



UNIVERSIDADE DE BRASÍLIA – UNB
Faculdade de Biologia
Curso De Licenciatura Em Ciências Biológicas

**DESENVOLVIMENTO DE PLATAFORMA WEB PARA ANÁLISE E INTEGRAÇÃO DE
DADOS PROTEÔMICOS**

WALDEYR MENDES CORDEIRO DA SILVA

FORMOSA-GO, BRASIL
2011

WALDEYR MENDES CORDEIRO DA SILVA

**DESENVOLVIMENTO DE PLATAFORMA WEB PARA ANÁLISE E INTEGRAÇÃO DE
DADOS PROTEÔMICOS**

Monografia apresentada, como requisito parcial para a obtenção do grau de Licenciado em Ciências Biológicas, na Universidade de Brasília, sob a orientação do Prof. Dr Wagner Fontes

**FORMOSA-GO, BRASIL
2011**

WALDEYR MENDES CORDEIRO DA SILVA

**DESENVOLVIMENTO DE PLATAFORMA WEB PARA ANÁLISE E INTEGRAÇÃO DE
DADOS PROTEÔMICOS**

Trabalho de conclusão de curso apresentado como requisito parcial para
obtenção do grau de Licenciado em Ciências Biológicas da Universidade de Brasília.

Aprovado em __ de _____ 2011.

Prof. Dr. Wagner Fontes
Universidade de Brasília
Orientador

Prof. Ms. Paula Marcela Duque Jaramillo

Prof. Ms. Lanuse Caixeta Zanotta

Prof. Dr. Wagner Fontes
Universidade de Brasília
Coordenador do Curso de Licenciatura em Biologia

**FORMOSA-GO, BRASIL
2011**

RESUMO

O conhecimento da sequência completa de todos os genes proporcionado pelo projeto genoma foi um grande avanço para o estudo dos seres vivos. O proteoma, conjunto de proteínas expressas numa célula ou tecido, vai além do genoma pois o processo é bem mais dinâmico, não sendo possível determinar, apenas com o genoma, que proteínas serão expressas num dado momento sob uma determinada condição. A proteômica está relacionada à necessidade investigar o controle da expressão gênica e seus impactos no metabolismo celular. As proteínas são as biomoléculas mais abundantes e ocorrem em grande diversidade, podendo agir como enzimas, anticorpos, hormônios, componentes estruturais e receptores celulares. Devido a essa diversidade de possibilidades, as proteínas exercem papel fundamental em quase todos os fenômenos biológicos, como produção de energia, defesa imunológica, contração muscular, atividade neuroquímica e reprodução. Durante as pesquisas de proteômica, um grande desafio é a rastreabilidade dos dados referentes à pesquisa e sua análise. O PIAS (Proteomics Analysis and Integration Software) proporciona essa organização através de um sistema web, com banco de dados relacional e algoritmos de análise e integração dos dados proteômicos e outras informações periféricas.

Palavras-chave: proteoma, software, análise e integração proteômica.

SUMÁRIO

INTRODUÇÃO.....	6
OBJETIVOS	6
MATERIAL E MÉTODOS.....	6
REFERENCIAL TEÓRICO	6
RESULTADOS E DISCUSSÃO	12
CONCLUSÕES.....	14
REFERÊNCIAS BIBLIOGRÁFICAS	6
ANEXOS.....	8

INTRODUÇÃO

A descoberta do DNA, do RNA mensageiro e a identificação do código genético humano levaram à determinação do fluxo da informação genética conhecido como Dogma Central da Biologia Molecular que apesar de servir de base para inúmeros progressos biotecnológicos, não é suficiente para determinar que proteínas podem ser expressas numa célula num dado momento sob uma determinada condição(KOSHINO; BARROSO, 2010). Proteoma, termo sugerido pela primeira vez por Wilkins no Encontro 2-DE de Siena em 1994, é o conjunto de proteínas expressas pelo genoma. O número limitado de genes revelado pelo sequenciamento do genoma em comparação ao extenso conjunto de proteínas em uma célula ou organismo mostra ainda a relevância do splicing alternativo na expressão gênica como um mecanismo saídas de genes combinados(ROBERT; SMITH, 2002), além de mudanças pós tradicionais, como fosforilações, glicosilações substituições ou remoção de alguns aminoácidos.

A Bioinformática possibilita estudos como identificação de proteínas, análise de imagens, de localização, função e interação. Há softwares, por exemplo, que capturam as imagens provenientes de géis resultantes de experimentos de eletroforese e fazem nele uma varredura das proteínas apresentadas nas imagens sob a forma de spots. Contudo, a integração, análise e a correlação entre os experimentos e: os resultados, as amostras analisadas, os indivíduos, os responsáveis por coletas, entre outros, nem sempre são tarefas possíveis ou automatizadas nos softwares disponíveis. Tal nível de integração dos resultados é fundamental em projetos de análise proteômica, de forma que os dados obtidos a partir de uma determinada amostra possam ser rastreados durante todo o processamento experimental.

O PIAS (Proteomics Analysis and Integration Software), cujo desenvolvimento é alvo do presente trabalho, possibilita através de cadastros o arquivamento e mantém um histórico dos fatores relacionados aos experimentos; proporciona ainda uma análise comparativa de seus resultados através de relatórios customizáveis numa interface web amigável.

OBJETIVOS

Construir uma plataforma web que seja capaz de integrar as informações acerca de experimentos e possa receber os dados oriundos de softwares de Bioinformática, analisá-los, correlacioná-los e produzir relatórios customizáveis para suprir as necessidades dos pesquisadores.

MATERIAL E MÉTODOS

Pesquisa em livros e artigos científicos relacionados com bioquímica, biologia molecular, síntese de proteínas e separação de proteínas por eletroforese em gel bidimensional. Observação de pesquisa no laboratório de biologia molecular da UnB. Entrevistas com os pesquisadores envolvidos. Levantamento de requisitos em arquivos XML(Extensible Markup Language) exportados do ImageMaster®, dados referentes a experimentos proteômicos. Documentação do software utilizando a UML e alguns artefatos do RUP. Desenvolvimento de um sistema web com a linguagem PHP5.2, ZendFramework, Banco de dados MySQL5.1.

REFERENCIAL TEÓRICO

Tecnologia Computacional

A UML, (Unified Modeling Language – Linguagem de Modelagem Unificada) tornou-se nos últimos anos uma linguagem padrão adotada internacionalmente pela indústria de engenharia de software (GUEDES, 2009). A UML não é uma linguagem de programação e sim uma linguagem de modelagem cujo objetivo é auxiliar os engenheiros a definirem características do software tais como requisitos, comportamento, estrutura lógica e dinâmica dos processos, e necessidades físicas. A UML não é um processo de

desenvolvimento e tampouco está ligada a um processo específico, podendo ser usada para documentar softwares independente do processo de desenvolvimento adotado.

RUP(Rational Unified Process) é um processo de desenvolvimento de software criado pela IBM e disponibilizado para a comunidade desenvolvedora de softwares. O RUP é um amplo processo que fornece as práticas da indústria para software abrangendo gestão, implementação, testes e entrega de sistemas. É um dos muitos processos contidos na Biblioteca Process Rational, que oferece práticas adequadas ao desenvolvimento particular ou necessidade do projeto.

PHP é uma linguagem de script amplamente utilizada na internet que atualmente está em sua versão 5.3. Para esta linguagem diversos frameworks - que são conjuntos de códigos baseados em melhores práticas com propósitos genéricos – estão disponíveis. Entre eles o ZendFramework oferece ampla documentação, e através da sua licença MIT, liberdade de desenvolvimento de software baseados em sua estrutura de forma que o software desenvolvido não precise ser disponibilizado de forma gratuita ou sequer seja disponibilizado, flexibilizando seu uso para aplicações comerciais.

O MySQL é um software gerenciador de banco de dados recentemente adquirido pela empresa SUN que está em sua versão 5.5.13, porém a versão utilizada foi a 5.1. O MySQL tornou-se o mais popular banco de dados de código aberto do mundo pelo seu alto desempenho confiabilidade e facilidade de uso. O MySQL pode ser executado em mais de 20 plataformas, como Linux, Windows, Mac, Solaris, HP-UX e IBM-AIX. Há ainda uma gama extensa de ferramentas para manipulação do MySQL disponível no mercado de forma gratuita e paga.

Tecnologia biológica

Um aminoácido possui um grupo carboxila e um amino em comum, ligados a um carbono central chamado centro quiral, uma cadeia lateral (ou grupo R) único, que é diferente em estrutura, tamanho, polaridade e solubilidade em água. Este grupo R diferencia os aminoácidos entre si. Aminoácidos podem se unir em ligações covalentes com sequências próprias e originar peptídeos, por meio das chamadas ligações peptídicas, onde há a remoção de uma molécula de H₂O, originada do OH⁻ do grupo carboxila e o H⁺ do grupo amina. Esta reação ocorre no sistema ribossomal, durante o processo de tradução da informação genética, onde irão formar dipeptídeos, tripeptídeos ou até mesmo polipeptídeos, dependendo do número de aminoácidos participantes. Os

aminoácidos podem se encontrar de forma monomérica, ou seja, não associados à proteínas, como precursores para neurotransmissores ou como os próprios neurotransmissores. O conjunto de aminoácidos unidos por ligações peptídicas forma uma cadeia polipeptídica de uma proteína, também chamado de estrutura primária(BERG; TYMOCZKO; STRYER, 2007).

"O sequenciamento do genoma humano foi uma imensa tarefa porque ele contém aproximadamente 3 bilhões de pares de bases. A estrutura do DNA revelou como a informação é armazenada nas sequências de bases ao longo de um filamento de DNA, cujo papel mais fundamental é codificar as sequências de proteínas." (BERG; TYMOCZKO; STRYER, 2007).

Mas este conhecimento é análogo a um conjunto de peças e engrenagens que não explica o funcionamento do equipamento em si. Assim como o DNA, as proteínas são polímeros lineares, entretanto diferem do DNA de dois modos importantes: as proteínas são feitas a partir de 20 blocos estruturais, os aminoácidos, em vez de apenas quatro como estão presentes no DNA, fato que por sua complexidade química permite que as proteínas desempenhem ampla gama de funções; em segundo lugar as proteínas enovelam-se em estruturas tridimensionais elaboradas, determinadas por suas sequências de aminoácidos e pelo auxílio de proteínas que auxiliam no dobramento de outras durante o processo de síntese. Essas estruturas espaciais de uma proteína podem se tornar bastantes complexas e para facilitar o seu estudo foram divididas em níveis organizacionais, de primária a quaternária (BERG; TYMOCZKO; STRYER, 2007).

Nas estruturas primárias as proteínas se unem por ligações peptídicas formando Cadeias Peptídicas; Nas estruturas secundárias as Cadeias Peptídicas podem dobrar-se em estruturas regulares como a Hélice Alfa, Folha Beta, Voltas e Alças; Nas estruturas terciárias as proteínas hidrossolúveis enovelam-se em estruturas compactas com o interior apolar; por fim as estruturas quaternárias têm Cadeias Peptídicas que podem associar-se estruturas de múltiplas subunidades. A diversidade da estrutura espacial, mais um nível de complexidade desse grupo de biomoléculas, associa-se diretamente à diversidade funcional dessas proteínas, ampliando as possibilidades em relação à estrutura primária (BERG; TYMOCZKO; STRYER, 2007).

No processo de tradução, íntrons são sequências não codificáveis que devem ser removidas antes que ocorra a tradução. O processo de remoção do íntron é chamado de splicing. Com a remoção dos íntrons, os éxons emendam-se formando uma nova molécula de mRNA. Os genes são mais complexos que o modelo simples utilizado para investigar os mecanismos de splicing, eles geralmente contém múltiplos íntrons e em muitos casos os éxons podem ser unidos sob mais de uma forma gerando vários mRNAs (ROBERT; SMITH, 2002). Além do splicing alternativo, podem haver modificações pós-traducionais que interferem nas proteínas sintetizadas.

O termo proteoma é derivado de proteínas expressas pelo genoma, o qual fornece uma lista de produtos gênicos possíveis, mas apenas uma parte dessa lista será realmente expressa num determinado contexto biológico. Ao contrário do genoma o proteoma não é estático, pois quase todos os produtos gênicos são proteínas que podem sofrer modificações químicas de vários modos; além disso as proteínas não existem em estado isolado, ao contrário, interagem umas com as outras formando complexos com propriedades funcionais específicas.

O estudo da proteômica pode levar a três vertentes básicas com implicações para vários campos da biologia e da biotecnologia conforme a Tabela 01 (EMBRAPA, 2005):

Tabela 01 -Tabela de vertentes e implicações do estudo proteômico.

Vertente 1	Vertente 2	Vertente 3
Vias metabólicas nas diversas etapas celulares.	Identificar biomoléculas bioativas.	Identificação e caracterização de marcadores biológicos.
Conhecimento em biologia celular e bioquímica.	Desenvolvimento de novos medicamentos.	Diagnóstico precoce de doenças e acompanhamento do tratamento.

O reconhecimento de uma proteína depende do processo de purificação da mesma que pode acontecer de acordo com sua solubilidade, tamanho, carga e afinidade de ligação. Há várias técnicas de purificação disponíveis como Salting Out (perda de solubilidade em altas concentrações de sal), Diálise (separação através de membrana semipermeável), Cromatografia de Filtração em Gel (Separação por penetração numa coluna de espaços ou matriz porosa), Cromatografia de Troca Iônica (Separação com base na carga global), Cromatografia de Afinidade (Usa uma coluna de grãos para separar por afinidade por grupamentos químicos específicos), Cromatografia Líquida de Alta Pressão (Uma combinação das cromatografias anteriores com maior poder de

resolução), Eletroforese em Gel (Movimento das moléculas de acordo com a sua carga elétrica). Estes últimos casos (cromatografias e eletroforese) são os métodos de separação mais aplicados à proteômica, por permitirem a separação de diversas proteínas em um menor número de experimentos e, por essa razão, são os métodos de onde os resultados serão cadastrados e analisados pelo software descrito no presente projeto com atenção especial.

A separação de proteínas por eletroforese é possível através da focalização isoelétrica, O princípio da focalização isoelétrica consiste em estabelecer um gradiente de pH em um gel: uma amostra é inserida e então aplicada uma voltagem; as proteínas migram até atingir o pH correspondente ao seu ponto isoelétrico, o local onde o somatório das cargas elétricas da proteína é nulo.

Outra estratégia de separação eletroforética de proteínas é a eletroforese em gel de poliacrilamida na presença de dodecilssulfato de sódio (SDS). Trata-se de um método muito usado para análise massas de proteínas, até mesmo as oligoméricas, aquelas formadas por mais de uma cadeia de polipeptídeos. O gel é uma matriz porosa, cuja porosidade pode variar conforme a concentração de acrilamida. As proteínas são revestidas por um composto anfipático, geralmente o detergente SDS e são aplicadas a uma das extremidades do gel. A seguir aplica-se uma diferença de potencial elétrico a esse gel, que causará a migração das proteínas em direção ao pólo positivo. De acordo com Weber e Osborn, 1975: “As mobilidades eletroforéticas de muitas proteínas em géis de poliacrilamida-SDS são inversamente proporcionais aos seus logaritmos de massa”.

Atualmente as principais técnicas usadas na proteômica são a eletroforese em gel de duas dimensões e a espectromia de massa. Para a eletroforese bidimensional, uma amostra de proteína é inicialmente fracionada em uma focalização isoelétrica e em seguida o gel da focalização é preso a um SDS-poliacrilamida e é feita a eletroforese na dimensão perpendicular à separação original(BERG; TYMOCZKO; STRYER, 2007). Assim, as proteínas antes separadas pelo seu *pI* agora são separadas por massa, esse processo gera um padrão bidimensional de pontos. Proteínas isoladas de células em diferentes condições e submetidas à eletroforese apresentam aumento ou diminuição de sua concentração em resposta ao estado fisiológico (trauma).

No caso de géis SDS 2D, ao final da separação os géis são submetidos a um processo de coloração, que permite a ligação de sais de cobre ou de prata às proteínas e sua conseqüente visualização. A seguir esses géis são digitalizados em scanner para que as imagens possam ser processadas com o auxílio de programas de computador.

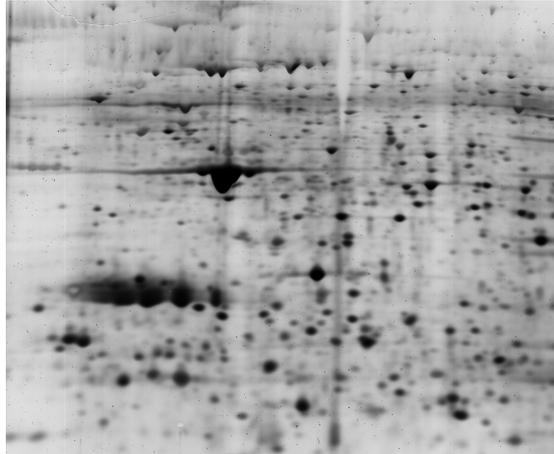


Figura 01 – Gel de Eletroforese – Imagem cedida pelo prof. Dr. Wagner Fontes.

O software ImageMaster da GE recebe as imagens dos géis e é capaz de identificar os spots em cada gel. Em seguida pode comparar os spots dos diferentes géis e estes dados são salvos num formato proprietário do software, o formato .mel. Essas informações podem ser exportadas no formato universal XML que é um formato muito flexível derivado do SGML (ISO 8879). Originalmente concebido para responder aos desafios das grandes publicações eletrônicas, XML desempenha um papel cada vez mais importante na troca de uma ampla variedade de dados entre arquiteturas diferentes. Os dados representados nesse formato obedecem a uma hierarquia representada por tags que possibilita a organização do conteúdo como mostra a figura 02.

A análise dos dados entre géis é limitada aos recursos do ImageMaster, porém, os dados exportados em XML permitem que outro software obtenha as informações mapeadas nos géis e possa fazer novas análises individuais ou comparativas gerando novas informações além daquelas que o ImageMaster oferece.

```

1  <?xml version="1.0" encoding="iso-8859-1"?>
2  <Gels>
3    <Fileinfo>
4      <Type>Export</Type>
5      <Version>1.0</Version>
6      <User>user</User>
7      <Date>19/04/2011 15:20:56</Date>
8      <Application_Name>ImageMaster</Application_Name>
9      <Application_Version>6.0.0.0</Application_Version>
10     <Organisation></Organisation>
11   </Fileinfo>
12   <Gels_Data>
13     <Version>1</Version>
14     <Gel Id="f54bd78f-c50a-4ff0-952d-d6dd51c06b83" Cols="2112" Ref="0" Rows="2122" Class="" Spots="1234" IsDige="0" Caption="a" GelName="F:\ImageMaster\
15     <Gel_Properties/>
16     <Spots>
17       <Spot X="975" Y="36" Id="48" Vol="44.3181" Area="2.35844" Flag="0" X_Align="0" Y_Align="0" Saliency="3.41545" ChainCode="970 30 5 4 4 4 6 6 6
18       <Spot X="1550" Y="57" Id="52" Vol="54.9537" Area="2.22224" Flag="0" X_Align="0" Y_Align="0" Saliency="0" ChainCode="1554 46 5 4 4 4 5 4 5 5
19       <Spot X="886" Y="929" Id="1508" Vol="494.519" Area="13.5341" Flag="0" X_Align="0" Y_Align="0" Saliency="0" ChainCode="883 906 6 6 5 5 4 5
20       <Spot X="787" Y="919" Id="1509" Vol="849.525" Area="21.8424" Flag="0" X_Align="0" Y_Align="0" Saliency="0" ChainCode="792 901 5 4 4 4 5 4 4
21     </Spots>
22   </Gels_Data>
23 </Gels>
24

```

Figura 02 – Exemplo de arquivo XML com dados de um gel proveniente do ImageMaster

RESULTADOS E DISCUSSÃO

A exploração do proteoma é importante para aprofundar os conhecimentos em biologia molecular bioquímica, desenvolvimento de novos medicamentos, diagnóstico precoce de doenças e tratamentos mais eficientes. O software web de análise e integração de dados proteômicos batizado como PIAS(Proteomics Analysis and Integration Software) proporciona basicamente três principais colaborações com a proteômica:

- 1 – Organização e indexação do material relacionado às pesquisas e dados dos experimentos.
- 2 – Gerar, a partir do cadastro dados de experimentos, relatórios que colaborem com a pesquisa.
- 3 – Funciona como um arquivo dos experimentos e mantém um histórico das atualizações ocorridas.

O PIAS prevê: cadastro de instituições e seus cursos, de responsáveis, de indivíduos, de amostras sanguíneas, de protocolos utilizados, de experimentos e seus resultados sob o formato XML quando oriundos do software ImageMaster®. O Documento de Visão, arquivo anexo de documentação de software no padrão UML apresenta uma visão geral do sistema.

O software é capaz de persistir num repositório os dados de experimentos de forma que guardem entre si relações. Essas relações são baseadas na teoria da álgebra relacional sob a qual é construído o banco de dados relacional usado: MySQL. A partir desse relacionamento entre os dados cadastrados de experimentos é possível realizar consultas direcionadas. O resultado dessas consultas é apresentado sob a forma de relatórios. No momento da consulta pode haver a necessidade de processamento dos dados obtidos, como comparação entre dados entre diferentes experimentos. Esse processamento é sempre guiado pelas escolhas do usuário na interface do PIAS. Os algoritmos criados para o PIAS varrem o banco de dados baseados nos parâmetros escolhidos pelo usuário, cruzam as informações e disponibilizam o resultado das análises sob a forma de um relatório personalizado. O conteúdo dos relatórios varia na proporção em que novos dados são cadastrados para os experimentos envolvidos na consulta.

No caso específico de arquivos XML de resultados de experimentos, cada XML contém dados organizados de forma hierárquica de onde é possível extrair informações específicas de algum aspecto do resultado dos experimentos. Um exemplo, com adaptações, de XML resultante de gel de eletroforese é abaixo apresentado para que possamos abstrair melhor o conceito:

```
<GelData>
  <Spots>
    <Spot X="975" Y="36" Id="48" Vol="44.3181" Area="2.35844" / >
    <Spot X="1550" Y="57" Id="52" Vol="54.9537" Area="2.22224" />
  </Spots>
</GelData>
```

No exemplo acima podemos ver a cadeia hierárquica onde *<Spot>dados do spot</Spot>* pertence a *<Stops>dados de cada spot</Stops>* que por sua vez pertence a *<GelData></GelData>*. Assim, podemos ver que para um determinado gel existem dados, entre eles, spots; e a definição de cada spot com atributos como: “X e Y” que identificam no gel a posição onde cada spot se encontra no plano cartesiano, “Vol” que indica o volume ocupado pelo spot. “Area” que indica a área ocupada pelo spot.

Esse resultado de experimento apresentado no arquivo XML gerado pelo software de análise de imagens não está ligado de forma alguma a nenhum protocolo utilizado, ou a uma amostra de sangue de um indivíduo que a forneceu, ou ainda, não é possível deduzir qual a instituição e curso do responsável pelo experimento. Neste ponto o PIAS apresenta uma solução, onde a partir de cadastro prévio é possível vincular o XML às demais informações do experimento que o originou.

O PIAS está sendo desenvolvido neste ano de 2011, em que a cada período estipulado em cronograma há pequenas entregas de funcionalidades que representam valor para o negócio. É previsto em cronograma que toda parte de cadastros esteja pronta até meados de outubro do ano corrente, quando os algoritmos de análise proteômica poderão ser desenvolvidos baseando-se nos dados cadastrados e a partir deles os relatórios possam ser gerados.

CONCLUSÕES

A pesquisa proteômica é complexa pois o proteoma é dinâmico enquanto o genoma é estático. As proteínas podem sofrer modificações químicas de vários modos além de interagirem umas com as outras formando complexos com propriedades funcionais específicas. Os progressos científicos gerados por essas duas pesquisas são extensos, mas no caso do proteoma há muito por ser descoberto e as possibilidades são amplas.

A biotecnologia disponível para proteômica ainda não é suficientemente eficaz em todos os sentidos. Um dos pontos sensíveis é a possibilidade de cruzamento dos dados de diferentes experimentos, gerando através de análise, resultados novos. Além disso a quantidade de dados cresce muito numa pesquisa e outra necessidade surge: a de organizá-los.

O PIAS(Proteomics Analysis and Integration Software) proporciona essa organização através de um sistema com interface amigável, banco de dados relacional e algoritmos de análise e integração dos dados proteômicos que possibilitam nova ótica para resultados de experimentos.

REFERÊNCIAS BIBLIOGRÁFICAS

BERG, Jeremy M; TYMOCZKO, John L.; STRYER, Lubert. Bioquímica. 6. ed. New York: Guanabara Koogan, 2007.

BURNS, George W. ; BOTTINO, Paul J.; Genética. Ed. Guanabara Koogan, Rio de Janeiro, 1991.

Embrapa. ROCHA, Thales Lima; et. al. Comunicado Técnico 136: Eletroforese Bidimensional e Análise de Proteomas. ISSN 9192-0099, Brasília-DF, Outubro de 2005.

FONTES, Wagner; CASTRO, Mariana de Souza. Estrutura e metabolismo das biomoléculas I: Aminoácidos e proteínas. Consórcio Setentrional, Processos de Manutenção da Vida. SEED, 2007.

GAVIN C. Roberts, SMITH W. J. Christopher. Alternative Splicing: combinatorial output from the genome. Current Opinion Chemical Biology, 2002.

GUEDES, Gileanes T. A. UML 2: Uma abordagem prática. Novatec Editora, São Paulo, 2009.

KOSHINO, Livia Luzia do Nascimento; BARROSO, Edson Wagner de Souza. A Proteômica Forense no Brasil: Estado Atual e Perspectivas. Ifar/PUC, 2010.

ROBERT, Gavin C; SMITH, Christopher W J. Alternative Splicing: combinatorial output from the genome. Current Opinion on Chemical Biology 6:375-383, 2002.

ROEPSTORFF, P. Mass spectrometry in protein studies from genome to function. Current Opinion in Biotechnology 8: 6-13, 1997.

SILVA, Adriana Moreira da Silva e; CORRÊIA, Gustavo Coelho; REIS, Emerson Moreira. Proteômica, Uma abordagem funcional do estudo do genoma. Saúde&Ambiente em Revista, v2, n2, págs 01-10, Duque de Caxias-RJ, 2007.

Venter JC, Adam MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA et al.: The sequence of the human genome. Science 2001.

WEBER, K. e; OSBORN, M. The Proteins 3ª edição Vol. 1, pág. 179, Academic Press, Nova York, 1975.

SOUZA, Marcelo Valle de; FONTES, Wagner; RICART, Carlos André Ornellas. Análise de Proteomas – o despertar da era pós-genômica. Revista on-line Biotecnologia – Ciência e desenvolvimento, pgs 12,13,14; Ano 2, número 7, 1999.

FONTES, Wagner. Aminoácidos. Disponível em:
<http://www.bioq.unb.br/htm/textos_explic/moleculas-intro/aminoacidos_intro.htm>.
Acesso em: 05 abr. 2011.

Url:<http://www.w3.org/XML> acessado em 21/04/2011.

Url:http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=44681 acessado em 20/01/2011.

Url:http://www.iso.org/iso/catalogue_detail.htm?csnumber=16387 acessado em 20/01/2011.

Url:<http://www-01.ibm.com/software/awdtools/rup/> acessado em 22/02/2011.

Url: <http://framework.zend.com/>

Url: <http://www.mysql.com/why-mysql/>

ANEXOS

MER (Modelo Entidade-Relacionamento)

Tela do Protótipo

Documento de Visão



Cadastro de Responsáveis

24 de junho de 2011

[Início](#) [Instituições](#) [Responsáveis](#) [Experimentos](#) [Protocolos](#) [Indivíduos](#)

Nome do Responsável:

Cpf:

Identidade:

Data de Nascimento:

Telefone 1 :

Telefone 2 :

Celular :

Email 1:

Email 2:

Cadastro de Amostras

24 de junho de 2011

[Início](#) [Instituições](#) [Responsáveis](#) [Experimentos](#) [Protocolos](#) [Indivíduos](#)

Data do Experimento:

Indivíduo:

Protocolo Utilizado:

Responsável pela Coleta:

Contagem:

Viabilidade (%):

Pureza (%):

Volume de Sangue Coletado (ml):

PIAS (Proteomics Analysis and Integration Software)

Documento de Visão

Versão 1.0

Histórico de Revisão

Data	Versão	Descrição	Autor
18/04/2001	1.0	Elaboração da versão inicial do artefato.	Waldeyr Mendes



Documento de Visão

1 INTRODUÇÃO

A proposta deste documento é coletar, analisar e definir as necessidades e funcionalidades gerais do PIAS. Seu foco está nas necessidades dos usuários e no motivo da existência destas necessidades.

Seu escopo engloba a definição dos gestores do sistema, dos representantes dos usuários, seus problemas, necessidades e das características essenciais do sistema para o atendimento destes requisitos.

Termos e abreviaturas específicos podem ser encontrados no Glossário do respectivo projeto.

2 DESENVOLVEDORES

Representante	Waldeyr Mendes
Responsabilidades	Coordenar o projeto, garantir a memória das reuniões e todos os requisitos especificados, garantir agenda juntos os usuários, desenvolver e o projeto, homologar os produtos entregues.
Função / Unidade	Aluno UnB
E-mail	contato@waldeyrmendes.com
Observações	Não aplicável

3 REPRESENTANTES DOS USUÁRIOS E GESTOR

Representante	Wagner Fontes
Responsabilidades	Especificar os requisitos técnicos do projeto, acompanhar o desenvolvimento, homologar os produtos entregues e prover o aceite final dos produtos.
Função / Unidade	Professor UnB
E-mail	wagnerf@unb.br
Observações	Não aplicável

4 PÚBLICO ALVO

O público alvo deste projeto são as instituições e pesquisadores da área de proteômica.

5 AMBIENTE DO USUÁRIO

5.1 Ambiente Físico

Não se aplica.

5.2 Ambiente Computacional

O ambiente computacional existente é composto por:

Ambiente de Produção:

Servidor de aplicação	Sistema Operacional	Linux Ubuntu 1010
	Application Server	Apache 2.2
	Hardware	Processador 2 x Xeon 3.6GHz, Memória: 8G, Dell PowerEdge 2850
Servidor de Dados	Gerenciador de banco de dados SGBD	MySQL 5.1
	Sistema Operacional	Linux Debian



	Hardware	Processador 2 x Xeon 3.6GHz, Memória: 8G, Dell PowerEdge 2850
--	----------	---

Ambiente de Homologação:

Servidor de aplicação	Sistema Operacional	Linux Ubuntu 1010
	Application Server	Apache 2.2
	Hardware	Máquina Virtual
Servidor de Dados	Gerenciador de banco de dados SGBD	MySQL 5.1
	Sistema Operacional	Linux Ubuntu 1010
	Hardware	Máquina Virtual

Ambiente de Desenvolvimento:

Servidor de aplicação	Sistema Operacional	Leopard OsX
	Application Server	Apache 2.2
	Hardware	Macbook
Servidor de Dados	Gerenciador de banco de dados SGBD	MySQL 5.1
	Sistema Operacional	Leopard OsX
	Hardware	Macbook



6 NECESSIDADES DO USUÁRIO

Necessidade 01

Descrição da Necessidade	Criar uma base de dados de experimentos.
Solução Proposta	Criar cadastros para os requisitos levantados
Prioridade	Alta.
Situação Atual	Requisitos levantados e banco de dados modelado.

Necessidade 02

Descrição da Necessidade	Processamento dos dados cadastrados.
Solução Proposta	Criar algoritmos capazes de manipular os dados cadastrados e relacioná-los à luz da proteômica.
Prioridade	Alta
Situação Atual	Os algoritmos estão sendo desenvolvidos.

Necessidade 03

Descrição da Necessidade	Gerar relatórios a partir da análise dos dados cadastrados.
Solução Proposta	Criar mecanismos de interação entre usuário e algoritmos de análise proteômica. Criar interface gráfica para a interação do usuário. Criar interface gráfica para a apresentação dos relatórios.
Prioridade	Alta.
Situação Atual	Não se aplica.

7 ESCOPO

Elaboração do Sistema, produzindo os seguintes artefatos:

- Documento de Visão;
- Documento de Requisitos e de Regras de Negócio;
- Projeto Arquitetural;
- Especificações de Casos de Uso;
- Protótipo Visual (Wireframe);
- Modelo de Banco de Dados (modelo de entidade e relacionamento);
- Código-fonte com testes nível 1 (unidade) executado pela equipe de desenvolvimento;
- Código-fonte com testes nível 2 (integridade) executado pela equipe de homologação;
- Código-fonte com testes nível 3 (aceitação) executado pela equipe de usuários do PIAS;
- Planejamento das atividades a serem desenvolvidas pela equipe do projeto (Planejamento Inicial);
- Execução e controle das atividades realizadas no âmbito do projeto;
- Encerramento do projeto.



- Criação de arquivos de propriedade para quaisquer idiomas;
- Criação de política de backup;
- Criação e manutenção da infra-estrutura de ambiente em relação ao banco de dados, servidores e serviços de comunicação internet;
- Implantação do sistema no ambiente de produção;
- Realização de treinamento dos usuários do sistema ou do corpo técnico que dá suporte ao sistema;
- A análise e a implementação das seguintes funcionalidades:
 - Manter Instituições.
 - Manter Cursos.
 - Manter Responsáveis.
 - Manter Indivíduos.
 - Manter contatos.
 - Manter Ativadores.
 - Manter amostras.
 - Manter Tipos de experimentos.
 - Manter Experimentos.
 - Manter Protocolos.
 - Manter Tipos de protocolos.
 - Manter Géis.
 - Manter Cromatografias.
 - Analisar e gerar relatórios.

8 PREMISSAS DO PROJETO

- Inserir dados de experimentos de forma integrada.
- Analisar os dados cadastrados e gerar relatórios.
- Permitir o acesso distribuído via browser.

9 RESTRIÇÕES DO PRODUTO

- A linguagem de programação a ser utilizada será PHP5 com o framework Zend.
- Linguagem auxiliar javascript com o framework jQuery e alguns plug-ins derivados jQuery.
- O sistema gerenciador de banco de dados será o MySQL 5.1.
- O sistema deverá garantir a integridade e confidencialidade dos dados das pesquisas através de mecanismos de segurança da informação.
- O sistema deverá suportar os browsers Mozilla Firefox 3.x ou superior.

10 VISÃO GERAL DO PRODUTO

10.1 Perspectivas do Produto

O PIAS tem por objetivo permitir o cadastro, integração e análise de dados proteômicos oriundos de experimentos em laboratório.



11 REQUISITOS DE TREINAMENTO

O treinamento dos usuários do sistema e do corpo técnico que dá suporte ao sistema será de responsabilidade do desenvolvedor Waldeyr Mendes, sendo não escopo do projeto.

12 ESTIMATIVAS DO PROJETO

O tamanho do projeto foi estimado em **240,13 (Duzentos de quarenta) Pontos de Função Ajustados**.

A duração do projeto foi estimada em **12 (doze) meses**.