Mentored Professional Enrichment Experience

Applicant:

Name of Project/Experience: A Searchable, Hyperlinked, Dynamic, Metalloprotein Clinical Database

Location where Project/Experience will take place:
SIU School of Medicine
Carbondale, IL 62901

Mentor Name and Contact Information:
Eric C. Niederhoffer, Ph.D.
SIU School of Medicine
Lindegren 112, MC 6503
Carbondale, IL 62901
eniederhoffer@siumed.edu
618-453-6467

RATIONALE

Online databases have revolutionized the way we store and retrieve information. They provide an organized and efficient way to search for information while also being flexible enough to allow for the expansion and addition of new information. In general, online databases save both time and effort for the end user. Since the database is searchable, less time is spent looking for information. Additionally, since the database contains a variety of information hyperlinked together, less effort is spent trying to gather data.

Database websites, such as the Protein Data Bank (PDB), have become useful tools for scientists and allow for the quick and easy access of information on protein structure and function. However, few databases exist which connect the information found in the PDB to the clinical world. Additionally, there is a plethora of information on the internet which is scattered among a multitude of databases. Much of this information is isolated to its specific database and not hyperlinked to other databases. This lack of organization makes finding information difficult and time consuming. Therefore, I propose to create an online, searchable, protein database that will link a variety of factors such as molecular function, nutrition, and clinical data. Primarily, this system would focus on metalloproteins, and would center on the following: structural, functional, and active site information from the PDB website, 3D structures viewable via Jmol (an open source molecular viewer) genetic information from OMIM (database on genes and genetic disorders), hyperlinked proteins with a common function and structure, nutritional information based on the specific metal prosthetic group, and clinical aspects of the protein defects with links to the disease database, eMedicine and other scholarly sources.

The PDB currently has over 55,000 protein structures [1]. I realize that in eight weeks, I can't create a database containing the number of proteins in the PDB. However, if I focus on a specific type of protein, such as metalloproteins, I believe I could create a useful database linking many proteins with their clinical significance. Currently, the Metalloprotein Site Database (MDB) is one of the only online databases specifically cataloguing metalloproteins. However, the MDB focuses mainly on the structural and thermodynamic information, and leaves out clinically relevant information. Furthermore, the MDB website hasn't been updated since 2003[2]. In effect, I'm attempting to create a "pilot program" which lays the foundation for an extensive clinical protein database. This database can be easily expanded and improved well after the 8 week MPEE program.

The proposed database would also have multiple audiences. It would be useful to clinical researchers wanting to identify protein structural changes in relation to nutritional deficiencies. This database would also be useful to students and healthcare professionals wanting to understand the biochemical etiology of diseases and nutritional deficits involving metalloproteins. Thus, the database I'm proposing is novel, up to date, and could benefit a variety of audiences.

GOALS

The goals of this project are:

1. Create an online, searchable database for metalloproteins focusing on clinically important information
2. Link each protein together with a series of searchable tags
3. Link together proteins with common structure or function
4. Provide clinical and nutritional information pertinent to the protein of interest
5. Provide Jmol representations of protein structures
6. Provide a stable database which could be used by multiple users at once
7. Integrate these features into a user-friendly website
8. Personally, gain a greater understanding about the biochemical etiology of many diseases

By the end of the eight weeks, I expect to have a stable, online database that will provide clinically relevant information about a variety of metalloproteins.

METHODS

The most useful feature of the proposed online database is the ability to quickly and easily search for information by using a variety of criteria. To accomplish this, each protein will be associated with searchable criteria called "tags." For example, Hemoglobin might possess the "heme/iron" tag, the "porphyrin" tag, the "globin" tag, the "oxygen storage/ transport" tag, the "tetramer" tag, and several clinically relevant tags. By searching under any of these tags, Hemoglobin would appear in the user's search results. Searching under multiple tags at the same time would refine the search results.

The clinically relevant tags are the key to making this database useful for both students and healthcare providers. Clinically relevant tags for hemoglobin would include hemoglobinophathies such as sickle-cell disease and thalassemia. Each one of these

diseases could be linked to the OMIM genetic database, the disease database, eMedicine and also to various other scholarly sources. Nutritional tags, such as iron deficiency anemia and megaloblastic anemia, would also be included. Additionally, the database would link to the Jmol 3D structures provided by the PDB. Thus, the user would have easy access to a large amount of organized information.

On a more technical note, the website will use HTML, Java, a MySQL database, and the PHP scripting language. HTML will be used for basic web design and Java will be used to integrate the Jmole structures into the website. The database will be built using MySQL, which is currently the most popular open source relational database management system [3]. PHP, an open source scripting language which embeds into HTML[4], will be used for the creation of dynamic content. Any other programming, such as the creation of hash tables, will be done using the Perl computer language. The website will be designed using tools from Adobe CS4, such as Dreamweaver and Fireworks. The hardware needed would be a PHP MySQL compatible server.

ANALYSIS

Since this project involves the creation of a novel database, the database's completion would mark the end of the project. Additionally, this database will only be useful if it is easily accessible to the public via search engines. Thus, the project wouldn't be completely finished until it was easily accessed in the public domain. Furthermore, my project would undergo extensive beta testing, making sure the database remains stable when in use. This would include testing the stability of the site with multiple users and assessing the site's accessibility from various locations. Since the hyperlinked information and searchable content are paramount to this project, the database and search engine would also undergo rigorous beta testing with multiple users attempting to find and access information. Finally, I will exceed my goals if this database becomes a useful tool for healthcare providers, clinical researchers, and students who want to learn more about nutritional deficiencies and the clinical aspects concerning metalloproteins.

SUPPORT

1. Do you request support funds? Yes

2. Would you be able to participate if a scholarship is not available? No

REFERENCES:

1. RSCB Protein Data Bank. Research Collaboratory for Structural Bioinformatics. <http://www.rcsb.org/pdb/home/home.do>.

2. Metalloprotein Site Database & Browser, TSRI. The Scripps Research Institute. <http://metallo.scripps.edu/>.

3. MySQL:: Why MySQL? Sun microsystems. <http://www.mysql.com/why-mysql/>.

4. PHP: Hypertext Preprocessor. The PHP Group. <http://www.php.net/>.