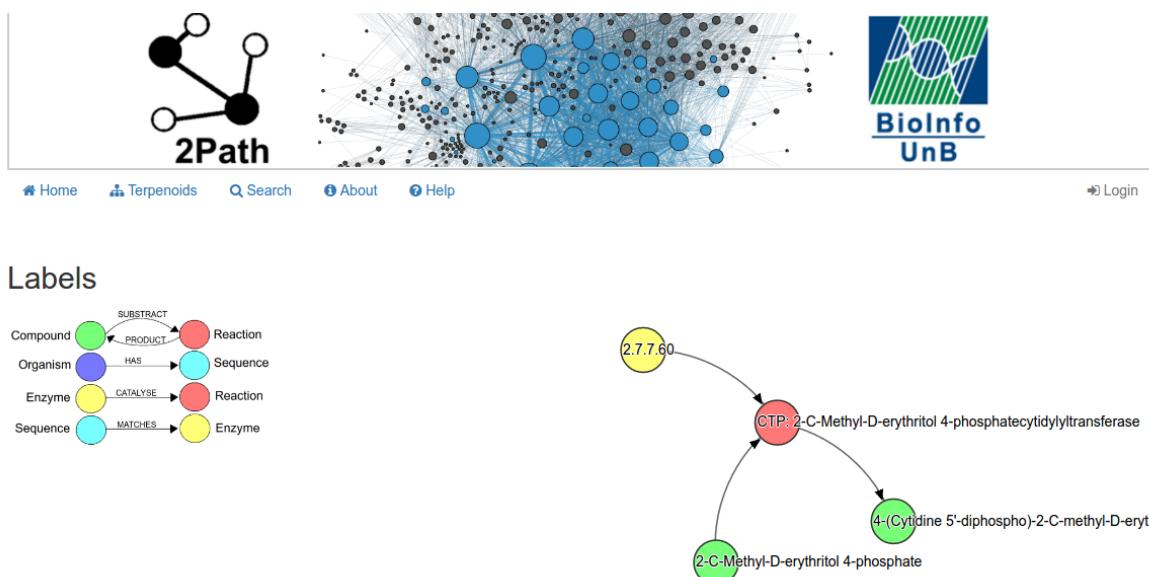


Figura 4.10: Página de apresentação da reação bioquímica catalisada pela enzima buscada.

A Figura 4.11 apresenta um exemplo de busca por via metabólica entre os compostos *1-Deoxy-D-xylulose 5-phosphate* e *2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-D-*

erythritol no organismo *Arabidopsis thaliana*. O grafo gerado para esta via é representado na Figura 4.12.

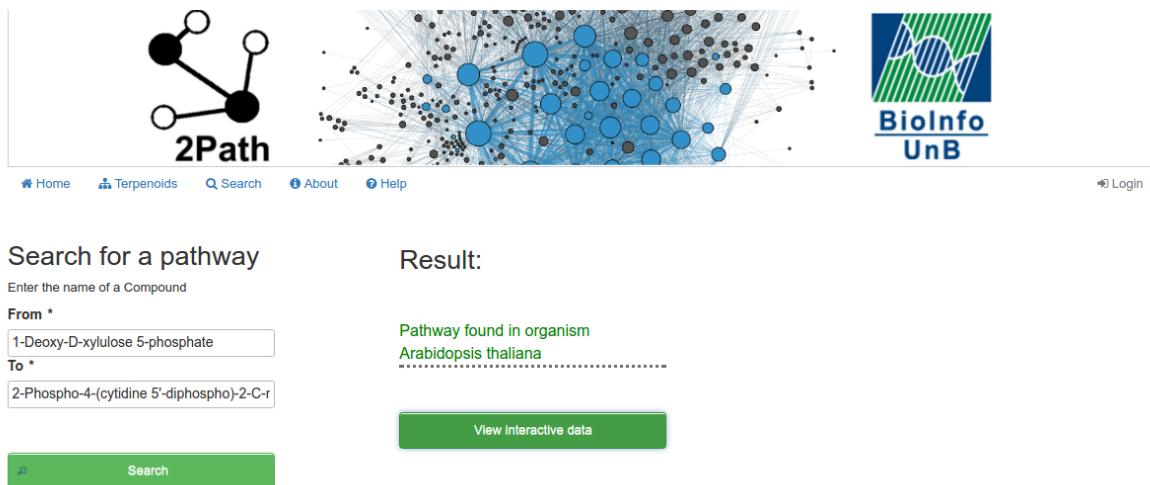


Figura 4.11: Página de busca por via metabólica no banco de dados formado para o organismo *Arabidopsis thaliana*.

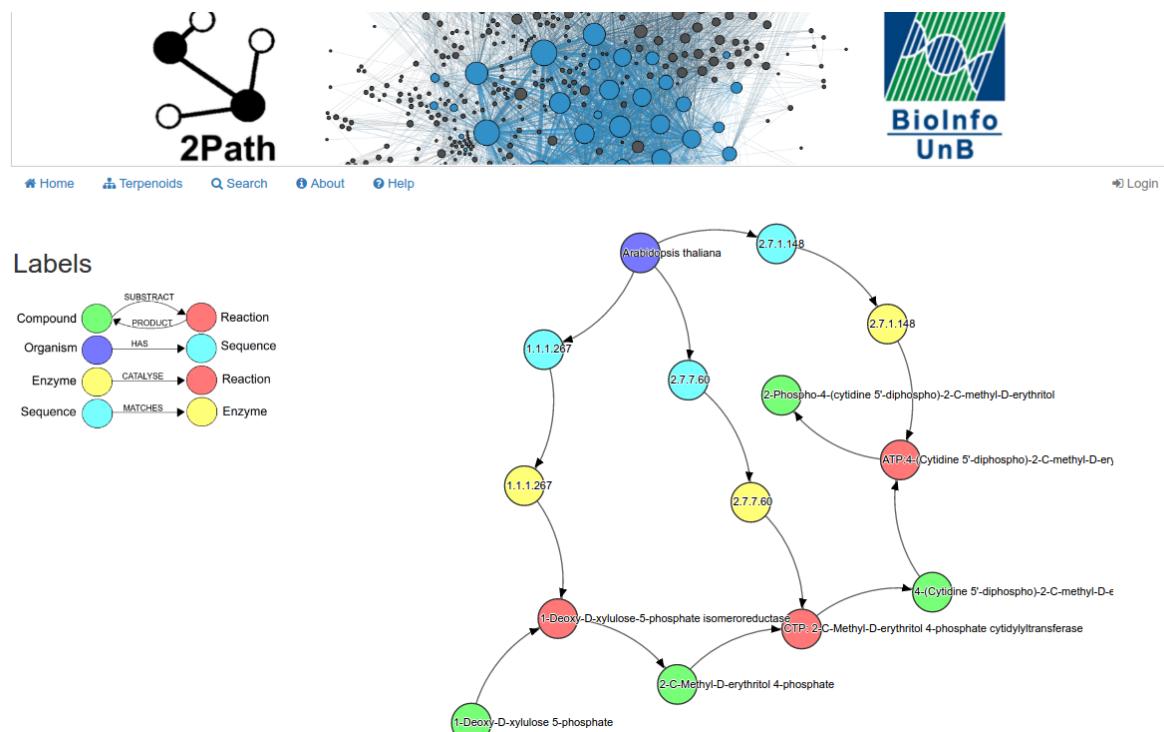


Figura 4.12: Página de apresentação da via metabólica de reações que ocorrem no organismo *Arabidopsis thaliana*.

4.3 Projeto de Avaliação

Os testes serão realizados no Laboratório de Bioinformática do Departamento de Biologia Celular, Instituto de Ciências Biológicas da Universidade de Brasília. Três biólogos foram selecionados para participar do teste descritos pelo Método de Avaliação de Comunicabilidade no capítulo 2. A identidade dos mesmos será preservada por questão de privacidade.

Primeiramente, cada biólogo deverá responder ao questionário do Anexo A. Ele é necessário para analisar se o perfil do usuário tem alguma relação com a facilidade com que o mesmo possui ao realizar um tarefa proposta no sistema. Para isso, é importante conhecer a formação acadêmica do usuário, bem como sua atuação profissional. Além disso, são solicitadas informações sobre a frequência em que ele utiliza ferramentas interativas e o nível de conhecimento a respeito de redes metabólicas.

Após responderem o questionário, cada usuário fará o teste separadamente, sem acesso às respostas de outro usuário para que o teste não perca a legitimidade. Apenas um avaliador guiará os usuários e apenas um computador será utilizado para os testes. Assim, serão realizados três testes em sequência. Será utilizado o computador pessoal do avaliador para a realização das tarefas, ao invés de um computador do Laboratório de Bioinformática, uma vez que o ambiente de teste já está configurado corretamente no computador pessoal. Configurar o ambiente em um outro computador poderia levar em torno de 2 horas, levando em consideração que seria necessário fazer *download* da IDE Eclipse e do banco de dados Neo4j.

Durante a realização das tarefas, serão filmadas a tela do computador, o *mouse / mousepad*, o teclado e as mãos do usuários. Serão apresentadas aos usuários, um de cada vez, as 13 etiquetas descritas no capítulo 2 para que eles entendam o propósito do teste. Não haverá comunicação entre avaliador e usuário durante a realização completa da tarefa, ou até que o usuário desista de cumprir o objetivo proposto.

Serão realizadas 5 tarefas, descritas no Anexo B. As primeiras três envolvem buscar por enzimas em um organismo. Considerou-se que o organismo já havia sido processado pelo sistema do 2Path e as sequências de seu arquivo FASTA já haviam sido ligadas às enzimas correspondentes do banco de dados. Assim, o usuário deverá apenas selecionar o organismo listado na página de busca e pesquisar pelas enzimas. As outras duas tarefas envolvem buscar por vias metabólicas no banco de dados inteiro do 2Path.

Após a tarefa, os usuários devem responder ao questionário do Anexo C, que visa coletar a opinião dos usuários a respeito do projeto de interface. Estes dados, juntamente com as etiquetas listadas para cada usuário, fornecem uma base de informação para análise e a re-implementação aperfeiçoada do sistema.

Por fim, será apresentado a nova proposta de interface do banco de dados 2Path, de acordo com as necessidades dos usuários e contemplando novos recursos que devem aumentar os quatro critérios de qualidade: usabilidade, experiência de usuário, acessibilidade e comunicabilidade.

Capítulo 5

Resultados dos testes

5.1 Avaliação da Interface

questionario 1, perfil de cada usuario. Definir como user 1 2 e 3; resultado etiquetas, pra user 1, 2 e 3. resultado questionario pra user 1, 2 e 3.

Tabela 5.1: Contabilização das etiquetas para cada usuário. Cada etiqueta é registrada no tempo do vídeo em que ela ocorreu.

Etiquetas	Usuário 1	Usuário 2	Usuário 3
Desisto.	-	-	-
Para mim está bom...	-	-	-
Não, obrigado.	-	-	-
Vai de outro jeito.	-	-	-
Cadê?	(01:22), (07:49)	(04:55)	-
Ué, o que houve?	(09:30)	-	-
E agora?	-	(02:01)	(01:46)
Onde estou?	-	-	-
Epa!	(09:01)	(00:41), (09:30), (10:45)	-
Assim não dá.	(5:10)	(01:29)	(01:05)
O que é isto?	(01:26), (03:50)	(02:20), (08:18)	(01:59)
Socorro!	(01:55), (07:53), (10:15)	(07:55), (10:01)	(06:00), (07:25)
Por que não funciona?	-	-	(04:20), (07:20)

Usuario 1

01:22 Usuário não encontrou as reações que a enzima catalisada, mas logo percebeu que deveria seguir para uma outra página para obter detalhes sobre a busca;

01:26 Usuário se deparou com um grafo em movimento, com 6 nós e cinco arestas, e não entendeu seu significado, a princípio;

01:55 Usuário precisou interagir com o grafo de maneira forçada, pois o mesmo não parava de se mover;

03:50 Usuário fala "Nossa, isso é muito ruim; Não dá pra saber se o substrato da enzima é verde (Composto) ou o rosa (Reação)";

- 05:10 Usuário tenta entender a legenda do grafo, mas discorda totalmente do proposto;
- 07:49 Usuário não encontra informação que buscava, até descobrir que deveria seguir para página de detalhes de busca, novamente;
- 07:53 Usuário expressa desgosto pelo grafo fechado que aparece ao iniciar a página de detalhes da via metabólica;
- 09:30 Usuário esqueceu de selecionar *Search* e percebeu que havia selecionado em ver detalhes da busca anterior;
- 10:15 Usuário logo clica no grafo para ele parar de se mover.

Usuario 2

- 00:41 Usuário percebeu que havia selecionado um botão que não o levava ao seu objetivo;
- 01:29 Usuário se confunde com função auto-complete do campo de enzimas;
- 02:01 Usuário demora alguns segundos para perceber que deve seguir para a página de detalhes de busca;
- 02:20 Usuário demora alguns segundos para entender, a partir da legenda, o significado do grafo que aparece na página de detalhes de busca.
- 04:55 Usuário não percebe que página retornou sucesso, pois não parece que ela foi atualizada;
- 07:55 Usuário tenta impedir grafo de se mover para fora de seu campo de visão;
- 08:18 Usuário tem dificuldade para entender significado biológico das arestas e nós do grafo;
- 09:30 Usuário esqueceu de selecionar *Search* e percebeu que havia selecionado em ver detalhes da busca anterior;
- 10:01 Usuário tenta impedir grafo de se mover para fora de seu campo de visão;
- 10:45 Usuário percebeu que havia errado a tarefa anterior e voltou para refazê-la.

Usuario 3

- 01:05 Usuário se confunde com função auto-complete do campo de enzimas. Ainda percebe que o problema não é o teclado, mas sim a função que aparece sem necessidade e atrapalha a busca;
- 01:46 Usuário diz "É só essa a informação?", pois não percebeu, a princípio, que para realizar a tarefa deveria seguir para a página de detalhes da busca;
- 01:59 Usuário não entende grafo que aparece na tela de detalhes e reclama muito por ele não ficar parado;
- 04:20 Usuário reclama da falta de respostas da interface;
- 06:00 Usuário acha um absurdo o tamanho do grafo que aparece na tela e pede a lista dos detalhes que ele gostaria de saber sobre a busca;

07:20 Usuário reclama da falta de respostas da interface;

07:25 Usuário acha um absurdo o tamanho do grafo que aparece na tela e pede a lista dos detalhes que ele gostaria de saber sobre a busca.

Conclusão:

Aplicação do método da avaliação da comunicabilidade

Pontos fortes:

- Os botões do menu ao topo do site facilitaram a navegação pelo site;
- A função auto-complete dos compostos facilitou muito a busca;
- A mensagem de sucesso ou fracasso sempre era reparada.

Pontos fracos:

- Busca

- 1 O botão de voltar um página, do navegador, não fazia o que o usuário esperava. Ele voltava para a pesquisa anterior ao invés de ir para a página central de buscas;
- 2 A função auto-complete das enzimas não foi uma boa decisão de projeto, pois os usuários ficavam confusos com os números que apareciam;
- 3 Resultado de busca não estava sendo inicializado toda vez que página é carregada, portanto apresentava um resultado anterior, o que causava confusão;
- 4 Botão "View Interative Data", para visualizar a via metabólica, esta sempre ativo. Isso fez com que os usuários não notassem que existe informação extra sobre enzima ou via quando a pesquisa retorna sucesso;
- 5 Botões "Search" e "View Interative Data" estavam muito perto um do outro. Muitos usuários não selecionavam o botão *Search* antes de clicar em "View Interative Data", pois a segunda coluna da página estava mal localizada, para um formulário;
- 6 A página de busca central dá mais foco em busca por elementos no organismo, e isso faz com que a seção menor à esquerda passe despercebida.

- Grafo

- 1 A reação não deveria ser representada por um nó, pois para os biólogos, ela é uma ação, e não um elemento. Assim, faz mais sentido biológico representar COMPOUND-[SUBSTRACT]->ENZIME e ENZIME-[PRODUCTOF]->COMPUND. A aresta "CATALISE" não é necessária. Houve muita confusão, pois pensaram que reação era representada por aresta invés de nó;
- 2 Não é necessário apresentar os nós do organismo e sequências que o organismo possui. Ao apresentar a enzima na busca de informações no organismo, já é implícito que aquela enzima é produzida pelo organismo;

- 3 Quanto maior o grafo, mais complexo é entendê-los;
- 4 Grafos muito grandes podem ficar cortados e alguns elementos podem sair do campo de visão do usuário.
- 5 É mais desejável um grafo que apareça primeiramente de maneira estática. Um grafo que aparece se movendo levemente é irritante. Caso o usuário queira mover e clicar nos elementos, ele o faz com mais facilidade com um grafo estático;
- 6 É mais agradável um grafo que apareça aberto;
- 7 As labels poderiam apresentar só as bolinhas e explicação a respeito do que elas representam. A informação sobre as arestas já é informação extra desnecessária;

Foi encontrado um *erro* de implementação em um dos testes. Uma das tarefas era encontrar uma via metabólica entre dois compostos no grafo completo do 2Path, porém um dos usuários utilizou a busca por via em um organismo. A busca deveria retornar que não há via, porém ela retornou que existia, e apresentou um grafo disjunto¹. Um grafo era do organismo com todas as enzimas que ele produzia e o outro grafo representava a via entre os compostos pesquisados.

Mudanças no projeto de interface após os testes:

- Busca

- 1 Foi adicionado um botão de voltar nas páginas específicas de busca por enzima e vias metabólicas, assim o usuário pode clicar nele e voltar diretamente para a página de buscas gerais;
- 2 A função *auto-complete* do campo de busca por enzima foi removido;
- 3 As mensagens de sucesso e erro são apresentadas no canto superior direito da tela e são atualizadas sempre que uma nova busca é feita. Ela aparece por 5 segundos e depois desaparece.
- 4, 5 O grafo que apresenta os detalhes sobre a enzima ou via metabólica buscada aparece na mesma página que o formulário de pesquisa. Se a busca retorna verdadeira, o grafo aparece, se retorna falsa, não aparece. Assim, o usuário não precisa seguir mais um passo na interface para visualizar os dados biológicos que procura;
- 6 As colunas de busca no banco completo e no organismo foram configuradas no mesmo tamanho e ícones que simbolizam elementos públicos e privados foram posicionados ao lado dos títulos de cada busca.

- Grafo

- 1 NÃO DÁ: Uma enzima pode catalisar mais de uma reação. Assim, não teria como representar o nó da enzima recebendo todos os possíveis substratos e produzindo todas os possíveis produtos. Seria mais fácil pedir pro waldeyr adicionar uma propriedade no nó da reação falando qual enzima catalisa a reação.

¹Dois grafos que não se conectam.

- 2, 3, 4 Para diminuir o tamanho do grafo, foram omitidos os nós referentes ao organismo e às sequências do genoma que ele possui.
- 5 O grafo aparece estático, porém o usuário pode clicar nos nós e movê-los como desejar;
- 6 Os nós do grafo não se sobrepõem, portanto ele tem a aparência de que está aberto;
- 7 A legenda apresenta apenas informações sobre as cores de três nós: Compostos, Reações e Enzimas;

5.2 Nova Interface do Sistema

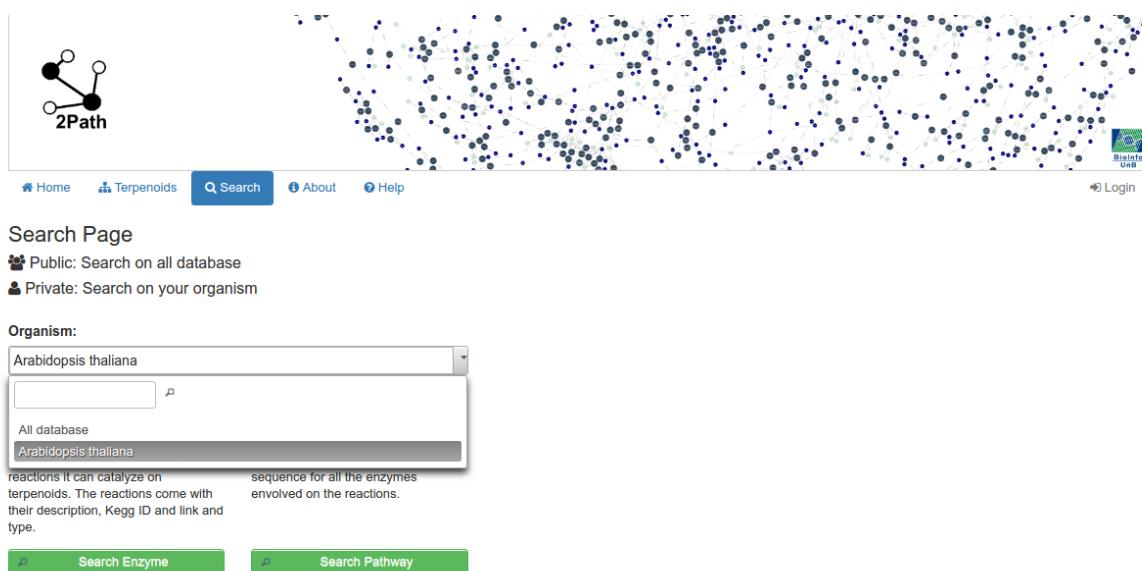


Figura 5.1

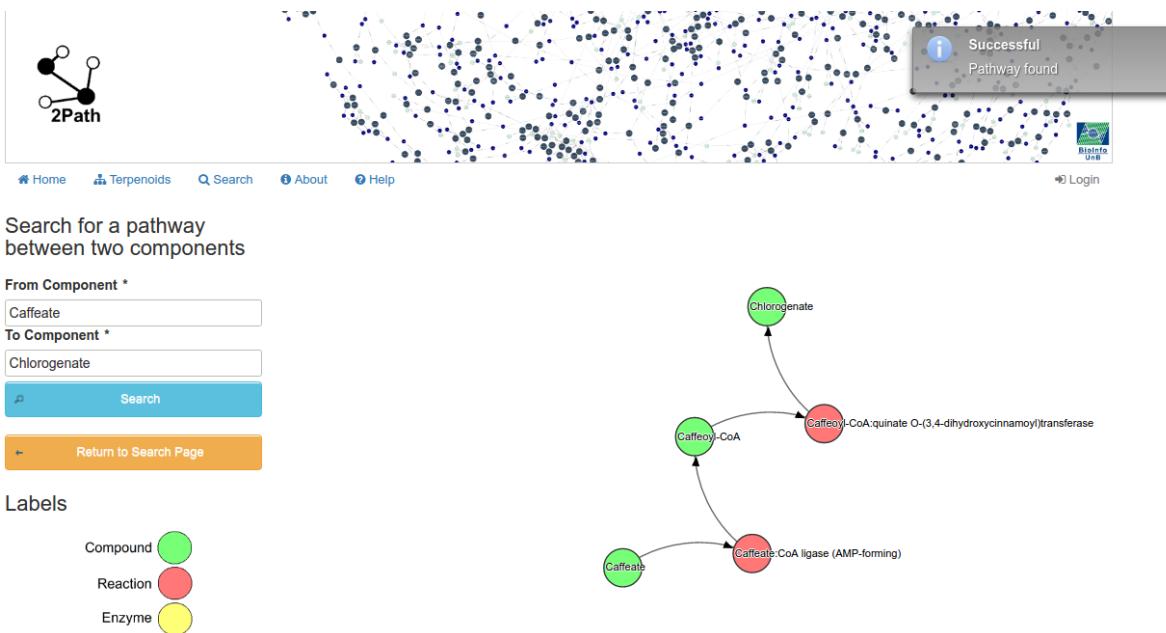


Figura 5.2



Figura 5.3

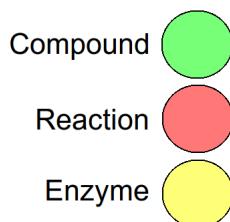
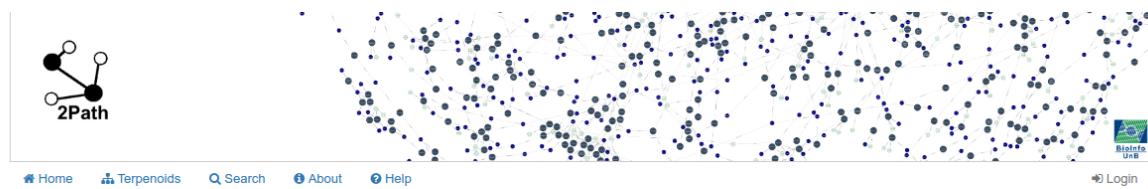


Figura 5.4



Search for a pathway
between two components

From Component *

1-Deoxy-D-xylulose 5-phosphate

To Component *

2-Phospho-4-(cytidine 5'-diphospho)-2-C-

Labels

Compound

Reaction

Enzyme

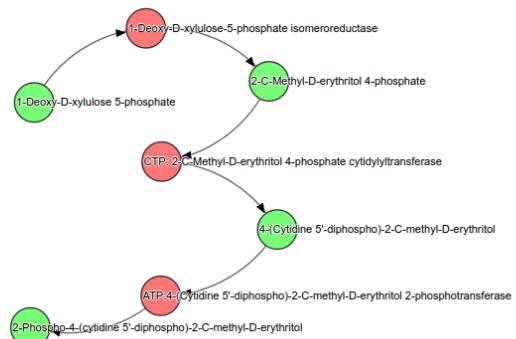
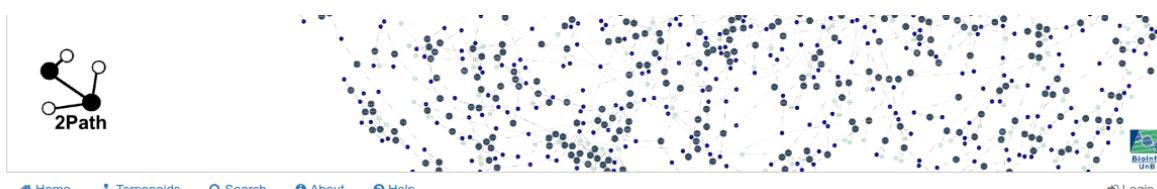


Figura 5.5



Search for an enzyme

EC Number *

2.3.1.133

Labels

Compound

Reaction

Enzyme

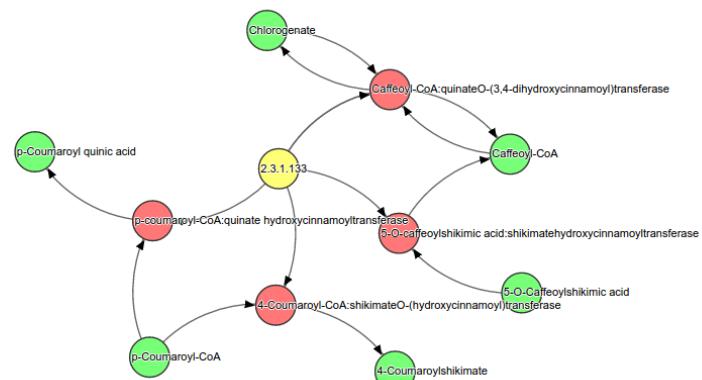


Figura 5.6: PROBLEMA: ENZIMA PODE CATALIZAR MAIS DE UMA REACAO

Capítulo 6

Conclusão e Trabalhos Futuros

Conclusão

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

Trabalhos Futuros

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

- x

Referências

- [1] Introduction to biocyc. <http://biocyc.org/intro.shtml>, visitado em 2016-10-01. 31
- [2] Kegg overview. <http://www.kegg.jp/kegg/kegg1a.html>, visitado em 2016-10-01. 28
- [3] Protein synthesis steps. <http://www.proteinsynthesis.org/protein-synthesis-steps/>, visitado em 2016-08-19. vi, 20
- [4] Usersguide. <http://wiki.reactome.org/index.php/Usersguide>, visitado em 2016-10-01. 25
- [5] Protein structure, 2009. <http://www.particlesciences.com/news/technical-briefs/2009/protein-structure.html>, visitado em 2016-01-02. vi, 18, 19
- [6] Márcio Santos Aleixo. Divisão celular. <http://www.infoescola.com/citologia/divisao-cellular/>, visitado em 2016-11-19. vi, 23
- [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, 10, 11, 12, 13, 15
- [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. 22
- [10] Ron Caspi. Metacyc pathway: bacterial bioluminescence. <http://metacyc.org/META/new-image?object=PWY-7723>, visitado em 2016-10-13. vi, 22
- [11] Krukemberghe Divino Kirk da Fonseca Ribeiro. Enzimas. <http://mundoeducacao.bol.uol.com.br/biologia/enzimas.htm>, visitado em 2016-11-20. 21
- [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/semiode.html>, visitado em 2016-10-06. 7
- [14] Luiz Antonio Gallo. Bioquímica - enzimas. <http://docentes.esalq.usp.br/luagallo/enzimas.html>, visitado em 2016-12-08. vi, 21
- [15] Laura González-Torres, Alfredo Téllez-Valencia, José Guadalupe Sampedro, and Hugo Nájera. Las proteínas en la nutrición. *Resbyn - Revista Salud Pública y Nutrición*, 8, 2007. 18
- [16] 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. 23
- [17] 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. 21, 22
- [18] 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. vi, 5
- [19] Gerhard Michal and Dietmar Schomburg. *The Cell and Its Contents*. John Wiley & Sons, Inc., 2012. 21
- [20] David Mount. *Bioinformatics : sequence and genome analysis*. Cold Spring Harbor, N.Y. Cold Spring Harbor Laboratory Press, 2001. 1, 24
- [21] Luca Pireddu, Duane Szafron, Paul Lu, and Russell Greiner. The path-a metabolic pathway prediction web server. *Nucleic Acids Research*, 34(Web-Server-Issue):714–719, 2006. 25
- [22] Leslie Pray. Discovery of dna structure and function: Watson and crick, 2008. <http://www.nature.com/scitable/topicpage/discovery-of-dna-structure-and-function-watson-397>, visitado em 2016-01-15. vi, 17, 18
- [23] João Carlos Setubal and João Meidanis. *Introduction to computational molecular biology*. PWS Publishing Company, 1997. 1, 16, 17, 18, 19, 21
- [24] Waldeyr Silva, Marcelo Brígido, Danilo Vilar, Daniel Souza, Maria Emilia Walter, and Maristela Holanda. 2path: a terpenoid metabolic network modeled as graph database. International Workshop on Biological Network Analysis and Integrative Graph-Based Approaches, 2016. 33, 35
- [25] Röbbe Wünschiers, Martina Jahn, Dieter Jahn, Ida Schomburg, Susanne Peifer, Elmar Heinze, Helmut Burtscher, Julia Garbe, Annika Steen, Max Schobert, Dieter Oesterhelt, Josef Wachtveitl, and Antje Chang. *Metabolism*. John Wiley & Sons, Inc., 2012. 22, 23

Apêndice A

Questionários de dados pessoais dos biólogos



Questionário de Avaliação da Interface do banco de dados de redes metabólicas 2Path

Parte 1: Dados pessoais do usuário

Nome completo: _____

Formação acadêmica: Graduação Período: de _____ à _____
 Mestrado Período: de _____ à _____
 Doutorado Período: de _____ à _____
 Pós-Doutorado Período: de _____ à _____

Atuação profissional atual: _____

Em seu trabalho, com que frequência você utiliza ferramentas interativas de visualização de dados?

Nunca utilizei [1] [2] [3] [4] [5] Frequentemente

Qual seu conhecimento a respeito dos seguintes assuntos:

I. Banco de dados biológicos (*KEGG, BioCyc, Reactome*, etc)

Nunca utilizei [1] [2] [3] [4] [5] Tenho pleno conhecimento

II. Redes metabólicas

Nunca utilizei [1] [2] [3] [4] [5] Tenho pleno conhecimento

III. Ferramentas de visualização de redes metabólicas (*Reactome Pathway Browser, KEGG Metabolic pathways*, etc)

Nunca utilizei [1] [2] [3] [4] [5] Tenho pleno conhecimento

Apêndice B

Tarefas realizadas pelos biólogos no sistema 2Path



Realização de tarefas na interface do banco de dados de redes metabólicas 2Path

Nome completo: _____

Será utilizado o Método de Avaliação de Comunicabilidade para avaliar a interface de acordo com os dados coletados nesta etapa.

O usuário é incentivado a usar a técnica *think aloud*, onde ele deverá falar em voz alta tudo o que está pensando ao realizar as tarefas propostas. As seguintes expressões são mais convenientes para a análise de dados:

“Cadê”	“Onde estou?”	“Vai de outro jeito”	“Socorro!”
“E agora?”	“Ué, o que houve?”	“Não, obrigado(a)!”	“Desisto”
“O quê é isso?”	“Por que não funciona?”	“Pra mim está bom”	
“Epa!”	“Assim não dá”		

Tarefa 1

Verifique se as seguintes enzimas estão no organismo *Arabidopsis thaliana* e, se estiverem, quais reações catalisam (pode ser o nome abreviado):

- 1.1.1.18 [] Não está
[] Está e catalisa a reação _____
[] Não encontrei a informação

2.7.7.60 [] Não está
[] Está e catalisa a reação _____
[] Não encontrei a informação

1.1.1.267 [] Não está
[] Está e catalisa a reação _____
[] Não encontrei a informação

Tarefa 2

Verifique se existe uma via metabólica entre os seguintes compostos. Se existe, quantas reações químicas ocorrem entre eles?

Acetyl-CoA → Oleandolide

[] Existe via metabólica com ___ reações
[] Não existe via metabólica
[] Não encontrei a informação

Apêndice C

Questionários de interface conforme o Método de Avaliação de Comunicabilidade



Questionário de Avaliação da Interface do banco de dados de redes metabólicas 2Path

Parte 2: Análise da Interface

Nome completo: _____

I. Em relação às tarefas realizadas, responda:

Foi fácil realizar a tarefa 1

Concordo plenamente [] [] [] [] [] Discordo totalmente
1 2 3 4 5

Explique o motivo: _____

Foi fácil realizar a tarefa 2

Concordo plenamente [] [] [] [] [] Discordo totalmente
1 2 3 4 5

Explique o motivo: _____

II. Em relação à interface do 2Path, responda:

Qual o grau de dificuldade da utilização do sistema, no geral?

Muito difícil [] [] [] [] [] Muito Fácil
1 2 3 4 5

Existem elementos nas páginas do *site* que não estão claros? Se sim, quais?

Houve algum momento em que você ficou em dúvida em relação à o que fazer? Se sim, por que e o que você fez para atingir seu objetivo?

O que você mudaria ou adicionaria no sistema? Por quê? (Opcional)