

A Web-Based Antibody Database

Thesis Proposal

For the degree of Master of Science in Computer Science

At Southern Connecticut State University

Baorong Shi

August 2003

Thesis Advisor: Dr. Hrvoje Podnar

Major-Field Approval-The advisor and the department chairperson

Advisor

Date

Chairperson

Date

A. Title

A Web-Based Antibody Database

B. Statement of Purpose

Substances foreign to the body, such as disease-causing bacteria and viruses and other infectious agents, are known as *antigens*. They are generally of high molecular weight and are commonly proteins or polysaccharides. Our natural defenses against these infectious agents are *antibodies*, which are proteins that seek out the antigens and help destroy them. There is a significant number of antibodies (in the range of hundreds of thousands) that are developed and used by today's research community. Research groups and labs generally depend on their own sources of antibodies necessary for their research, and keep the antibody information according to their needs. The antibody information is stored using different data formats, and the search for the correct antibody is sometimes impossible. We would like to propose a web based database solution to this problem, by creating a database framework that can be easily used by any research group or a lab.

The type of information to be stored in the database will follow after a short overview of the role of antibodies and their general properties.

Antibodies (or immunoglobulins) can be regarded as universal adapter molecules (Sirch and Lennox, 1995). They have two very useful characteristics:

1. Each antibody binds to and attacks only one particular antigen.
2. Some antibodies, once activated by the occurrence of a disease, continue to confer resistance against that disease. Classic examples are the antibodies to the childhood diseases: chickenpox and measles.

Figure 1 shows the antibody structure. Each antibody consists of four polypeptides: two

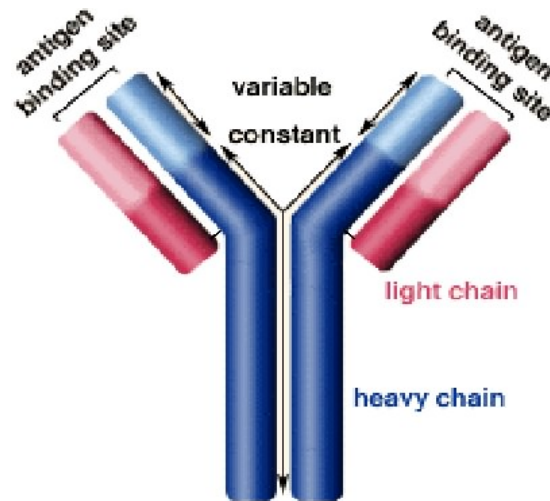


Figure 1. Antibody Structure

identical heavy chains and two identical light chains joined to form a "Y" shaped molecule. Each light chain is bound to a heavy chain via disulfide bridges to form a heterodimer (HL). The exact positions and numbers of the interchain bonds vary with the species and isotype. The variable region, which includes the ends of the light and heavy chains, is composed of 110-130 amino acids. This region gives the antibody its specificity for antigen binding. The constant region determines the mechanism used to destroy the antigen. Based on their constant region structure and immune function, antibodies are divided into five major classes: IgA, IgD, IgE, IgG, and IgM. IgG. (The Antibody Resource Page, 2003)

In the study of antibody related phenomenon, two types of antibody samples are used: *Polyclonal* and *Monoclonal*. *Polyclonal* antibodies can be obtained by immunizing an animal, such as a goat, mouse or a rabbit. After immunization, blood is removed and the antibodies can be purified directly from the serum. *Monoclonal* antibody is derived from a more complex process. An animal, almost always an inbred mouse, is immunized with an antigen. After repeated immunizations, the spleen of the animal is removed. The spleen cells contain the genetic information that gives rise to antibody production. These cells, which do not divide in culture, are fused with myeloma (a tumor cell of the bone marrow). The fused cells are called hybridoma

cells, which will continually produce antibodies. These antibodies are called *Monoclonal* because they come from only one type of cell: hybridoma cell. Depending on the type of the experiment, either a Polyclonal or a Monoclonal antibody approach may be warranted.

Antibodies are the most important part of the immune system. Not only can antibodies be used to protect against diseases, but can also help to diagnose a wide variety of illnesses, detect antigens in cells and tissues, or to suppress some special biological reactions. The most promising use of antibodies is in cancer treatment and in bone marrow transplants. Given their significance, it is not a surprise that they have been receiving a lot of attention from research community.

There are two common methods to obtain antibody products for research purposes:

- The first method is to buy antibody products from already existing antibody companies. These antibodies are widely used to detect already known antigens. Purchasing the antibody products could be rather costly.
- The other method is to produce antibodies. This is done in a case where a researcher needs a large volume of antibodies, or when the antibodies are used to detect novel antigens.

Once they are produced and tested, the investigators may provide their antibody products along with the relevant information to their collaborators for further research. Although they can exchange antibody information and the research results by personal communication means, the process is still not convenient.

Therefore, for a particular research group, there is a strong need for a reliable online antibody database. The database can be used as a research sharing system by researchers who can easily check the availability of present antibodies, search their desired antibodies, submit or update antibody information. In the case when there is a need to purchase antibody products, the database will provide the desired company link. This is the main motivation of this project.

C. Literature Review And Current State -Of-The Art

Currently, the basic information about antibodies can be only accessed through the company's catalog or from the published articles. Some of the companies are *Santa Cruz Biotechnology Inc*, *MSRS Corporation*, *BIOTREND Company*, *Antibody Solutions Inc* and *HLDA workshop*. *Santa Cruz Biotechnology Inc* provides superior, innovative primary antibodies and supports products. *MSRS Corporation* offers an interactive searchable database containing a list of 165,000 primary antibodies. *BIOTREND Company* provides more than 11,000 antibodies available for research and diagnosis purpose. *Antibody Solutions Inc* offers a user-friendly searchable antibody catalog. *HLDA Antibody Database* (Hadam, 1998) was developed by Martin R. Hadam. It covers all monoclonal antibodies that have ever been part of a prior HLDA workshop. The users can obtain all the antibodies information by giving the first letter or the number of the clone name. But, the main intention of all these companies is to sell their antibody products. All the mentioned databases list only the antibody products of the corresponding companies. A researcher would need to search a number of the databases in order to obtain their desired antibody product. This can certainly be time consuming and inefficient.

Majority of research groups maintain their own antibody databases using Microsoft Excel or Access. But, because they are not online, a researcher needs to contact the database administrator in order to search, submit, exchange or update antibody information. This problem affects research efficiency and slows the collaboration.

Therefore, there is a strong need for a unifying online antibody database research system that can be easily used by members of one particular research group.

D. Methodology

The proposed antibody database research system will provide user-friendly Web interface for a research group. One part of the group works exclusively on antibody production (*providers*),

while the other part does not produce antibodies, but uses them in their research (*users*). Every member of the group will have access to the same main page of this system. The main page (as shown in Figure 2) will include six parts: HOME, VIEW LIST, SEARCH, SUBMIT, UPDATE and COMPANY LINKS .

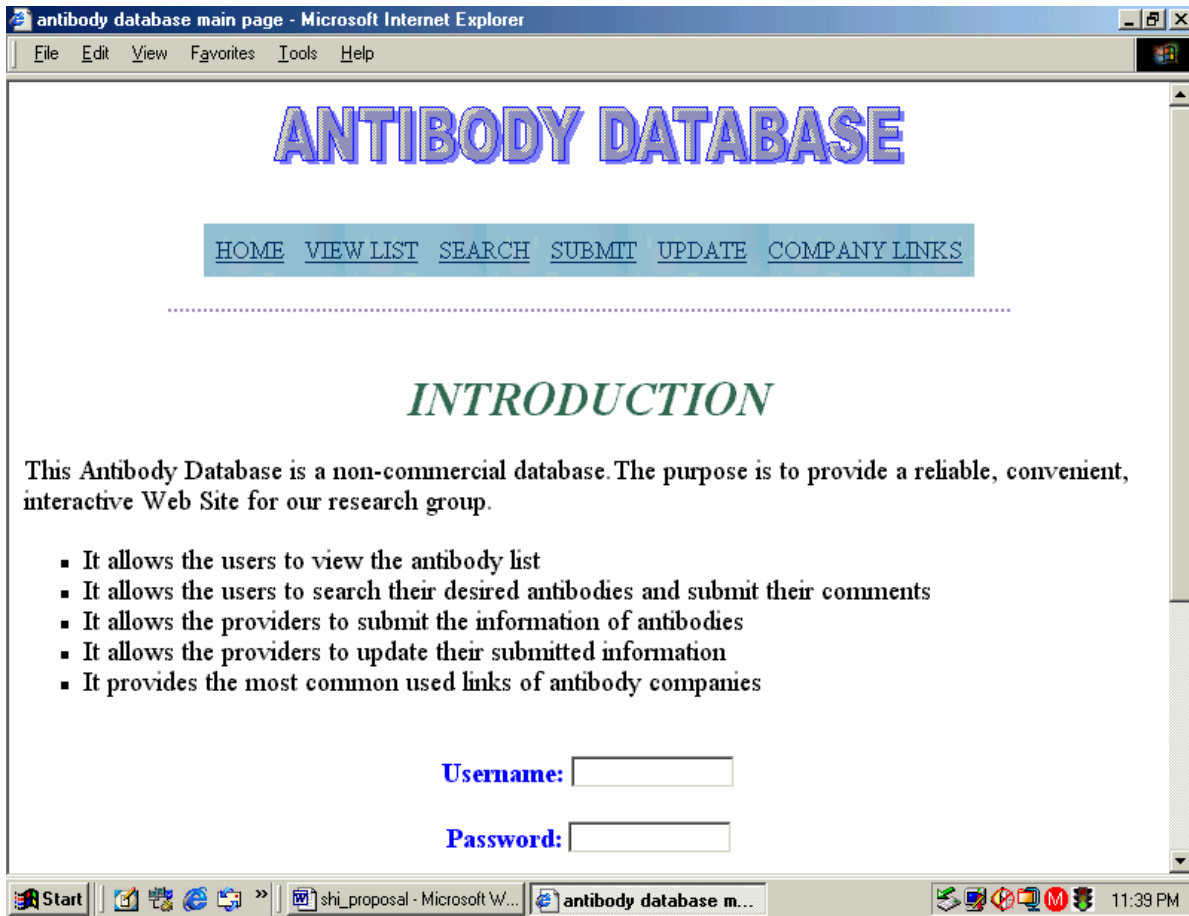


Figure 2. Main page

- HOME: explains the basic features and the purpose of this system.
- VIEW LIST: allows users to view the list of all antibody products in this group.

- SEARCH: allows users to search their desired antibodies by different properties, such as antibody name, antibody type, antigen name, isotype, provider name, submit date etc. It also allows users to submit their comments.
- SUBMIT: allows providers to submit antibody information. A username and a password will be assigned to each provider.
- UPDATE: allows providers to update their submitted information using their usernames and passwords.
- COMPANY LINKS: provides the most common used links of antibody companies for this group.

In order to secure the antibody information for the particular research group, a password and a username are needed for access. Each group member will obtain the password and the username from the database administrator.

The information of antibodies and providers will be stored in the database server. It will include five tables: PROVIDER_INF, LAB_INF, ANTIBODY_INF, ANTIBODY_USAGE, CLONE_INF.

- PROVIDER_INF: stores the information of the providers, such as the name, address, phone number, e-mail address, etc.
- LAB_INF: stores lab name, lab location, etc.
- ANTIBODY_INF: contains antibody name, antibody type (*Polyclonal* or *Monoclonal*), antigen name, species, storage, quantity information (if available), etc.
- ANTIBODY_USAGE: contains antibody usage and the corresponding dilution.

- CLONE_INFOR: includes clone name and isotype (IgA, IgD, IgE, IgG, IgM). This table is only for Monoclonal antibodies.

Like other Web-based database systems, two important issues must be considered:

Security and Performance.

1. Security: Database security issue is of very high importance, especially for the Web-based databases. This project will address the main threats to the proposed system and apply the corresponding techniques or methods against the crackers, such as Encrypted Passwords, Access Privileges, Server Security, SSH (Secure Shell), SSL (Secure Sockets Layer), Database Backup, etc. (Garfinkel, 2001; Jeff, 2000; Hoffer, Prescott and Mcfadden, 2002).
2. Performance: Some issues about query speed will be considered.

This project will be implemented using PHP and MySQL software packages.

PHP is a server-side scripting language. It originally stood for *Personal Home Page*, now stands for PHP *Hypertext Preprocessor* (Welling and Thomson, 2001). It is a widely used Open Source general-purpose scripting language that is especially suited for Web development and can be embedded into HTML (Bakken et al. 2003). Besides that PHP is available at no cost, an advantage of PHP is its cross-platform compatibility. In other words, it will not matter if the users are running Windows, Macintosh, or a version of Unix, since there is no need for any additional software in order to see PHP's dynamic content. This is because the dynamic content is processed on the server side, and then sent as if it were static. Another significant feature of PHP is its Database Integration. PHP has native connections available to many database systems. Using ODBC (*Open database Connectivity*), you can connect to any database that provides an ODBC driver. PHP has also support for most of the web servers including Apache, Microsoft Internet Information Server, Personal Web Server, Netscape and iPlanet servers, and many others. PHP

has many built in functions for performing many useful Web-related tasks, and can support millions of hits per day.

MySQL will be used as the database server in this project. Like PHP, MySQL has also provided us with its cross-platform compatibility and huge cost savings. It is a very fast, robust, *relational database management system* (RDBMS) (Matthew and Stones, 2002). MySQL is supported by Structured Query Language (SQL), an industry standard fourth generation language interacting with the RDBMS data (Linebaugh, 2001). The MySQL server controls access to the data to ensure that multiple users can work with it concurrently, to provide fast access to it, and ensure that only authorized users can obtain access. It is noted mainly for its speed, reliability, and flexibility.

Figure 3 shows the typical Web Database System Architecture in PHP and MySQL. When a user visits the database-driven Web Site, the following will occur: (Widenius and Axmark, 2002)

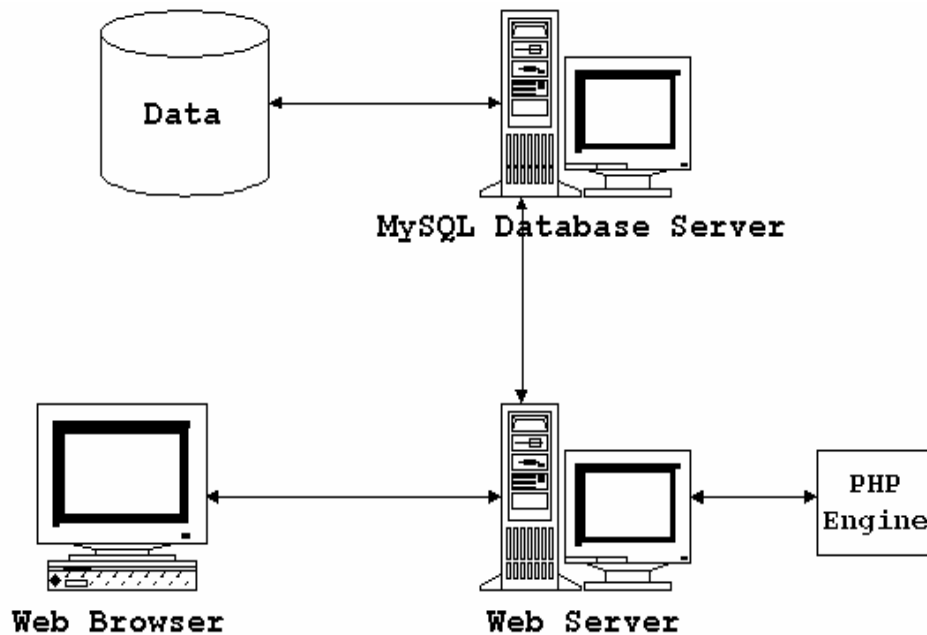


Figure 3. Web Database System Architecture

- The visitor's Web browser asks for the Web page using a standard URL.
- The Web server recognizes that the requested file is a PHP script, and so interprets it using its PHP plug-in before responding to the page request.
- Some PHP commands connect to the MySQL database and request the content that belongs in the Web page.
- The MySQL database responds by sending the requested content to the PHP script.
- The PHP script uses *echo* function to output it as part of the Web page.
- The PHP plug-in finishes up by handing a copy of the HTML it has created to the Web server.
- The Web server sends the HTML to the Web browser as it would a plain HTML file.

As discussed earlier, some research groups maintain their own antibody databases using MS Access or Excel. In order to easily populate data into the database, we will use MyODBC (*Open database Connectivity for MySQL*). The ODBC is a very powerful database solution and it greatly simplifies the conversion process. The following steps illustrate how the MyODBC works:

- Client tells the ODBC Manager what he/she wants to do
- ODBC Manager forwards the client's actions to the MyODBC Driver
- MyODBC Driver performs the operations on the MySQL database

This proposed database system will be tested with some data obtained from a research group in the department of Cell Biology, Yale University School of Medicine. It is expected that this system will be run and used by the same group.

The project documentation will include five chapters: INTRODUCTION, SYSTEM STRUCTURE, IMPLEMENTATION, SYSTEM SECURITY and CONCLUSION.

To make the documentation more complete, it will include a variety of screen shots. Complete program source code will be provided.

E. Contributions

The goal of this project is to develop and implement a user-friendly, convenient and reliable Web-based Antibody Database research sharing system for a research group.

This system will enable a user in this group to easily view the antibody product list, obtain his/hers requested products by contacting the antibody providers, send his/hers comments to the antibody providers, and find the desired company link quickly. On the other hand, the antibody providers can also easily submit or update product information. It will significantly improve scientific research efficiency and collaboration within the group.

F. References

Sirch, J.R., Lennox, E.S. (1995). *Monoclonal Antibodies, Principles and Applications* (1st ed.).

Wiley-Liss Inc.

The Antibody Resource Page (2000). *The Study of Antibody Recognition*, from

<http://www.antibodyresource.com/antibody.html>

Hadam, M.R. (1998). *HLDA Antibody Database*, from

<http://www.mh-hannover.de/aktuelles/projekte/hlda7/hldabase/alphaind.htm>

Garfinkel, S. (2001). *Web Security, Privacy & Commerce* (2nd ed.). O'REILLY

Jeff, F. (2000). *Maintaining Secure Web Applications*. Network Computing, Vol. 11 Issue 5.

Hoffer, J.A., Prescott, M.B. and Mcfadden, F.R. (2002). *Modern Database management* (5th ed.).

Prentice Hall

Welling, L., Thomson, L. (2001). *PHP and MySQL Web Development*. SAMS.

Bakken, S.S. et al. (2003). *PHP Manual*, from <http://www.php.net/manual/en/>

Matthew, N., Stones, R. (2002). *Databases with MySQL*. Wrox Press Ltd.

Linebaugh, B. (2001). *The Complete MySQL Tutorial*, AbriaSoft Company.

Widenius, M., Axmark, M. (2002). *MySQL Reference Manual*. O'REILLY.