ENGR 102 - Winter 2011

Freshman Engineering Design Lab

"BioGUI: Protein Circle Plot Visualization"

Project Proposal

Section 19, Group 275, Project 126	Date: March 9, 2011

Submitted to:	Fred Douglas Allen
	Ahmet Sacan
Group Members:	Brian Tighe
	Ethan Sena
	Henry Kuns
	Gavin Rapp

Abstract:

This proposal investigates a potential solution to the lack of a unified, modern GUI system to display information from BioGUI. The proposed program is a C++ application that is given a list of protein structures and subsequently converts the data to an image that features connections between different alpha carbon groups. Potential challenges include the interaction between C++ and the various libraries (BioGUI, libgd) that the team needs to learn to create a working program with the aforementioned goals. At the end of the term, deliverables include a working application that is able to read in, process, and display protein structures that may also include web functionality.

1. Introduction

In the field of bioinformatics there is not a single unifying GUI system. The goal of this project is to work on small program sections to be integrated into the main BioGUI. The first stage of work is to develop a program for taking a protein file and outputting a protein circle graph. After the completion of this small part work will be continued on different parts of the GUI as needed. The overall goal is to gain experience in programming within a group environment and apply it to a realistic task. The desired outcome is to have submitted a good portion of code for use in the larger BioGUI project.

2. Technical Approach

The primary aspect of this project will be to develop a program that will be able to produce an image of various protein structures. The data for these graphics comes from files in a database that the program will be able to access. Once the program has read in the required values, it will process the data and convert it into an image. The graphic will consist of a circle, with the name of different alpha carbon groups around the circumference. In addition, if these groups are within a certain range of each other, a connection will be made between them. In order to keep the graphic simple, the connection will initially be made by drawing a line.

The graphing library that was selected to accomplish this task is called cairo. This is a library which is designed to render 2D images. Cairo comes equipped with specific functions to draw arcs, lines, and points, as well as functions that can control the color and opacity of each of these lines. The cairo library was selected primarily because it was easy to use where other libraries were difficult or unavailable. Cairo provides all the tools necessary to complete the objective, and is compatible with both Python and C++.

2.1 Project Timeline

Table 1: Freshman Design Project Timeline

	Week									
Task	1	2	3	4	5	6	7	8	9	10

Circle Plot Program Design	x	x	x							
Secondary Structure Visualization				x	x	x				
Connecting with splines							x	x		
Final report preparation									x	x

2.2 Facilities and Resources (if appropriate)

There are only a few resources that will be needed for the completion of this project. The preliminary resources that will be needed for the actual programming itself come in the form of the libraries that we will need to download, and apply. For instance, in order to do the graphical/ visual portion of the programming, we will need to use a specific library. Additionally, we will use a library that comes with a variety of functions specializing in interacting with files containing information about proteins. We will also need to use SVN, a program designed for groups of people to work on the same code simultaneously.

3. Deliverables

By the end of the term, our objective is to have a working application that will read a user file containing information about a particular protein, and then map out various information about the amino acid sequences contained within. The exact information that the application will display at this point is unclear, it depends on how quickly we can program as a group, and how quickly we can learn the required programming. The program would at least be capable of producing a circular representation of the protein, and would show the connections between the alpha carbons on various parts of the protein. At the end, it is also a goal to have this program functioning through a website.

4. Expertise

This particular project requires good programming skills from the start. The code will be written in C++, a language which all members are familiar with. Basic knowledge of biological structures will be learned throughout the project. Members will also be expected to learn and use SVN repositories for group coding. Graphical libraries such as "libgd" will be learned for generation of protein graphs and future work.

5. References

[1] C. Jarabek and S. von Mammen, "Protein Painting: Aesthetic 2D Visualization of Amino Acid Sequences," University of Calgary., Calgary., Alberta, Jan. 10, 2011.

[2] G. Vreind and C. Sander, "Detection of Common Three-Dimensional Substructures in Proteins," European Molecular Biology Laboratory, Heidelberg, Germany, Feb. 23, 1991.

[3] H. Liu, X. Liu, Y. Yao, "Identification of Secretory Proteins Based on Similarity of Amino Acid Sequences," 2010 3rd International Conference on Biomedical Engineering and Informatics (BMEI), Nov. 18, 2010.

[4] R. Vivanco, N. Pizzi, "Computational performance of Java and C++ in processing large biomedical datasets" Canadian Conference on Electrical and Computer Engineering, 2002. IEEE CCECE 2002, Aug. 7, 2002.