

Project Name

Protein Fluorescence and Structure Toolkit (PFAST)

Proposal Outline

1. Background
2. Goals
3. Requirements
4. People
5. Project Planning
 - 5.1 Responsibilities
 - 5.2 Software Design
 - 5.3 Prototype Development
 - 5.4 Milestones
 - 5.5 Testing
 - 5.6 Problem/Risk Analysis
6. Resources Required
7. Client Interviews
8. Final Report
9. References

1. Background

Fluorescence spectroscopy is a traditional method for studying how protein structure varies in different environments, such as temperature and acidity. One of the amino acids in proteins, tryptophan, naturally fluoresces and is highly sensitive to its surrounding environment.

The first step is to obtain fluorescence spectra from the protein. Each spectrum is the sum of the emissions from all tryptophans within the protein. The next step is to break the total spectra into its component parts to determine the locations of the tryptophans in the protein. To do this, Dr. Reshetnyak has developed two algorithms: SIMS and PHREQ [1].

The second step is to find a correlation between this component spectral data and the structural properties of the protein. Dr. Reshetnyak's work has shown that there is a statistically significant correlation between the tryptophan fluorescence characteristics of a protein and some structural characteristics of the protein [2]. For this purpose, Dr. Reshetnyak has written an analysis program using the Statistica, a software statistical analysis program.

More details about this research can be found online in three of her papers hosted on the Biophysics Journal website.

2. Goals

The general purpose of this project is to apply our knowledge of computer science to aid Dr. Reshetnyak's research in protein fluorescence and structure.

Dr. Reshetnyak has developed some software tools and algorithms for analyzing fluorescence and structural data of proteins. For finding correlation, she also has collected many files which represent data and analysis of previous experiments. Finally, she has also written software for performing statistical analysis. The software tools were developed a long time ago, and are not in a form that is conveniently usable by other researchers.

Our goals are to improve the format of storing experimental data and to provide other researchers better accessibility to this data as well as her software analysis tools. To this end, we will use database technology for data storage and provide web interfaces for the analysis tools and the database.

3. Requirements

The successful project must meet the following requirements:

- The database must contain the fluorescence and structural parameters of proteins.
- The fluorescence spectra analysis programs must interface with the above database.
- The structural analysis programs must interface with the above database.
- Design and implement a new program that performs statistical analysis of the spectral and structural data, based on an existing program.
- All programs and databases must be accessible via the web.
- The database design should accommodate database growth (new data in the future).

4. People

Dr. Yana Reshetnyak, reshetnyak@mail.uri.edu, Department of Physics

The goal of the team is to help Dr. Reshetnyak's research. She will provide the team with background and other relevant information for the project.

Dr. Joan Peckham, joan@cs.uri.edu, Department of Computer Science

Dr. Peckham is one of the professors for this bioinformatics class. She will provide general guidance for the project.

The student team is composed the following graduate students with various skills:

- **Nidhi Bansal**, nidybansal@gmail.com, 569-9908 (C), 874-5833 (O)
Web development, databases, programming
- **Neelima Guduru**, neelima_guduru@yahoo.com, 874-5833 (O), 273-5762 (H)
Programming, web development, databases
- **Stephen Jaegle**, jaegles@cs.uri.edu, 419-0281 (C), 823-8672 (H)
Project management, programming, web development, graphic design, databases
- **Rajiv Menon**, rajiv.menon@gmail.com, 874-2898 (O), 709-9447 (H)
Web development, programming, database design
- **Chi Shen**, shenc@cs.uri.edu, 792-3572 (H), 874-5568 (O)
Project management, information systems, web development, programming, interface design, graphic design, database design

5. Project Planning

5.1 Responsibilities

Dr. Reshetnyak will be responsible for providing the team background and other information relevant to the project. There isn't a need for specialization, so student members will share responsibility for all aspects of the project. The one specific role is:

- **Project Leader: Chi Shen**

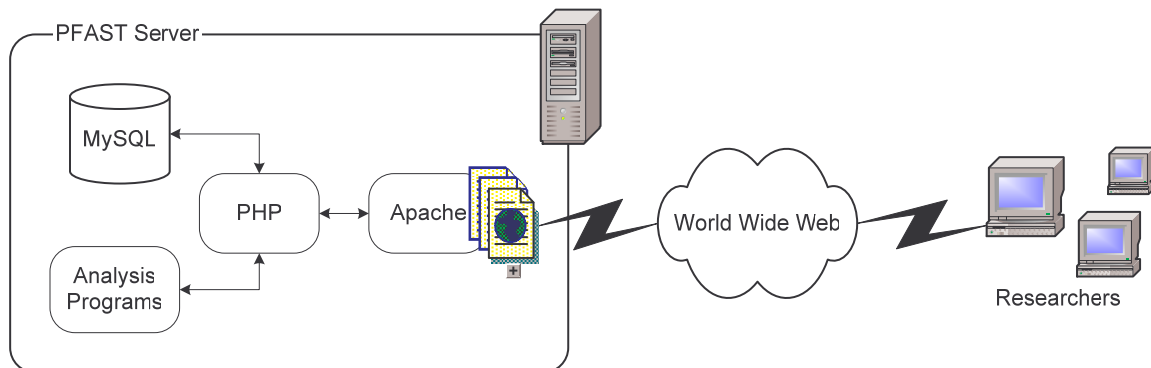
Organize meetings, maintain project records, plan and provide direction to the efforts of the team, and otherwise facilitate the overall progress of the project.

5.2 Software Design

To meet the requirements, the system is composed of many different components. To avoid monetary costs, we have selected a typical collection of free and open-source software for this project:

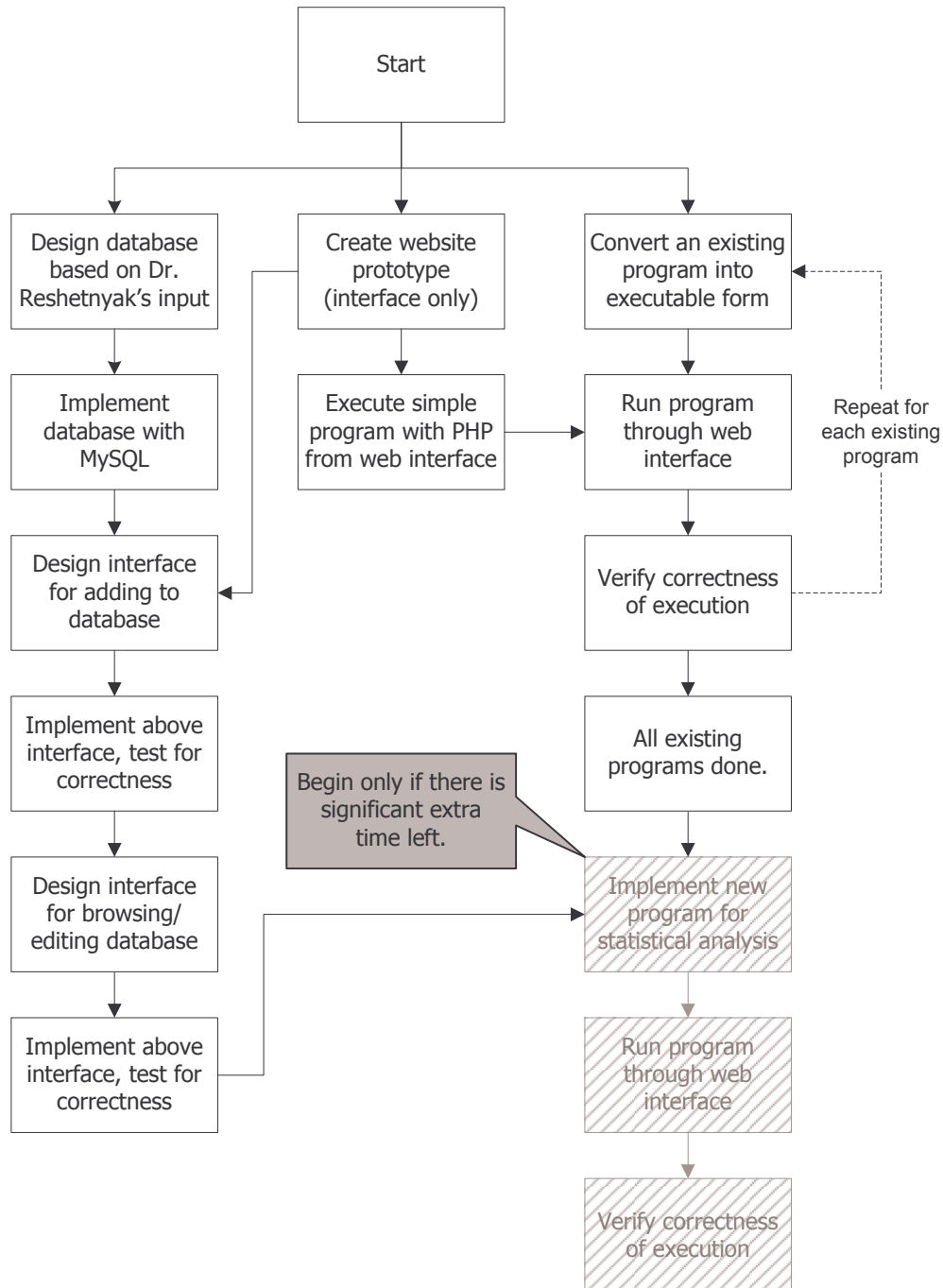
- Apache web server, for providing the web interface
- PHP scripting, for Apache to connect to the database and analysis programs
- MySQL database, for storing the fluorescence and structural parameters of proteins

The overall system architecture is as follows:



5.3 Prototype Development

We will proceed with developing our system prototype as follows:



5.4 Milestones

The following is a list of the major milestones of the project:

- March 16, 2005: Prototype design of database. Prototype of web interface (no functionality)
- March 23, 2005: Implementation of database.
- March 30, 2005: One existing program is accessible through the web.
- April 6, 2005: Database functionality is complete (browse, add, edit). All programs accessible through web interface.
- April 20, 2005: Testing completed, system is complete, and can be launched.
- April 27, 2005: Draft documentation complete.
- May 4, 2005: Documentation complete.

5.5 Testing

For testing our database, we will perform all related functions through the web interface and verify the correctness by directly inspecting the database records.

For testing the analysis programs, we will use combinations of input files and output files that are known to be correct. We will feed the input files into our system, and compare the system output to the known output files.

The web interface will be tested to meet basic usability standards informally by soliciting feedback from some representative users, Dr. Martin, Dr. Peckham, and Dr. Reshetnyak.

5.6 Problem/Risk Analysis

The following are problems we foresee and how we plan to address them:

- **Visual Basic / Fortran / Statistica experience**

The team has minimal experience with Visual Basic, FORTRAN, and Statistica. However, Dr. Reshetnyak is the author of the programs to be rewritten, and has a clear understanding of the algorithms. We will rely on her knowledge to aid us in rewriting the programs, and use other resources as necessary.

- **Biophysics background**

All team members are computer science graduate students with strong technical backgrounds, but little or no background in Dr. Reshetnyak's area of research. We do not currently understand well the data to be stored in the database nor the algorithms and programs used for analysis. Dr. Reshetnyak is willing to provide us with the background knowledge that we need in order to properly design the database and implement the algorithms.

- **PDB format**

Most of the team has no knowledge of the content and format of PDB files. The team plans to meet with another student group who is currently working extensively with the conversion of PDB formatted data to "Universal Research Interchange" XML formatted data.

- **Project Scope**

It is infeasible to fully implement all features specified in the project scope. Within the framework of the typical software development cycle, we feel that it is reasonable for us to complete the planning, design, and one to two rounds of iterative implementation. Our target end product will be good design documents and a working prototype that includes the basic functionality that meets the project requirements.

6. Resources Required

Below is a list of the resources the team needs to complete the project.

- Source code of existing analysis programs
- Statistics package Statistica
- Protein fluorescence and structure data
- A computer with a good Internet connection to provide a technical working environment and also be the web server.

7. Client Interviews

Dr. Reshetnyak is willing to provide us with the background protein structural and spectral analysis knowledge that we need in order to properly design the database. In addition, she is the author of the analysis programs for which we will provide a web interface. She can tell us about the characteristics of these programs so that we can construct an appropriate web interface.

8. Final Report

The final report will consist of the following content:

- Executive Summary: Review project goals and requirements, and describe the project in terms of meeting them.
- System Overview: Give a high-level description of how the different parts of the project interact with each other.
- Database Design: Provide design documents related to the database.
- Interface Design: Provide design documents related to the web interface.
- Implementation Status: List the state of the programs and the web interface.
- Defect Report: List known problems with the programs and the web interface.

9. References

- [1] Edward A. Burstein, Sergei M. Abornev, and Yana K. Reshetnyak, **Decomposition of Protein Tryptophan Fluorescence Spectra into Log-Normal Components. I. Decomposition Algorithms**, Biophys. J. 2001 81: 1699-1709, <http://www.biophysj.org/>
- [2] Yana K. Reshetnyak, Yuly Koshevnik, and Edward A. Burstein, **Decomposition of Protein Tryptophan Fluorescence Spectra into Log-Normal Components. III. Correlation between Fluorescence and Microenvironment Parameters of Individual Tryptophan Residues**, Biophys. J. 2001 81: 1735-1758, <http://www.biophysj.org/>